Article

Improving Building Change Detection in VHR Remote Sensing Imagery by Combining Coarse Location and Co-Segmentation

Jie Chen [1,*], Haifei Liu [1,2], Jialiang Hou [2], Minhua Yang [1] and Min Deng [1]

[1]   School of Geosciences and Info-Physics, Central South University, Changsha 410083, China; haifeiliu@csu.edu.cn (H.L.); yangmh@csu.edu.cn (M.Y.); dengmin@csu.edu.cn (M.D.)
[2]   The Third Surveying and Mapping Institute of Hunan, Changsha 410004, China; qzhb213@csu.edu.cn
[*]   Correspondence: cj2011@csu.edu.cn; Tel.: +86-13607487755

check for
updates

Abstract: Building change detection based on remote sensing imagery is a significant task for urban construction, management, and planning. Feature differences caused by changes are fundamental in building change detection, but the spectral and spatial disturbances of adjacent geo-objects that can extensively affect the results are not considered. Moreover, the diversity of building features often renders change detection difficult to implement accurately. In this study, an effective approach is proposed for the detection of individual changed buildings. The detection process mainly consists of two phases: (1) locating the local changed area with the differencing method and (2) detecting changed buildings by using a fuzzy clustering-guided co-segmentation algorithm. This framework is broadly applicable for detecting changed buildings with accurate edges even if their colors and shapes differ to some extent. The results of the comparative experiment show that the strategy proposed in this study can improve building change detection.

Keywords: change detection; differencing method; coarse location; co-segmentation; fuzzy clustering

1. Introduction

Land use or cover change is a direct reflection of human activity, such as land use, urban expansion, and architectural planning, on the earth's surface caused by urbanization [1]. Remote sensing images are important data sources that can efficiently detect land changes. Meanwhile, remote sensing image-based change detection is the change identification of surficial objects or geographic phenomena through the remote observation of two or more different phases [2]. Change detection emphasizes the change in spectral response to pixels of remote sensing images at different moments, which may denote the increase, disappearance, movement, and transformation of geographic features [3]. Satellite-based multispectral images with very-high spatial resolution (VHR) have recently become available, thus making change detection a highly relevant topic in this field of research [4]. Buildings are important products of urbanization during human civilization. Thus, obtaining building change information is of great importance in scientifically promoting urban planning, and it can help realize the sustainable development of the economy and the environment [5].

The method of change detection based on remote sensing images is usually divided two types: pixel-based methods and object-based methods [1]. The former is widely used with medium or low spatial resolutions, but it is not suitable for VHR remote sensing images [6]. VHR remote sensing images can supply abundant geo-objects and details, which increase the internal spectral variability (intra-class variability) of each land cover class and decrease the spectral variability between different classes (inter-class variability) [7]. However, compared with traditional pixel-based change detection

methods, the increased variabilities of VHR remote sensing images often result in too many changes being detected. Moreover, modeling the contextual information from a pixel-based change detection method is difficult to conduct because spatial relationships and arrangements of real-world objects are often neglected [8]. In addition, change detection based on VHR remote sensing images is sensitive to geometric registration accuracy and radiation correction accuracy [1].

To meet the challenges of change detection based on VHR remote sensing images, the techniques from object-based image analysis (OBIA), which have drawn great attention in recent years, have gradually been proposed as the mainstream solution in this research field. OBIA treats adjacent homogenous pixels as the unit of analysis for change detection, and the unit called "object" is usually gained through segmentation. Given OBIA's rich auxiliary information, including texture, shape, and spatial relations, the unit may be regarded more relevant in understanding scenes [9]. Moreover, by extracting and using the information, correctly distinguishing the complex objects in real world from VHR remote sensing images is therefore feasible. In general, change detection accuracy can be improved by integrating multi-features [10]. Many methods to detect changed buildings can be devised based on objects and by implementing combined methods on pixel-based detection and object-based recognition [5,8,11–16]. With the development of new strategies, the results of building change detection can achieve satisfactory improvements. Some typical problems in this field have attracted attention, and the extant solutions have contributed to relatively better detection results. For example, both [13] and [14] can lessen false detection caused by observing angles to some extent.

However, object-based building change detection from VHR remote sensing images are constrained by certain limitations that are mainly associated with mixed or diverse objects. The first problem on mixed objects implies that some spatial adjacent objects are easily misidentified as mixed changes. If the feature differences in two or more images at different phases are similar, then the magnitude of pixels of changed buildings and other surrounding changed objects (i.e., shadows, roads, and bare land) all appear large as a result of the differencing method, which then lead to the emergence of mixed objects. Some existing differencing methods initially identify the change information by its entirety, after which the changed building objects are identified [5]. However, due to the presence of mixed objects, the false-alarm problem becomes serious. The mixed changed objects in the final change detection result always render the shapes and edges of changed buildings misaligned with reality.

The second problem on diverse objects indicates that the feature differences (e.g., shapes, colors, sizes, and orientations) of urban buildings are neither easy to model nor integrate into a single framework. Usually, a tradeoff manifests between the macro-clustering of objects and the generation of micro details (e.g., boundary extraction). In this respect, some well-known indexes have been considered for building change detection by using VHR remote sensing images. For example, the detection methods of [5,13–15] likely used the morphological building index (MBI). However, buildings of diverse types and the impact of low-contrast ground environment are both often neglected. The change detection of urban buildings with different colors, textures, and structures is a common limitation, and this scenario makes it difficult for different methods, even the object-based ones, to obtain a satisfactory result. Therefore, developing a new change detection method is essential when identifying changed buildings of different types within areas with exact boundaries.
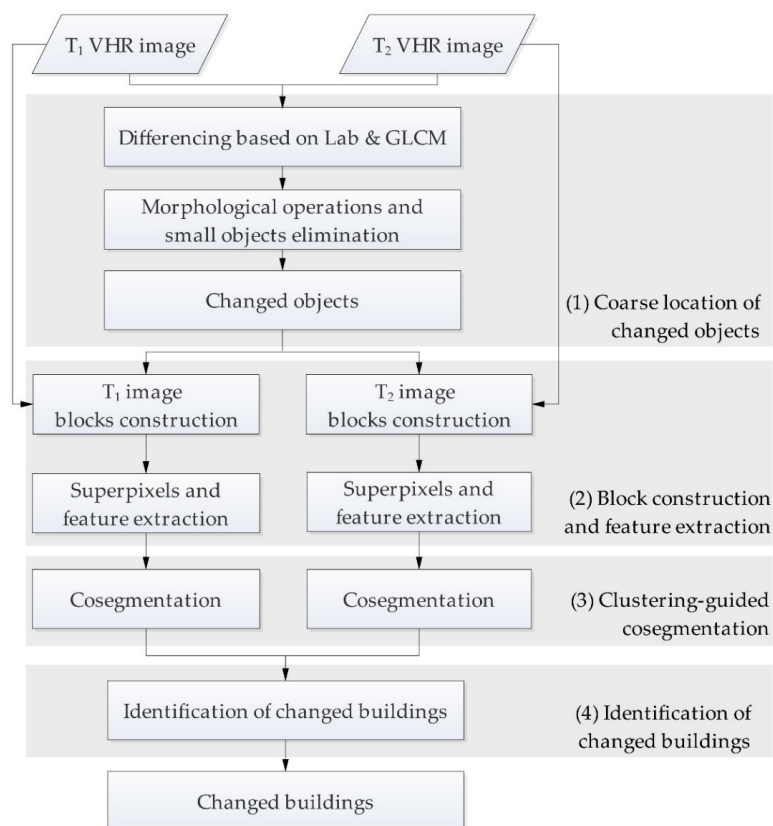
The development of a city is gradual and the constant changing of buildings is usually regarded a small event in a region, but this scenario renders the larger area to be considered unchanged for a certain period, i.e., a small number of changed building pixels and a higher number of changed non-building pixels may both exist in VHR scenes. The unchanged class may also exhibit dominantly more samples than the changed class in numerous applications, and the distribution of classes, an unknown priori, may be heavily imbalanced when unsupervised techniques are utilized [17]. Moreover, when adopting a learning approach, the building change detection will likely confront a small sample problem. From the arithmetic viewpoint, on the one hand, searching for changed buildings in each corner of the two-phase images is unnecessary; on the other hand, extracting the scattered changed buildings over a wide area is difficult if the changed non-building objects and unchanged objects coexist.

In this study, by viewing the change from a macro-point, the changed pixels are initially extracted to determine the coarse locations of potentially changed buildings by using Lab and grey level co-occurrence matrix (GLCM) for the differencing method. Then, combined with the coarse locations, this study improves building change detection by using collaborative segmentation (co-segmentation) to gain an accurate identification of various changed buildings. Co-segmentation assumes that two or more images have the same or similar foreground objects, after which they are segmented [18–20]. In this study, co-segmentation is used to capture urban building information across blocks according to their similar features. The improvements can contribute to building change detection in the following aspects: For buildings with large intra-class spectral differences, the differencing method proposed in this study can effectively capture the location of changed buildings. In addition, the multi-constraint for co-segmentation provided by the weak supervised approach and the fuzzy classification can exclude most non-building objects and accurately extract the shape and boundary information of the changed buildings.

This article is arranged as follows: The second section introduces the main steps of the proposed method. The third section presents the experimental data, results analyses, and results of the method comparison. The last section draws the conclusion and offers prospect for future work.

2. Methodology

The proposed method mainly includes three parts: coarse location of changed objects, block sequence construction, and fine extraction of changed buildings. The overall process is shown in Figure 1. First, the location of the change is approximated based on the results of the differencing method. Second, local blocks with the changed objects at their centers are constructed. Then, for the blocks at each phase, each of them is divided into superpixels, and the spectral and texture features of each superpixel are extracted. On this basis, a fuzzy clustering-guided co-segmentation is implemented. Finally, the changed building is obtained by combining the building objects between local blocks from two phases.

Figure 1. Flowchart of the proposed change detection framework.

2.1. Coarse Location of Changed Objects

The first step of the proposed method is to approximate the position of the change. To this end, improvements are applied to a typical differencing method to remove as much changed pixels of non-buildings as possible while retaining most of the changed pixels of buildings. Then, some morphological operations are applied to the changed pixels, and these pixels are aggregated into objects. Finally, small objects are removed in succession.

2.1.1. Differencing Based on Lab and GLCM

To approximate the positions of potentially changed buildings, this study initially determines the changed objects by improving the change detection method proposed in [5], in which the pixel-based spectral and the textural differences on each band of multi-temporal high-spatial resolution images are calculated. The spectral feature is represented by the pixel brightness of each band, while the texture information is given by GLCM [21]. Then, the variance measure is validated to show the largest difference in the mean values (i.e., selected from among several measures) between built-up land and cropland [5]. Thus, the GLCM variance measure is selected as the textural feature for change detection.

$$\text{GLCM } \textit{Variance} = \sum_{s=1}^{N} \sum_{t=1}^{N} P_{\theta,d}(s,t)(s-\mu)^2 \tag{1}$$

where $P_{\theta,d}(s,t)$ represents the probability of pairwise gray levels (s, t) for all adjacent pixels in the selected direction θ and distance d; N is the total number of gray levels; and μ is the GLCM mean. The differential matrix is calculated as follows:

$$f_i(x,y) = \left| f_i^2(x,y) - f_i^1(x,y) \right|, \tag{2}$$

where i is the number of spectral or texture band; $f_i^1(x,y)$ and $f_i^2(x,y)$ are the spectral or texture feature values at coordinate (x, y) in two phases; and $f_i(x, y)$ is a value at (x, y) in the differential matrix. For each differential matrix, the changed pixels are identified as follows:

$$f_i'(x,y) = \begin{cases} 1, & if(f_i(x,y) \geq \mu_i + T \cdot \sigma_i) \\ 0, & otherwise \end{cases}, \tag{3}$$

where μ_i and σ_i are the mean and standard deviation of all pixels in the ith differential matrix respectively; T is the threshold parameter of the manual setting, in which T_S and T_T are for the spectral and textural layers, respectively; and $f_i'(x,y)$ is the change label in the ith differential matrix (1 for changed pixels and 0 for unchanged pixels).
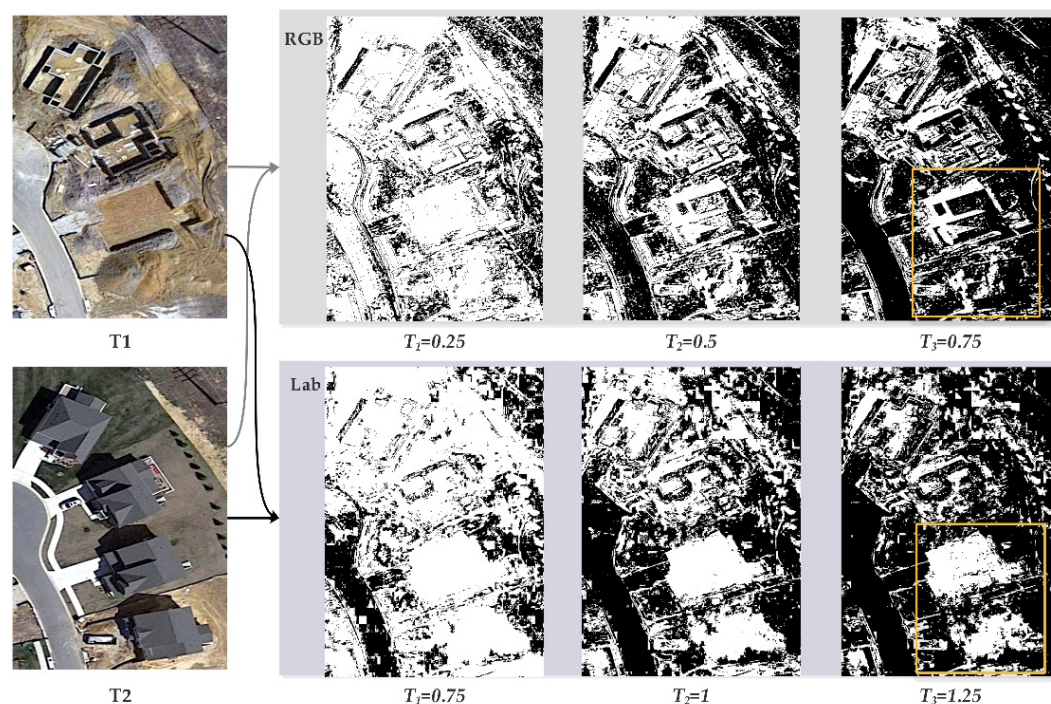
As shown by Equations (2) and (3), knowing whether a pixel has changed or not can be directly determined by comparing the differential values with a threshold, i.e., the comparison of the two differential values between changed pixels and unchanged pixels definitively affects the result. If the differential value of the changed building pixels is bigger than that of the other objects when T is increased, then the changed pixels of the other objects will disappear and the changed building pixels will remain. By contrast, if the differential value of the changed building pixels is smaller than that of other objects, then threshold T should be low to retain more changed building pixels; consequently, some unchanged building pixels will emerge in the change map. If the changed buildings are thickly surrounded by pixels, the ultimately changed object cannot fully describe the reality of individually changed buildings.

For the purpose of determining the coarse location in this study, the presence of many other features (i.e., bare land and roads) is not recommended to be near the changed building pixels. If the inputted feature of the differential matrix can highlight the changed building pixels, then it is more possible for the differencing image to highlight the locations of the potentially changed buildings.

We noticed that the contrast between the differential values of changed building pixels and that of other changed feature pixels is inconsistent in different color spaces. Lab, one of several color models created by the Commission Internationale d'Eclairage (CIE) [22], was developed to simulate the human perception of color. Lab is composed of three components, namely, lightness L and color-related a and b. Recky et al. [23] argued that building surfaces are usually composed of large areas of uniform colors or a small number of different colors. However, because lighting often changes (usually caused by shadows) in an image, the same color may appear with different brightness levels. They also found that the standard color descriptors (i.e., RGB or HSV histograms) cannot sufficiently and effectively distinguish different colors. By contrast, in Lab space, the Euclidean distance between two different colors is directly proportional to the visual similarity of colors.

Figure 2 shows the contrast in RGB and Lab color spaces. When *T* is increased in the RGB color space, the building pixels and other pixels within the orange rectangle box disappear. However, in Lab color space, the pixels of building are preserved while those of other objects disappear. A possible reason may be that the color space of Lab is much wider than RGB, and Lab contains not only all the color gamut of RGB but also the colors perceived by human eyes [22]. Therefore, to enhance the ability for identifying different colors, this study calculates the differencing image by using the Lab color space.

Figure 2. Changed pixels in RGB and Lab color spaces obtained with different *T* values. Although the increment of *T* is similarly set to 0.25, the contrasting images exhibit different degrees in the two-color spaces.

2.1.2. Morphological Operation and Small Object Elimination

To derive accurate locations, a union operation is initially conducted on all the differential matrices generated from the spectral and texture bands. Then, morphological operations (close and open operations and hole filling) are sequentially performed on the changed pixels, after which the pixels are grouped into objects. Considering that the changed buildings may represent new construction, expansion, and demolition, the small objects called "non-buildings" are eliminated by using threshold T_A.

2.2. Block Construction and Feature Extraction

Although the improved differencing method can remove many changed pixels of non-buildings, a few inevitable changed objects of non-buildings can still be observed. In other words, the changed objects are not pure buildings but a mixture of different objects (i.e., roads, bare ground, and other buildings). To derive more accurate information of the changed buildings, an additional extraction process is needed to improve the detection results at the pixel level. On the basis of these considerations, utilizing the attributes of geo-objects is necessary to distinguish the changed buildings and refine their boundaries. Therefore, from the initially obtained changed objects, blocks are constructed for each image in two phases. Then, the features of the superpixels in all blocks are extracted. This process is an important prerequisite of co-segmentation.

2.2.1. Block Construction

Image blocks are constructed based on the coarse locations in two phases. All blocks at each phase are arranged as the sequence Seq_t ($t = 1, 2$). The steps of block sequence Seq_t construction (Figure 3) are as follows:
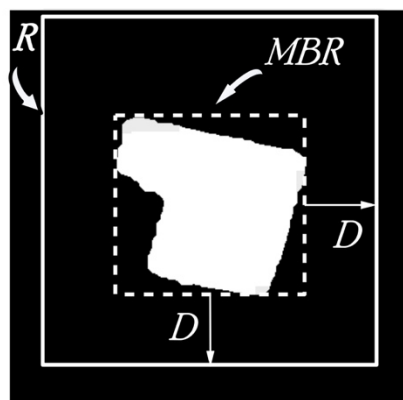
Step 1: Designate a number to all changed objects and build the minimum bounding rectangle (MBR) for each changed object by bordering their contours.

Step 2: Move four sides of the MBR in distance D to obtain an extended rectangular R.

Step 3: Record the row and column indexes of the four vertices of R in the entire image.

Step 4: Fetch the corresponding parts of R as a block from the tth original image and record the location information.

Step 5: Sequence all blocks into $Seq_t = \left\{ I_1^t, \ldots, I_n^t \right\} (t = 1, 2)$.

Figure 3. Schematic of block construction.

2.2.2. Superpixels and Feature Extraction

To precisely achieve the building boundary, we use simple linear iterative clustering (SLIC) [24] to segment each block in Seq_1 and Seq_2 and obtain superpixels with high homogeneity. The superpixel usually represents a single feature, such as a road, a meadow, or a local homogeneous patch of complex feature (i.e., a modern building). The operators in Table 1 are used to extract the color and texture features from each superpixel. The color features represent the mean of RGB, Lab, and HSV, and the texture features are calculated by using the local binary pattern (LBP) [25] and the Lee–Malik (LM) filter [26]. The color and texture features are combined to form a feature vector for each superpixel.

Table 1. Color and texture description operators.

Symbol	Feature Description Operator	Dimension
a1	RGB mean	3
a2	Lab mean	3
a3	HSV mean	3
r1	Absolute response of LM filter	15
r2	Maximum response of LM filter	15
h1	RGB histogram	256
h2	Lab histogram	256
h3	HSV histogram	256
h4	LBP histogram	256

Given that a certain correlation exists in the feature components, reducing the dimension of the feature is necessary to remove redundant information. Principal component analysis (PCA) [27] can be used to remove the linear correlation of attributes, but it does not work well for nonlinear correlation data. The nonlinear function can transform the original features into high-dimensional linear separable feature bases [28]. Therefore, we reduce the dimension of the features by using kernel PCA (KPCA) based on the Gaussian radial basis function [29]. In addition, given that the characteristic value λ_i generated by KPCA contains the reconstructed feature information, the characteristic values are recorded as $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ (i.e., arranged from the largest to the smallest). To select the optimal k-former feature, the cumulative information Inf_k is estimated as follows:

$$Inf_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_n} \geq T_{\text{inf}}, \tag{4}$$

where T_{inf} is a threshold and k equals the minimum value when Inf_k is greater than T_{inf}.

2.3. Clustering-Guided Co-Segmentation

Various types of urban buildings, such as residential, commercial, and factory buildings, normally have common features in terms of structure, color, texture, and contrast. A weak supervised methodology was used for the global constraint of the co-segmentation model to effectively deal with the change characteristics of different targets [30–32]. In the present study, a weak supervised co-segmentation method is also used to extract individual buildings (foreground) from blocks and capture the common features of buildings. Generating priori probabilities of the foreground or background is beneficial to the co-segmentation model, in which priori information can be inputted as parameters [32]. Aimed at effectively segmenting similar foreground objects from complicated background, the proposed method initially collects foreground or background superpixel samples via a scribbling process. In the proposed method, some of the blocks from Seq_1 and Seq_2 selected for the scribble are called I_{label}, and the foreground and background pixels are scribbled on I_{label} in advance. Then, the block is segmented to obtain superpixels. If a superpixel contains scribbled pixels, then it is considered a foreground/background sample, as shown by the samples of scribbled superpixels in Figure 4. The collected background or foreground superpixels are added to $U = \{U_0, U_1\}$, where U_0 and U_1 represent the background and foreground superpixel samples, respectively. However, although the priori information derived from foreground and background samples is essential for co-segmentation, too few or too many scribbled samples is not advisable.

The foreground probability in the model used in [32] was estimated based on the RGB feature of the superpixel by using the maximum likelihood. However, if this estimation is used, then the non-buildings whose colors are similar to those of the buildings will likely obtain high foreground probability. To enhance the ability of distinguishing buildings from non-buildings, the foreground

probability used for the co-segmentation model is improved in the proposed method via a fuzzy clustering process (Figure 4). First, to determine the optimal clustering number K for each Seq_t, the superpixels in Seq_t are pre-clustered by using K-means, and the number at the elbow position of the error curve is designated as the optimal number (i.e., see next paragraph for the discussion). Second, fuzzy clustering means (FCM) [33] is performed to determine the probability of each superpixel to be K subclasses. Finally, each superpixel in the block sequence is assigned the probability vector $P = \{P^1, \ldots, P^i, \ldots, P^K\}$, where P^i represents the probability of the superpixel to belong to subclass i.

The optimal K is determined by performing a pre-clustering on the superpixels in Seq_t and by calculating the cost function J for different cluster numbers of K. The cost function J is the sum of the distortions of the individual classes.

$$J = \sum_{j=1}^{k} \sum_{i \in C_j} |x_i - \mu_j|^2, \tag{5}$$

where x_i represents the feature vector of sample i; C_j and μ_j are the sample sets of category j and category *center*, respectively. If the members within the class are more compact, then the distortion degree J of the class becomes smaller. By contrast, if the members within the class are more dispersed, then the distortion degree J of the class becomes greater. At this point of the process, the number of clusters can be estimated by the elbow rule [34].
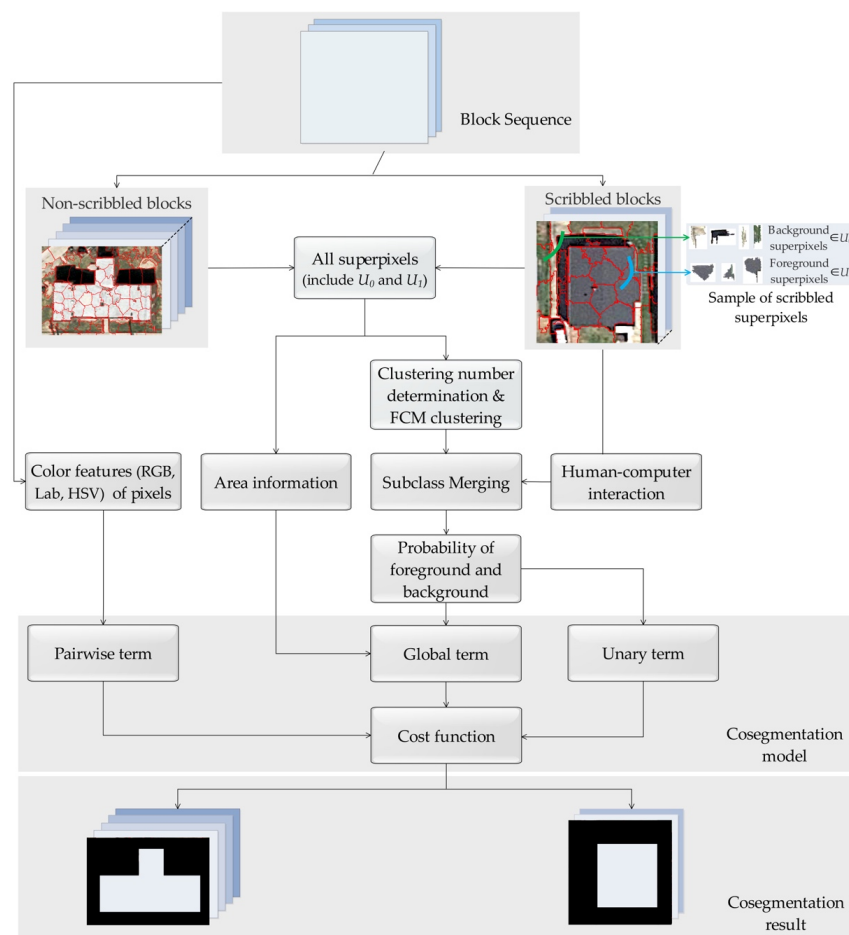

Figure 4. Clustering-guided co-segmentation.

As described above, the probability vector P of all superpixels is obtained by FCM with optimal K. Thus, through human–computer interaction implemented on I_{label}, the K subclasses can be divided into foreground set (buildings) $C_1 = \{C_1^1, \ldots, C_1^m\}$ and the background set (non-buildings) $C_0 = \{C_0^1, \ldots, C_0^n\}$, where m and n are the numbers of subclasses (i.e., $m + n = K$). The subscript 1 represents a subclass belonging to the foreground while 0 represents that belonging to the background. On the basis of human–computer interaction, the rule of each subclass (i.e., either foreground or background) is established. Consequently, a probability vector to denote the foreground or the background can be obtained by $P = \{P_1^1, \ldots, P_1^m, P_0^1, \ldots, P_0^n\}$ for each superpixel. In terms of the sth superpixel, its priori probabilities $P_{s,1}$ and $P_{s,0}$ for the foreground and background can be achieved by separately combining the probabilities of relevant subclasses, such that $P_{s,1} = P_{s,1}^1 + \cdots + P_{s,1}^m$, $P_{s,0} = P_{s,0}^1 + \cdots + P_{s,0}^n$, and $P_{s,1} + P_{s,0} = 1$.

With regard to the guidance of the fuzzy clustering, on the one hand, the features used for clustering has a stronger ability than the color feature but only to distinguish buildings from non-buildings. On the other hand, by subclass merging, different types of buildings have high chances to obtain high foreground probabilities. Consequently, the superpixel probabilities for the foreground and background can be inputted to initiate the energy function of the co-segmentation as follows:

$$E = \sum_{i=1}^{n} \left(\varepsilon_u^i E_u^i + \varepsilon_p^i E_p^i \right) + E_g(I_1, \ldots, I_n, I_{label}), \tag{6}$$

where E_u^i and E_p^i represent the unary term and the pairwise term of the ith block I_i, respectively. The ε_u^i and ε_p^i are the corresponding weights of these terms, such that $\varepsilon_u^i + \varepsilon_p^i = 1$.

The unary term E_u and the pairwise term E_p are defined as follows:

$$E_u^i = \sum_k - \log(p_{k,1}^i) \cdot \phi(x_k^i) - \log(p_{k,0}^i) \cdot (1 - \phi(x_k^i)), \tag{7}$$

where x_k^i represents the kth pixel of block I_i; $\varnothing(x_k^i)$ is a binary function for assigning x_k^i to background or foreground; $p_{k,1}^i$ represents foreground probability, which is an estimated value of x_k^i classified as a building; and $p_{k,0}^i$ represents the background probability. The foreground and background probabilities $p_{k,1}^i$ and $p_{k,0}^i$ of x_k^i equal the building and non-building probabilities $P_{s,1}^i$ and $P_{s,0}^i$ of the sth superpixel, of which x_k^i is also a component.

$$E_p^i = \sum_{k,k' \in N} \| v_k^i - v_{k'}^i \| \cdot \left| \phi(x_k^i) - \phi(x_{k'}^i) \right|, \tag{8}$$

where N is the set that contains all four adjacent pixel pairs in a block; k and k' denote two adjacent pixels in N; and v_k^i is the color feature vector (RGB, Lab, and HSV) of x_k^i.

From Equation (7), we can infer that the larger foreground probability is, the more likely that a pixel will be assigned as a foreground. This phenomenon is mainly due to the constraint of the unary term E_u. On the one hand, E_p can be used to enhance the smoothness of the intra-image label by constraining the label consistency. If two adjacent pixels have similar features, then the model will be more inclined to be assigned with the same label. On the other hand, E_p can be used to refine candidate boundary pixels and obtain a more accurate depiction of the objects' boundaries.

The global term $E_g(I_1, \ldots, I_n, I_{label})$ is a high-order energy function that contains a high-order cluster function, and it is defined as follows:

$$\begin{cases} E_g(I_1, \cdots, I_1, I_{label}) = \sum\limits_{i=1}^{n} E_{high}(R_i, U) \\ \sum\limits_{i=1}^{n} E_{high}(R_i, U) = \sum\limits_{s=1}^{N(R_i)} N(v_s^i) \cdot k_s^i \end{cases}, \tag{9}$$

where $\{I_1, \dots, I_n\}$ are blocks and I_{label} is the scribbled block; U is the set of scribbled foreground and background superpixels; R_i represents the superpixels segmented from the ith block; n is the amount of block R; $E_{high}(R_i, U)$ is a higher-order term; $N(R_i)$ is the number of superpixels in block I_i; k_s^i is the matching coefficient of the energy function that takes into account its probability to belong to the foreground or the background (the details are supplied in [32]); and $N(v_s^i)$ is the number of pixels in the high-order cluster v_s^i that comprise three parts (foreground U_1, background U_0, and superpixel sets R_i). The high-level clusters are defined as follows:

$$v_s^i = \left\{ r_s^i, u_1(r_s^i), u_0(r_s^i) \right\}, \tag{10}$$

where $r_s^i \in R_i$ represents the sth superpixel in the ith block I_i; $u_1(r_s^i) \in U_1$ represents the foreground superpixels nearest to superpixel r_s^i in feature space; and $u_0(r_s^i) \in U_0$ represents the background superpixels nearest to superpixel r_s^i in feature space.

Consequently, the high-order energy term can take advantage of the priori knowledge derived from the scribbled superpixels. The high-order energy term can also effectively extract buildings, including those with great visual diversity, by constraining foreground feature distribution. Intuitively, the more similar the features of the superpixels with the foreground (labeled by scribbling) are, the smaller the global item E_g will be and vice versa. In addition, considering that the unary term and the pairwise term can be effectively calculated by the graph cut, the higher-order energy can also be implemented optimally by using this traditional method.

2.4. Identification of Changed Buildings

By using the improved co-segmentation model on Seq_1 and Seq_2, each superpixel in a block can be assigned as the building (foreground) or the non-building (background). To generate the changed buildings, the following post-processing actions are considered: First, when two blocks with the same location appear in the two-phase images, the building masks are overlapped to produce the basic changed building. Second, a hole–fill process is conducted on the basic changed building to maintain its integrity. Third, all blocks that contain the changed buildings are returned to their locations in the original image. A logical OR operation is used to identify the changed buildings if neighboring blocks overlap. Finally, threshold T_A is used to discard small objects. Consequently, the final detection results are achieved.
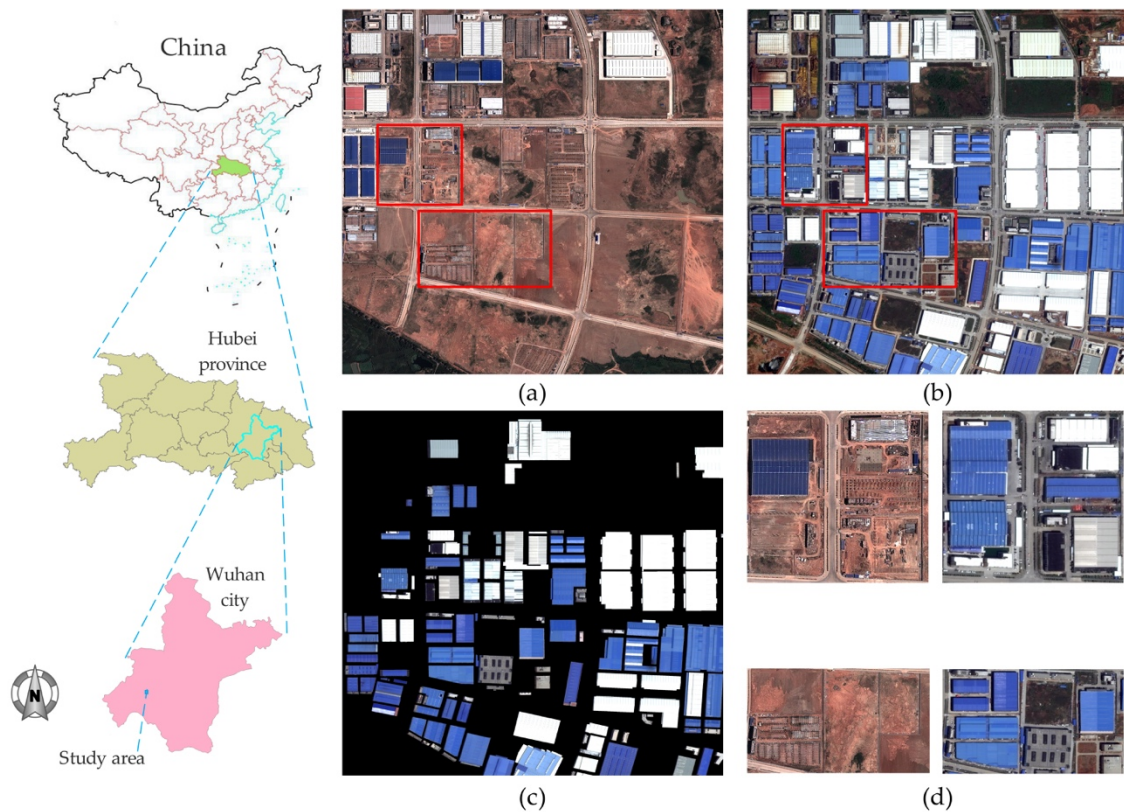
3. Results and Discussion

To verify the validity of the proposed method, two VHR images at two phases are used for the experimental analysis. Considered the core stage of the proposed method, the effect of the three terms of co-segmentation is highlighted in this paper. Some crucial parameters, such as T_s, T_t, and D that can potentially influence the results of differencing and co-segmentation, have been discussed from the perspective of sensibility. Consequently, having stressed the relevance of the typical pixel-based and object-based approach in the introduction of this paper, the method in [5] is compared with the proposed method. Then, the effects of changed building extraction and false alarm removal procedures are analyzed.

3.1. Data and Parametric Setting

The bi-temporal WorldView-1 images with spatial resolutions of 0.5 m acquired on 16 August 2013 and 28 April 2017 were used as the experiment data. A part of a district in the suburb of Wuhan, Hubei, China with latitude and longitude range of 114.0450 E–114.0611 E and 30.4267 N–30.4423 N is shown in both images (Figure 5a,b). The images show three bands in red, green, and blue, and the sizes of which are 2876×3000 pixels. Some fundamental image pre-processing operations similar to those in [35] were performed to reduce the discrepancies between the bi-temporal images. In particular, the present work conducted radiometric calibration to eliminate reflectance

differences and co-registration to ensure the obtainment of the bi-temporal image pixels with the same location. A comparison of the two images shows that the spectral characteristics of the newly built buildings are inconsistent (i.e., see blue, white, and gray bands). Apart from the changed buildings, many other changed geo-objects, such as roads, bare land, and grasslands, among others, are observable. The results of the visual interpretation, as depicted by the ground truth in Figure 5c, show that 133 buildings, including newly built or extended ones, have changed from 2013 to 2017.

Figure 5. Study data: (**a**) image captured in 2013; (**b**) image captured in 2017; (**c**) ground truth of changed buildings; and (**d**) pertinent details of the two images.

According to the methodology described in Section 2, the proposed method mainly includes the determination of the coarse location of the changed objects and the fine extraction of the changed buildings. The former is conducted by the differencing approach and the latter by the co-segmentation method. Several parameters should be set to run the algorithms and influence the final result at different extents. The parameters that affect coarse location are related to texture feature extraction, spectral and texture thresholds, and area threshold. Meanwhile, the relevant parameters that affect fine extraction are related to the extension distance of image blocks, determination of feature dimensions, and use of unary and pairwise terms in co-segmentation. The settings of these parameters are explained as follows:

(1) Window size, θ, d, and N are used to calculate the texture feature (GLCM variance) of VHR remote sensing images in this study. Considering that the texture feature is not as essential as the color feature in terms of affecting the differencing result (see Section 3.4.1), their settings are selected by following the optimal parameters of Xiao [5]. Hence, the window size was set to 7×7, θ was set to $45°$, d was set to 1, and N was set to 8.

(2) T_S and T_T are important in controlling the changed pixels obtained by the differencing method, and they affect the location of the changed buildings. For the study data, the optimal values of T_S and T_T were set to 0.75 and 3.5. The sensitivity analysis of these parameters is presented in Section 3.4.1.
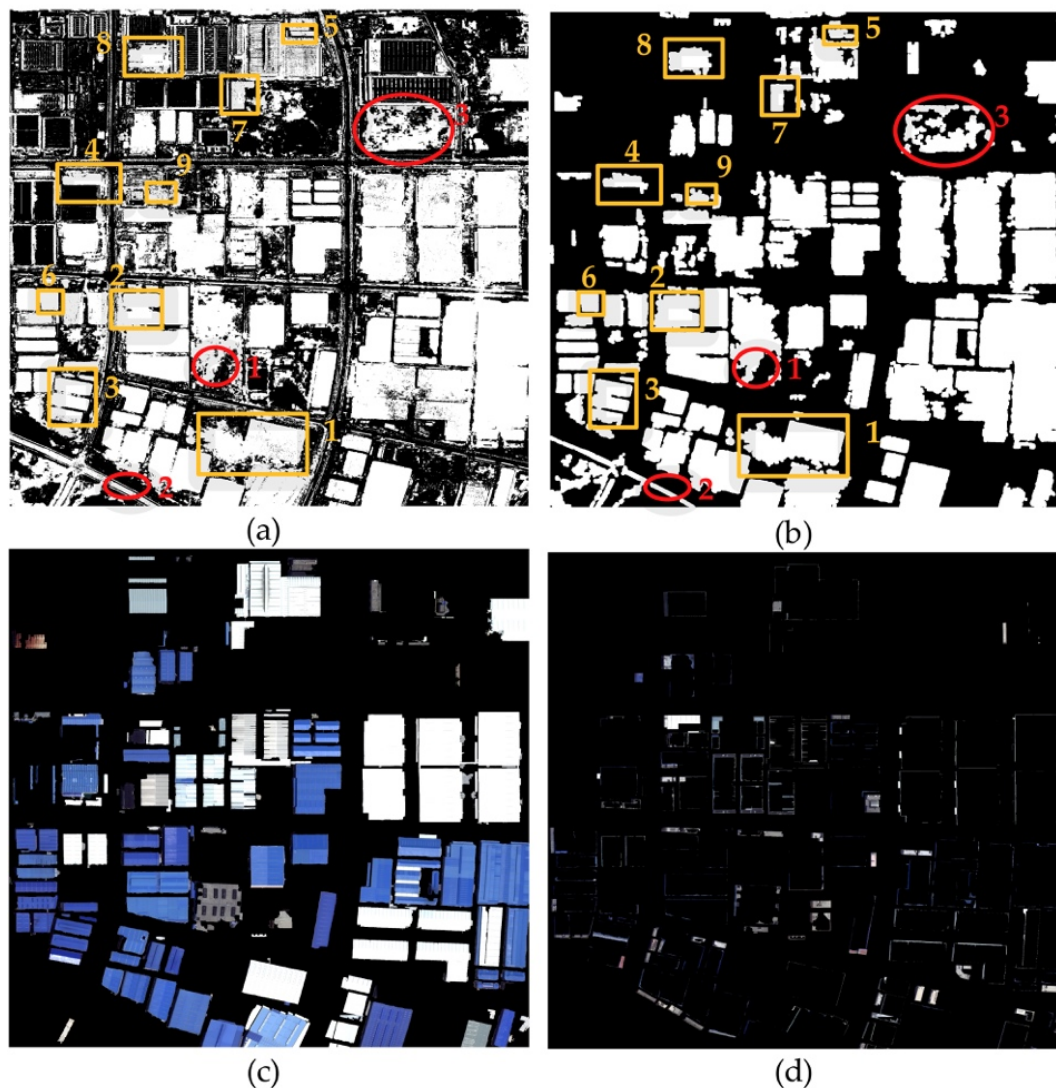
(3)	T_A represents the area threshold for building object differentiation. It is the number of pixels contained in the object and used as basis to remove objects that are too small in terms of size (i.e., considered non-buildings). The spatial resolution of the study data is 0.5 m. Considering that the number of pixels of a building object with this resolution cannot be too small, T_A was set to 300 in this study.

(4)	D is the extension distance of the image block that is generated based on the changed object. Considering that the changed object obtained by the differencing method simply gives the coarse location, its edge will deviate from the actual object. However, even a small number of missing parts may occur. Therefore, this parameter is generally set to a suitable value to cover the changed building area. Taking into account the spatial resolution of remote sensing image, D was set to 30 in this study. The sensitivity of D will be analyzed in Section 3.4.2.

(5)	The number of scribbled blocks is the number of image blocks selected for scribbling from the block sequence. The superpixels in the scribbled blocks are used to guide the subclasses produced by FCM for merging into the foreground and background classes and obtain the foreground probability. Preferably, the selected scribbled image blocks cover all foreground and background ground features. Given that an image block may contain many types of ground features, the number of scribbled images need not be very large. Therefore, in this study, 8 and 13 images were selected from the block sequences for scribbling by using the images from 2013 and 2017, respectively.

(6)	T_{inf} determines the new feature vector dimension obtained from the KPCA-transformed features. The larger the value is, the higher the dimension of the new feature vector will be. Samples that belong to a certain class in the low dimension can be subdivided into different subclasses in the high dimension, and this approach enables for a larger best K value to be obtained by the elbow rule. In general, most of the information of the original feature component is contained in a small part of components in front of the KPCA-transformed features. Therefore, after T_{inf} is increased to a certain extent, the value of K becomes relatively stable. In this study, before K became stable, T_{inf} was set to 0.55 when it was observed that the features could distinguish the building objects from the other objects.

(7)	ε_u and ε_p are the weights of the unary and pairwise terms in the co-segmentation, respectively. They affect the result of the co-segmentation by adjusting the proportion of the unary and pairwise items in the energy function. The unary term is mainly determined by the foreground probability of pixels. The higher the probability of the foreground is, the more likely the pixels will be assigned a foreground label. The pairwise term is the dissimilar punishment between adjacent pixels. On the one hand, if the spectral characteristics of the adjacent pixels are similar, then the energies generated when assigned to the same label become smaller, and vice versa. In this study, the superpixel has higher internal homogeneity and distinct boundaries. Adjustments to ε_u and ε_p may impact the limited number of pixels around superpixels' boundaries. On the other hand, to highlight the effect of the probability of the foreground, ε_u should be set as a large value. Therefore, ε_u^i and ε_p^i were set to 0.8 and 0.2 for the block co-segmentation in this study, respectively.

3.2. Experimental Results

The experimental results are shown in Figure 6. The changed pixels generated by the improved differencing method are shown in Figure 6a, in which each changed building can be identified based on an approximate location. The changed objects (Figure 6b) were determined consecutively by morphological processing based on the close-and-open and hole-filling operation and area filtering according to T_A, as discussed in Section 2.1.2. As shown in Figure 6b, false detection and omission are highly noticeable. As depicted by the red ellipse 1, some changed buildings cannot be detected. The road and bare land covered by red ellipses 2 and 3 are also incorrectly detected. In addition,

although the shape of the changed building object was restored by the morphological process, the boundaries were inaccurate.

Figure 6. Experimental results of the proposed method: (**a**) changed pixels; (**b**) changed objects; (**c**) final detection result; and (**d**) undetected building.

On the basis of coarse locations of the changed buildings, 126 blocks were established regardless of the number of buildings contained in one block. Then, on the basis of the block images, co-segmentation was implemented to extract the building objects and discard non-building objects. The process of refinement by using this method is beneficial in searching for the optimal results for a limited area. After T_A was used to remove the small objects, the binary objects in all blocks were returned to their original locations (i.e., recorded position information). The final detection result is shown in Figure 6c, in which 111 changed buildings are identified. However, as shown in Figure 6d, some gray and black buildings are not detected and some gray non-building objects are falsely detected by both steps of differencing and co-segmentation. The limitation can be mainly explained by the highly similar color features of these buildings with those of the roads, concretes, or bare land. Nonetheless, compared with the results of the differencing method in Figure 6b, the extracted changed buildings in Figure 6c have more accurate boundaries without any non-building object attached to them.

3.3. Analysis of Co-Segmentation

To further investigate the improved scheme on co-segmentation, the detection results of the first five objects in Figure 6b (orange rectangles) are shown again in Figure 7. In the two gray maps of foreground probability, a larger gray value denotes a relatively higher foreground probability of the pixel. Even if the building superpixels show different probability values, and even if there is a high probability that part of the road superpixels will have similar colors as that of the buildings, most of the building areas have higher foreground probabilities than their surrounding objects. Priori knowledge was generated for the co-segmentation model, and thus, different types of buildings can obtain higher foreground probability. For instance, although some differences are observed in the spectral and texture characteristics of the buildings in 2017, these buildings have higher foreground probabilities than their surroundings.

The foreground probability provides priori knowledge not only to the unary term but also to the global term in the co-segmentation model. The higher the foreground probability is, the more likely the objects will eventually be assigned as a foreground, and this condition can partly guarantee a potential satisfactory result. Apart from the unary and global terms, the pairwise term in the co-segmentation model also played an important role in improving the results. In general, a pairwise term is used to control the smoothness of foreground pixels after comparing the similar features of adjacent pixels. The adjacent pixels with similar features can thus be rendered more easily by allocating them to the same foreground or background, whereas the adjacent pixels with different features are assigned with different labels. As shown in Figure 7, even if some building pixels have lower foreground probabilities (orange rectangles) and even if some non-building pixels have similar foreground probabilities compared with the buildings in green rectangles, all these pixels are assigned the correct labels in the final co-segmentation. The accurate results can be attributed to the effect of the pairwise term. In other words, the risk of attaining uncertain probabilities in the clustering step can be minimized by using pairwise terms in the co-segmentation. The co-segmentation results shown in the fourth and seventh columns of Figure 7 have good boundaries.

An overlapping judgment was conducted for the blocks in the two phases. Subsequently, different foregrounds can be easily determined, as depicted by the changed buildings in the eighth column of Figure 7. In this column, the first to fourth rows show the newly built and expanded buildings with clear edges. Considered a special case, the buildings in the fifth row are unchanged objects, which are likely detected as false alarms because of the large spectral difference and low radiance correction. Overall, most false alarms were effectively identified in the improved co-segmentation by countering the non-apparent differences of the spectral characteristics.
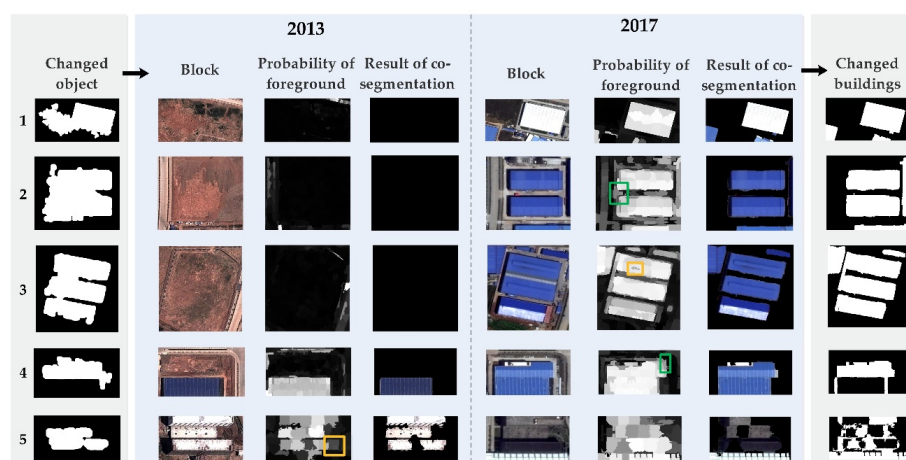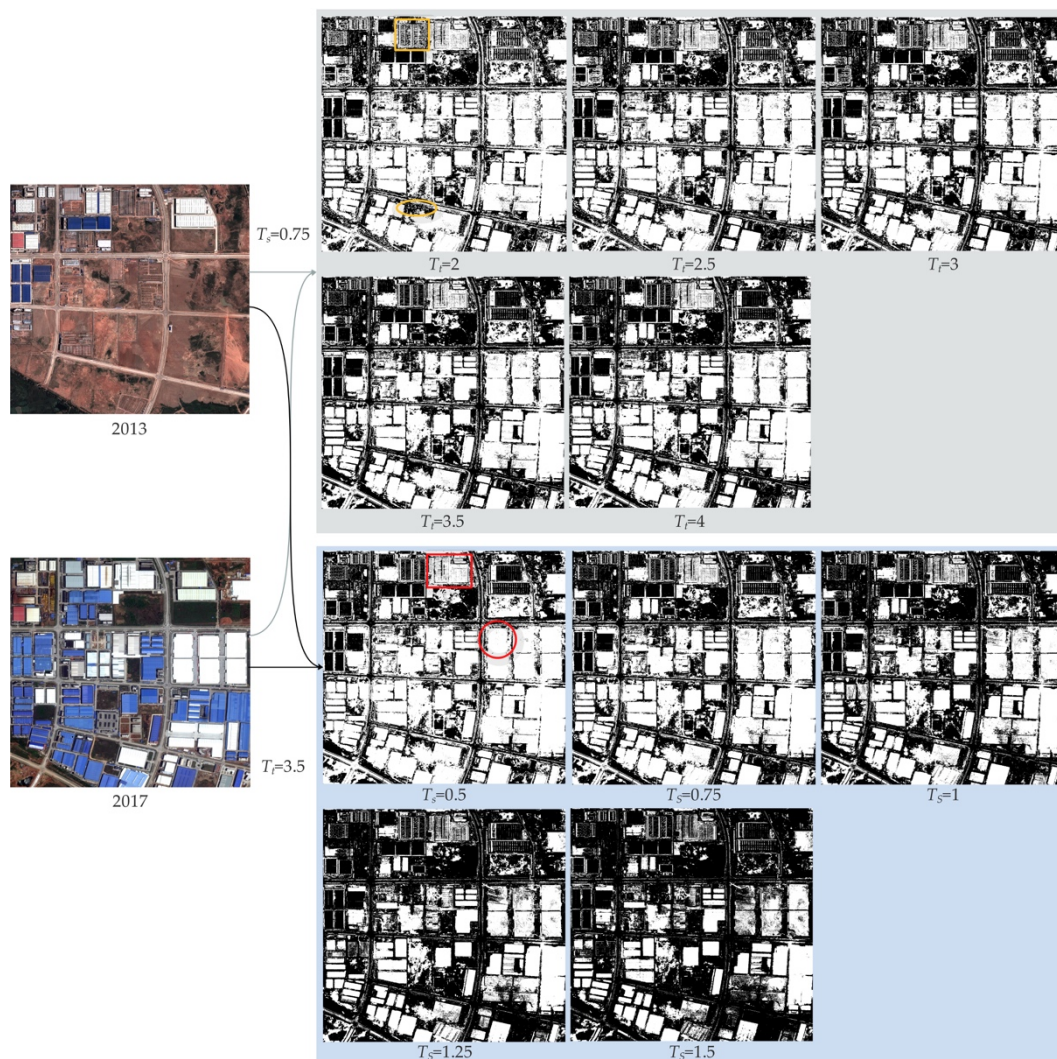

Figure 7. Co-segmentation results of some blocks.

3.4. Parametric Sensitivity Analysis

3.4.1. Parameters T_s and T_t

On the basis of the methodology described in Section 2, the changed pixels obtained by the differencing method are aggregated into objects after a simple post-processing procedure. The aggregations provide an approximate positioning for the subsequent extraction of the changed buildings via co-segmentation. The effect of the changed pixels obtained by the differencing method is mainly determined by T_s and T_t. The extraction results (Figure 8) of the changed pixels for the two images with different T_s and T_t. values are discussed in the next paragraphs.

Figure 8. Differencing results with different T_s and T_t values.

On the basis of images in the gray rectangle in Figure 8, the effect of different T_t values on the differencing results was investigated. The T_s value used in the experiment was 0.75. Under different T_t conditions, the extraction effects of the internal body of real changed buildings were essentially the same and only some feature edges or stripe-like features exhibited alterations. Examples of these stripes for some buildings are highlighted by the orange rectangle and circle in the first row of Figure 8. When T_t increased, the changed pixels on these edges gradually disappeared. When T_t exceeded a certain value, the obtained changed pixels became stable. In other words, regardless of the change in

T_t, the main body of the building will be retained. This phenomenon may be attributed the texture feature used as the variance of GLCM. The change in the texture feature is relatively more responsive at the edge of the object, and the change within the object is generally zero (i.e., Xiao [5]). Therefore, if a pixel is located inside the object (i.e., any object type) before and after the change, the pixel will be unlikely affected by the change in T_t. Given that the method proposed in this study is used for the coarse positioning of changed pixels, there is no need for object edges to be highly accurate in the differencing phase. Therefore, the value of T_t can be set slightly larger for experiment.

The blue rectangle of Figure 8 depicts the effect of different T_s values on the results of the differencing method. In the experiment, T_t was 3.5. The change in T_s can seriously influence the extraction effect of the internal pixels of a changed building (i.e., see red rectangle and circle). Therefore, the designated value of T_s has greater influence on the result of the differencing method compared with that of T_t. Moreover, the spectral change can serve as the main basis of the differencing method. Given that the color and spectral feature of Lab was adopted, a relatively higher T_s (0.75) was used to extract most of the changed building pixels without having to affect the several non-building pixels. However, despite the advantage of using the Lab color space, an even higher T_s tended to dissipate all changed pixels in the differencing method result.

3.4.2. Parameter D

The constructed block based on the coarse location is the search scope of the co-segmentation, and the block size is highly essential in determining the final accuracy of building change detection. The extension parameter D directly controls the size of each block (i.e., D affects the amount of buildings and the number and shape of superpixels in a block). Under this condition, the sensitivity of parameter D is elaborated in this paper.

Take the changed buildings labeled sixth to ninth in Figure 6b as examples. The co-segmentation results of the blocks extracted from the image in 2017 with different sizes of D are shown in Figure 9. As shown by the second column of each gray rectangle in Figure 9, an increase in the size of D leads to a relatively higher number of superpixels. Too large blocks will cover many other surrounding objects, and this will not only influence the computation efficiency of the co-segmentation but also lead to the risk of false or missing detection. In Figure 9, several unchanged buildings can be observed around the changed objects (i.e., see 7 and 8), and false detections are likely to happen. Incidentally, the difference in D only affected a small part of the buildings, and most of the building pixels were segmented as the foreground.
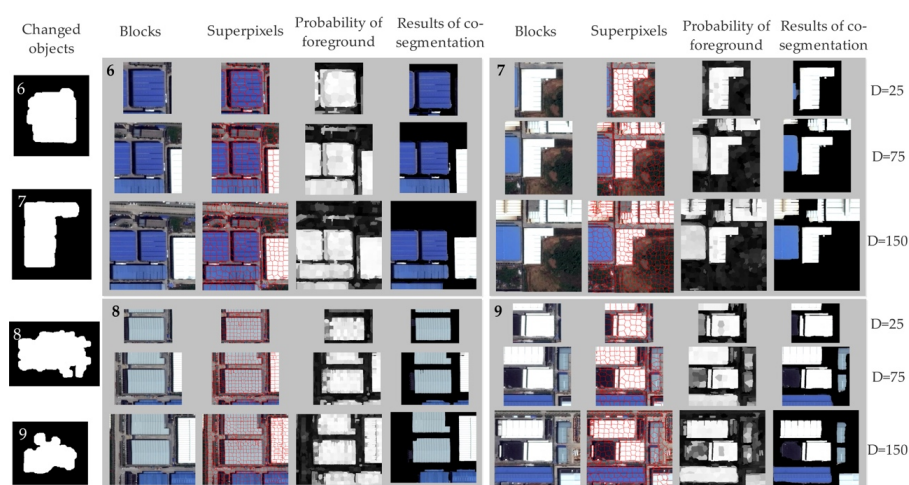

Figure 9. Co-segmentation results with different D values.

The accurate boundaries of buildings were captured by the over-segmentation based on SLIC. In addition, because the foreground probability was generated via the strategy of combining few scribbled blocks and fuzzy clustering, various types of buildings with different spectral features achieved higher foreground probability than the non-buildings. As shown in Figure 9, although the spectral characteristics of the buildings in the four patches differ, the building pixels ultimately obtained high foreground probability.

Therefore, even if the smaller D rendered the block to contain only the main building, or even if the larger D rendered the block to contain other surrounding objects (including other buildings), the improvement strategy of combining coarse location and co-segmentation can extract different kinds of buildings. Changes have been roughly positioned in the differencing step, and thus, D with extended distance should be set slightly larger. For example, D was 30 in this study considering that the spatial resolution of the study data was 0.5 m.

3.5. Comparative Experimental Evaluation

To evaluate the strategy proposed in this study, a comparative experiment was conducted between the combining pixel-based detection and object-based recognition (CPDOR) [5] and the proposed method. First, CPDOR was used to calculate the difference in the MBI mean values of the objects detected by the differencing method. Then, threshold value T_{MBI} was used to compare and subsequently obtain the ultimately changed building object. The threshold of texture feature difference T_T in the first step was 2.5, whereas that of the spectral difference T_S were the two values of 0.5 and 1.0 (for comparison). The results generated by the differencing method are shown in Figure 10a,b. The findings show that when T_S increased, the labeled buildings disappeared relative to the roads and bare land (i.e., varied RGB spectral values were used for differencing). Therefore, to detect as many changed buildings as possible, T_S was 0.5 for the CPDOR in this study. The above limitation can be rectified by conducting differencing with the Lab color space. Then, the changed pixels in Figure 10a were compared with those in Figure 6a. Most of the changed building pixels were retained, but the changed pixels of non-buildings in RGB color space were much larger than those in Lab color space. After morphological processing and area filtering, the MBI difference threshold T_{MBI} was set to 0.1 to determine whether the changed objects are buildings for CPDOR.
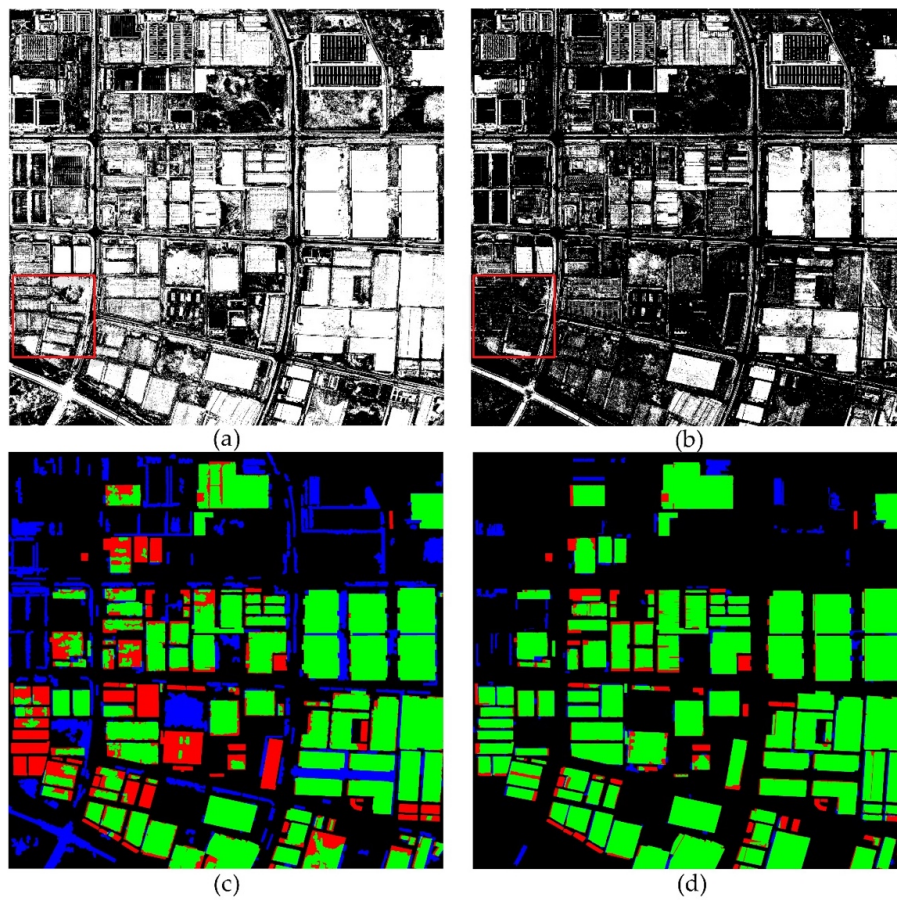
On the basis of the ground truth in Figure 5c, the comparative results generated by CPDOR and our method were further analyzed (Figure 10c,d, respectively). False detection and missing detection are also shown in the figures. In Figure 10, although some large buildings appear to have changed pixels after differencing, they are missed in the result of CPDOR. The difference may be attributed to the MBI values of the 2017 image that were extremely close to the ones of bare land in the 2013 image. In addition, relative to the final results of MBI for the differencing method, those of CPDOR showed some weaknesses in terms of distinguishing buildings from their adjacent changed objects (i.e., grass, shadow, etc.). Subsequently, their edges may appear greatly deviated from reality.

Change detection can be described as the binary classification of images wherein pixels in an image can be divided into two categories (i.e., change in buildings and those of others in this study). Therefore, image classification can be used an evaluation tool to assess the change detection results. Evaluation indexes, such as recall rate, false positive rate, and overall accuracy [36], can be established by constructing a confusion matrix (Table 2). The corresponding formulas are also presented.

$$\text{RecallRate} = \frac{Tp}{Tp + Fn} \tag{11}$$

$$\text{FalsePositiveRate} = \frac{Fp}{Fp + Tn} \tag{12}$$

$$\text{OverallAccuracy} = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \tag{13}$$

Figure 10. Comparison evaluation: (**a**,**b**) changes obtained by combining pixel-based detection and object-based recognition (CPDOR) when $T_S = 0.5$ and $T_S = 1$; (**c**) overlay between CPDOR result and ground truth, and (**d**) overlay between results of this study and ground truth. Green represents correct detection, red represents missing detection, and blue represents false detection.

Table 2. Change detection confusion matrix.

	Changed Building	Others
Detected Changes	TruePositive (*Tp*)	falsePositive (*Fp*)
Detected Non-changes	falseNegative (*Fn*)	TrueNegative (*Tn*)

The comparative statistical data of the results at the pixel level are shown in Table 3. The recall rate, false positive rate, and overall accuracy of the proposed method are better than the CPDOR. The proposed method is therefore effective.

Table 3. Statistical accuracy comparison at the pixel level between the results produced by two approaches.

Metrics	Proposed Method	CPDOR Method
Actual changed building (*Tp* + *Fn*)	2,985,301 (133 objects)	2,985,301 (133 objects)
Detected changed building (*Tp*)	2,706,866 (111 objects)	2,241,789(102 objects)
False alarm units (*Fp*)	220,856	949,662
Recall rate	90.67%	75.09%
False positive rate	3.91%	16.83%
Overall accuracy	94.21%	80.38%

4. Conclusions and Prospect

The change detection method based on differencing and co-segmentation was elaborated in this paper. The difference in spectral features was apparent when the Lab color space was used, and more change information of buildings were retained in the process. The change information generated by the differencing method objectively derived the coarse locations of the changed buildings. This first part of the proposed method can be regarded an important basis for subsequent refinement processes. The image blocks were constructed based on the derived locations of the changed objects. Another key component was co-segmentation, which was adopted to derive similar information from candidate buildings and extract individual buildings from blocks. In particular, FCM was performed to derive the foreground or background probability of superpixels, and some blocks were scribbled to provide priori information to the co-segmentation. The improvements proposed in this study were apparent, and the objects were refined to generate buildings with highly accurate shapes and edges. The experimental results verified the effectiveness of the proposed method in terms of precisely detecting changed buildings in VHR remote sensing images.

In future studies, we will focus on the learning approach-based feature representation for diverse buildings. Furthermore, the learning feature of VHR remote sensing image will be considered in co-segmentation to render it more robust and fit for complex urban environments.

Author Contributions: J.C. participated in the concepts, design, definition of intellectual content. H.L. performed literature search, data acquisition, and manuscript preparation. J.H. carried out the study and provided assistance for manuscript preparation. M.Y. collected important background information and carried out manuscript editing. M.D. performed manuscript review.

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

References

1. Hussain, M.; Chen, D.M.; Angela, C.; Wei, H.; Stanley, D. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS J. Photogramm. Remote Sens.* **2013**, *80*, 91–106. [CrossRef]
2. Singh, A. Digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* **1989**, *10*, 989–1003. [CrossRef]
3. Ingram, K.; Knapp, E.; Robinson, J.W. *Change Detection Technique Development for Improved Urbanized Area Delineation*; Computer Science Corporation: Silver Spring, MD, USA, 1981.
4. Bruzzone, L.; Bovolo, F. A novel framework for the design of change-detection systems for very-high-resolution remote sensing images. *Proc. IEEE* **2013**, *101*, 609–630. [CrossRef]
5. Xiao, P.F.; Zhang, X.L.; Wang, D.G.; Yuan, M.; Feng, X.Z.; Kelly, M. Change detection of built-up land: A framework of combining pixel-based detection and object-based recognition. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 402–414. [CrossRef]
6. Chen, G.; Geoffrey, J.H.; Luis, M.T.C.; Michael, A.W. Object-based change detection. *Int. J. Remote Sens.* **2012**, *33*, 4434–4457. [CrossRef]
7. Bruzzone, L.; Carlin, L. A multilevel context-based system for classification of very high spatial resolution images. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2587–2600. [CrossRef]
8. Peng, D.F.; Zhang, Y.J. Object-based change detection from satellite imagery by segmentation optimization and multi-features fusion. *Int. J. Remote Sens.* **2017**, *38*, 3886–3905. [CrossRef]
9. Aguirre-Gutiérrez, J.; Seijmonsbergen, A.C.; Duivenvoorden, J.F. Optimizing land cover classification accuracy for change detection, a combined pixel-based and object-based approach in a mountainous area in Mexico. *Appl. Geogr.* **2012**, *34*, 29–37. [CrossRef]
10. Le Hégarat-Mascle, S.; Seltz, R. Automatic change detection by evidential fusion of change indices. *Remote Sens. Environ.* **2004**, *91*, 390–404. [CrossRef]

11. Taubenböck, H.; Esch, T.; Felbier, A.; Wiesner, M.; Roth, A.; Dech, S. Monitoring urbanization in mega cities from space. *Remote Sens. Environ.* **2012**, *117*, 162–176. [CrossRef]

12. Tian, J.; Cui, S.; Reinartz, P. Building change detection based on satellite stereo imagery and digital surface models. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 406–417. [CrossRef]

13. Xiao, P.F.; Yuan, M.; Zhang, X.L.; Feng, X.Z.; Guo, Y.W. Cosegmentation for object-based building change detection from high-resolution remotely sensed images. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 1587–1603. [CrossRef]

14. Tang, Y.Q.; Huang, X.; Zhang, L.P. Fault-tolerant building change detection from urban high-resolution remote sensing imagery. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1060–1064. [CrossRef]

15. Huang, X.; Zhang, L.P.; Zhu, T.T. Building change detection from multitemporal high-resolution remotely sensed images based on a morphological building index. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *7*, 105–115. [CrossRef]

16. Hou, B.; Wang, Y.H.; Liu, Q.J. A saliency guided semi-supervised building change detection method for high resolution remote sensing images. *Sensors* **2016**, *16*, 1377. [CrossRef] [PubMed]

17. Leichtle, T.; Gei, C.; Lakes, T.; Taubenböck, H. Class imbalance in unsupervised change detection—A diagnostic analysis from urban remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *60*, 83–98. [CrossRef]

18. Dong, S.C.; Figueiredo, M.A.T. Cosegmentation for image sequences. In Proceedings of the 14th International Conference on Image Analysis and Processing, Modena, Italy, 10–14 September 2007; pp. 635–640.

19. Hochbaum, D.S.; Singh, V. An efficient algorithm for co-segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 269–276.

20. Rother, C.; Minka, T.; Blake, A.; Kolmogorov, V. Cosegmentation of image pairs by histogram matching—Incorporating a global constraint into mrfs. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006; pp. 993–1000.

21. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *6*, 610–621. [CrossRef]

22. Connolly, C.; Fleiss, T. A study of efficiency and accuracy in the transformation from RGB to CIELAB color space. *IEEE Trans. Image Process.* **1997**, *6*, 1046–1048. [CrossRef] [PubMed]

23. Recky, M.; Leberl, F. Windows detection using k-means in CIE-lab color space. In Proceedings of the 2010 20th International Conference on Pattern Recognition (ICPR), Istanbul, Turkey, 23–26 August 2010; pp. 356–359.

24. Achanta, R.; Shaji, A.; Smith, K.; Lussichi, A.; Fua, P.; Susstrunk, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [CrossRef] [PubMed]

25. Ojala, T.; Pietikainen, M.; Harwood, D. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In Proceedings of the 12th IAPR International Conference on Pattern Recognition, Jerusalem, Israel, 9–13 October 1994.

26. Leung, T.; Malik, J. Representing and recognizing the visual appearance of materials using three-dimensional textons. *Int. J. Comput. Vis.* **2001**, *43*, 29–44. [CrossRef]

27. Jolliffe, I.T. *Principal Components Analysis*; Springer-Verlag: Berlin, Germany, 1986.

28. Shao, R.P.; Hu, W.T.; Wang, Y.Y.; Qi, X.K. The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform. *Measurement* **2014**, *54*, 118–132. [CrossRef]

29. Schölkopf, B.; Smola, A.; Müller, K.R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **1998**, *10*, 1299–1319. [CrossRef]

30. Batra, D.; Kowdle, A.; Parikh, D.; Chen, T. Interactively co-segmentating topically related images with intelligent scribble guidance. *Int. J. Comput. Vis.* **2011**, *93*, 273–292. [CrossRef]

31. Collins, M.D.; Xu, J.; Grady, L.; Singh, V. Random walks based multi-image segmentation: Quasiconvexity results and gpu-based solutions. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 1656–1663.

32. Wang, W.G.; Shen, J.B. Higher-order image co-segmentation. *IEEE Trans. Multimed.* **2016**, *18*, 1011–1021. [CrossRef]

33. Dunn, J.C. Some recent investigations of a new fuzzy partitioning algorithm and its application to pattern classification problems. *J. Cybern.* **1974**, *4*, 1–15.

34.   Bholowalia, P.; Kumar, A. EBK-means: A clustering technique based on elbow method and k-means in WSN. *Int. J. Comput. Appl.* **2014**, *105*, 17–24.

35.   Wang, X.; Liu, S.C.; Du, P.J.; Liang, H.; Xia, J.S.; Li, Y.F. Object-Based Change Detection in Urban Areas from High Spatial Resolution Images Based on Multiple Features and Ensemble Learning. *Remote Sens.* **2018**, *10*, 276. [CrossRef]

36.   Jensen, J.R. *Introductory Digital Image Processing: A Remote Sensing Perspective*; Prentice-Hall: Upper Saddle River, NJ, USA, 2004; p. 382.