



Article Building Change Detection Based on 3D Co-Segmentation Using Satellite Stereo Imagery

Hao Wang ^{1,2,3}, Xiaolei Lv ^{1,2,3,*}, Kaiyu Zhang ^{1,2,3} and Bin Guo ⁴

- Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Chinese Academy of Sciences, Beijing 100190, China; wanghao197@mails.ucas.ac.cn (H.W.); zhangkaiyu18@mails.ucas.ac.cn (K.Z.)
- ² Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
- ³ School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
- ⁴ Beijing Capital International Airport Group, Beijing Daxing International Airport, Beijing 102604, China; guo.b@bdia.com.cn
- * Correspondence: lvxl@aircas.ac.cn

Abstract: Building change detection using remote sensing images is significant to urban planning and city monitoring. The height information extracted from very high resolution (VHR) satellite stereo images provides valuable information for the detection of 3D changes in urban buildings. However, most existing 3D change detection algorithms are based on the independent segmentation of two-temporal images and the feature fusion of spectral change and height change. These methods do not consider 3D change information and spatial context information simultaneously. In this paper, we propose a novel building change detection algorithm based on 3D Co-segmentation, which makes full use of the 3D change information contained in the stereoscope data. An energy function containing spectral change information, height change information, and spatial context information is constructed. Image change feature is extracted using morphological building index (MBI), and height change feature is obtained by robust normalized digital surface models (nDSM) difference. 3D Co-segmentation divides the two-temporal images into the changed foreground and unchanged background through the graph-cut-based energy minimization method. The object-to-object detection results are obtained through overlay analysis, and the quantitative height change values are calculated according to this correspondence. The superiority of the proposed algorithm is that it can obtain the changes of buildings in planar and vertical simultaneously. The performance of the algorithm is evaluated in detail using six groups of satellite datasets. The experimental results prove the effectiveness of the proposed building change detection algorithm.

Keywords: building change detection; 3D change detection; stereo satellite imagery; morphological building index; digital surface models; Co-segmentation

1. Introduction

Automatic building change detection using remote sensing imagery is significant for city monitoring, disaster assessment, map updating, etc. For the very high resolution (VHR) satellite remote sensing imagery, the spatial resolution is better than 1 m, which brings the complex details of ground features. It makes the results of two-dimensional (2D) change detection vulnerable to minor changes on the ground surface. Furthermore, as the resolution improves, slight changes in the viewing angle may cause severe geometric distortions, making 2D change detection more challenging. With the development of dense stereo image matching technology [1–3] and the launch of more and more VHR stereo observation satellites, we can extract accurate height information from the satellite stereo images, which is very useful for building change detection. The introduction of the height features makes three-dimensional (3D) change detection insensitive to the image's detailed



Citation: Wang, H.; Lv, X.; Zhang, K.; Guo, B. Building Change Detection Based on 3D Co-Segmentation Using Satellite Stereo Imagery. *Remote Sens.* 2022, 14, 628. https://doi.org/ 10.3390/rs14030628

Academic Editors: Mi Wang, Hanwen Yu, Jianlai Chen and Ying Zhu

Received: 15 December 2021 Accepted: 25 January 2022 Published: 28 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). change information. It can correct the geometric distortion caused by different observation angles. Therefore, 3D change detection appears to have obvious advantages for change detection in urban areas.

However, there are still many challenges in the 3D change detections [4]. With different sensors, algorithms, and resolutions, 3D data have varying degrees of uncertainty. Elevation data differs from image data in terms of properties and characteristics, so special considerations must be made to effectively integrate these two types of data. Besides, the perspective distortion, occlusion, and incompleteness in the data also bring great challenges.

In recent years, Xiao et al. [5] applied the graph-cut algorithm to 2D change detection and proposed a co-segmentation algorithm. They simultaneously used spectral change and spatial context information to solve the global energy minimizing problem. The co-segmentation can directly divide the image into the changed object and unchanged areas. Afterward, Zhang et al. [6] further analyzed and compared multiple multi-temporal segmentation methods in detail and co-segmentation obtained the best change detection performance. Chen et al. [7] detected building changes by combining coarse location and co-segmentation. Gong et al. [8] generated the changed building proposals based on graph-cut and then the changed buildings were detected using a patch matching approach and conditional random field (CRF) optimization. Then they proposed a roof-cut algorithm [9] to detect change between imagery and existing footprint map. Zhu et al. [10] applied the superpixels algorithm to co-segmentation, which solved the problem of slow segmentation due to the complexity of graph structure. Hao et al. [11] used the support vector machine (SVM) algorithm to calculate the prior probability of change for graph-cutbased segmentation. Zhang et al. [12] proposed a multitemporal building change detection method based on co-segmentation using time-series synthetic aperture radar (SAR) images.

For 3D change detection, Qin et al. [4] summarized the existing 3D change detection algorithms in detail. Gstaiger et al. [13] concluded that using 3D change information combined with 2D information for change detection can deliver valuable indications. For the pixel-based method, Tian et al. [14,15] used digital surface models (DSM) difference and threshold segmentation to detect building change. Then they proposed a 3D change detection algorithm based on Dempster–Shafer (DS) theory [16], which fused the height change feature and spectral change feature using DS theory at the decision level. Subsequently, they continued to improve the algorithm in [17–21] by using extended DS theory named Dezert-Smarandache Theory and reliability discounting method, which considers DSM quality. As for the object-based method, Tian et al. [22] used the separate segmentation and region combination for region-based automatic building change detection. Qin et al. [23] proposed an object-based 3D building change detection method using synergic mean-shift with the constraints of the DSM. They also proposed a 3D change detection algorithm based on the LOD2 (Level of Detail 2) building model [24]. But this algorithm relies on the existing LOD2 3D building model, which is difficult and expensive to be acquired. Gharibbafghi et al. [25] used the edge-based simple linear iterative clustering (ESLIC) superpixel segmentation method to refine the 3D building model and detect building change. Wen et al. [26] proposed an object-based image-to-map change detection method to generate time-series 3D building maps. Huang et al. [27] proposed an object-based time-series newly constructed building areas detection method using both planar and vertical features. Supervised classification was also introduce into the 3D change detection algorithm, including SVM [28], decision-tree [23], and random forest [29]. However, the training label was generated by a rule set method, which has certain limitations. Tian and Qin et al. [30,31] used the random forest to extract building probability map and then combined multi-temporal building probability map with the 3D bilateral filter. Yuan et al. [32] extracted the building probability map using deep learning and integrated it with the 3D change feature based on DS fusion theory. Pang et al. [33] used a graph-cuts-based algorithm to extract changed objects from two-temporal DSMs and used the corresponding aerial images to remove non-building objects with a structural feature. Then, they used semantic

segmentation based on deep learning to extract building areas and integrated the result into the graph-cuts-based segmentation algorithm [34]. There are also some methods based on post-classification [23,35–38] in the 3D building change detection.

In this paper, we propose a novel 3D change detection algorithm for urban buildings based on 3D co-segmentation. We construct an energy function containing spectral change information, height change information, and spatial context information. By taking the 3D change feature as the prior knowledge, combining it with the spatial context information, 3D co-segmentation is performed via graph-based energy minimization [39]. It segments the two-temporal images into the changed foreground and unchanged background and obtains the change detection results with spatial correspondence. Subsequently, we can get the corresponding relationship between the two-temporal changed buildings through overlay analysis. Finally, the quantitative height change values are obtained using robust statistics.

The main contributions of the proposed algorithm are as follows:

- 1. A novel energy function is proposed that can fully mine the information contained in stereo image pairings by taking height information, spectral information, and spatial neighborhood information into consideration;
- 2. Using a sigmoid function to map the change indicator to the energy value, which can better deal with the gross error in the change features;
- 3. A 3D co-segmentation algorithm for building change detection is proposed based on the proposed energy function. The algorithm considers height change, spectral change, and spatial neighborhood information, and uses segmentation to directly obtain change detection results;
- 4. The algorithm can obtain the quantitative height change value of each changed building.

This paper is organized as follows. Section 2 describes the details of the proposed 3D co-segmentation based building change detection approach. We provide some experimental results of the proposed method on six groups of actual satellite stereo images in Section 3. Section 4 discusses the influence of adjustable parameters on the performance of the proposed algorithm. Finally, Section 5 draws the conclusion.

2. Materials and Methods

This section will introduce the proposed 3D-co-segmentation-based urban buildings change detection algorithm in detail. The workflow of the proposed method is shown in Figure 1. The inputs of the algorithm are two pairs of satellite stereo images with changes. After the dense image matching, we can obtain the DSMs as well as the orthorectified images of the two-temporal and extract the nDSM from the DSM. Then, the spectral change feature is extracted through MBI [40,41] difference, and the height change feature is extracted through the robust height difference of nDSM. The 3D co-segmentation is then done using the same 3D change feature as the prior knowledge and the spatially associated segmentation results corresponding to two-temporal orthoimages are acquired. After that, the overlay analysis is applied to the two-temporal segmentation results, and the object-to-object correspondences of the changed objects are determined. Finally, the height change values of each changed building are calculated using the robust height means.



Figure 1. Flowchart of the proposed methodology.

2.1. Stereo Reconstruction

For the input two-temporal stereo images, the block adjustment [42–44] is first performed using the automatically extracted tie-points (TP) and the external ground control points (GCP). After the adjustment, the error of the positioning model is greatly reduced, so that the DSM elevation calculation and planar positioning are more accurate. After that, the semi-global-matching (SGM) [1,3] algorithm, which employs the CENSUS transform and hamming distance as the matching cost, is used to densely match the stereo images. The DSM elevation is then calculated based on the matching result. Finally, the orthoimage is obtained using the produced DSM as the reference elevation.

Since the objects of interest are the changed buildings, the nDSM containing only the height above ground surface is extracted from the DSM using white top-hat transformation based on morphological open-by-reconstruction operation [45].

$$nDSM = DSM - R_{DSM}(DSM \ominus SE), \tag{1}$$

where \ominus represents morphological erosion operation, $R_G(B)$ represents the morphological reconstruction of *G* with *B* as the marker, and SE represents the structural element for morphological reconstruction. Here, we select the square structural element with the size of $w \times w$. The size of the structural elements needs to be greater than the maximum short axis of the buildings in the area of interest to ensure that all buildings are preserved.

2.2. 3D Change Feature

Change features provide the prior change information for subsequent co-segmentation. The main characteristics of building changes are spectral changes and height changes. We make full use of the information in stereo images to extract the 3D change features in orthoimages and nDSMs.

2.2.1. Spectral Change Feature

As the bare ground changes to buildings, the gray value of each pixel in the image often changes. Using the brightness, local contrast, size, and directionality characteristics, MBI can extract buildings based on gray-level morphological processing. The basic idea of MBI is to build the relationship between the spectral structural features of buildings and the morphological operators. These spectral structural characteristics of buildings are represented using the open-by-reconstruction with a series of linear structural elements. The change of MBI can reflect the change of buildings in the image. When the MBIs of two-temporal orthoimages are calculated, the spectral change index can be obtained by the difference between the two MBIs. Then, a 5×5 median filter is used to reduce noises.

$$MBI^{dif} = median(|MBI(I_2) - MBI(I_1)|),$$
(2)

where MBI^{dif} represents the MBI difference value between two-temporal and MBI(I) represents the MBI of image *I*. The MBI can be calculated according to [41].

$$MBI = \frac{\sum_{d,s} DMP_{WTH}(d,s)}{D \times S},$$
(3)

where DMP is the differential morphological profiles as described in [41]; *D* and *S* denote the numbers of directionality and scale of the profiles, respectively. Four directions (D = 4) are enough, and the length *s* should be determined according to the spatial resolution. In this paper, $s_{min} = 10$, $s_{max} = 210$ and $\Delta s = 50$.

2.2.2. Height Change Feature

As a significant feature of building changes, the building height changes can significantly improve change detection accuracy and reduce the impact of uninterested objects on the ground surface. The height change values can be obtained from the nDSM difference of the two temporal images. However, the nDSM produced from stereo matching cannot obtain precise boundaries in regions of elevation leap, especially at building edges. This will result in a large number of false alarms at building edges with the direct difference of the nDSM, so we use a robust nDSM difference algorithm [16] to calculate the height change values.

$$H^{dif} = robust_difference(nDSM_2, nDSM_1).$$
(4)

The robust nDSM difference calculates the minimum value of height difference within a certain window range. The robust difference between $nDSM_1$ at time T1 and $nDSM_2$ at time T2 is defined as the minimum value of the height difference between a pixel in $nDSM_2$ and all pixels within the specific neighborhood range (windows of radius *w*) of the corresponding pixel in $nDSM_1$. That is, only the difference with the smallest absolute value in all positive changes and the smallest absolute value in all negative changes are considered. The robust positive difference and the robust negative difference are defined as follows:

$$X_{Pdif}(i,j) = \min_{\substack{p \in [i-w,i+w], \\ q \in [j-w,j+w]}} \{ (x_2(i,j) - x_1(p,q)), (x_2(i,j) - x_1(p,q)) > 0 \},$$
(5)

$$X_{Ndif}(i,j) = \max_{\substack{p \in [i-w,i+w], \\ q \in [i-w,i+w]}} \{ (x_2(i,j) - x_1(p,q)), (x_2(i,j) - x_1(p,q)) < 0 \}.$$
(6)

The window size *w* depends on the quality of the nDSM. A smaller window can be selected for a higher quality nDSM and a larger window for a lower quality nDSM.

Figure 2 shows the schematic diagram of robust difference and the profile of the result. The schematic diagram in Figure 2a shows the direct difference results of two buildings with certain height differences and different boundaries. The height difference at the edge is higher than the actual buildings change due to different edge positions. The schematic

diagram in Figure 2b shows the result of robust difference. For each point, the point with the smallest difference in elevation within a certain range is found for differencing. The height difference at the edge is effectively removed, and the part really higher than the building is successfully retained. Figure 2c,d show the profile of nDSM and height difference. There is a certain degree of inconsistency at the boundary between the two nDSM due to the performance of the matching algorithm. It will lead to a large number of spikes in the direct difference results while the robust difference results effectively avoid this problem.



Figure 2. The schematic diagrams of (**a**) directed difference and (**b**) robust difference, and the profiles of (**c**) nDSMs and (**d**) height differences.

2.3. 3D Co-Segmentation

According to the principle of global energy minimization, 3D co-segmentation employs the above 3D change features as the prior, combining the spatial neighborhood similarity of the images to divide the image into foreground and background, that is, changed and unchanged.

2.3.1. Energy Function

In order to integrate the height change information, the spectral change information, and the spatial context information, we construct the following energy function:

$$E = \lambda E_{change} + (1 - \lambda) E_{smooth}$$

= $\lambda (\alpha E_{change}^{H} + (1 - \alpha) E_{change}^{I}) + (1 - \lambda) E_{smooth}$
= $\lambda \alpha E_{change}^{H} + \lambda (1 - \alpha) E_{change}^{I} + (1 - \lambda) E_{smooth}$, (7)

where $0 < \lambda \leq 1$ indicates the relative importance of the data item to the smoothing item and $0 \leq \alpha \leq 1$ indicates the relative importance of the height change feature to the spectral change feature. E_{change}^{H} , E_{change}^{I} , and E_{smooth} represent the energy item of height change, spectral change, and spatial smoothness, respectively.

Let $\lambda_H = \lambda \alpha$, $\lambda_I = \lambda (1 - \alpha)$ represent the weights of height change energy and spectral change energy, then

$$\lambda = \lambda_H + \lambda_I,$$

$$\alpha = \lambda_H / (\lambda_H + \lambda_I),$$
(8)

and

$$E(L) = \lambda_H E_{change}^H + \lambda_I E_{change}^I + (1 - \lambda_H - \lambda_I) E_{smooth}$$

= $\lambda_H \sum_{p \in P} D_p^H(l_p) + \lambda_I \sum_{p \in P} D_p^I(l_p) + (1 - \lambda_H - \lambda_I) \sum_{(p,q) \in N} V_{p,q}(l_p, l_q),$ (9)

where $L = \{l_p | l_p \in \{l_{fg}, l_{bg}\}, p \in P\}$ is a possible label for all pixels p in image P. l_{fg} and l_{bg} represent foreground and background labels, respectively. $D_p^H(l_p)$ and $D_p^I(l_p)$ are the energy of the feature items and represent the cost of assigning label l_p to pixel p according to the height change or spectral change. $V_{p,q}$ is the smoothing item energy, which represents the penalty when pixel p and pixel q allocates different labels. N represents the set of all possible neighborhood pixel pairs.

2.3.2. Graph Construction

As show in Figure 3, for each image, we construct the weighted graph

$$\begin{cases}
G = (V, E), \\
V = P \cup \{s, t\}, \\
E = N \cup \{(s, p), (p, t)\}.
\end{cases}$$
(10)

The nodes set *V* of the graph contains two kinds of nodes: pixel nodes and terminal nodes. The pixel nodes set *P* compose all pixels of the image, and each pixel represents a node. Besides, there are two terminal nodes. The source node *s* represents the background that didn't change, and the sink node *t* represents the foreground that changed. Similarly, the edges set *E* contains two kinds of edges, neighbor edges N (*n*-links) that connect two-pixel nodes, and terminal edges *T* (*t*-links) that connect a pixel node with a terminal node. Each pixel node *p* has two *t*-links (*s*, *p*) and (*p*, *t*) connected to the two terminal nodes and several *n*-links connected to its neighboring pixel nodes.



Figure 3. The graph structure of 3D co-segmentation and the cut (adapted from [46]). The yellow edge represents *n*-links, and its weight depends on the image. The red and blue edges represent *t*-links, and their weight depend on the change index. The green dashed line represents the minimum cut. Edge weights are reflected by their thickness.

2.3.3. Edge Weights

The weight setting for each edge is listed in Table 1.

Table 1. Graph edge weight setting.

Edge Type	Value	Condition
<i>n</i> -link (<i>p</i> , <i>q</i>)	$(1-\lambda_H-\lambda_I)V_{p,q}$	$(p,q) \in N$
t-link (s, p)	$\lambda_H D_p^H(l_{fg}) + \lambda_I D_p^I(l_{fg})$	$p \in P$
<i>t</i> -link (<i>p</i> , <i>t</i>)	$\lambda_H D_p^H(l_{bg}) + \lambda_I D_p^I(l_{bg})$	$p \in P$

The edge weights correspond to the energy function items. The *n*-link weights correspond to the smoothing energy item and depend on the image neighborhood features. They represent the penalty when similar pixels are assigned different labels. The similarity measure of pixels is expressed by the Euclidean distance and the difference between the pixel's gray values. The weight of the edge connecting the two pixels is larger when their positions are closer, and their gray values are more similar. The weighting function of *n*-links is defined as follows

$$\begin{cases} V_{p,q} = \exp(-\frac{\|I_p - I_q\|^2}{2\sigma^2}) \cdot \frac{1}{d(p,q)}, \\ d(p,q) = \sqrt{(r_p - r_q)^2 + (c_p - c_q)^2}, \\ \sigma^2 = E(\|I_p - I_q\|^2), \end{cases}$$
(11)

where I_p , I_q denote the gray value of pixels p, q; r_p , c_p denote the row and column coordinates of pixel p; d(p,q) represents the Euclidean distance of pixels p, q in image coordinate and $E(\cdot)$ represents the expectation for all possible neighboring pixels pairs in N.

The weights of *t*-link correspond to the data energy item and depend on the change features. They represent the prior knowledge of whether a pixel has changed or not. For the edge (s, p) connecting the source node *s* to a pixel node *p*, the weight of the corresponding edge should be lower when the probability of change is higher. Similarly, for the edge (p, t) connecting a pixel node *p* to the sink node *t*, the weight of the corresponding edge should be higher when the probability of change is higher. Many existing works [5,6,8,12]

of literature use the negative logarithm as the weighting function. It can map the change index in the range of [0, 2T] to the edge weight, where *T* is the threshold value. When the change index is greater than 2*T*, it is considered that the pixel must have changed. Due to errors, some pixels in the 3D building change detection based on stereo satellite images may have a change index of greater than 2*T*, therefore we cannot infer that this pixel has changed. Therefore, to preserve some prior probabilities even for the pixels with large changes, we use the sigmoid function as the weighting function. The sigmoid function has an "S" shape and maps the change values to the intervals (0, 1). The specific weighting function is defined as follows:

$$D_{p}(x, l_{p}; T, \tau) = \begin{cases} \frac{1}{1 + \exp\left(\frac{x-T}{\tau}\right)}, & \text{if } l_{p} = l_{fg} \\ \frac{1}{1 + \exp\left(-\frac{x-T}{\tau}\right)}, & \text{if } l_{p} = l_{bg} \end{cases}$$
(12)

$$\begin{cases} D_{p}^{H}(l_{p}) = D_{p}(H_{p}^{dif}, l_{p}; T_{H}, \tau_{H}), \\ D_{p}^{I}(l_{p}) = D_{p}(MBI_{p}^{dif}, l_{p}; T_{I}, \tau_{I}), \end{cases}$$
(13)

where *x* represents the change index of pixel *p*. The greater the change index, the higher the probability of building change. T is the threshold for change index and controls the symmetry point of the sigmoid function. Figure 4a shows the sigmoid functions with different parameters *T*. Since the height change has a specific physical meaning, we can determine the threshold value T_H manually according to the corresponding application and the precision of the nDSMs. The threshold value T_I of spectral change can be determined by an automatic threshold selection method such as OTSU [47]. From now on, we ignore the coefficients λ and the other item in the *t*-link weight and only discuss one $D_p(l_p)$. When the change index is T, it is considered that the probability of change is 50%. At this time, the weights of edge (s, p) and edge (p, t) are the same. When the change value is greater than T, the change probability increases, the weight of edge (s, p) approaches 0, and the weight of edge (p, t) approaches 1. At this time, the pixels tend to be classified as changed pixels, vice versa. However, the weight of all *t*-links will always not be 0. The parameter τ controls the shape of the sigmoid function. Figure 4b shows the sigmoid functions with different parameters τ . A larger τ will make the weight function smoother, and the change of weight with the change index will be slower. On the contrary, a smaller τ will make the weight function steeper, and the change of weight with change index is severer. After determining the threshold T, the parameter τ can be obtained using a given sampling point on the sigmoid function.



Figure 4. Sigmoid functions with (**a**) different *T* and (**b**) different τ .

The weight coefficients λ_H and λ_I determine the relative importance between height change feature, spectral change feature, and spatial neighborhood feature. Too large λ means that the change feature plays a leading role. It will make the results over-segmented and produce more false alarms. Too small λ means the spatial neighborhood smoothing feature plays a leading role. The segmentation result will be too smooth, multiple changed buildings will be connected as a whole, and some buildings with small areas and small changes will be smoothed out. When $\lambda = 1$, the segmentation result depends on the change feature, and the result is the same as the pixel-based threshold method. λ_H and λ_I cannot be set to 0 at the same time because the changes cannot be detected without the change feature.

2.3.4. Energy Minimization

At this point, the problem has been converted to a standard graph-cut problem which can be solved by the standard max-flow/min-cut [48] algorithm. After the segmentation, all the pixels connected to the *s*-node are the unchanged pixels, and all the pixels connected to the *t*-node are the changed pixels.

2.4. Spatial Correspondence

Using the same prior knowledge of the change, we segment the two-temporal orthoimages separately and obtain two spatially related segmentation results. Firstly, we perform morphological closing and morphological opening operations on the segmented results to fill the small gaps and remove the speckle noises. Fragmentation areas smaller than a certain threshold are then removed, which is often due to other changes on the ground surface and stereo matching errors. Subsequently, overlay analysis [5] is performed on the two-temporal segmentation results. With this step, we get the object-to-object correspondence relationship between the two-temporal segmentation results. Because two-temporal image features are added to the co-segmentation process, the segmentation results are related to specific images, which allows the geometric change of the building to be detected.

2.5. Height Change Determenation

After obtaining the correspondence relationship of the changed objects, the heights of each building at two-temporal are calculated based on the nDSM. There will be some coarse pixels in the height values of each area because the building edges in the nDSM are inaccurate and inconsistent with the segmentation results. So, we use a robust height mean method to calculate the average height of each building.

As shown in Figure 5, for each building, the nDSM height values of all the pixels are sorted. The maximum and minimum 10% pixels are removed, and then the average height of the remaining intermediate pixels is calculated as the final height. After overlay analysis, the object-to-object correspondence relationship between the two-temporal segmentation result is obtained. We can get the height change values of the changed buildings by subtracting the average heights of the corresponding objects.



Figure 5. The schematic diagram of robust height mean (adapted from [16]).

3. Results

To illustrate the effectiveness of the proposed algorithm, the experimental results using several real satellite datasets are shown below.

3.1. Study Area and Datasets

The first experimental area is located in Henan Province, China. It contains two pairs of GaoFen-7 stereo imagery data and the corresponding multi-spectral images collected on

11 December 2019, and 17 January 2021, respectively. The resolution of the panchromatic stereo imagery is 0.8 m, and the intersection angle is about 31 degrees. Because the satellite is in a sun-synchronous orbit and the data collecting seasons are the same, the sun altitude angle of two-temporal is almost the same, implying that the impact of illumination conditions on change detection is insignificant. The second experimental area is located in Sichuan Province, China. It contains two pairs of GeoEye-1 stereo imagery data and the corresponding multi-spectral images collected on 21 August 2021, and 21 November 2021, respectively. The resolution of the panchromatic stereo imagery is 0.5 m, and the intersection angle is greater than 30 degrees. As shown in Table 2, the two-temporal stereo images have different viewing angles, so the geometric distortion will result in the failure of 2D change detection. However, this distortion can be corrected by introducing the DSM to obtain a well-registered true orthorectified image.

		Т	`1	T2		
		Azimuth	Zenith	Azimuth	Zenith	
DataSet 1	BackWard	191.475671	5.528352	241.092863	9.128378	
	ForeWard	12.051169	28.249459	359.066534	28.711070	
DataSet 2	BackWard	290.6	17.1	348.3	15.3	
	ForeWard	221.2	32.9	208.2	23.4	

Table 2. Satellite view angles.

The input stereo pairs are processed using the stereo 3D reconstruction algorithm described above. The GCPs are acquired manually using an existing reference digital orthophoto map (DOM), and the TPs are obtained automatically using a phase-correlationbased matching method. For all images in each dataset, the block adjustment is performed using the GCPs and TPs mentioned above. The DSMs and corresponding orthoimages are generated using the adjusted positioning model, and nDSMs are extracted from DSMs. With the adjusted positioning model, all generated DSMs and orthoimages are co-registered to each other with errors of less than one pixel. The elevation accuracy of the DSM is better than 1.5 m. The reference change maps are manually labeled, and only buildings larger than 100 m² are considered. Figure 6 shows the orthoimages, generated DSMs, and manually labeled reference changed building maps of the first dataset. Both pairs of data are collected in winter, and the buildings are very high, so the long shadows make it difficult for traditional 2D methods to detect the correct building changes. Additionally, changes in plant height have less influence on the detection result because there is less vegetation in the winter. As a result, the changes in height are mainly caused by the building changes. Figure 7 shows the corresponding orthoimages, DSMs, and reference ground truth of the second dataset. This area is mainly composed of construction sites and also contains less vegetation. Because of the large off-nadir viewing angle of this dataset, the building will cause a large region of occlusion, making change detection more difficult.

Test area 1 contains a large number of buildings under-construction in 2019. In 2021, these buildings have been mostly completed. The majority of the completed buildings are more than 80 m tall, with a small number of lower buildings. Test area 2 is located in a construction area. Most buildings just started construction in 2019 and are still not completed in 2021, but the building heights have changed significantly. Test area 3 is a school district with four new dorms and relatively low building heights. In the center of the area, there is a large gymnasium with a single and smooth texture that cannot be reconstructed correctly by stereo matching, bringing significant challenges for follow-up detection. Test area 4 is a new buildings under construction. Although the acquisition interval of two-temporal data is only three months, the height of these under-construction buildings has changed significantly. Test area 5 contains many buildings that cannot obtain elevation due to the occlusion in T1 imagery, but they can in T2 imagery. These occlusion areas will be interpolated as the ground, which will produce a large number of false alarms.



Figure 6. Stereo reconstruction results of dataset 1, Test areas 1–4: (**a**) Orthorectified panchromatic image from T1; (**b**) Orthorectified panchromatic image from T2; (**c**) DSM from T1; (**d**) DSM from T2; (**e**) Ground truth.





3.2. Evaluation Method

3.2.1. Pixel Based Accuracy

For the pixel-based evaluation, the changed areas are usually relatively small compared with the unchanged areas. To reduce the impact of class imbalance, we use *F*-score to measure the result accuracy. *F*-score is defined as follows:

$$precision = \frac{TP}{TP + FP'},$$
(14)

$$recall = \frac{TP}{TP + FN'}$$
(15)

$$F\text{-score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}},$$
(16)

where TP is true-positive, indicating the number of correctly detected changed pixels, FP is false-positive, indicating the number of unchanged pixels mistakenly detected as changed pixels, and FN is false-negative, indicating the number of changed pixels that are mistakenly detected as unchanged pixels. Precision is negatively correlated with the false alarm rate. The higher the precision, the lower the false alarm rate. The recall rate is negatively correlated with the missed detection rate. The higher the recall rate, the lower the missed detection rate. *F-score* is the harmonic average of precision and recall, which means a compromise between the false alarm and missed detection. It can provide a more reasonable accuracy evaluation when the numbers of samples in the two classes are unbalanced. In addition, the true-negative (TN) is not considered in the calculation of *F-score*, which reduces the influence that most pixels in the change detection result are TN.

3.2.2. Object Based Accuracy

In the application of building change detection, we are more concerned about whether changed buildings are detected correctly rather than whether the edge of the detection results is accurate. Therefore, we introduce the object-based accuracy evaluation index. As a high-level evaluation index, we ignore the size of the detected building and only count the number of detected buildings. When the overlap area between the detected object and the reference map is higher than 40%, we believe that a changed building has been correctly detected. The following four indicators [16] are used to evaluate the accuracy of the experiment results

- True detected number (TDN): The number of changed buildings detected correctly;
- True detected rate (TDR): The ratio of the number of correctly detected buildings to the total number of changed buildings in the reference map, TDR = TDN / $N_R \times 100\%$;
- False detected number (FDN): The number of incorrectly detected buildings with unchanged as changed;
- False detected rate (FDR): The ratio of the number of buildings that are incorrectly detected as changed to the total number of buildings detected in the result, FDR = $FDN/N_D \times 100\%$.

3.2.3. Height Change Accuracy

When calculating the accuracy of height change value, we only consider the correctly detected changed buildings. The height change accuracy is evaluated by the root-mean-square error (RMSE) between the detected height change value and the actual height change.

RMSE =
$$\sqrt{\frac{1}{\text{TDN}} \sum_{i=1}^{\text{TDN}} (H_d^i - H_r^i)^2}$$
, (17)

where TDN is the number of correctly detected buildings mentioned above, H_d^i represents the height change value of the *i*th detected changed building, H_r^i represents the reference height change of the corresponding building.

3.3. Parameter Setting

The structure element size of nDSM extraction is set to 100 pixels. The DSM obtained by the SGM algorithm is difficult to get the accurate building boundary, so the window size w of nDSM robust difference is set to 19 pixels. Set the threshold T_H for calculating t-links weight to 5 m. That is, when the height changes by 5 m, it is considered that there is a 50% probability of building change. The threshold T_I for MBI change is selected by the OTSU method. As for the shape parameter τ , we consider that even when the MBI and building height do not change, the probability of actual building change is 20%. The value of parameter τ_H and τ_I can be obtained by substituting the point (0,0.2) into the sigmoid function. The parameters λ in this paper are determined according to Table 3. The λ at time T1 is slightly larger than that at time T2 because it needs a higher weight for the change feature to preserve the changed object for a smoother image. We set the weight λ_I of the spectral change feature to 0 for Test areas 1 and 2 since the tops of the buildings are too dark to see changes in MBI. We pick 100 m² as the overlay analysis area threshold. We believe that buildings smaller than 100 m² are tiny houses, which can be ignored.

Table 3. λ setting for Test areas 1–6.

	λ	Test Area 1	Test Area 2	Test Area 3	Test Area 4	Test Area 5	Test Area 6
T1	λ_H	0.5	0.5	0.4	0.3	0.3	0.2
TO	λ_{H}	0.0 0.4	0.0	0.3	0.2	0.2	0.15 0.2
12	λ_I	0.0	0.0	0.1	0.1	0.1	0.15

3.4. Experiment Result

The results of proposed algorithm are shown in Figures 8 and 9. The height difference values calculated using the robust difference of nDSM are shown in Figures 8a and 9a. The red color denotes a height difference greater than 20 m, whereas the blue color denotes a height change less than 0 m. Only the positive height changes are considered because there are only new buildings and no demolition in the study area. The influence of ground surface change has been almost eliminated through nDSM extraction operation. Test areas 4, 5, and 6 contain a large number of errors caused by DSM interpolation. These errors are difficult to eliminate without the help of image information. Figures 8b,c and 9b,c show the object-to-object correspondence of the detected results between two-temporal, overlaid with the panchromatic image. The same color indicates the corresponding changed buildings. Almost all the changed buildings have been correctly detected, and only a few areas have missed detection and false alarm. Because the same change features are utilized, the results are almost identical, with only a few differences at the border. Most areas with obvious height changes are saved through segmentation, whereas those with slight height or spectral changes are successfully eliminated. The boundaries of changed buildings are more fit reality through the segmentation. After overlay analysis, some fragmented false alarm areas are further removed. But there are still some false alarms due to the large area of height error during the stereo match.



Figure 8. Experiment results of dataset 1, Test areas 1–4: (**a**) Robust difference value of nDSM; (**b**,**c**) Detected changed buildings at T1 & T2, same colors represent corresponding changed buildings; (**d**) Height change values of changed buildings between T1 and T2.



Figure 9. Experiment results of dataset 2, Test areas 5–6: (**a**) Robust difference value of nDSM; (**b**,**c**) Detected changed buildings at T1 & T2, same colors represent corresponding changed buildings; (**d**) Height change values of changed buildings between T1 and T2.

To illustrate the accuracy of the detected result, the comparison between the detected binary change maps and the reference values are shown in Figures 10d and 11d, in which the green represents correct detection, the red represents false alarm, and the blue represents missed detection. The corresponding accuracy evaluation can be seen in Table 4. Except that the *F*-score of test area 5 is 0.79, the *F*-score of all other test areas is more than 0.8. In the final result, 33 out of the 35 changed buildings are correctly detected for Test area 1, with only two missed detected buildings and one false alarm. Since two buildings in the occlusion area of nearby buildings are not visible in the stereoscopic forward view image, the actual elevation in the DSM cannot be obtained. So, the two missed detection occur. The false alarm occurs because there is one detected building with the wrong position. For Test area 2, 54 out of the 60 changed buildings are successfully detected. Here, the reason why N_D is greater than TDN plus FDN is that some buildings are successfully detected, but the overlaps with the reference map are less than the threshold value. Therefore, they do not belong to TDN or FDN but belong to DN. Many miss detections occur in this area due to a large number of under-constructed buildings with slight height and spectral changes that are difficult to detect. The one false alarm is caused by a tower crane at that position. For Test areas 3-6, all changed buildings are successfully detected. The false alarm is also caused by a tower crane for Test area 4. Figures 8d and 9d shows the height change value of each detected changed buildings. We can observe the location and degree of building changes through this figure. Most of the changed buildings with different height changes are detected. The maximum height change is about 100 m, and the minimum height change is only about 10 m. Other buildings that have not changed will not appear on this figure. The accuracy evaluation of height change value will be carried out later.



Figure 10. Result evaluation of dataset 1, Test areas 1–4. (a) Results of threshold segmentation of height change map using threshold T_{H} ; (b) Results of belief function-based DS fusion method; (c) Results of graph-cut based DSM segmentation method; (d) Results of the proposed method. The green color represents true detection. The red color represents false alarms. The blue color represents missed detected.



Figure 11. Result evaluation of dataset 2, Test areas 5–6. (a) Results of threshold segmentation of height change map using threshold T_{H} ; (b) Results of belief function-based DS fusion method; (c) Results of graph-cut based DSM segmentation method; (d) Results of the proposed method. The green color represents true detection. The red color represents false alarms. The blue color represents missed detected.

Table 4. Accuracy Evaluation of various methods. dnDSM-T: Direct threshold segmentation of nDSM difference value. DS Fusion: Belief function-based DS fusion method. DSM-GC: Graph-cut based DSM segmentation method. The best performance is highlighted in bold.

DataSet	Method	Pixel Based			Object Based			
		Precision	Recall	F-Score	TDN/N_R	TDR	FDN/N_D	FDR
	dnDSM-T	0.75353	0.85181	0.79966	/	/	/	/
Test succ 1	DS-Fusion	0.77981	0.83344	0.80573	33/35	94.29%	0/35	0%
lest area 1	DSM-GC	0.79883	0.83449	0.81627	33/35	94.29%	1/35	2.86%
	Proposed	0.78533	0.83475	0.80929	33/35	94.29%	1/34	2.94%
	dnDSM-T	0.74105	0.83150	0.78367	/	/	/	/
Test area 2	DS-Fusion	0.76292	0.80187	0.78191	54/60	90%	3/57	5.26%
lest died 2	DSM-GC	0.77557	0.82755	0.80072	57/60	95%	4/61	6.56%
	Proposed	0.81698	0.79966	0.80823	54/60	90%	1/57	1.75%
	dnDSM-T	0.75395	0.74629	0.75010	/	/	/	/
Test area 2	DS-Fusion	0.82399	0.73545	0.77720	4/4	100%	1/5	20%
lest area 5	DSM-GC	0.75901	0.69289	0.72445	4/4	100%	3/7	42.86%
	Proposed	0.88972	0.89174	0.89073	4/4	100%	0/4	0%
	dnDSM-T	0.68805	0.90771	0.78276	/	/	/	/
Test area 4	DS-Fusion	0.65317	0.91346	0.76169	17/17	100%	1/17	5.88%
lest alea 4	DSM-GC	0.67037	0.88906	0.76438	17/17	100%	3/19	15.79%
	Proposed	0.73124	0.88523	0.80090	17/17	100%	1/18	5.56%
	dnDSM-T	0.53095	0.95516	0.68251	/	/	/	/
Test area 5	DS-Fusion	0.60928	0.92995	0.73621	19/19	100%	1/20	5%
lest area 5	DSM-GC	0.56258	0.96028	0.70950	19/19	100%	8/27	29.63%
	Proposed	0.72754	0.87491	0.79445	19/19	100%	0/19	0%
Test area 6	dnDSM-T	0.55767	0.92747	0.69653	/	/	/	/
	DS-Fusion	0.67196	0.88452	0.76373	7/7	100%	3/9	33.33%
	DSM-GC	0.63604	0.91741	0.75124	7/7	100%	6/12	50%
	Proposed	0.86920	0.87839	0.87377	7/7	100%	0/7	0%

3.5. Accuracy Evaluation

To further verify the effectiveness of our proposed algorithm, we compare the proposed method with several existing state-of-the-art 3D building change detection algorithms. The first algorithm directly performs threshold segmentation on the difference value of nDSM (dnDSM-T) as the change detection result. It serves as the baseline for the comparison of subsequent algorithms. The second algorithm is the DS fusion method [21] (DS-Fusion) based on belief function. It models 3D change detection as a feature decision fusion problem which fuses the change information in the spectrum and DSM using DS evidence theory. The detection results are obtained without setting a threshold. Furthermore, it incorporates spectral and DSM confidence into the reliability discounting process, which greatly reduces false alarms caused by DSM interpolation errors. The third method segments the two-temporal DSMs using the graph-cut method [33] (DSM-GC). It builds a graph using DSM, nDSM, and differential DSM (dDSM) to extract changed buildings. Since the original paper uses aerial images, we will adjust the algorithm parameters based on the data resolution and quality in our experiment to achieve the best performance. All algorithms use the robust difference of the same window to calculate the height change value. Except for the dnDSM-T, all detection results undergo the same fragmentation removal procedure, where regions with an area less than 100 m² are removed.

The accuracy evaluation of the detection results of all algorithms are shown in Table 4. The corresponding change detection results can be seen in Figures 10 and 11. Compared with other algorithms, our proposed algorithm obtains better results in all Test areas. Especially for Test areas 4–6 with more DSM interpolation errors, the improvement of the proposed algorithm is significant. Only in Test area 1, the *F*-score is slightly lower than the DSM-GC method. As the baseline of detection performance, dnDSM-T can retain most of the changed buildings, but there are many false alarm areas due to the existence of DSM errors. The recall rate is very high but the precision is very low, so the performance of the *F-score* is ordinary. However, it still reaches more than 0.68. Especially for Test areas 1–3 with few errors, the performance is only a little lower than the other three algorithms. We can foresee that for high-quality DSMs, good results can be obtained only through dDSM threshold segmentation. The DS-Fusion method can effectively reduce mistakes caused by shadows and DSM interpolation errors because of the incorporation of image and DSM reliability. The DS-Fusion algorithm can effectively combine the advantages of the spectral and height change features. This makes the precision of the detection results improved at the cost of a slight decrease in recall. However, the detected building boundary cannot be optimized due to the lack of spatial neighborhood information, resulting in inaccurate building boundaries and some holes inside the building. The DSM-GC algorithm uses nDSM to indicate the existence of buildings and dDSM to indicate the changes of buildings. Graph cut is used to detect building changes using DSM smoothing as the constraint. The incorporation of a spatial smoothing constraint resulted in a smoother border in the detection results, as well as a high recall value. The algorithm achieves good performance in Test areas 1 and 2 which have high DSM quality. It effectively maintains buildings with slight height changes in Test area 2 and obtains the highest TDN. However, because only DSM data is used for segmentation, the algorithm cannot deal with a large area of DSM error. This makes many false alarms in the result. Although some false alarms can be removed through post-processing, the FDN value of the final result is still very high. As for the proposed method, with the co-segmentation process and subsequent overlay analysis, the precision of the detected result is greatly improved with a slight decrease of the recall. In comprehensive, the *F*-score of the detection result is improved. This is mainly due to the inclusion of the MBI change and image neighborhood information. Through segmentation, a large number of noises are removed, but the real changes are retained. The false alarm caused by the wrong building height is successfully removed by introducing the spectral change feature. The detected building boundary is more smooth and more consistent with the actual boundary in the image than the pixel-based method with the consideration of neighborhood information in co-segmentation. With the advantage of height features, almost all algorithms can correctly detect changed buildings. However, our algorithm can effectively reduce the false alarm in the result while maintaining the correct detection of changed buildings through the comprehensive consideration of height change features, spectral change features, and spatial smoothing features.

For the accuracy of the detected height change value, we only obtain the height reference values of some buildings in Test areas 1 and 2 in 2021. The data comes from the public information of Zhengzhou Bureau of natural resources and planning. Since the building was under construction in 2019, the building height cannot be obtained. The height change value accuracy is reflected by the height accuracy of the detected changed building at time T2. Table 5 shows the detected building height and the reference value. The heights of most buildings are close to the reference value and the overall RMSE is 1.435 m, which is close to the accuracy of DSM. Since the building HanS1# has not been completed at time T2, its height is far lower than the reference height. After removing this building, the overall RMSE value is 1.066 m.

Building ID	Detected T2	Ref	Bias	Building ID	Detected T2	Ref	Bias
Xin1	98.87	99.9	-1.03	Hong1#	70.8	71.4	-0.6
Xin2	99.33	99.9	-0.57	Hong2#	71.46	71.4	0.06
Xin3	100.1	99.9	0.2	Hong3#	71.62	71.4	0.22
Xin4	98.95	99.9	-0.95	Hong5#	72.38	74.3	-1.92
Xin5	100.5	99.9	0.6	Hong6#	81.12	79.5	1.62
Xin6	100.5	99.9	0.6	Hong7#	80.71	79.5	1.21
Han1#	100.3	99.8	0.5	Hong8#	80.89	79.5	1.39
Han2#	97.43	97.5	-0.07	Hong9#	80.01	79.95	0.06
Han3#	96.72	97.5	-0.78	Hong10#	80.44	79.95	0.49
Han5#	95.45	97.5	-2.05	Hong11#	79.86	79.5	0.36
Han6#	96.21	97.5	-1.29	Hong12#	79.2	79.5	-0.3
Han7#	98.07	99.8	-1.73	Ze1#	99.65	97.6	2.05
HanS1#	13.94	20.4	-6.46	Ze2#	96.88	97.6	-0.72
HanS2#	12.77	14.7	-1.93	Ze3#	96.53	97.6	-1.07
Shao1#	100.4	99.8	0.6	Ze5#	11.02	12.1	-1.08
Shao2#	99.34	99.8	-0.46	Ze6#	77.68	77.3	0.38
Shao3#	99.76	99.8	-0.04	Ze7#	96.74	97.6	-0.86
Shao5#	95.98	97.5	-1.52	Ze8#	74.17	74.4	-0.23
Shao6#	93.04	94.6	-1.56	Ze9#	96.85	97.6	-0.75
Shao7#	93.34	94.6	-1.26	Ze10#	8.925	8.65	0.275
ShaoS1#	11.81	13.35	-1.54	Ze11#	99.2	97.6	1.6
ShaoS2#	12.19	13.35	-1.16	Ze12#	60.16	59.9	0.26

Table 5. Height accuracy evaluation (unit: meter). RMSE = 1.435 m.

4. Discussion

To clarify the influence of the adjustable parameters of the algorithm on the results, we do three groups of experiments to analyze the influence of the weight parameters λ_H and λ_I and the threshold parameter T_H in detail.

First, we set the height threshold T_H to 5, the spectral change feature weight λ_I to 0, and height change feature weight λ_H from 0.1 to 1 in steps of 0.1. The same λ is used for the two segmentations. Figure 12 shows the effect of the weight parameter λ_H on the algorithm result. When $\lambda_H = 1$, the result is the same as direct threshold segmentation. With the increase of λ_H , the performance of the algorithm improves rapidly and then decreases very slowly until it is the same as the result of direct threshold segmentation. The precision increases gradually with λ , while recall decreases gradually. When the λ_H is small, the segmentation result is too smooth, and many changed buildings are filtered out, resulting in many missed detections. When λ_H reaches a certain threshold, we get the best results. The correct changed buildings are retained, and the noises are successfully removed. Finally, as λ_H increases continuously, more noises are retained. Because the noise pixels account for only a few in the pixel-based evaluation indicators, the *F-score* decreases slowly. However, the number of false alarm objects is gradually growing, as seen by the object-based evaluation indicators. Here is the result of post-processing, so much of the noises can still be successfully removed.



Figure 12. Influence of λ_H on co-segmentation accuracy. (**a**,**c**,**e**,**g**,**i**,**k**) Influence of λ_H on TDR and FNR for Test areas 1–6; (**b**,**d**,**f**,**h**,**j**,**l**) Influence of λ_H on precision, recall and *F*-score for Test areas 1–6.

Figure 13 shows the influence of parameter λ_I on the result. In this experiment, we set the same λ_H as Table 3 and λ_I from 0 to 0.5 in steps of 0.05. When the weight of the MBI change feature is slightly increased, the accuracy of Test areas 1 and 2 is almost unchanged, but TDR decreases slightly. Some buildings are in shadow or incomplete construction, resulting in small height changes and almost unchanged MBI. These buildings will be missed detected with the increase of λ_I . As for Test area 3, after introducing the spectral change feature, the false alarm caused by the height error of the gymnasium in the center is successfully removed, and the four changed buildings become complete. Therefore, both precision and recall are significantly improved, resulting in a large improvement of F-score. The accuracy of Test area 4 is slightly improved, and the false alarm caused by the surface change is further eliminated. The accuracy of Test areas 5 and 6 are improved significantly due to the elimination of false alarms caused by DSM errors. When λ_I increases further, the performance of the algorithm decreases rapidly. It is difficult to extract accurate change information from the image containing complex details. So, the error information in the spectral change feature is much more than that in the height change feature. When the weight of the spectral change is higher than the height change, the result of the algorithm will deteriorate rapidly. Therefore, we should choose λ_I smaller than λ_H in practical application.



Figure 13. Influence of λ_I on co-segmentation accuracy. (**a**,**c**,**e**,**g**,**i**,**k**) Influence of λ_I on TDR and FNR for Test areas 1–6; (**b**,**d**,**f**,**h**,**j**,**l**) Influence of λ_I on precision, recall and *F*-score for Test areas 1–6.

Regarding the height threshold T_H , we set parameter T_H from 1 to 20 in steps of 1. The parameter λ_H is the same as in Table 3. λ_I is set to 0 and τ_H is set to constant 3. Figure 14 shows the influence of height threshold T_H on the algorithm results. The setting of parameter T_H depends on the actual height of the region of interest and the target of interest. The results show that the algorithm can achieve good results in a large range. When T_H is too small, the segmentation result will retain small height changes. These changes may be real building changes or some other surface changes, shown as high recall rate and low precision in the result. When T_H is greater than the height of the interesting building, the corresponding building will be recognized as unchanged. It will lead to many missed detections so that the evaluation index will drop rapidly. The target of interest in this paper is buildings in urban areas, so the threshold $T_H = 5$ is a good choice, corresponding to the change of buildings with at least two floors.



Figure 14. Influence of T_H on co-segmentation accuracy. (**a**,**c**,**e**,**g**,**i**,**k**) Influence of T_H on TDR and FNR for Test areas 1–6; (**b**,**d**,**f**,**h**,**j**,**l**) Influence of T_H on precision, recall and *F*-score for Test areas 1–6.

5. Conclusions

Height information is essential for urban area monitoring, especially building change detection. The development of very high-resolution stereo satellite imagery and dense stereo matching technology provides convenience for building change monitoring in largescale urban areas. A novel change detection algorithm based on 3D co-segmentation is proposed in this paper to combine the spectral change information, the height change information, and the spatial context information. The proposed algorithm can make full use of the 3D change information contained in the stereoscope data. Using the spectral change and height change information as the prior, we construct graphs for two-temporal images to obtain the spatial corresponded two-temporal segmentation results. The object-to-object detection results are obtained through overlay analysis, and the quantitative height change values are calculated according to this correspondence. The algorithm is unsupervised and fully automatic. The performance of the algorithm is evaluated in detail through experiments with several groups of real satellite data. Our proposed algorithm achieves satisfactory results for pixel-based, object-based, and numerical-based evaluation metrics when compared to the state-of-the-art algorithms. Problems caused by DSM errors are successfully eliminated with the incorporation of the MBI change feature. The experimental results are consistent with the reference map in terms of the detected building's position, number, and height change accuracy.

For future work, we will consider the following aspects: (1) The MBI extracted in this paper is affected by other surface changes, especially the shadow changes, and contains

a lot of noise. We will introduce the multi-spectral image attached to the data to extract better image-based change features. (2) The integration of spectral change features and height change features is simple weighting, making it difficult to adjust the parameters. More novel feature fusion methods will be used to give full play to their advantages. (3) We will consider other types of changes and obtain the detection of multiple types of changes simultaneously through one processing.

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

Funding: This research was funded by the LuTan-1 L-Band Spaceborne Bistatic SAR Data Processing Program, grant number E0H2080702.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: All authors would like to thank the anonymous reviewers whose insightful suggestions have improved the paper significantly.

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

References

- Hirschmuller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* 2007, 30, 328–341. [CrossRef] [PubMed]
- d'Angelo, P.; Lehner, M.; Krauss, T.; Hoja, D.; Reinartz, P. Towards automated DEM generation from high resolution stereo satellite images. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. Int. Soc. Photogramm. Remote Sens. 2008, 37, 1137–1342.
- d'Angelo, P.; Reinartz, P. Semiglobal matching results on the ISPRS stereo matching benchmark. In *High-Resolution Earth Imaging for Geospatial Information*; Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover: Hanover, Germany, 2011.
- Qin, R.; Tian, J.; Reinartz, P. 3D change detection–approaches and applications. *ISPRS J. Photogramm. Remote Sens.* 2016, 122, 41–56. [CrossRef]
- Xiao, P.; Yuan, M.; Zhang, X.; Feng, X.; Guo, Y. Cosegmentation for object-based building change detection from high-resolution remotely sensed images. *IEEE Trans. Geosci. Remote Sens.* 2017, 55, 1587–1603. [CrossRef]
- Zhang, X.; Xiao, P.; Feng, X.; Yuan, M. Separate segmentation of multi-temporal high-resolution remote sensing images for object-based change detection in urban area. *Remote Sens. Environ.* 2017, 201, 243–255. [CrossRef]
- Chen, J.; Liu, H.; Hou, J.; Yang, M.; Deng, M. Improving Building Change Detection in VHR Remote Sensing Imagery by Combining Coarse Location and Co-Segmentation. *ISPRS Int. J. Geo-Inf.* 2018, 7, 213. [CrossRef]
- Gong, J.; Hu, X.; Pang, S.; Li, K. Patch Matching and Dense CRF-Based Co-Refinement for Building Change Detection from Bi-Temporal Aerial Images. Sensors 2019, 19, 1557. [CrossRef]
- 9. Gong, J.; Hu, X.; Pang, S.; Wei, Y. Roof-Cut Guided Localization for Building Change Detection from Imagery and Footprint Map. *Photogramm. Eng. Remote Sens.* **2019**, *85*, 543–558. [CrossRef]
- Zhu, L.; Zhang, J.; Sun, Y. Remote Sensing Image Change Detection Using Superpixel Cosegmentation. *Information* 2021, 12, 94. [CrossRef]
- 11. Hao, M.; Zhou, M.; Cai, L. An improved graph-cut-based unsupervised change detection method for multispectral remote-sensing images. *Int. J. Remote Sens.* 2021, 42, 4005–4022. [CrossRef]
- 12. Zhang, K.; Fu, X.; Lv, X.; Yuan, J. Unsupervised Multitemporal Building Change Detection Framework Based on Cosegmentation Using Time-Series SAR. *Remote Sens.* **2021**, *13*, 471. [CrossRef]
- Gstaiger, V.; Tian, J.; Kiefl, R.; Kurz, F. 2d vs. 3d change detection using aerial imagery to support crisis management of large-scale events. *Remote Sens.* 2018, 10, 2054. [CrossRef]
- Tian, J.; Chaabouni-Chouayakh, H.; Reinartz, P. 3D building change detection from high resolution spaceborne stereo imagery. In Proceedings of the 2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping, Xiamen, China, 10–12 January 2011; pp. 1–7.. [CrossRef]

- Tian, J.; Reinartz, P. Multitemporal 3D change detection in urban areas using stereo information from different sensors. In Proceedings of the 2011 International Symposium on Image and Data Fusion, Tengchong, China, 9–11 August 2011; pp. 1–4. [CrossRef]
- Tian, J.; Cui, S.; Reinartz, P. Building change detection based on satellite stereo imagery and digital surface models. *IEEE Trans. Geosci. Remote Sens.* 2013, 52, 406–417. [CrossRef]
- 17. Tian, J.; Reinartz, P. Dempster-Shafer fusion based building change detection from satellite stereo imagery. In Proceedings of the 17th International Conference on Information Fusion (FUSION), Salamanca, Spain, 7–10 July 2014; pp. 1–7.
- Tian, J.; Reinartz, P.; Dezert, J. Building change detection in satellite stereo imagery based on belief functions. In Proceedings of the 2015 Joint Urban Remote Sensing Event (JURSE), Lausanne, Switzerland, 30 March–1 April 2015; pp. 1–4. [CrossRef]
- Tian, J.; Dezert, J.; Reinartz, P. Refined building change detection in satellite stereo imagery based on belief functions and reliabilities. In Proceedings of the 2015 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), San Diego, CA, USA, 14–16 September 2015; pp. 160–165. [CrossRef]
- Tian, J.; Dezert, J.; Qin, R. Time-Series 3D Building Change Detection Based on Belief Functions. In Proceedings of the 2018 21st International Conference on Information Fusion (FUSION), Cambridge, UK, 10–13 July 2018; pp. 1–5. [CrossRef]
- Tian, J.; Dezert, J. Fusion of multispectral imagery and DSMs for building change detection using belief functions and reliabilities. Int. J. Image Data Fusion 2019, 10, 1–27. [CrossRef]
- Tian, J.; Reinartz, P.; d'Angelo, P.; Ehlers, M. Region-based automatic building and forest change detection on Cartosat-1 stereo imagery. *ISPRS J. Photogramm. Remote Sens.* 2013, 79, 226–239. [CrossRef]
- 23. Qin, R. An Object-Based Hierarchical Method for Change Detection Using Unmanned Aerial Vehicle Images. *Remote Sens.* 2014, 6, 7911–7932. [CrossRef]
- 24. Qin, R. Change detection on LOD 2 building models with very high resolution spaceborne stereo imagery. *ISPRS J. Photogramm. Remote Sens.* **2014**, *96*, 179–192. [CrossRef]
- Tian, J.; Gharibbafghi, Z.; Reinartz, P. Superpixel-Based 3D Building Model Refinement and Change Detection, Using VHR Stereo Satellite Imagery. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–17 June 2019; pp. 493–495. [CrossRef]
- Wen, D.; Huang, X.; Zhang, A.; Ke, X. Monitoring 3D Building Change and Urban Redevelopment Patterns in Inner City Areas of Chinese Megacities Using Multi-View Satellite Imagery. *Remote Sens.* 2019, 11, 763. [CrossRef]
- Huang, X.; Cao, Y.; Li, J. An Automatic Change Detection Method for Monitoring Newly Constructed Building Areas Using Time-Series Multi-View High-Resolution Optical Satellite Images. *Remote Sens. Environ.* 2020, 244, 111802. [CrossRef]
- 28. Qin, R.; Gruen, A. A supervised method for object-based 3d building change detection on aerial stereo images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 2014, 40, 259. [CrossRef]
- 29. Chen, B.; Chen, Z.; Deng, L.; Duan, Y.; Zhou, J. Building change detection with RGB-D map generated from UAV images. *Neurocomputing* **2016**, *208*, 350–364. [CrossRef]
- 30. Tian, J.; Qinb, R.; Cerra, D.; Reinartz, P. Building Change Detection in Very High Resolution Satellite Stereo Image Time Series. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* 2016 **2016**, *3*, 149–155. [CrossRef]
- Qin, R.; Tian, J.; Reinartz, P. Spatiotemporal inferences for use in building detection using series of very-high-resolution space-borne stereo images. *Int. J. Remote Sens.* 2016, 37, 3455–3476. [CrossRef]
- 32. Yuan, X.; Tian, J.; Reinartz, P. Building Change Detection Based on Deep Learning and Belief Function. In Proceedings of the Joint Urban Remote Sensing Event 2019, Vannes, France, 22–24 May 2019.
- 33. Pang, S.; Hu, X.; Cai, Z.; Gong, J.; Zhang, M. Building change detection from bi-temporal dense-matching point clouds and aerial images. *Sensors* **2018**, *18*, 966. [CrossRef]
- 34. Pang, S.; Hu, X.; Zhang, M.; Cai, Z.; Liu, F. Co-Segmentation and Superpixel-Based Graph Cuts for Building Change Detection from Bi-Temporal Digital Surface Models and Aerial Images. *Remote Sens.* **2019**, *11*, 729. [CrossRef]
- Nebiker, S.; Lack, N.; Deuber, M. Building Change Detection from Historical Aerial Photographs Using Dense Image Matching and Object-Based Image Analysis. *Remote Sens.* 2014, 6, 8310–8336. [CrossRef]
- Krauss, T.; d'Angelo, P.; Kuschk, G.; Tian, J.; Partovi, T. 3D-information fusion from very high resolution satellite sensors. ISPRS-Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2015, XL-7/W3, 651–656. [CrossRef]
- Mohammadi, H.; Samadzadegan, F. An Object Based Framework for Building Change Analysis Using 2D and 3D Information of High Resolution Satellite Images. *Adv. Space Res.* 2020, *66*, 1386–1404. [CrossRef]
- Tabib Mahmoudi, F.; Hosseini, S. Three-Dimensional Building Change Detection Using Object-Based Image Analysis (Case Study: Tehran). *Appl. Geomat.* 2021, 13, 325–332. [CrossRef]
- Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* 2001, 23, 1222–1239. [CrossRef]
- 40. Huang, X.; Zhang, L. A Multidirectional and Multiscale Morphological Index for Automatic Building Extraction from Multispectral GeoEye-1 Imagery. *Photogramm. Eng. Remote Sens.* **2011**, 77, 721–732. [CrossRef]
- 41. Huang, X.; Zhang, L. Morphological Building/Shadow Index for Building Extraction From High-Resolution Imagery Over Urban Areas. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 161–172. [CrossRef]
- Fraser, C.S.; Hanley, H.B. Bias Compensation in Rational Functions for Ikonos Satellite Imagery. *Photogramm. Eng. Remote Sens.* 2003, 69, 53–57. [CrossRef]

- 43. Fraser, C.S.; Hanley, H.B. Bias-Compensated RPCs for Sensor Orientation of High-Resolution Satellite Imagery. *Photogramm. Eng. Remote Sens.* 2005, *71*, 909–915. [CrossRef]
- Grodecki, J.; Dial, G. Block Adjustment of High-Resolution Satellite Images Described by Rational Polynomials. *Photogramm.* Eng. Remote Sens. 2003, 69, 59–68. [CrossRef]
- Arefi, H.; Hahn, M. A morphological reconstruction algorithm for separating off-terrain points from terrain points in laser scanning data. In Proceedings of the ISPRS WG III/3, III/4, V/3 Workshop "Laser scanning 2005", Enschede, The Netherlands, 12–14 September 2005.
- 46. Boykov, Y.; Funka-Lea, G. Graph Cuts and Efficient N-D Image Segmentation. Int. J. Comput. Vis. 2006, 70, 109–131. [CrossRef]
- 47. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. Syst. Man, Cybern. 1979, 9, 62–66. [CrossRef]
- 48. Boykov, Y.; Kolmogorov, V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1124–1137. [CrossRef]