Cosegmentation for Object-Based Building Change Detection From High-Resolution Remotely Sensed Images

Pengfeng Xiao, *Member, IEEE*, Min Yuan, Xueliang Zhang, Xuezhi Feng, and Yanwen Guo

***Abstract*— This paper presents a cosegmentation-based method for building change detection from multitemporal high-resolution (HR) remotely sensed images, providing a new solution to object-based change detection (OBCD). First, the magnitude of a difference image is calculated to represent the change feature. Next, cosegmentation is performed via graph-based energy minimization by combining the change feature with image features at each phase, directly resulting in foreground as multitemporal changed objects and background as unchanged area. Finally, the spatial correspondence between changed objects is established through overlay analysis. Cosegmentation provides a separate and associated, rather than a separate and independent, multitemporal image segmentation method for OBCD, which has two advantages: 1) both the image and change features are used to produce foreground segments as changed objects, which can take full advantage of multitemporal information and produce two spatially corresponded change detection maps by the association of the change feature, having the ability to reveal the thematic, geometric, and numeric changes of objects and 2) the background in the cosegmentation result represents the unchanged area, which naturally avoids the problem of matching inconsistent unchanged objects caused by the separate and independent multitemporal segmentation strategy. Experimental results on five HR datasets verify the effectiveness of the proposed method and the comparisons with the state-of-the-art OBCD methods further show its superiority.**

***Index Terms*— Cosegmentation, graph cut, high-resolution (HR) remotely sensed image, multitemporal segmentation, object-based change detection (OBCD).**

I. Introduction

CHANGE detection is the process of identifying and quantifying differences in the state of an object or phenomenon by observing it at different phases [1]. It is of great significance to a diverse range of applications, such as the investigation of land cover and land use [2]–[4], urban development [4]–[6], vegetation change [7]–[12], and damage monitoring [13], [14].

Pixel-based change detection methods have been commonly used to detect abrupt changes from low- or medium-resolution images [15], [16]. However, the results of pixel-based methods are often limited when applied to high-resolution (HR) remotely sensed images. Because of the increased high-frequency components in HR images, a large amount of small spurious changes are possibly produced [17]. Furthermore, pixel-based methods are sensitive to the accuracy of geometric registration and radiometric correction [18].

Object-based change detection (OBCD) has evolved from the concept of object-based image analysis [19], [20] that uses image segmentation techniques, which can not only employ the spectral, texture, and transformed values that pixel-based methods use as the measure of change, but also exploit extra information about the shape features and spatial relations of objects. Since image objects are used as the basic units in OBCD, the important issue is how to define changes between them [17]. One solution is to compare image objects from different images, which is similar to pixel-based methods. In general, two strategies are developed [18]. In the first, objects from the image at one phase (e.g., T1 image) are extracted, and are assigned to or searched from the image at the other phase (e.g., T2 image) without segmentation [21]. In the second strategy, objects from multitemporal images are independently extracted and then compared for change detection [13], [22]–[27]. Another solution of OBCD has managed to derive image objects by segmenting all multitemporal states of the scene in one step [7]–[11], [14], [28].

Therefore, the multitemporal image segmentation for OBCD can be performed with three different strategies depending on the input data [27]: 1) on the basis of the multitemporal images, i.e., using a data stack that consists of both images; 2) based on the image of one phase; the generated object boundaries are then simply assigned to the image of another phase without segmentation; and 3) separately for two images. For 1) and 2), a single segmentation is produced, which makes the comparisons of image objects straightforward, because the image objects are consistent in size, shape, and location over time. However, the single segmentation leads to the limitation that the geometric changes, e.g., size and shape changes, cannot be detected by 1) and 2). With respect to 3), it produces image objects separately at different phases and allows using

Manuscript received June 13, 2016; revised October 17, 2016; accepted November 8, 2016. Date of publication January 2, 2017; date of current version February 23, 2017. This work was supported in part by the Qinglan Project of Jiangsu Province under Grant 201423, in part by the National Natural Science Foundation of China under Grant 41601366, 61373059, and 61321491, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20160623 and BK20150016, and in part by the China Postdoctoral Science Foundation under Grant 168353. *(Corresponding authors: Pengfeng Xiao and Xueliang Zhang.)*

P. Xiao, M. Yuan, X. Zhang, and X. Feng are with the Department of Geographic Information Science, Nanjing University, Nanjing 210023, China, the Collaborative Innovation Center of South China Sea Studies, Nanjing 210023, China, and also with the Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China (e-mail: xiaopf@nju.edu.cn; zxl@nju.edu.cn).

Y. Guo is with the State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2016.2627638

all the object features for change analysis, making it able to detect the geometric changes of objects. Theoretically, the separate segmentation seems the most powerful strategy for image object comparison, because both the thematic and geometric changes can be analyzed [29]. The challenge of 3) is to extract similar image objects in unchanged area and then correctly link image objects over time [17], [18], [27], [29]. However, the separate segmentations are usually performed independently. As multitemporal images are acquired at different phases, the images tend to be different due to variations in the sun angle, sensor angle, and atmospheric conditions, making it difficult to independently generate similar image object boundaries for the same landscape feature in multitemporal images [27]. Hence, the separate and independent segmentations of multitemporal images would produce image objects of different sizes and shapes for both changed and unchanged areas. The inconsistence of unchanged objects between multitemporal images makes it difficult to search spatially corresponded objects [17], [18], [29] and thus difficult to distinguish whether the object difference is due to real change or geometric inconsistence caused by independent segmentation [30].

Several strategies were proposed to overcome the difficulty of searching spatially corresponded objects caused by the inconsistent unchanged objects from separate segmentations. The multiple segmentation results can be fused into single segmentation by the raster intersection operation [31], but this loses the advantage of comparing geometric changes. The method in [22] was to segment one image as the reference and assign it to the other image, and then resegment the potentially changed regions in the other image. A similar work was found in [32] but taking the land cover map as reference and reclassifying the potentially changed regions. This kind of strategy can easily build the spatial correspondence, but it relies on the high accuracy of the reference map and the resegmentation or reclassification result. Another strategy was to detect or select specific objects of interest first and then search the spatial correspondence [13], [24], [26], which would highly depend on the object detection or selection performance. In this paper, we propose the cosegmentation method by separate and associated, rather than separate and independent, segmentation of multitemporal images, which can produce spatially corresponded changed objects and allow the comparison of object geometry.

Cosegmentation [33], which simultaneously segments the common parts of an image pair, was first introduced in the field of computer vision and has been applied to concurrent foreground extraction tasks, such as segmenting image sequences [34]. Cosegmentation can take full advantage of associations between multiple images over the same scene. Usually, cosegmentation is performed through energy minimization to obtain the same foreground object. Such energy functions usually consist of each image features and a similarity feature, e.g., a similarity measure between the foreground histogram of both images [33], [35]–[39].

If we view the process of land cover change as a movement of nature, multitemporal remotely sensed images that depict land cover changes could be regarded as an image sequence. Consequently, change detection from multitemporal images may be performed by cosegmentation. Usually, the aim of cosegmentation is to find common parts in multiple images, and the association is the similarity between images. On the contrary, when applied to change detection, cosegmentation aims to discover regions of change at different phases, and the association is the difference between images.

Building is an important man-made object and a significant factor of urbanization. The number of buildings has increased dramatically in conjunction with the urban expansion of China since the country adopted an economic reform and openness policy in 1978 [40]. Based on the concept of cosegmentation, we propose a novel method for building change detection from multitemporal HR remotely sensed images. In particular, we first generate the change feature by the change vector composed of the morphological building index (MBI) [41] and spectral features of multitemporal images. It serves as both the prior information of building changes and the association of cosegmentation. Cosegmentation is then performed via graph-based energy minimization under the guidance of the change feature, combined with the image features at each phase, directly resulting in multitemporal changed objects with accurate boundaries and spatial correspondence.

The proposed method is a "concurrent segmentation and detection" pattern, rather than the "sequential segmentation and detection" pattern widely employed by existing OBCD methods. In the sequential pattern, image segmentation and object change analysis are performed sequentially, facing the problem of inconsistent unchanged objects caused by separate segmentation strategy. While in the concurrent pattern, both image and change features are combined in an energy function to segment multitemporal images into foreground and background. Cosegmentation has two advantages.

1) The foreground segments are viewed as changed objects. Instead of independent segmentation of images at different phases, the cosegmentation procedures are associated with the same change features, producing multitemporal changed objects that are spatially corresponded. By using the image features at each phase, cosegmentation can take full advantage of multitemporal information and delineate accurate boundaries for changed objects at each phase. Thus, the cosegmentation results have the ability to reveal the thematic, geometric, and numeric changes of objects.
2) The background represents the unchanged area, which spontaneously avoids the problem of matching inconsistent unchanged objects caused by the separate but independent multitemporal segmentation strategy.

The rest of this paper is organized as follows. In Section II, the cosegmentation-based building change detection method is described. Section III presents the approach to assess the accuracy of the proposed method. Section IV is the description of the datasets and experimental results, and Section V presents the discussion. Finally, Section VI draws the conclusion of this study.

II. Methodology

The proposed building change detection methodology consists of three steps as shown in Fig. 1. First, the change

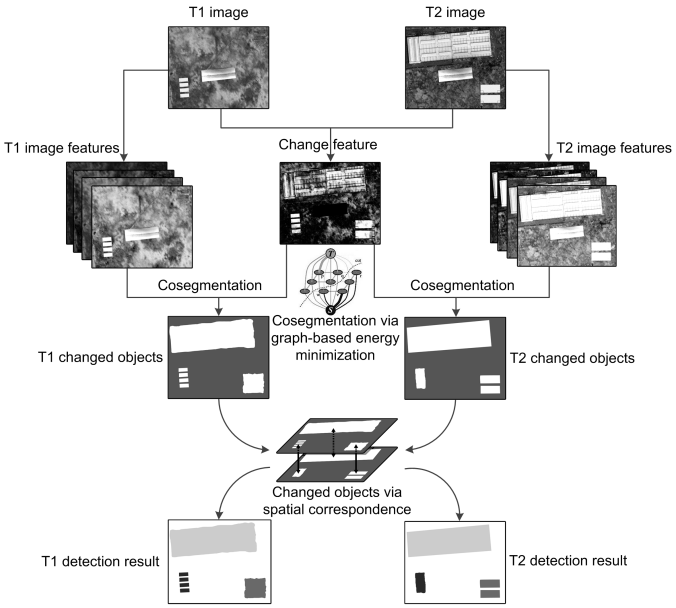

Fig. 1. Flowchart of the proposed cosegmentation-based building change detection methodology.

feature is generated from multitemporal images to provide the prior information and association for successive cosegmentation. Second, combining the same change feature and the image features at each phase, cosegmentation is performed via graph-based energy minimization. It segments multitemporal images separately under the association of the common change feature, producing spatially corresponded multitemporal changed objects by foreground segments. Third, the spatial correspondence between changed objects is established to complete the change detection.

A. Change Feature

The change feature is used to provide the prior information and association for the subsequent cosegmentation process. Target-driven change detection should be related to the characteristics of the target of interest. For building change detection, we adopt MBI, the basic idea of which is to build a relationship between the spectral-structural characteristics of buildings and morphological operators [41]–[43]. Spectral features and the MBI feature are then used in conjunction to calculate the magnitude of the difference image representing the change feature of buildings. Let $\boldsymbol{I}1$ and $\boldsymbol{I}2$ represent two images acquired at different phases T1 and T2, respectively. Each is composed of N bands, including spectral bands and MBI band. The change vector [44] is adopted to calculate the magnitude of the difference image as

$$\boldsymbol{I}^C = \sqrt{\sum_{i=1}^{N} (\boldsymbol{I}1^i - \boldsymbol{I}2^i)^2} \tag{1}$$

where i represents the serial number of bands. A larger value in the difference image indicates a higher possibility of change, and the opposite for a smaller value.

B. Cosegmentation

The aim of cosegmentation is to separately and associatively divide multitemporal images into foreground and background segments representing the changed and unchanged objects, respectively. Both the change feature and the image features at each phase are combined in a graph-based energy function to perform cosegmentation. The change feature serves as the prior information and association for cosegmentation. On the one hand, the change feature represented by the magnitude of the difference image provides the prior information of building changes in the process of cosegmentation. A larger value indicates a higher possibility of change, restraining the cosegmentation scope to changed area to generate foreground segments as changed objects. On the other hand, the change feature plays an important role as an association for separate segmentation procedures to produce spatially corresponded changed objects. The image features are considered to produce complete and accurate objects at each phase.

Cosegmentation is accomplished through minimizing the graph-based energy function. To achieve a globally optimal solution, there are several approaches, such as dynamic programming, simulated annealing, and graph cuts. Dynamic programming can solve a function with a restricted form, and the huge space complexity is the bottleneck of the approach. Simulated annealing can minimize an arbitrary energy function. However, as a result of its generality, it is inefficient in practice. A function to be minimized via graph cuts needs a necessary and sufficient condition of regularity [45], which is exactly satisfied in this study. Therefore, the graph cuts method is used to optimize the binary energy function intuitively represented by a graph and produce the optimum segmentation of the foreground and background [46].

In particular, we first construct an energy function made up of the change feature and image features. A graph is then defined to represent the energy function, comprising a set of nodes and edges. Each edge in the graph is assigned a nonnegative weight based on energy terms of the energy function. The cost of a cut for the graph is equal to the involved energy value. Accordingly, a cut with minimum cost implies energy minimization and generates an optimal segmentation of foreground and background [47]. The general flow of a segmentation based on graph cuts is shown in Fig. 2. The details of cosegmentation are presented as follows.

1) Energy Function: The energy function is defined as

$$E = \lambda E_{\text{change}} + (1 - \lambda) E_{\text{image}} \tag{2}$$

where coefficient $\lambda > 0$ specifies the relative importance of the change feature potential E_{change} versus image feature potential E_{image}.

The change feature is represented by the magnitude of the difference image, which guides multitemporal segmentations to jointly focus on potential changes, thereby generating changed objects with a spatial correspondence. We define

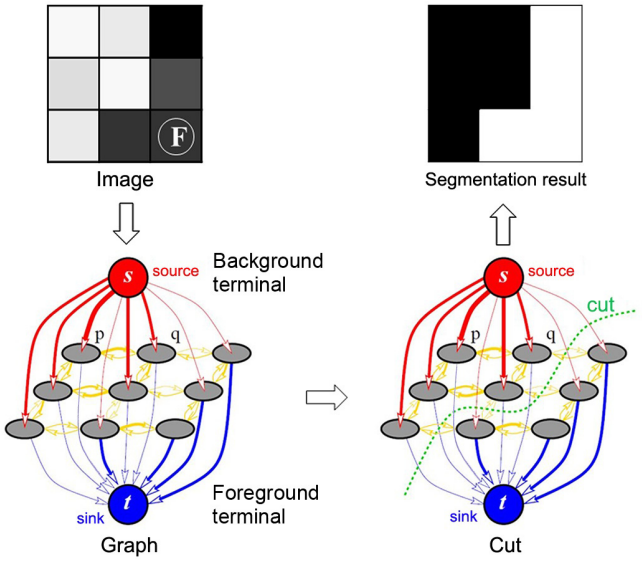

Fig. 2. Example graph cut segmentation for a simple 3×3 image (adapted from [48]). The pixel labeled "F" is a hard constraint. The image feature term in (2) defines the weights of the n-links (yellow edges) while the change feature term defines the weights of the t-links (red and blue edges). The weight (cost) of each edge is reflected by its thickness.

E_{change} as

$$E_{\text{change}} = \sum_{p \in P} D_p(l_p) \tag{3}$$

$$D_p(l_p) = \begin{cases} -\ln \dfrac{I_p^C}{2T}, & \text{if } l_p = \text{fg} \\ -\ln\left(1 - \dfrac{I_p^C}{2T}\right), & \text{if } l_p = \text{bg} \end{cases} \tag{4}$$

where P represents the set of all pixels in an image, l_p indicates a label that is either "fg" or "bg" (abbreviations of "foreground" and "background"), $D_p(l_p)$ represents the penalty of assigning label "fg" or "bg" to pixel p, I_p^C represents the magnitude of pixel p in difference image $\boldsymbol{I}^C$, and T is a threshold parameter, which is determined by the Bayesian theory based on an expectation–maximization algorithm [49].

In general, a pixel p with large I_p^C has a high probability of belonging to the foreground, whereas p with small I_p^C tends to be background. Given the condition that p belongs to the foreground when $I_p^C > 2T$, the pixels satisfying this condition are called hard constraints and labeled "F," as shown in Fig. 2, providing clues of what the user intends to segment. When $I_p^C \leq 2T$, according to (4), a pixel p with large I_p^C has a small penalty $D_p(\text{fg})$ if it is labeled "fg." In contrast, a pixel p with small I_p^C has a small penalty $D_p(\text{bg})$ if it is labeled "bg," according to (4). Consequently, dividing all pixels into foreground or background correctly achieves the minimum energy. However, if the energy function E only contains the change feature potential ($\lambda = 1$), the energy function is $E = E_{\text{change}}$. In this case, cosegmentation is performed only depending on the change feature without image features and it is degraded into a threshold-based change detection method at pixel level. For example, when $I_p^C = T$, $D_p(\text{fg}) = D_p(\text{bg})$, i.e., penalties of assigning labels "fg" and "bg" to pixel p are equal. In other words, the difference image is segmented into two parts only according to the threshold T, i.e., when $I_p^C > T$, p belongs to the foreground; when $I_p^C \leq T$, p belongs to the background.

Hence, in addition to the change feature, image features are integrated into the energy function to relate generated objects to specific images. Image feature potential E_{image} is defined as

$$E_{\text{image}} = \sum_{\{p,q\} \in N} V_{\{p,q\}} \tag{5}$$

$$V_{\{p,q\}} = \exp\left(-\frac{\|\boldsymbol{I}_p - \boldsymbol{I}_q\|^2}{2\sigma^2}\right) \cdot \frac{1}{d(p,q)} \tag{6}$$

where N represents the set of all pairs of adjacent pixels $\boldsymbol{I}_p$ and $\boldsymbol{I}_q$ represent the spectral feature vectors of pixels p and q, respectively, $d(p, q)$ represents the Euclidean distance between pixels p and q, and $V_{\{p,q\}}$ represents the penalty for a discontinuity between pixels p and q, which can also be interpreted as a similarity. According to (6), $V_{\{p,q\}}$ is large when pixels p and q are similar. In this case, both p and q tend to belong to the same segment. On the contrary, $V_{\{p,q\}}$ is small when p and q are different, and such pixels tend to belong to edges between the foreground and the background. Furthermore, σ^2 is a normalization factor. Assuming that $||\boldsymbol{I}_p - \boldsymbol{I}_q||^2$ is invariable, a large σ^2 results in a large $V_{\{p,q\}}$, and it is difficult to separate two pixels into two parts, resulting in coarse segments. Contrarily, a small σ^2 makes $V_{\{p,q\}}$ small, and two pixels are easily separated, leading to fine segments. Hence, σ^2 can be considered a scale parameter, which is set as suggested in [50]

$$\sigma^2 = \langle \|\boldsymbol{I}_p - \boldsymbol{I}_q\|^2 \rangle \tag{7}$$

where $\langle \cdot \rangle$ denotes the average value over the whole image.

2) Graph Construction: For each image, we create a graph $G = (V, E)$, as shown in Fig. 2. In the graph, a node represents a pixel p ($p \in P$) in the image. There are two additional nodes: a foreground terminal (sink t) and a background terminal (source s). Therefore

$$V = P \cup \{s, t\}. \tag{8}$$

The set of edges E consists of two types of edges: n-links (neighborhood links) and t-links (terminal links). Each pixel p has two t-links (s, p) and (p, t) that connect to each terminal. The eight-neighborhood system is used for n-links and each pair of neighboring pixels (p, q) is connected by an n-link. Therefore

$$E = N \cup \{(s, p), (p, t) | p \in P\}. \tag{9}$$

The edge weights are calculated according to the energy function. Table I gives the edge weights in the graph. For n-links, weights are related to the image feature potential E_{image}, i.e., the similarity $V_{\{p,q\}}$ in (6) between neighboring pixels. For t-links, weights depend on the change feature potentialE_{change}, specifically the penalty of $D_p(l_p)$ in (4). If a cut severs (s, p), pixel p belongs to the foreground. Accordingly, edge (s, p) is assigned a weight $\lambda D_p(\text{fg})$, which is the penalty of assigning "fg" to p. In a similar

TABLE I
EDGE WEIGHTS IN THE GRAPH

Edge type	Weight	Condition
n-link (p, q)	$(1-\lambda)V_{\{p,q\}}$	$(p, q) \in N$
t-link (s, p)	$\lambda D_p(\text{fg})$	$p \in P, I_p^c \leq 2T$
	0	$p \in P, I_p^c > 2T$
t-link (p, t)	$\lambda D_p(\text{bg})$	$p \in P, I_p^c \leq 2T$
	W	$p \in P, I_p^c > 2T$

way, edge (p, t) is assigned a weight $\lambda D_p(\text{bg})$. Specially, when $I_p^C > 2T$, the weight of edge (s, p) is assigned zero because of the hard constraint condition, and the weight of edge (p, t) is assigned W, as shown in (10), which is a little greater than the maximal sum of all n-link weights for a pixel

$$W = 1 + \max_{p \in P} \sum_{q:\{p,q\} \in N} V_{\{p,q\}}. \tag{10}$$

3) Energy Minimization: A cut is a subset of edges $C \subset E$ that partitions the nodes into two disjoint parts: the foreground connecting to sink t and the background connecting to source s. The cost of a cut is defined as the sum of the edge weights that it severs

$$|C| = \sum_{e \in C} w_e. \tag{11}$$

The minimum cut problem on a graph is to find the cut that has the minimum cost.

In combinatorial optimization [51], the minimum cut problem can be solved by finding the maximum flow from source s to sink t. The maximum flow is the maximum "amount of water" that can be sent from s to t by regarding graph edges as "pipes" with edge weights equal to capacities [52]. The theorem of Ford and Fulkerson [53] states that the maximum flow from s to t saturates a set of edges in the graph to divide the nodes into two disjoint parts corresponding to the minimum cut. Thus, min-cut and max-flow problems are equivalent, and the maximum flow value is equal to the cost of the minimum cut.

There are two types of standard min-cut/max-flow algorithms: augmenting paths [53] and push–relabel [54], both of which are inefficient to implement as shown by the experiments in [52]. A new min-cut/max-flow algorithm that significantly outperforms the standard ones was presented in [52], which is adopted for energy minimization in this study.

C. Spatial Correspondence of Changed Objects

Multitemporal changed objects are generated through cosegmentation and represented as the foreground segments. Because of the association of the change feature in the segmentation procedure, the spatial correspondence can be inherently yielded between changed objects and needs to be apparently established to reveal the object-to-object changes.

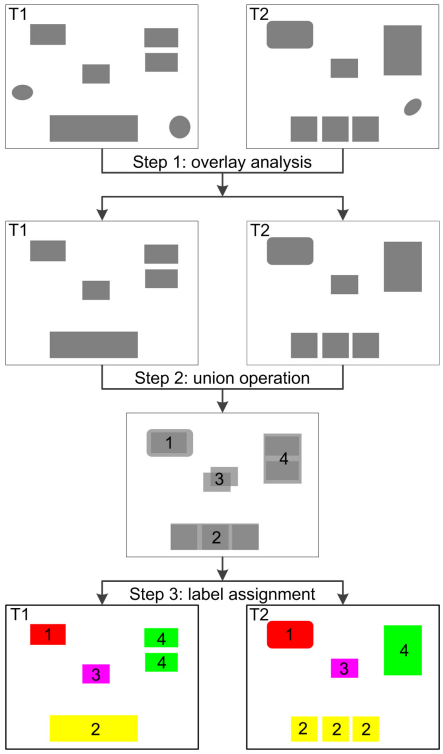

Fig. 3. Steps of establishing spatial correspondence between multitemporal changed objects.

1) Fragmentation Removal: Before establishing a spatial correspondence, we process the segmented objects to address the fragmentation that can occur. For instance, the heterogeneity of buildings may result in several small gaps within detected objects. Besides, small insignificant patches may be produced in the detection results because of confounding factors, such as different imaging angles, registration error, and image noise. This fragmentation of the results does not contribute to the establishment of a spatial correspondence between multitemporal changed objects. As a consequence, in order to make changed objects more complete and meaningful, we adopt a three-step fragmentation removal procedure: 1) a morphological closing operation with a square structural element of 3×3 pixels to fill gaps; 2) a morphological opening operation with a square structural element of 3×3 pixels for smoothing; and 3) an eliminating operation for small patches with a region area less than 100 m^2.

2) Correspondence Establishment: Using the fragmentation-free results, a spatial correspondence between changed objects is established according to the following three steps, as shown in Fig. 3.

Step 1: Performing an overlay analysis of the two maps and delete the objects of each map that have no overlaid objects in the other map. Deleted objects are generally false alarms, such as isolated pixels, resulting from shadows, obstruction, and image noise.

Step 2: Conducting a union operation on the two maps and assign labels for each object in the newly generated map.

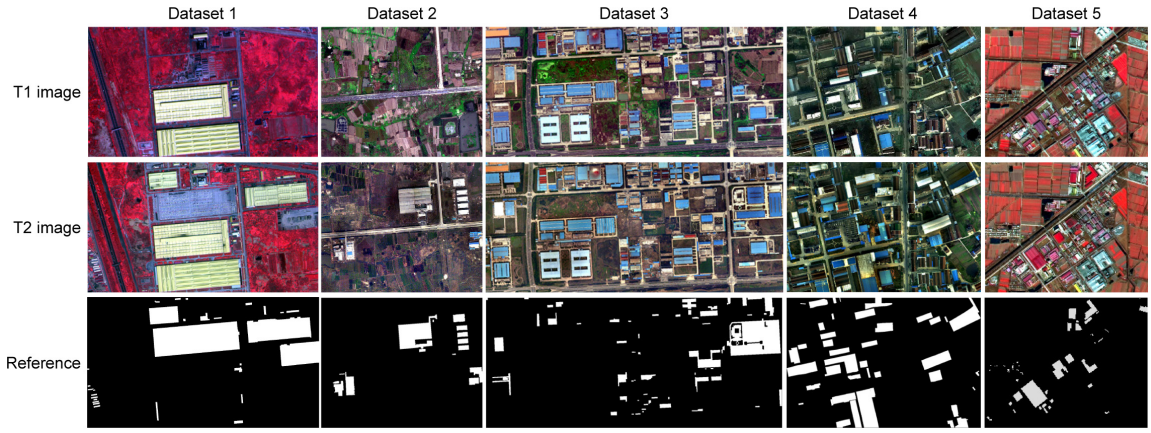

Fig. 4. Datasets and reference maps. The IKONOS images (dataset 1) and GF-1 images (dataset 5) are composited with false color and aerial photographs (datasets 2–4) are composited using true color. The first to third rows are T1 image, T2 image, and reference map, respectively.

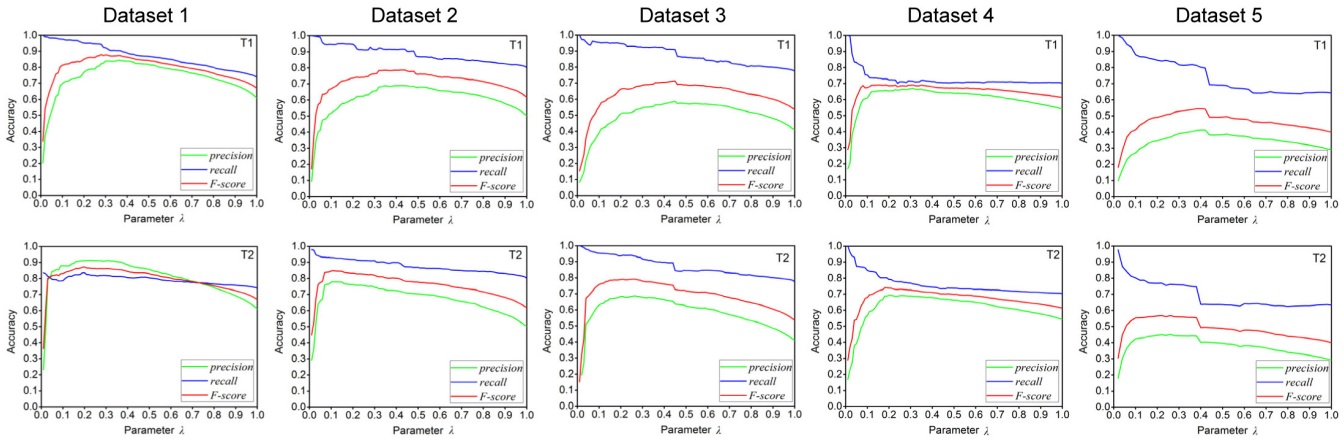

Fig. 5. Influence of parameter λ in the energy function on cosegmentation accuracy, i.e., *precision*, *recall*, and *F-score*.

Step 3: Assigning these labels to the corresponding objects of each map and thus establishing a spatial correspondence between the multitemporal changed objects.

Because image features are added to the energy function in the cosegmentation process, generated objects are directly related to specific images, allowing the geometric and numeric changes to be detected. For instance, as shown in Fig. 3, the expansion of an object is detectable by the change of boundaries at different phases, such as the red polygons. The displacement of an object will result in a detectable change of the object's position, such as the magenta polygons. In addition, the numeric changes of objects can also be exploited. One object in an earlier phase could be transformed to several objects at a later phase, such as the yellow polygons. This could happen when several buildings are newly built on a bare field. On the other hand, several objects in an earlier phase could be transformed to a single object at a later phase, such as the green polygons. This could happen when several buildings are demolished, and then, a new building is reconstructed at the same place.

III. Accuracy Assessment

This study assesses the accuracy by comparing the detection result with a manually delineated reference. Overlaying the change detection result with the reference map, we can calculate three measures, *precision*, *recall*, and *F-score*, which are defined as

$$precision = \frac{tp}{tp + fp} \tag{12}$$

$$recall = \frac{tp}{tp + fn} \tag{13}$$

$$F\text{-}score = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{14}$$

where tp is the number of changed pixels that are correctly detected, fp is the number of unchanged pixels that are incorrectly detected as change, and fn is the number of changed pixels that are incorrectly detected as unchanged. *Precision* is related to false alarms, and a high *precision* value indicates a small number of false alarms. *Recall* is related to missed detections, and a high *recall* value indicates a small number of missed detections. *F-score* is evenly balanced between false alarms and missed detections, revealing the overall detection performance.

TABLE II
LIST OF DATASETS

Dataset	Platform	Size (pixel)	Spatial resolution (m)	Available band	Year (T1)	Year (T2)	Location	Number of changed buildings	Number of pixels of changed buildings
1	IKONOS	900 × 1600	1	NIR, R, G, B	2000	2009	Jiangning District, Nanjing, China	29	249,963
2	Aerial	800 × 1000	1	R, G, B	2008	2012	Jintan City, Changzhou, China	14	56,955
3	Aerial	700 × 1600	1	R, G, B	2008	2012	Jintan City, Changzhou, China	57	91,844
4	Aerial	800 × 1200	0.5	R, G, B	2008	2012	Xinbei District, Changzhou, China	39	161,883
5	GF-1 PMS	800 × 1000	2	NIR, R, G, B	2013	2016	Pizhou City, Xuzhou, China	44	43,412

In addition to the above-mentioned three measures based on pixels, we adopt two measures based on geometric properties to evaluate the edges and positions of changed objects. In the reference map, the set of r reference objects is defined as $R = \{R_1, R_2, \ldots, R_r\}$. The set of s changed objects in the detection map is defined as $O = \{O_1, O_2, \ldots, O_s\}$. In total, there are d pairs of matched reference objects and changed objects $\{(R_i, O_i)|i = 1, 2, \ldots, d\}$. Given a pair (R_i, O_i), we can calculate local measures $accur_i^h$ where h indicates the measure of *edge* similarity or *position* similarity as 1 and 2, respectively.

Edge similarity measures the accuracy of the edges of the changed objects with regard to those of the reference objects. Considering a tolerance in the recognition of object edges, we use an operator $e(\cdot)$ that extracts edge pixels of objects with a width of five pixels [55]. This measure is defined as

$$edge = \frac{|e(R_i) \cap e(O_i)|}{|e(R_i)|}. \tag{15}$$

This measure calculates the ratio between the overlapping area of the edges of (R_i, O_i) and the edge area of the reference object. The value is equal to 1 if all edges of O_i and R_i are coincident.

Position similarity refers to the similarity of the centroid position between O_i and R_i, which is defined as

$$position = 1 - \frac{d_{\text{cent}}(R_i, O_i)}{d_{\text{cac}}(R_i \cup O_i)}. \tag{16}$$

This measure is evaluated by first computing the Euclidean distance between the centroids of O_i and R_i (cent) and then normalizing by the diameter of a combined area circle, i.e., a circle whose area is equal to the sum of O_i and R_i [56]. The value is 1 in the optimal case where there is no offset between the centroids of O_i and R_i.

The global measure can be calculated by averaging d local measures, which is expressed as

$$accur^h = \frac{1}{d} \sum_{i=1} accur_i^h. \tag{17}$$

Considering that the proposed method produces two change detection maps, ideally we need to produce two reference maps to evaluate the two detection results separately. The geometric and numeric changes can then be quantitatively evaluated. However, in order to compare with other OBCD methods, we only produce a single reference map for accuracy assessment. We assess the results of the proposed method in three ways based on the aforementioned measures: 1) comparing the T1 change detection result with the reference map; 2) comparing the T2 change detection result with the reference map; and 3) jointly comparing T1 and T2 change detection results with the reference map. For 3), we calculate the values of the five measures (*precision*, *recall*, *F-score*, *edge*, and *position*) of both the T1 and T2 changed objects, and then choose the maximum value as the local measure for every pair of spatially corresponding changed objects. After that, the global measure is calculated by averaging the local measures, as shown in (17).

IV. EXPERIMENTAL RESULTS

In the experiment, five HR remotely sensed datasets are used to show the effectiveness of the cosegmentation method. We first analyze the sensitivity of parameters T and λ in the energy function. Then, we analyze the influence of misregistration on cosegmentation results. Next, we show multitemporal detection results and evaluate the accuracy of the proposed method. Finally, we compare the proposed method with the state-of-the-art OBCD methods both quantitatively and qualitatively.

A. Study Area and DataSets

The study area is located in Jiangsu Province, China. As part of the Yangtze River Delta Econimic Zone, Jiangsu has experienced extensive development over the past years, which has led to rapid clustering and growth of emerging industries. Correspondingly, the land uses in the cities have undergone significant changes, with rapid increases in the number of buildings used for industrial and residential purposes. Within the study area, five regions are selected to show the change detection results. Correspondingly, five datasets, including a pair of IKONOS images, three pairs of aerial photographs, and a pair of GF-1 images are used in the experiment, as shown in Fig. 4. GF-1 was the first satellite of the China's HR Earth

Observation System launched in 2013. There are four wide-field-of-view sensors and two panchromatic and multispectral (PMS) sensors aboard on GF-1. The details of datasets are listed in Table II. Dataset 1 is composed of multitemporal IKONOS images in Nanjing City. The changes mainly come from newly built buildings and a few buildings that are under construction. Dataset 5 is a set of GF-1 PMS images in Xuzhou City. The main changes are composed of rebuilt buildings and a few newly built ones. The two datasets were preprocessed in three steps: 1) pan-sharpening the multispectral bands by the Gram–Schmidt method; 2) geometric registration by a polynomial function based on 25 ground control points where the residual misregistration is less than 0.5 pixels; and 3) radiometric correction using pseudo-invariant features [57]. The three datasets in Changzhou City consist of multitemporal aerial photographs. The main changes consist of newly built buildings and a few rebuilt ones. In the preprocessing stage, the geometric registration and radiometric correction were applied, similar to those for dataset 1 and dataset 5.

B. Parameter Analysis

Parameter λ in (2) is a weight coefficient that measures the relative importance of the change feature versus the image features in the energy function. A larger value of λ indicates more consideration of the change feature and a smaller λ indicates more consideration of the image features in the cosegmentation process. We set λ from 0.01 to 1 in the increments of 0.01. When λ is equal to 1, cosegmentation entirely depends on the change feature and degrades into the threshold-based change detection method at pixel level. However, λ cannot be set to 0, because changes will not be detected without the guidance of the change feature.

The influence of λ on the *precision*, *recall*, and *F-score* values on cosegmentation results is shown in Fig. 5. It can be seen that the general influence of the five datasets is similar. As λ increases, *precision* increases quickly and then decreases slowly, while *recall* experiences a gradual descending trend. *F-score* also increases rapidly and then decreases slowly with the increment of λ, which shows a tradeoff between the change feature and image features. The balance point is strictly reached at which the *F-score* achieves its highest value. The optimal λ of T2 image tends to be smaller than that of T1 image. This may be attributed to the newly built buildings in the image. On the one hand, changed regions in T1 image commonly have image features that are similar to their adjacent unchanged regions so that a larger weight of the change feature is required to segment these changed regions. On the other hand, newly built buildings in T2 image may be internally heterogeneous and a larger weight is needed for image features in order to make generated objects complete.

Taking dataset 1 as an example, local details of cosegmentation results with different λ values are enlarged in Fig. 6 to show the differences. Because of the guidance of the same change feature in the process of cosegmentation, changes are almost reflected in both results. However, due to different image features of each image, the two results are distinct

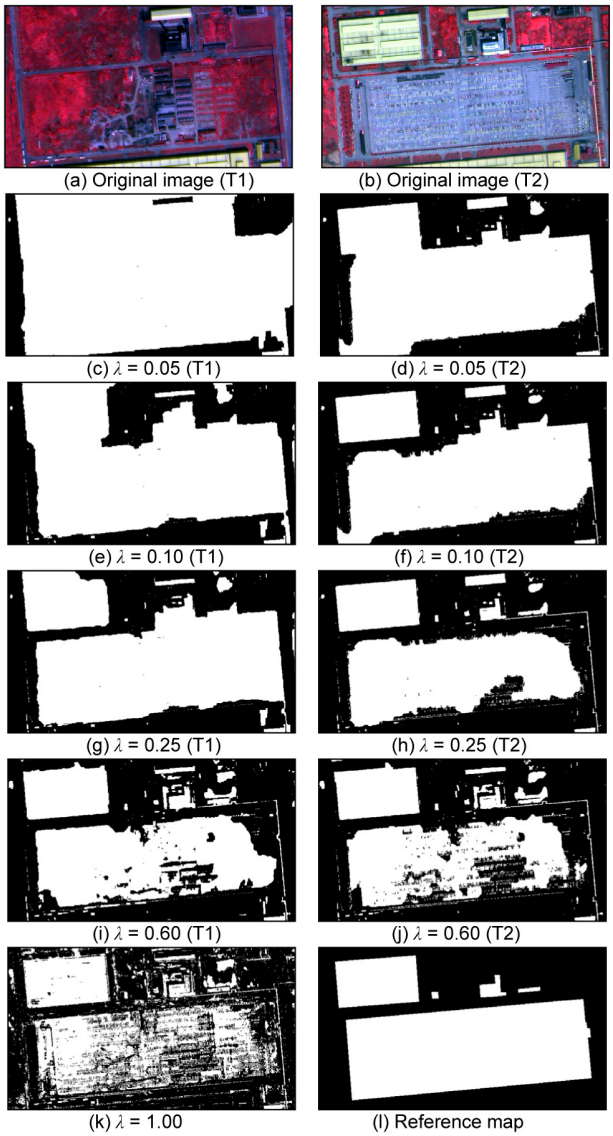

Fig. 6. Local details of cosegmentation maps from dataset 1 for different λ values. (a) Original image (T1). (b) Original image (T2). (c) $\lambda = 0.05$ (T1). (d) $\lambda = 0.05$ (T2). (e) $\lambda = 0.10$ (T1). (f) $\lambda = 0.10$ (T2). (g) $\lambda = 0.25$ (T1). (h) $\lambda = 0.25$ (T2). (i) $\lambda = 0.60$ (T1). (j) $\lambda = 0.60$ (T2). (k) $\lambda = 1.00$. (l) Reference map.

with each other even if the same parameter λ is applied. Specially, cosegmentation produces the same result when λ is equal to 1, since image features are not involved, as shown in Fig. 6(k). In general, if λ is large, the change feature plays a relatively important role in the energy function and pixels with larger I_p^C are easily divided into changed regions. Thus, detected change information tends to be trivial and incomplete, as shown in Fig. 6(i)–(k). On the contrary, if λ is small, image features play an important role in cosegmentation and adjacent regions tend to be merged into a larger one, resulting in too large changed regions, as shown in Fig. 6(c) and (d). The results in Fig. 6 exactly correspond to the quantitative influence of parameter λ in Fig. 5. When λ is set too small, e.g., close to 0, generated large changed regions signify a large number of false alarms and a small number of missed detections, leading to a very small *precision* value and a very

TABLE III
WEIGHT PARAMETER λ IN THE ENERGY FUNCTION

λ	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
T1 image	0.3	0.4	0.4	0.3	0.4
T2 image	0.2	0.1	0.2	0.2	0.2

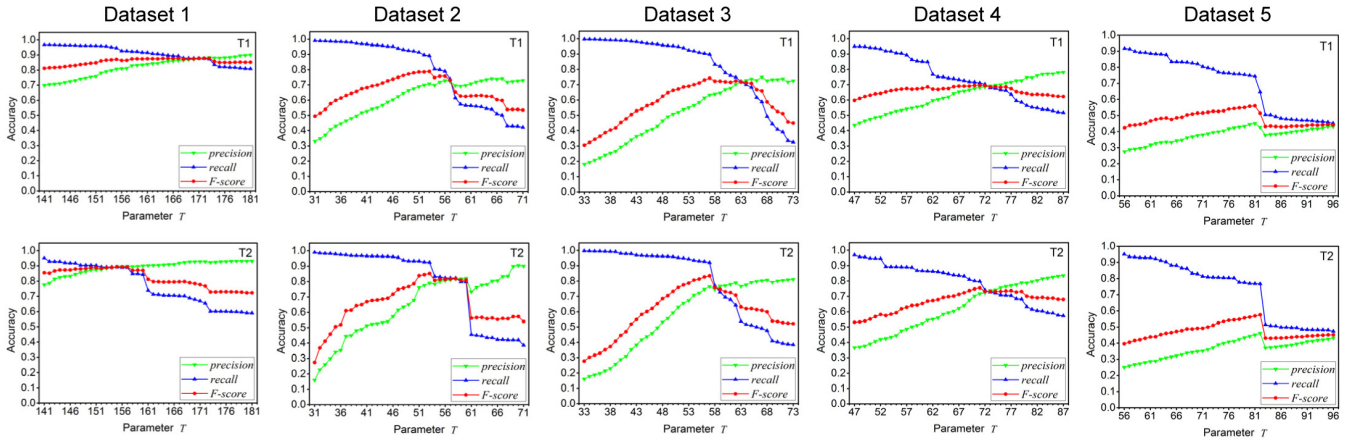

Fig. 7. Influence of parameter T in energy function on cosegmentation accuracy, i.e., *precision*, *recall*, and *F-score*, given the parameter λ as in Table III for each image.

large *recall* value. When λ is set too large, the changed regions would be fragmented and incomplete, resulting in the decrease of F-score. Hence, λ cannot be set too large or too small. In a real application, no reference data are provided and it is difficult to precisely determine the optimal λ. However, it can be seen from Fig. 5 that *F-score* is relatively stable within a certain range near the highest point, which may provide some reference for the setting of parameter λ. For the main changes come from newly built buildings, we recommend to set λ for T1 and T2 images within the ranges [0.3, 0.4] and [0.1, 0.2], respectively. The weight parameter λ in this study is given in Table III.

With regard to the threshold parameter T in (4), unlike the "hard" threshold to distinguish foreground regions from background, it serves as "soft" threshold only to determine the constrain condition, and then, the foreground regions are segmented by combining the change feature and image features. The "soft" threshold T may have lower sensitivity to generated foreground segments, but it still influences the cosegmentation result, because too many foreground segments would be produced by a small T and, on the contrary, too few foreground segments by a large T. Accordingly, the segmentation accuracy increases first, achieves the highest point for the optimal T, and then decreases as the increment of T value. Hence, the parameter T should be set neither too large nor too small.

We calculate the optimal T based on the expectation–maximization algorithm and the Bayesian theory, which is 161, 51, 53, 67, and 76 for datasets 1–5, respectively. In order to show the sensitivity of the threshold, we vary it within the range $[T-20,\ T+20]$ in an increment of 1, which covers the highest point, increasing range, and decreasing range of segmentation accuracy. Given the parameter λ in Table III, the influence of T on *precision*, *recall*, and *F-score* values for cosegmentation results is shown in Fig. 7. On the one hand, it can be seen that even though the calculated T does not exactly achieve the highest *F-score* in the curve, it is very near to the highest point, showing the effectiveness of the calculation of T. On the other hand, the *F-score* curves for datasets 1, 4, and 5 are near flat, and there are few fluctuations, indicating that cosegmentation results are not highly sensitive to threshold parameter T for these datasets. This would be caused by the effectiveness of "soft" threshold. While for datasets 2 and 3, there exists relatively larger *F-score* fluctuations than the other three datasets. It is then shown in Fig. 8 to lower the sensitivity of T by setting optimal λ values rather than a given λ value for each T. We can visually see the smaller *F-score* fluctuations by setting optimal λ values for each T, and the decreases of the standard deviation of the *F-score* values (*STD_F*) further demonstrate this.

C. Misregistration Analysis

Multitemporal image registration is an essential preprocessing for change detection [58]. The accuracy of image registration is typically related to two factors: image spatial resolution and the structure of geographic objects of interest [17]. Misregistration usually results in a decrease of change detection accuracy. In the experiments of [59], the registration error of less than one-fifth of a pixel was required in order to detect 90% of true changes. However, it is difficult to perform an accurate image registration in practical work. Thus,

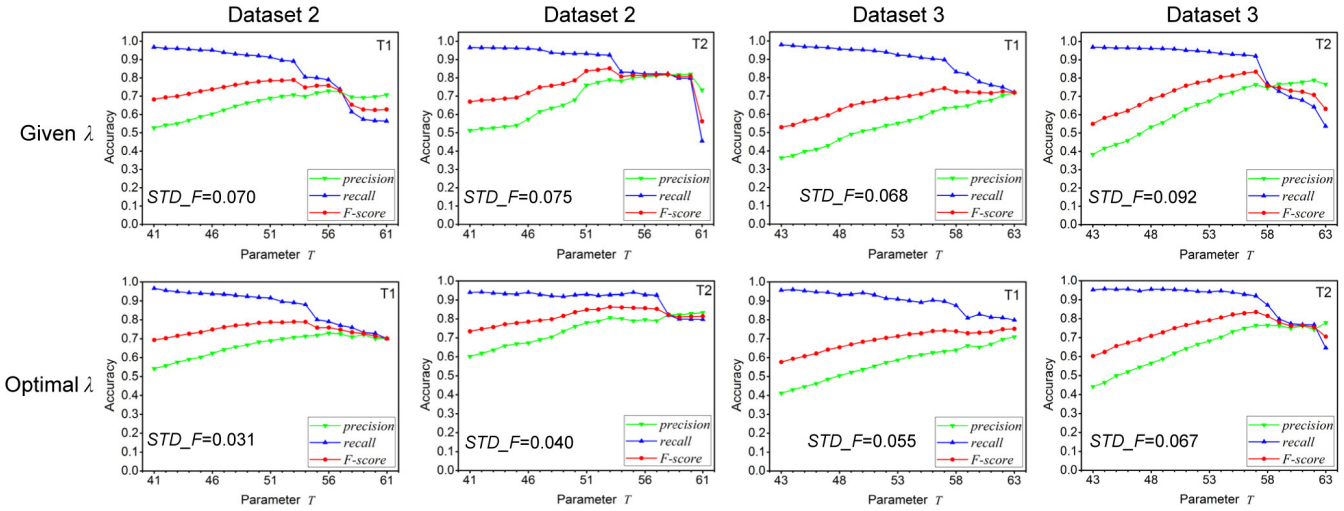

Fig. 8. Examples of lowering the sensitivity of parameter T in energy function on cosegmentation accuracy, i.e., *precision*, *recall*, and *F-score*, by giving the optimal parameter λ for each T value. The *STD_F* is the standard deviation of the *F-score* values, indicating the fluctuation range of accuracies.

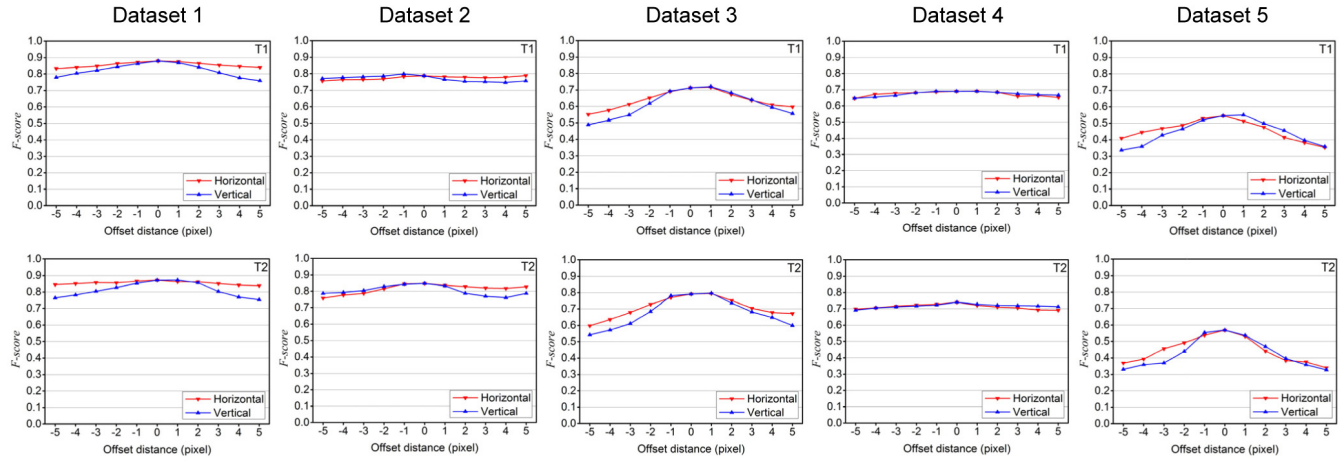

Fig. 9. Influence of misregistration on cosegmentation accuracy *F-score*. Multitemporal images that have been registered are overlapped, one of which keeps unmoved and the other of which is shifted manually pixel by pixel along the horizontal and vertical directions, respectively.

it is necessary to analyze the influence of misregistration on cosegmentation results.

The misregistration analysis results are presented in Fig. 9. The horizontal axis is the offset distance, representing registration errors between multitemporal images. When the offset distance is 0, it is the case where multitemporal images are geometrically registered. It can be seen that registration errors have little influence on cosegmentation results for datasets 2 and 4, while have relatively greater influence for datasets 1, 3, and 5, since the *F-score* decreases slowly as the offset distance increases. The difference is because there are several small objects in the references of datasets 1, 3, and 5, as shown in Fig. 4, which are easier to be affected by registration error than large objects. However, it can be seen that the registration error of one pixel has little effect on cosegmentation accuracy for all the five datasets, through which we conclude that the proposed cosegmentation-based change detection does not require very high precision of registration and can accept an error of at least one pixel.

D. Accuracy Assessment

Multitemporal detection results by conducting fragmentation removal and spatial correspondence establishment based on cosegmentation results are shown in Fig. 10. Spatially corresponded changed objects are labeled with the same color. In addition to showing the thematic changes of buildings, the cosegmentation-based detection results can provide more change information, e.g., geometric and numeric changes of objects. In Fig. 11, local details of multitemporal detection results are enlarged to show the geometric and numeric changes clearly. In Fig. 11(a), we can intuitively see that one object changes into another object. The object in T2 result has a more regular shape than that in T1 result, because the newly built building in T2 image has obvious boundary that can be distinguished from adjacent unchanged regions. Although the building is internally heterogeneous, the detection result is absolutely complete. This is because the cosegmentation comprehensively considers both change and

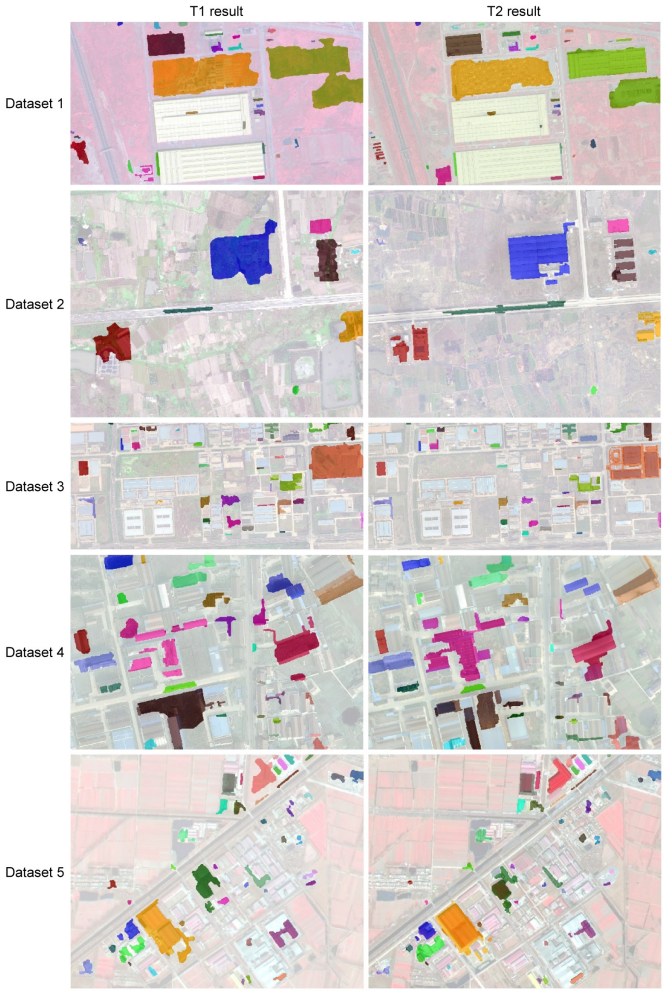

Fig. 10. Multitemporal change detection results by conducting fragmentation removal and establishing spatial correspondence between multitemporal changed objects based on cosegmentation. The spatially corresponded change objects between two phases are marked with the same color.

image features, which can not only discover changed regions, but also make the detected objects complete and significant. In Fig. 11(b), where there are four newly built buildings in T2 image, we can see a pattern where one object transforms to four objects. Regions with large I_p^C in T1 image have similar image features to adjacent regions, while regions with large I_p^C in T2 image have quite different image features from adjacent regions. Therefore, different image features make the cosegmentation produce different detection results, allowing the comparison of one object to multiple objects. Fig. 11(c) is an instance of the demolition and reconstruction of buildings. Buildings in T1 and T2 image are distinct in size, shape, and position.

The accuracies of the multitemporal detection results with and without fragmentation removal procedure are reported in Table IV. It can be seen that the *precision*, *F-score*, and *position* values with fragmentation removal procedure are higher than those without that procedure, which indicates that the post-processing removes several false alarms and improve the position accuracy of changed objects. However, fragmentation removal may eliminate some small changed objects, which

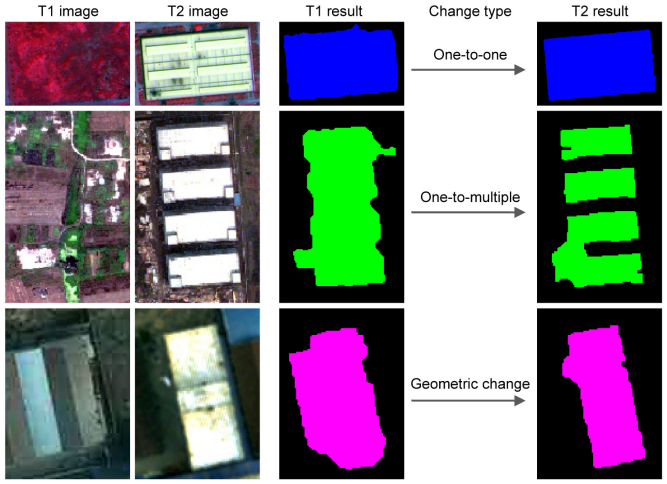

Fig. 11. Local details of multitemporal change detection results in Fig. 10 showing numeric and geometric changes. The presented regions are from datasets 1, 2, and 4 from top to bottom, respectively.

leads to a small decrease in the *recall* value. In addition, morphological operations may slightly decrease the *edge* values of changed objects. We can also see that *precision* values are generally lower than *recall* values for all datasets, which shows that most true changes have been detected, and at the same time, a number of patches are falsely detected. For every dataset, we can conclude that T1 assessment way results in the lowest accuracies, and T1 and T2 assessment way leads to the highest accuracies.

E. Comparison With Other Methods

In order to further validate the effectiveness of the proposed cosegmentation method, we compare the results with those yielded by several state-of-the-art object-based change vector analysis (OBCVA) [60], [61] methods by taking different multitemporal segmentation strategies. To make a fair comparison, the same spectral and MBI features are used. Since both the proposed method and the comparative OBCVA methods take change vector to find changes, then the differences of detection results would mainly due to the multitemporal segmentation methods. For OBCVA, the image segmentation is usually performed in the following four ways (see Fig. 12) to extract multitemporal image objects: 1) segmenting a stack image consisting of all bands of both images (OBCVA_Com); 2) just segmenting T1 image and then assigning boundaries to T2 image (OBCVA_T1); 3) just segmenting T2 image and then assigning boundaries to T1 image (OBCVA_T2); and 4) separately segmenting the two images and then using the raster overlay function [13], [31] to connect and compare image objects (OBCVA_Sep). In order to further show the difference caused by cosegmentation, we add other two segmentation strategies for OBCVA. One is to segment a stack image consisting of T1 image and the change feature used in cosegmentation (OBCVA_T1+C), and the other is to segment a stack image consisting of T2 image and the change feature (OBCVA_T2+C). Then, the inputs of OBCVA_T1+C and OBCVA_T2+C are exactly the same as that of cosegmen-

TABLE IV
ACCURACY OF MULTITEMPORAL CHANGE DETECTION RESULTS

Dataset	Assessment way	Fragmentation removal (with/without)	Accuracy				
			precision	*recall*	*F-score*	*edge*	*position*
1	T1	Without	0.82	0.94	0.88	0.61	0.86
		With	0.84	0.94	0.89	0.64	0.88
	T2	Without	0.90	0.89	0.89	0.84	0.89
		With	0.92	0.88	0.90	0.84	0.94
	T1 & T2	Without	0.90	0.95	0.93	0.84	0.90
		With	0.93	0.94	0.94	0.85	0.94
2	T1	Without	0.69	0.92	0.79	0.44	0.84
		With	0.73	0.93	0.81	0.50	0.86
	T2	Without	0.78	0.93	0.85	0.87	0.93
		With	0.83	0.92	0.87	0.85	0.94
	T1 & T2	Without	0.87	0.93	0.90	0.87	0.93
		With	0.91	0.92	0.92	0.85	0.94
3	T1	Without	0.59	0.91	0.71	0.67	0.79
		With	0.63	0.90	0.74	0.65	0.82
	T2	Without	0.68	0.94	0.79	0.80	0.85
		With	0.79	0.92	0.85	0.82	0.91
	T1 & T2	Without	0.75	0.94	0.83	0.86	0.84
		With	0.83	0.91	0.87	0.85	0.91
4	T1	Without	0.65	0.73	0.69	0.51	0.78
		With	0.69	0.73	0.71	0.53	0.86
	T2	Without	0.69	0.80	0.74	0.70	0.84
		With	0.72	0.80	0.76	0.71	0.90
	T1 & T2	Without	0.72	0.83	0.77	0.70	0.86
		With	0.75	0.80	0.78	0.73	0.92
5	T1	Without	0.41	0.80	0.55	0.67	0.78
		With	0.54	0.77	0.64	0.59	0.79
	T2	Without	0.46	0.77	0.58	0.70	0.79
		With	0.56	0.74	0.63	0.57	0.76
	T1 & T2	Without	0.49	0.81	0.62	0.73	0.80
		With	0.62	0.78	0.69	0.64	0.83

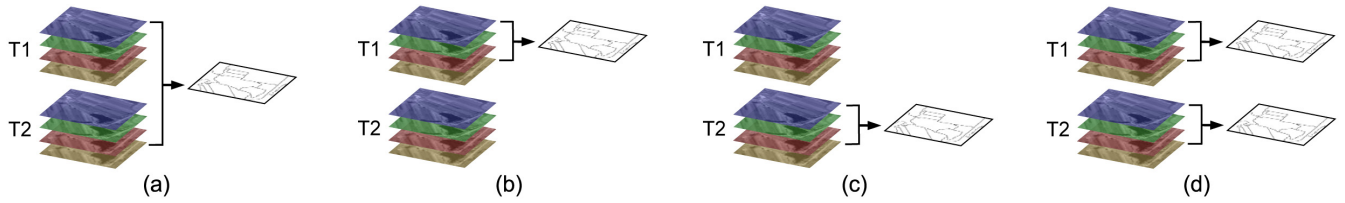

Fig. 12. Multitemporal segmentation strategies for comparison (adapted from [6]). (a) OBCVA_Com. (b) OBCVA_T1. (c) OBCVA_T2. (d) OBCVA_Sep.

tation; the differences between them would clearly show the effectiveness of cosegmentation. The widely used multiresolution segmentation technique [62] embedded in the commercial software eCognition is adopted for the comparative OBCVA methods. The scale and shape parameter are set as (50, 0.4), (30, 0.3), (30, 0.4), (30, 0.5), and (30, 0.4) for datasets 1–5, respectively. For comparative methods, change thresholds are determined by the Bayesian theory based on an expectation–maximization algorithm [49] and detection results undergo the same fragmentation removal procedure as the proposed method.

The quantitative comparison of change detection results produced by different methods is given in Fig. 13. It is clear that accuracies yielded by the proposed method (Coseg) are overall higher than those obtained by other methods for all datasets, especially for the *F-score* of Coseg_T2 and Coseg_T1&T2. The differences between Coseg and OBCVA_Ti+C (i = 1, 2) further demonstrate the superiority of cosegmentation, because the input features are the same for these two strategies. Meanwhile, there are differences among the results obtained by these comparative OBCVA methods because of the different multitemporal segmentation strategies. It can

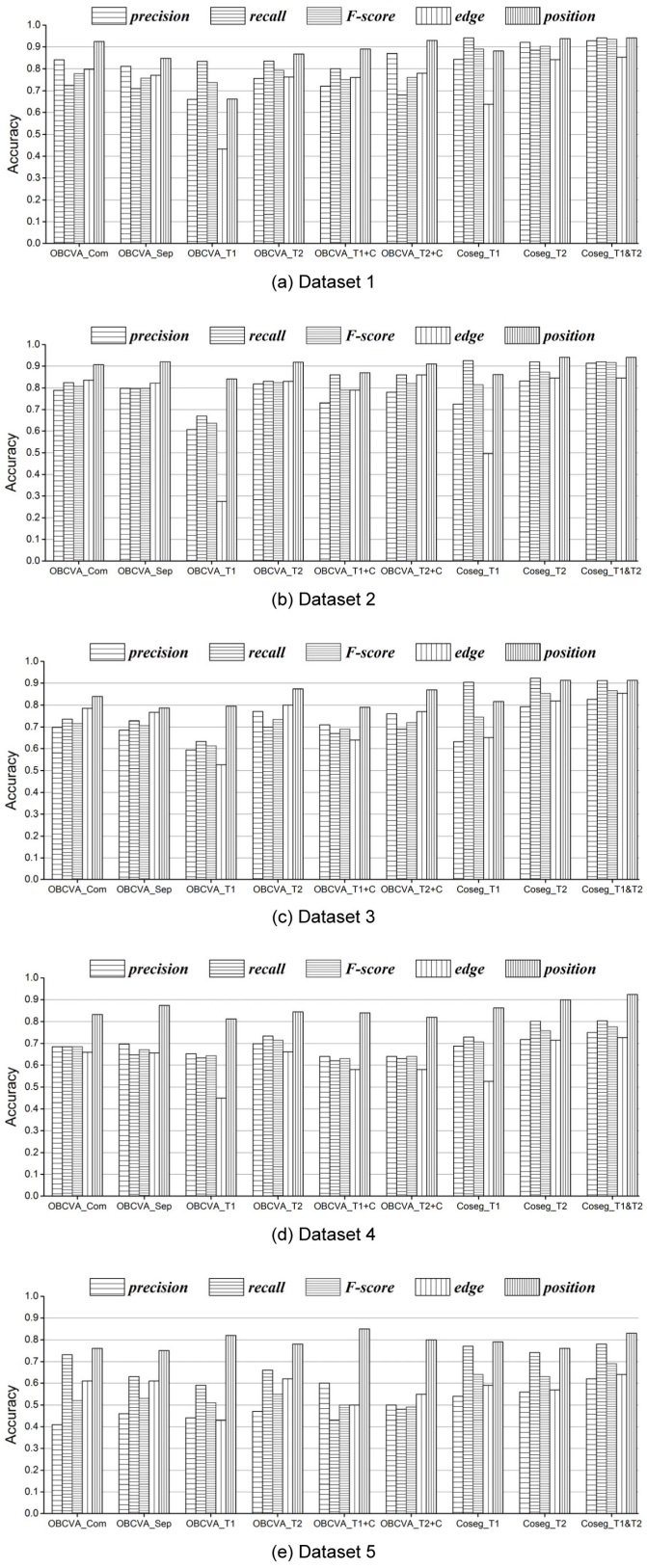

Fig. 13. Quantitative comparison of change detection accuracies for different methods on five datasets. (a) Dataset 1. (b) Dataset 2. (c) Dataset 3. (d) Dataset 4. (e) Dataset 5.

be seen that OBCVA_Com and OBCVA_Sep yield similar accuracies, which may be due to the image features at different phases that are involved in segmentation for both of them. The OBCVA_T2 method also has similar performance with OBCVA_Com and OBCVA_Sep, while outperforms the OBCVA_T1 method. This can be explained by the fact that most changes are newly built buildings contained in T2 image with obvious boundaries, and then, a segmentation based on this image can generate preferable object boundaries that are conducive to exact building detection. Therefore, the single image segmentation strategy needs to properly select T1 or T2 image for segmentation. By comparing OBCVA_Ti with OBCVA_Ti+C, we can see that OBCVA_T2 and OBCVA_T2+C have similar performance, while OBCVA_T1+C outperforms OBCVA_T1 especially for the *edge* measure; this is because the addition of change feature can help to separate the changed regions from the adjacent unchanged regions in T1 image.

The superiority of the proposed method with respect to other methods can be assessed also from a qualitative viewpoint by comparing the building change detection maps shown in Fig. 14. It can be seen that the maps obtained by the proposed method show a significantly lower number of false alarms, such as the region enclosed by the yellow line in dataset 1. Furthermore, the proposed method obtains a more complete representation of buildings than other methods, such as the example enclosed by the red line. This depends on the capacity of the "soft" threshold parameter and the addition of image features in cosegmentation. In particular, thresholds of these comparative methods are "hard" thresholds, and the detection result is entirely based on the change feature. Nevertheless, the threshold of the proposed method plays the role of a "soft" threshold and the cosegmentation depends on both the change feature and image features. That is to say, for the proposed method, pixels satisfying $I_p^C < T$ are still likely to be detected as change, because image features are considered to improve the completeness and significance of changed objects. However, the proposed method tends to expand false alarms because of the false hints in the difference image, such as the instance enclosed by a green line in Coseg_T2 result of dataset 1. As a whole, the qualitative comparison shows that the proposed method achieved the most similar results to the reference maps for all datasets.

V. Discussion

In the proposed method, cosegmentation, as defined in the field of computer vision, is introduced to provide a new solution to OBCD from multitemporal HR remotely sensed images. Cosegmentation is performed via graph-based energy minimization. The min-cut/max-flow algorithm is used to solve the global energy minimization problem. The energy is made up of two terms: 1) the change feature represented by the magnitude of the difference image and 2) the image features based on original spectral features of each image. By combining both the change feature and image features, cosegmentation-based change detection is a "concurrent segmentation and detection" pattern instead of the "sequential segmentation and detection" pattern used in general OBCD methods.

The change feature potential in the energy function serves as the prior information and association for cosegmentation.

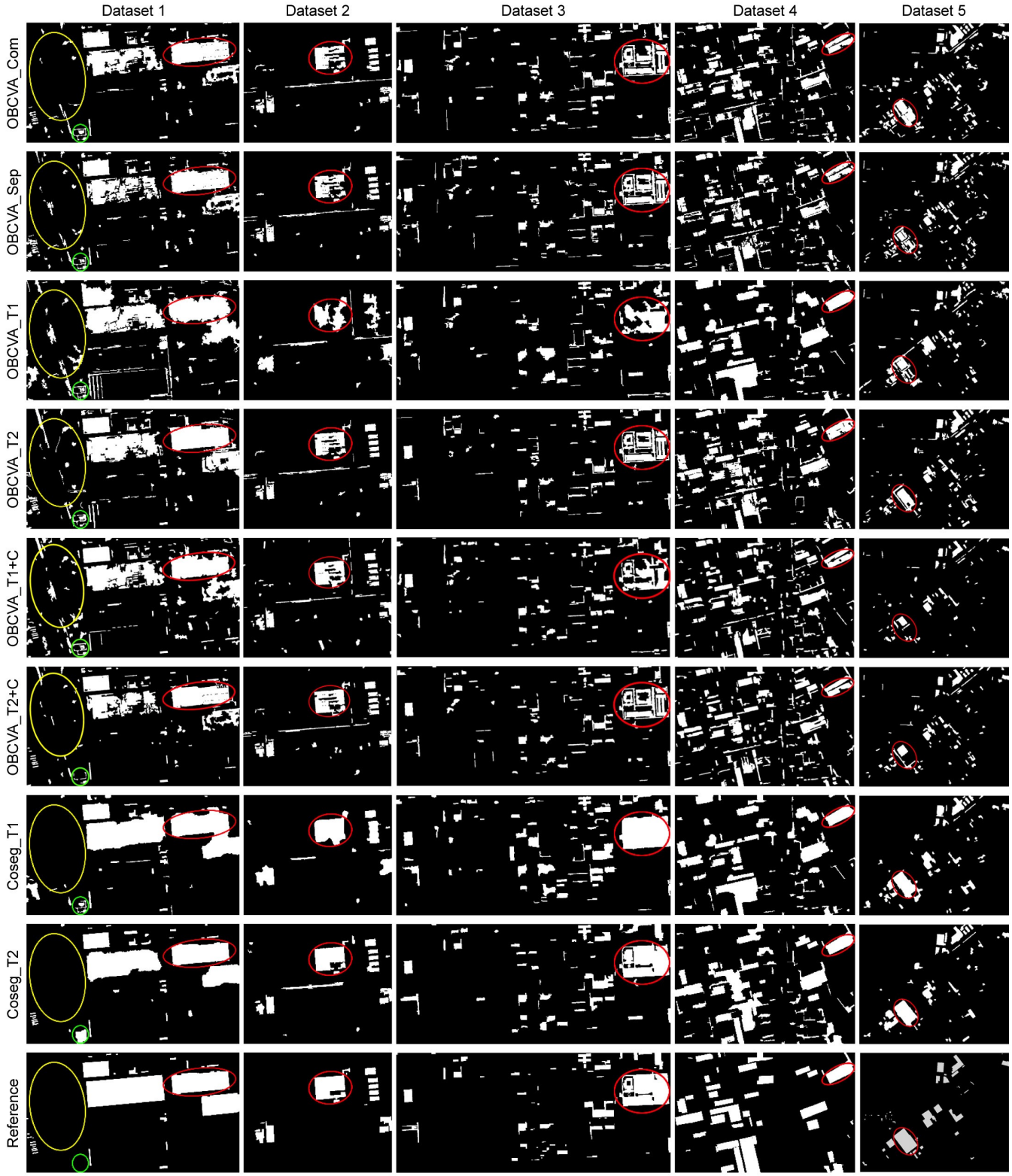

Fig. 14. Qualitative comparison of building change detection maps for different methods and corresponding reference maps on five datasets.

It provides the same clues about building changes during cosegmentation for each image in order that the segmentation procedure especially focuses on potential change regions. Therefore, cosegmentation can directly generate foreground as multitemporal changed objects and the spatial correspondence between the changed objects can be naturally yielded. Furthermore, the unchanged objects are represented as background in the cosegmentation results, which addresses the problem of discrepancies between unchanged objects in OBCD.

The image features are added to the energy function to generate complete and significant objects at each phase. Thus, the cosegmentation-based method can make full use of multitemporal information and produce two detection maps closely related to specific images at different phases, having the ability to reveal the attribute, geometric, and numeric changes of objects.

With regard to threshold parameter T, we determine it by the Bayesian theory based on an expectation–maximization algorithm. The threshold in this study plays the role of a "soft" threshold, which means that the detection does not entirely depend on the threshold. That is to say, pixels satisfying $I_p^C < T$ are still likely to be detected as change, because image features are considered to generate complete objects. Fig. 7 shows the sensitivity of T on cosegmentation results, which has low sensitivity to certain images. The sensitivity of T can be further lowered by optimizing the parameter λ.

Coefficient parameter λ in the energy function measures the relative importance of the change feature and image features. If λ is large, the change feature plays a relatively important role in cosegmentation and detected change information tends to be fragmented. On the contrary, if λ is small, image features play a dominant role, which results in change regions that are too large. Hence, λ should not be set too large or too small. Ideally, it should be set according to specific images and main change types. If the main changes come from newly built buildings, we recommend setting λ for T1 and T2 images in the [0.3, 0.4] and [0.1, 0.2] ranges, respectively.

The proposed cosegmentation-based change detection does not require highly precise registration between multitemporal images and can accept a registration error of at least one pixel. The low requirements of image registration can be attributed to the performance of the graph cut algorithm. We also found that small objects are more sensitive to misregistration than large objects, as shown in Fig. 9. It would be interesting to do a specific analysis on evaluating the influences of misregistration to objects of different sizes in the future.

In Figs. 13 and 14, cosegmentation shows better performance than other multitemporal segmentation strategies for OBCD. This may be because cosegmentation limits the scope to only those potentially changed areas to directly produce changed and unchanged objects. While other multitemporal segmentation strategies has the scope to the full image to generate image objects for further change analysis, this could increase the introduction of new errors compared with the limited scope of cosegmentation. A similar idea can be found in [32] by updating and backdating an existing land cover map with the scope of changed regions.

A significant advantage of cosegmentation is to provide two spatially corresponded changed maps to reflect the geometric and numeric changes of objects, which are justified by both theoretical analysis in Fig. 3 and real detection results in Figs. 10 and 11. As a future work, we can exploit the geometric and numeric changes to verify the spatially corresponded changed objects, which could further improve the detection accuracies. For the numeric change of objects, there may be a special case of multiple-to-multiple change of buildings. In this case, the proposed method of establishing spatial correspondence can also work by taking the multiple objects as a whole, as shown in Fig. 15. But it needs future work to distinguish the one-to-one correspondence within this multiple-to-multiple case.

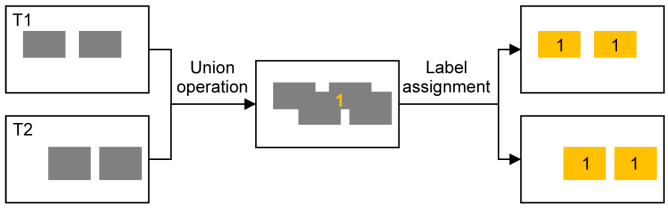

Fig. 15. Example of establishing the spatial correspondence between multiple-to-multiple changed objects.

The disadvantage is that the proposed method tends to expand false alarms because of the false hints in the difference image. Therefore, it is important to acquire correct change feature in order to provide accurate clues as to what the user intends to segment. Moreover, we focus on the change detection of buildings and obtain satisfactory results in this study. We consider that the cosegmentation method can be applied to change detection of other land cover types by inputting the specific change feature and then substituted for that of buildings.

VI. Conclusion

A novel method for building change detection from multitemporal HR remotely sensed images has been proposed based on the concept of cosegmentation. Five datasets, including a pair of IKONOS images, three pairs of aerial photographs, and a pair of GF-1 images, were used to verify the effectiveness and show the superiority of the proposed method.

In contrast to the conventional change detection methods, the proposed method results in two change detection maps with both connection and difference. The connection between the two maps results from the guidance of the same change feature in the cosegmentation process and the difference is due to specific image features at each phase added to the energy function. Comprehensive change information, i.e., attribute, geometric, and numeric changes of objects, can be exploited by the proposed method.

The proposed method achieves more satisfactory results than the state-of-the-art OBCD methods with respect to both quantitative accuracy measures of *precision*, *recall*, *F-score*, *edge*, *position*, and qualitative analysis in the experiment. This could be owing to that cosegmentation limits the scope to potentially changed areas to directly generate changed objects, while other multitemporal segmentation methods has the scope of the whole image to generate image objects for change analysis, which would increase the introduction of new errors.

As future developments of this study, we are considering the following: 1) attempting to detect changes of other land cover types (e.g., road, lake, and vegetation) using the cosegmentation-based method; 2) developing a robust algorithm to better delineate the change feature in order that the difference image can provide accurate hints for cosegmentation; and 3) exploiting the geometric and numeric changes to further verify the spatially corresponded changed objects.

REFERENCES

[1] A. Singh, "Digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989.

[2] C. He, A. Wei, P. Shi, Q. Zhang, and Y. Zhao, "Detecting land-use/land-cover change in rural–urban fringe areas using extended change-vector analysis," *Int. J. Appl. Earth Observat Geoinf.*, vol. 13, no. 4, pp. 572–585, Aug. 2011.

[3] L. D. Robertson and D. J. King, "Comparison of pixel- and object-based classification in land cover change mapping," *Int. J. Remote Sens.*, vol. 32, no. 6, pp. 1505–1529, Mar. 2011.

[4] D. Tiede, A. Wania, and P. Füreder, "Object-based change detection and classification improvement of time series analysis," in *Proc. Int. Conf. Geogr. Object-Based Image Anal. (GEOBIA)*, vol. 1. 2012, pp. 223–227.

[5] L. Durieux, E. Lagabrielle, and A. Nelson, "A method for monitoring building construction in urban sprawl areas using object-based analysis of SPOT 5 images and existing GIS data," *ISPRS J. Photogram. Remote Sens.*, vol. 63, no. 4, pp. 399–408, Jul. 2008.

[6] M. Bouziani, K. Goïta, and D.-C. He, "Automatic change detection of buildings in urban environment from very high spatial resolution images using existing geodatabase and prior knowledge," *ISPRS J. Photogram. Remote Sens.*, vol. 65, no. 1, pp. 143–153, Jan. 2010.

[7] B. Desclée, P. Bogaert, and P. Defourny, "Forest change detection by statistical object-based method," *Remote Sens. Environ.*, vol. 102, nos. 1–2, pp. 1–11, May 2006.

[8] S. Bontemps, P. Bogaert, N. Titeux, and P. Defourny, "An object-based change detection method accounting for temporal dependences in time series with medium to coarse spatial resolution," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 3181–3191, Jun. 2008.

[9] G. Conchedda, L. Durieux, and P. Mayaux, "An object-based method for mapping and change analysis in mangrove ecosystems," *ISPRS J. Photogram. Remote Sens.*, vol. 63, no. 5, pp. 578–589, Sep. 2008.

[10] G. Duveiller, P. Defourny, B. Desclée, and P. Mayaux, "Deforestation in Central Africa: Estimates at regional, national and landscape levels by advanced processing of systematically-distributed landsat extracts," *Remote Sens. Environ.*, vol. 112, no. 5, pp. 1969–1981, May 2008.

[11] D. Stow, Y. Hamada, L. Coulter, and Z. Anguelova, "Monitoring shrubland habitat changes through object-based change identification with airborne multispectral imagery," *Remote Sens. Environ.*, vol. 112, no. 3, pp. 1051–1061, Mar. 2008.

[12] P. Gärtner, M. Förster, A. Kurban, and B. Kleinschmit, "Object based change detection of Central Asian Tugai vegetation with very high spatial resolution satellite imagery," *Int. J. Appl. Earth Observat. Geoinf.*, vol. 31, pp. 110–121, Sep. 2014.

[13] J. Gong, H. Sui, K. Sun, G. Ma, and J. Liu, "Object-level change detection based on full-scale image segmentation and its application to Wenchuan earthquake," *Sci. China E, Technol. Sci.*, vol. 51, no. 2, pp. 110–122, Dec. 2008.

[14] N.-W. Park and K.-H. Chi, "Quantitative assessment of landslide susceptibility using high-resolution remote sensing data and a generalized additive model," *Int. J. Remote Sens.*, vol. 29, no. 1, pp. 247–264, Jan. 2008.

[15] D. Lu, P. Mausel, E. Brondízio, and E. Moran, "Change detection techniques," *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2401, Jun. 2004.

[16] A. Hecheltjen, F. Thonfeld, and G. Menz, "Recent advances in remote sensing change detection—A review," in *Land Use and Land Cover Mapping in Europe*, I. Manakos and M. Braun, Eds. Dordrecht, The Netherlands: Springer, 2014, pp. 145–178.

[17] G. Chen, G. J. Hay, L. M. T. Carvalho, and M. A. Wulder, "Object-based change detection," *Int. J. Remote Sens.*, vol. 33, no. 14, pp. 4434–4457, Jul. 2012.

[18] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogram. Remote Sens.*, vol. 80, pp. 91–106, Jun. 2013.

[19] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogram. Remote Sens.*, vol. 65, no. 1, pp. 2–16, Jan. 2010.

[20] T. Blaschke *et al.*, "Geographic object-based image analysis—Towards a new paradigm," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 180–191, Jan. 2014.

[21] X. Li, A. G.-O. Yeh, J.-P. Qian, B. Ai, and Z. Qi, "A matching algorithm for detecting land use changes using case-based reasoning," *Photogram. Eng. Remote Sens.*, vol. 75, no. 11, pp. 1319–1332, Nov. 2009.

[22] C. Listner and I. Niemeyer, "Object-based change detection," *Photogrammetrie, Fernerkundung, Geoinformation*, vol. 4, pp. 233–245, Aug. 2011.

[23] O. Hall and G. J. Hay, "A multiscale object-specific approach to digital change detection," *Int. J. Appl. Earth Observat. Geoinf.*, vol. 4, no. 4, pp. 311–327, Nov. 2003.

[24] O. Miller, A. Pikaz, and A. Averbuch, "Objects based change detection in a pair of gray-level images," *Pattern Recognit.*, vol. 38, no. 11, pp. 1976–1992, Nov. 2005.

[25] A. Lefebvre, T. Corpetti, and L. Hubert-Moy, "Object-oriented approach and texture analysis for change detection in very high resolution images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 4. Jul. 2008, pp. 663–666.

[26] J. Chen, Z. Mao, B. Philpot, J. Li, and D. Pan, "Detecting changes in high-resolution satellite coastal imagery using an image object detection approach," *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2454–2469, Apr. 2013.

[27] I. Niemeyer, P. R. Marpu, and S. Nussbaum, "Change detection using object features," in *Object-Based Image Analysis*, T. Blaschke, S. Lang, and G. J. Hay, Eds. Berlin, Germany: Springer-Verlag, 2008, pp. 185–201.

[28] J. Im, J. R. Jensen, and J. A. Tullis, "Object-based change detection using correlation image analysis and image segmentation," *Int. J. Remote Sens.*, vol. 29, no. 2, pp. 399–423, Jan. 2008.

[29] A. P. Tewkesbury, A. J. Comber, N. J. Tate, A. Lamb, and P. F. Fisher, "A critical synthesis of remotely sensed optical image change detection techniques," *Remote Sens. Environ.*, vol. 160, pp. 1–14, Apr. 2015.

[30] T. Blaschke, "Towards a framework for change detection based on image objects," *Göttinger Geographische Abhandlungen*, vol. 113, pp. 1–9, 2005.

[31] L. Bruzzone and D. F. Prieto, "An adaptive parcel-based technique for unsupervised change detection," *Int. J. Remote Sens.*, vol. 21, no. 4, pp. 817–822, Feb. 2000.

[32] J. Linke *et al.*, "A disturbance-inventory framework for flexible and reliable landscape monitoring," *Photogram. Eng. Remote Sens.*, vol. 75, no. 8, pp. 981–995, Aug. 2009.

[33] C. Rother, V. Kolmogorov, T. Minka, and A. Blake, "Cosegmentation of image pairs by histogram matching—Incorporating a global constraint into MRFs," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. Jun. 2006, pp. 993–1000.

[34] D. S. Cheng and M. A. T. Figueiredo, "Cosegmentation for image sequences," in *Proc. IEEE Int. Conf. Image Anal. (IVIAP)*, Sep. 2007, pp. 635–640.

[35] L. Mukherjee, V. Singh, and C. R. Dyer, "Half-integrality based algorithms for cosegmentation of images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 2028–2035.

[36] D. S. Hochbaum and V. Singh, "An efficient algorithm for co-segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Sep./Oct. 2009, pp. 269–276.

[37] S. Vicente, V. Kolmogorov, and C. Rother, "Cosegmentation revisited: Models and optimization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2010, pp. 465–479.

[38] L. Mukherjee, V. Singh, and J. Peng, "Scale invariant cosegmentation for image groups," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 1881–1888.

[39] Y. Mu and B. Zhou, "Co-segmentation of image pairs with quadratic global constraint in MRFs," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2007, pp. 837–846.

[40] P. Xiao, X. Wang, X. Feng, X. Zhang, and Y. Yang, "Detecting China's urban expansion over the past three decades using nighttime light data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Ses.*, vol. 7, no. 10, pp. 4095–4106, Oct. 2014.

[41] X. Huang and L. Zhang, "A multidirectional and multiscale morphological index for automatic building extraction from multispectral GeoEye-1 imagery," *Photogram. Eng. Remote Sens.*, vol. 77, no. 7, pp. 721–732, Jul. 2011.

[42] X. Huang and L. Zhang, "Morphological building/shadow index for building extraction from high-resolution imagery over urban areas," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Ses.*, vol. 5, no. 1, pp. 161–172, Feb. 2012.

[43] X. Huang, L. Zhang, and T. Zhu, "Building change detection from multitemporal high-resolution remotely sensed images based on a morphological building index," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Ses.*, vol. 7, no. 1, pp. 105–115, Jan. 2014.

[44] J. Chen, P. Gong, C. He, R. Pu, and P. Shi, "Land-use/land-cover change detection using improved change-vector analysis," *Photogram. Eng. Remote Sens.*, vol. 69, no. 4, pp. 369–379, Apr. 2003.

[45] V. Kolmogorov and R. Zabin, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004.

[46] Y. Boykov and O. Veksler, "Graph cuts in vision and graphics: Theories and applications," in *Handbook of Mathematical Models in Computer Vision*, N. Paragios, Y. Chen, and O. Faugeras, Eds. New York, NY, USA: Springer-Verlag, 2006, pp. 79–96.
[47] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Jul. 2001, pp. 105–112.
[48] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation," *Int. J. Comput. Vis.*, vol. 70, no. 2, pp. 109–131, Nov. 2006.
[49] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.
[50] W.-S. Chu, C.-P. Chen, and C.-S. Chen, "MOMI-cosegmentation: Simultaneous segmentation of multiple objects among multiple images," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2010, pp. 355–368.
[51] D. M. Greig, B. T. Porteous, and A. H. Seheult, "Exact maximum *a posteriori* estimation for binary images," *J. Roy. Statist. Soc. B (Methodol.)*, vol. 51, no. 2, pp. 271–279, 1989.
[52] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.
[53] L. Ford, Jr., and D. R. Fulkerson, *Flows in Networks*. Princeton, NJ, USA: Princeton Univ. Press, 1962.
[54] A. V. Goldberg and R. E. Tarjan, "A new approach to the maximum-flow problem," *J. Assoc. Comput. Mach.*, vol. 35, no. 4, pp. 921–940, Oct. 1988.
[55] C. Persello and L. Bruzzone, "A novel protocol for accuracy assessment in classification of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1232–1244, Mar. 2010.
[56] I. Lizarazo, "Accuracy assessment of object-based image classification: Another STEP," *Int. J. Remote Sens.*, vol. 35, no. 16, pp. 6135–6156, 2014.
[57] J. R. Schott, C. Salvaggio, and W. J. Volchok, "Radiometric scene normalization using pseudoinvariant features," *Remote Sens. Environ.*, vol. 26, no. 1, pp. 1–16, Oct. 1988.
[58] J. R. G. Townshend, C. O. Justice, C. Gurney, and J. McManus, "The impact of misregistration on change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 5, pp. 1054–1060, Sep. 1992.
[59] X. Dai and S. Khorram, "The effects of image misregistration on the accuracy of remotely sensed change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 5, pp. 1566–1577, Sep. 1998.
[60] O. Dubovyk, G. Menz, C. Conrad, F. Thonfeld, and A. Khamzina, "Object-based identification of vegetation cover decline in irrigated agro-ecosystems in Uzbekistan," *Quaternary Int.*, vol. 311, pp. 163–174, Oct. 2013.
[61] F. Bovolo, "A multilevel parcel-based approach to change detection in very high resolution multitemporal images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 1, pp. 33–37, Jan. 2009.
[62] M. Baatz and A. Schäpe, "Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation," in *Angewandte Geographische Informationsverarbeitung XII*, T. Strobl, T. Blaschke, and G. Griesebner, Eds. Karlsruhe, Germany: Wichmann Verlag, 2000, pp. 12–23.

Pengfeng Xiao (M'13) was born in Hunan, China, in 1979. He received the B.M. degree in land resource management from Hunan Normal University, Changsha, China, in 2002, and the Ph.D. degree in cartography and geographical information system from Nanjing University, Nanjing, China, in 2007.

From 2007 to 2009, he was a Lecturer with the Department of Geographic Information Science, Nanjing University. He was a Visiting Scholar with the Department of Geography, University of Giessen, Giessen, Germany, from 2011 to 2012, and the Department of Environmental Science, Policy, and Management, University of California at Berkeley, Berkeley, CA, USA, from 2014 to 2015. Since 2009, he has been an Associate Professor with Nanjing University. He has authored over 50 articles. His current research interests include remote sensing image processing, land use and land cover change, and remote sensing of snow cover.

Min Yuan was born in Shandong, China, in 1991. She received the B.E. degree in surveying and mapping engineering from the China University of Petroleum, Qingdao, China, in 2013, and the M.S. degree in photogrammetry and remote sensing from Nanjing University, Nanjing, China, in 2016.

She is currently an Examiner with the Patent Examination Cooperation Jiangsu Center of the Patent Office, State Intellectual Property Office, Suzhou, China. She has authored two articles. Her current research interests include remote sensing image processing and land use and land cover change.

Xueliang Zhang was born in Hunan, China, in 1987. He received the B.S. degree in geographical information system from Nanjing University, Nanjing, China, in 2010, and the joint Ph.D. degree in remote sensing of resources and environment from Nanjing University and the Informatics Institute, University of Missouri, Columbia, MO, USA, in 2015.

He is currently an Associate Researcher with the Department of Geographic Information Science, Nanjing University. He has authored 11 articles. His current research interests include remote sensing image processing and land use and land cover change.

Xuezhi Feng was born in Gansu, China, in 1953. He received the B.S. degree in cartography from Nanjing University, Nanjing, China, in 1979.

From 1980 to 1981, he was a Lecturer with the Department of Geography, Nanjing University. He was with the Lanzhou Institute of Glaciology and Cryopedology, Chinese Academy of Sciences, as an Assistant Professor from 1981 to 1990, an Associate Professor from 1990 to 1994, and a Professor from 1994 to 1997. Since 1997, he has been a Professor with the Department of Geographic Information Science, Nanjing University. He has authored over 200 articles. His current research interests include remote sensing of snow cover and remote sensing image processing.

Yanwen Guo was born in 1980. He received the Ph.D. degree in applied mathematics from the State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou, China, in 2006.

He was a Visiting Professor with the Department of Computer Science, The University of Hong Kong, Hong Kong, in 2008, 2012, and 2013, and a Visiting Scholar with the Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign, Champaign, IL, USA, from 2013 to 2015. He is currently a Professor with the State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China. His current research interests include image and video processing, computer vision, and computer graphics.