

Object-Based 3-D Building Change Detection on Multitemporal Stereo Images

Rongjun Qin, *Student Member, IEEE*, Xin Huang, *Senior Member, IEEE*,
Armin Gruen, and Gerhard Schmitt

Abstract—Due to the rapid process of urbanization, there is an increasing demand for detecting building changes over time using very high-resolution (VHR) images. Traditional two-dimensional (2-D) change detection methods are limited due to the image perspective variation and illumination discrepancies. One current trend for building detection combines the use of orthophotos and digital surface models (DSMs), because of its robustness against false changes, as well as its capability of providing volumetric information. In this paper, we propose an object-based three-dimensional (3-D) building change detection framework based on supervised classification, which makes use of the height, spectral, and shape information in a combined fashion with object-based analysis. The proposed method follows the following steps: First, a synergic mean-shift segmentation method is applied on the orthophoto with the constraints of the DSM, which derives segments with homogenous spectrum and height. In a second step, the segments are classified with a hybrid decision tree and SVM approach, and then the segments of the building class are merged as building objects for change detection. An initial change indicator (CI) is then computed for each building object concerning height and spectral information. Finally, an adaptive CI updating strategy based on segment overlapping is proposed and the traffic light system based on a dual threshold is used to identify the change status of each building as “change,” “no-change,” and “uncertain change”. The experimental results on scanned aerial stereo images have demonstrated that our proposed framework is able to achieve high-detection accuracy on images with limited spectral quality.

Index Terms—Change indicator (CI), classification, decision tree analysis, digital surface model, random forest, support vector machine, three-dimensional (3-D) change detection.

I. INTRODUCTION

MONITORING building changes are important tasks for building-related applications. On one hand, it serves as a critical step for updating the maps and three-dimensional (3-D) models in the geo-database; on the other hand, the building change itself is of much interest for applications such as urban growth analysis, building code compliance, and building material flow estimation.

Manuscript received September 30, 2014; revised February 15, 2015; accepted March 13, 2015. This work was supported in part by the Singapore-ETH Center for Global Environmental Sustainability (SEC), in part by the Singapore National Research Foundation (NRF), and in part by ETH Zurich.

R. Qin, A. Gruen, and G. Schmitt are with the Singapore ETH Center, Future Cities Laboratory, Singapore 138602 (e-mail: rqin@student.ethz.ch).

X. Huang is with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: huang_wuhu@163.com).

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/JSTARS.2015.2424275

Change detection techniques use two-dimensional (2-D) low-resolution images have been intensively studied to assess the change at the landscape level, which have achieved satisfactory results [1], [2]. However, such techniques applied to very high-resolution (VHR) images are limited by problems with illumination discrepancy and perspective variations [3], [4], as well as the increased spectral ambiguities. These techniques usually require images to have approximately the same viewpoint with similar spectral responses [5]. Thus, the increased resolution may not necessarily bring about positive effects on the change detection results. The recent development of dense matching techniques produces light detection and ranging (LiDAR) comparable results, which has substantially increased the availability of usable height information from optical images [6]–[8]. Therefore, a current trend is to incorporate the 3-D (height or depth) information for detecting changes from VHR images [9]–[12], which is supposed to be more robust and capable of providing additional information such as volumetric differences.

Several studies compared digital surface models (DSMs) by subtracting them for volumetric analysis [13], which is straightforward to avoid illumination and shadow problems from image-based 2-D change detection approaches. Such methods usually require high-quality DSMs (i.e., generated by LiDAR), as the results are purely dependent on the analysis of the geometry. However, a simple subtraction of DSMs may result in unwanted changes, such as disturbances from seasonal variation of trees, noise, and coregistration errors. This is even more problematic with DSMs generated with dense image matching (DIM) techniques, since the matched DSMs usually contain some noise and blunders. Moreover, researchers sometimes are interested in only one type of urban objects (i.e., buildings, roads, or vegetation). To reduce the ambiguities from unwanted changes, a postclassification method in [14] was proposed for building change detection by first truncating the height differences with a given threshold, and then applied supervised methods based on various shape features to eliminate false positives from trees. Rottensteiner [11] combined the multispectral images and DSM to update the geo-database of the buildings, where the multispectral images were used for extracting a vegetation index, and the resulting buildings were specified as “demolished,” “new,” and “changed”.

Pixel-based change detection is sensitive to noise and other disturbances, whereas object-based methods are more robust [15]. Tian *et al.* [12] proposed a region-based method to jointly compare the height and intensity differences under the constraints of the segmented regions from the orthophotos, and

then applied change vector analysis [16]. Similar object-based methods using Dempster–Shafer fusion [17] of spectral and height information can be found in [18]. To fully explore the use of object-based analysis, a 3-D change detection method proposed in [19] adopted several geometric and texture difference indicators using unsupervised self-organizing maps (SOM) in the context of 3-D model updating: buildings of date 2 are detected using morphological building detection algorithms, and new buildings of date 2 are highlighted by excluding the existing buildings of date 1.

Among the existing change detection methods using VHR images and DSMs for building detection, most of them first apply the robust interpretation of change indicators (CIs) derived from the height and textural/spectral differences, and then employ post-filtering techniques to eliminate unwanted changes. The performance of the change detection largely relies on the quality of the images and DSMs in the first place. It is also crucial to exclude the unwanted changes (blunders from DSMs and seasonal variation of the vegetation). The post-filtering strategy sometimes is limited by the quality and availability of the vegetation index, as well as the lack of urban class information. Therefore, we propose to preclassify the ground scene using the image and DSM, and then to apply an object-based change detection method on the building class, taking into account their height and textural differences, as well as the shape discrepancies. The advantage of this idea lies in the fact that, by knowing segments of the building classes, we are able to consider the shape differences of buildings. These building segments provide appropriate regions to robustly evaluate the height and texture difference. The features extracted from image and DSM can be learned from the data itself, which can reduce the dependence of the algorithm to some particular indexes requiring high-quality spectral information such as normalized difference vegetation index (NDVI) and shadow index. Moreover, this idea can easily be extended for detecting the change of different urban classes such as roads and trees.

A complicated workflow is proposed in this paper to detect building changes. It utilizes the combined height and spectral information of the scene and adopts supervised classification as a means of providing building primitives under an object-based framework. The shape differences of building segments are considered in addition to the height and spectral differences, leading to more robust performance of the algorithm and higher detection accuracy.

This paper is organized as follows. Section II presents the general workflow and methodological considerations. Section III briefly introduces the preprocessing steps including DSM generation and depth constrained image segmentation. Section IV describes the supervised classification and building segment merging. In Section V, the CI computation, considering height, spectral information, and shape are introduced. Section VI demonstrates the results and analyzes the parameter sensitivity. Section VII concludes this paper by discussing the pros and cons of the proposed method.

II. GENERAL WORKFLOW

The development of DIM techniques and high-resolution sensors with stereo capability increases the availability of 3-D

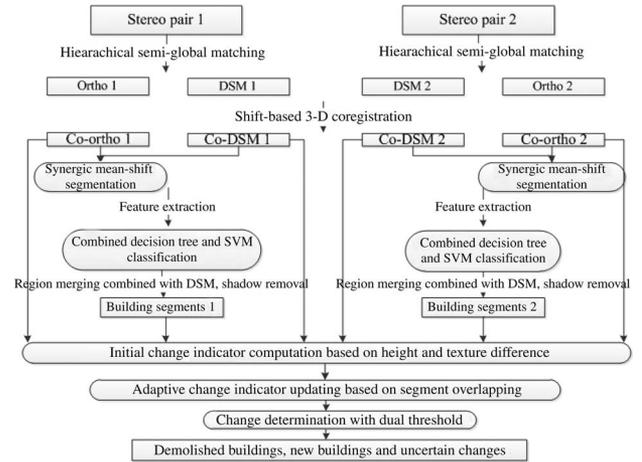


Fig. 1. Proposed workflow.

information. Our idea is to make use of the height information to localize building segments more accurately for building change detection. The overall workflow of our proposed method is shown in Fig. 1.

The DSMs of the stereo pairs were generated with hierarchical semi-global matching (H-SGM) [6], [20]. As the orientation parameters of the images might contain systematic errors, a 3-D coregistration procedure is used to minimize the distance between the DSMs [21] (least squares 3-D surface matching). In our case, only the shift parameters are considered, as the rotation and scale differences are not significant for the aerial images [22]. To derive meaningful segments, the segmentation procedure should consider the height information, especially when the spectral/radiometric information is weak, like in the case of scanned aerial photos. Therefore, the synergic mean-shift segmentation is employed to derive finer segments, which are used for the subsequent feature extraction and classification.

There are many techniques available for building detection, which usually adopt prior knowledge such as building heights, brightness, and NDVI [11], these methods rely on the quality of the vegetation index and sometimes the types of buildings. We propose to use supervised methods, as the prior information can be learned from the images and DSMs, regardless of data types (panchromatic, multispectral, normal RGB). By incorporating the height information, it is possible to have improved classification results [23], thus to provide more reliable building primitives for change detection. As building segments can be extracted with a higher accuracy, we consider the shape information (overlapping region) of the objects in addition to the height and spectral differences to further enhance robustness of the change detection. In the following sections, our proposed framework will be introduced in detail.

III. DATA PREPROCESSING

The data used in this study are scanned aerial stereo pairs of a part of the city of Zurich. Due to the character of analogue photographic recording, radiometric sampling and physical photo fade, the scanned photos usually have higher noise level than digital images. Therefore, a 3×3 bilateral filter [24] is applied to reduce its negative impact on the DIM, which smoothes the

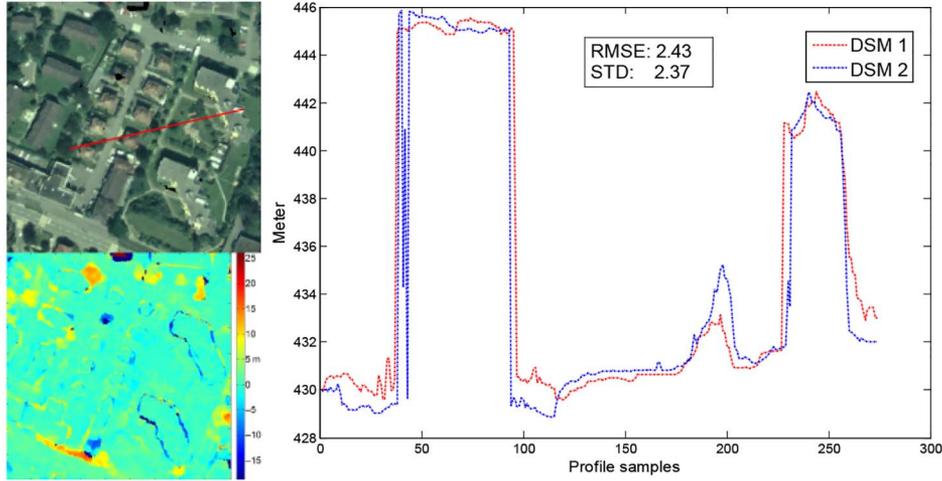


Fig. 2. Profile analysis of height differences of the DSMs. Upper left: an unchanged area; the red line denotes the analyzed profile; bottom left: height differences; right: DSM comparison of the profile (m).

homogenous areas of the images and at the same time keeps sharp edges.

A. DSM Generation and Coregistration

1) *DSM Generation*: As a basic geometric processing procedure, the interior orientation is established based on the fiducial marks of scanned images. For DIM, the epipolar images are generated using the exterior orientation parameters.

The semi-global matching (SGM) algorithm has been proved to be one of the leading DIM methods, which is able to leverage both the performance and efficiency [6]. In this study, we apply a hierarchical SGM method for the generation of the DSM, whose performance has been reported as being similar to the classic SGM algorithm, but has less memory requirement and computation load [20]. Given a pair of epipolar images, SGM tries to minimize the following matching energy with multipath dynamic programming [25]:

$$E(D) = \sum_p C(p, D_p) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1] \quad (1)$$

where the first term denotes the sum of the initial energy with disparity D , and the second term and third term regularize the smoothness disparity map. P_1 and P_2 are jump penalties, with the first one for pixels with disparity jump of 1 pixel to the neighboring pixels (N_p) and the latter for more than 1 pixel. P_2 can be adaptively chosen to be smaller for edge pixels and larger for pixels in the homogeneous area. With a nonparametric pixel-wise cost (Census) [26], these two parameters (P_1 and P_2) can be fixed as constants, thus this method can be seen as being parameter-less [27]. The hierarchical SGM adopts a coarse-to-fine strategy that dynamically determines the disparity range computed from a higher level of the pyramids to the bottom ones for finer search, which saves the computation times and was reported to have similar performance as the classic SGM algorithm.

2) *Coregistration*: The generated DSMs from both dates usually have coregistration errors due to systematic errors such as the control point measurement errors and modeling uncertainties from geo-referencing. Coregistration methods based on feature points and known orientation parameters [9] are able to reduce the misalignment according to the measurement redundancies, whereas such methods are related to the sensors, as for different models, the method needs to be formulated in a case-by-case fashion. Therefore, we apply the least squares 3-D (LS3D) coregistration [21] based on the direct minimization of the squared sum of Euclidean distances of the 3-D surfaces, being independent of sensor models. This method estimates the parameters of a spatial similarity transformation [rotation (ω, φ, κ), translation (x, y, z), and scale differences] between 3-D surfaces. In our case, we only estimate the translations, since the rotations and scale differences are not significant. Fig. 2 shows an example of a small unchanged area with their root-mean-square error (RMSE) of the height differences.

B. DSM Constrained Segmentation

We use object-based analysis for change detection; therefore, it is important to have meaningful and reliable segments. The image-based segmentation method groups pixels with similar colors, while in our case, the height information is available, which is used to further improve the performance of the segmentation. We adopt the synergic mean-shift (MS) segmentation [28], [29] to find segments that have both homogeneous color and height. The original formulation of the method imposes a very large weight on the edges of the color images for each iteration of MS process, so that the MS vector will not go beyond the edge pixels. In our formulation, we apply the Canny magnitude [30] of the DSM as the edge weight of the synergic MS segmentation, which restricts the segmentation within height-jump free areas. Fig. 3 shows an example of the synergic mean-shift segmentation constrained by height using a gray scale image. It can be seen that the ground and building annotated in Fig. 3(a) have very similar

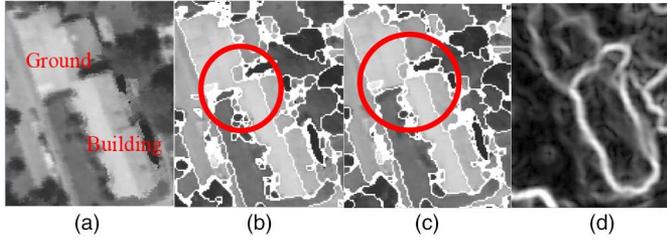


Fig. 3. Example of mean-shift segmentation: (a) the gray image; (b) classic mean-shift segmentation; (c) synergic mean-shift segmentation; and (d) Canny magnitude of the DSM as weight map.

gray values, and the classic mean-shift segmentation result links them as a one segment [Fig. 3(b), marked in the red circle], whereas the synergic mean-shift segmentation [in Fig. 3(c)] separates them due to the constraints of the weight map from the DSM.

The synergic mean-shift segmentation has an additional edge strength parameter β compared to those of the classic mean-shift segmentation (the spatial bandwidth H_s and the spectral bandwidth H_r), which is used for adjusting the weight of the edge maps. In our experiments, $H_s = 7$, $H_r = 4$, and $\beta = 0.1$ are set as constants due to empirical tests. The spatial parameter $H_s = 7$ is an empirical value that usually works for most of the images. Since the synergic mean-shift segmentation takes the DSM as a constraint, $H_r = 4$ is set as a relative large value to reduce the effect of over-segmentation.

IV. SUPERVISED CLASSIFICATION AND BUILDING DETECTION

The proposed workflow adopts a supervised method for object-based classification to derive building segments. Due to the available DSM, the geometric features can be extracted to improve the classification accuracy [23]. Shape features extracted from the radiometric information can be used to improve the separability of urban classes [31]. Some features are linearly linked to the urban classes, such as building and shadow indices [32], [33]. Therefore, we propose to use a combined decision tree analysis with support vector machine (SVM) (DT-SVM) for classifying the scene, aiming to examine useful indices in a hierarchical order. The basic idea is to first separate the off-terrain pixels from the ground using a top-hat by reconstruction of the DSM, and to detect shadows using the shadow index based on morphological reconstruction [32]. The other classes at a finer level (e.g., buildings and trees) are distinguished using the joint height and spectral information via a binary SVM. We consider only four basic urban classes “building,” “vegetation,” “ground,” and “shadow”. The following sections describe the feature extraction and the proposed DT-SVM classification.

A. Feature Extraction

The following features are extracted for each segment.

- 1) Mean value of the top-hat by reconstruction of DSM (THDSM) as an indicator for above-terrain points

The top-hat by reconstruction is a morphological operation that aims to find hat-shaped blobs. A top-hat by reconstruction Th_J^e of an image J is computed as

$$Th_J^e = J - BR_{J,\varepsilon(J,e)} \quad (2)$$

where $\varepsilon(J, e)$ is the gray level morphology erosion, and e is the structuring element. $BR_{J,I}$ is the morphological image reconstruction of J from I , where I is smaller than J pixel-wise [34]. I is computed as the erosion of J with a disk-shaped structuring element, and the radius of the disk is defined as the maximal value of the buildings, as further increasing the value will not necessarily improve the performance [33]. The top-hat by reconstruction on the DSM effectively extracts the hat-shaped blobs, which is a good indicator for above-terrain points.

- 2) Mean value of the top-hat of inverse image (THIIM) for shadow indicator

Shadows in an urban area are mainly casted by buildings and trees. A major characteristic of the shadow is its high luminance contrast to the surrounding area, thus the top-hat by reconstruction can be applied to the negative image of the luminance (in CIELAB space [35]) to extract dark blobs as shadows. Since these shadows usually create extended regions in one direction, with the similar size as the buildings in the other direction, the same radius as used in THDSM can be adopted for computing THIIM.

- 3) Variance of DSM (VDSM) and variance of image luminance (VIM)

As the vegetation (usually trees or bushes) surfaces are usually rougher than the impervious surfaces (roads, grounds, and building roofs) in terms of DSM and image radiometry, VDSM and VIM are adopted to distinguish them.

- 4) Mean value of the color (MC)

The spectral information is the driving force for urban classification in most of the cases, and it is particular helpful to distinguish vegetation from impervious surfaces.

- 5) Elongation (EL) and expansion (EP) of each segment

The EL and EP describe the geometric properties of the segments, which are computed as follows:

$$EL = \frac{MIN_LEN}{MAJ_LEN},$$

$$EP = \sqrt{(MAJ_LEN)^2 + (MIN_LEN)^2} \quad (3)$$

where the “MAJ_LEN” refers to “major axis length” and “MIN_LEN” refers to “minor axis length,” being those of the fitted ellipse of the segment. EL describes the thinness of the segments, and this can be used to discriminate the thin road from building segments. EP indicates the lateral expansion of segments, meaning how elongated the segments are, which is effective to distinguish large segments from small ones. Since our segmentation procedure in Section III-B tends to reduce the over-segmented regions, these shape features are useful to distinguish buildings from other land cover objects.

B. Combined Decision Tree and SVM Classification

The THDSM reveals a linear relationships to above-terrain objects (buildings and trees) and THIIM to shadows [32], [33].

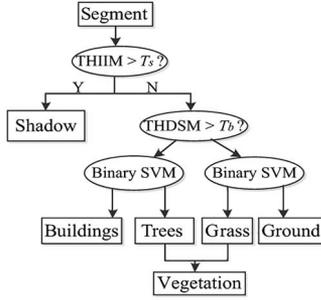


Fig. 4. DT-SVM method for classification.

A decision tree-based classification is considered to determine these two classes using the linear indicator, and then SVM is used for finer classification. Fig. 4 shows the proposed DT-SVM classification method.

Since heavy shadows casted onto buildings will cause matching errors, the shadow segments are first extracted based on the THIIM. The above-terrain segments (building and trees) and ground segments (grass and imperious grounds) are then separated using THDSM. For both of the groups (above-terrain and ground segments), a binary SVM is used to provide finer classes using the rest of the features (VIM, MC, EL, and EP) described in Section VI-A. The Gaussian radial basis function (RBF) is used as the kernel function for the SVM classification [36]. To ensure equivalent contribution of the rest features, these features are normalized across all the segments, thus the scaling factor of RBF is set to 1.

The threshold of THDSM and THIIM can be learned by computing the mean value of the given training samples. Since the shadows are determined by truncating the shadow index (THIIM), the training samples can be small and representative. Only heavy shadows are selected in order to avoid removing buildings under light shadows. The advantage of the DT-SVM is that it allows the feature vector to be hierarchically examined, so that more important and informative components of the feature will make more contributions.

C. Building Merging

It is possible that one actual building object may have been divided into several segments (over-segmentation) during the segmentation process. Neighboring segments with similar height have high chances of belonging to the same building; therefore, such building segments should be merged as building objects. For our experiment, the neighboring segments whose height differences are less than 1 m (2.5 pixels) that are merged into one building segment using a fast eight-neighborhood connectivity method [37]. The fast connectivity merging algorithm considers the height differences of each pair of neighboring segments instead of all the connected segments; therefore, it is also applicable to nonflat roofs (e.g., slant roofs and dome roofs), as the neighboring segments of such roofs usually have a smooth transition.

It should be noted that we tolerate misclassification of the ground segments to the buildings, as the main purpose of the supervised classification is to eliminate the vegetation, which is

the main disturbance for building change detection. Moreover, the ground segments usually have large size and are less likely to produce false positives. Therefore, the samples of the buildings should be sufficient to cover most of spectrums of the building roofs, and the number of vegetation and road samples can be very small.

V. OBJECT-BASED CHANGE DETECTION

The direct subtraction of DSM and color (or spectrum) results in many uncertainties and noise. We perform the change detection only on the building segments, and the advantage mainly lies in three aspects: 1) as the quality of data may vary between different dates, detecting building segments independently for each date reduces errors induced by noises of the data from the other date. 2) For data with poor quality, highly overlapping building objects (building segments detected from two dates) could provide additional shape information for the change evaluation. 3) The unwanted changes (i.e., in vegetation) can be removed in the classification stage.

To implement the shape information, our proposed method computes the CI for each building object in two steps: 1) initial CI based on robust height and texture difference and 2) adaptive CI updating based on segment overlapping. The first step computes an initial CI that indicates the change probability for each building segment. Based on the initial CI, the shape information (segment overlap between buildings detected from two dates) is used to adaptively update the CI for overlapping and nonoverlapping building segments.

A. Initial CI Computation

The initial CI is computed using the orthophotos and corresponding DSMs for each building object on each date. It should be designed to exploit both the spectral/textural and the height information while maintaining its robustness to the DSM and spectral uncertainties. Due to the temporal variation of the luminance, simple Euclidean differences between the color information from two dates may result in many false positives. Therefore, for each segment, we compute the normalized correlation coefficient (NCC) of its bounding box, as it is robust against radiometric differences. NCC is usually sensitive to misalignment between two image patches, which may occur due to the matching errors. Tian *et al.* [38] proposed to reduce the misalignment for DSM differencing by computing the minimal value of height differences over a small window. Based on the same idea, we compute the maximum of the NCCs by shifting the rectangular box with a 5×5 window to robustly compute the correlation between two patches.

The height differences are computed within each building segment. We adopt a histogram-based robust height difference as proposed in [19]: it first calculates the histogram of the height difference of all the pixels within the segment, and then excludes those values with frequent number less than 10% of the total number of pixels in this segment. The mean value is computed over the remaining pixels as the robust height difference. Since the NCC is a normalized value, the height difference should also be normalized so as to be fused with NCC.

The normalization of the height differences $nhdif$ employs a truncated form by computing the ratio of the height difference $hdif$ and a fixed base height m_b for each segment

$$nhdif = hdif/m_b. \quad (4)$$

The CI can be written in a weighted form combining both measures

$$C = \omega \times NCC + (1 - \omega)nhdif. \quad (5)$$

In our experiment, we take $m_b = 5, \omega = 0.2$. $m_b = 5$ is approximately two times of the RMSE of the height difference after coregistration (as shown in Fig. 2), meaning that segments with height difference larger than 5 m will have a very high probability of change. Since the DSM of the high-resolution data is more reliable, we set less weight to the NCC value, determined as an empirical value.

B. Adaptive CI Updating Based on Segment Overlapping

The initial CI reveals the change probability based on the height and texture differences. However, it might contain errors due to the texture ambiguity and matching uncertainties, which causes false positives. Most of the previous works focused on robust methods to compute the CI, yet ignored the fact that the shape coherence of the segments between two dates could provide more information to reduce such uncertainties. For example, if two building objects detected at each date have a high overlap, they are likely to be the same objects, so the CI should be lower to reduce their chance of having been changed.

The idea is to suppress the CI value for each segment when there is a highly overlapping building segment from the other date, and to increase the CI value when a building object at one date could not find a correspondence at the other date. For building segments from two dates that are partially overlapping, they may be a rebuilt area. Their common regions may reveal smaller height differences than the nonoverlapped regions. In this case, these building segments should be used to support each other for more robust change evaluation. We propose an adaptive process to update the CI of partially overlapping building segments based on their correlation, as described below.

For a building segment A at one date and B at the other date, we define the following values:

$$R(A, B) = \frac{\text{Area}(\text{Intersect}(A, B))}{\text{Area}(B)} \quad (6)$$

$$M(A, B) = \text{Min}(R(A, B), R(B, A)) \quad (7)$$

$$G(A, B) = \text{Max}(R(A, B), R(B, A)). \quad (8)$$

According to the formulation of $R(A, B)$, $M(A, B)$, and $G(A, B)$, it can be easily proved that they are all within $[0, 1]$. Different values of $M(A, B)$ and $G(A, B)$ can indicate an overlap status of these two segments: A high $M(A, B)$ implies that these two segments have a high agreement (correlation) in their 2-D shapes, which indicates a high possibility that these two segments are unchanged [Fig. 5(a)]. A low $G(A, B)$ indicates small overlaps, meaning a low correlation between these

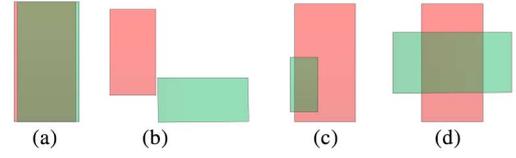


Fig. 5. Abstract examples of the segment interaction (red: segment A; green: segment B). (a) High $M(A, B)$; (b) low $G(A, B)$; (c) low $M(A, B)$ but high $G(A, B)$; and (d) both segments are partially overlapped.

two segments [Fig. 5(b)]. A high $G(A, B)$ shows a high correlation between two segments, as at least one segment is largely covered by the other [Fig. 5(c)]. Fig. 5(d) shows a common case of a rebuilt area, where buildings from different times are partially overlapping. For highly correlating buildings such as in Fig. 5(a), the CI values of both segments are similar, as they indicate the same area of the difference. In the case of Fig. 5(c), segment B is almost fully covered by A , and B gives no more information for computing the CI of A . On the contrary, the CI value of B is highly correlated to that of A , as the information for computing the CI is from a part of A . For uncorrelated segments shown in Fig. 5(b), both buildings are independent and should not be used to support each other. For partially correlated buildings as shown in Fig. 5(d), their overlapping area might have similar height, which reduces the CI value of each segment, thus the nonoverlapping regions should be used to support each other.

From an information point of view, either high correlation or no correlation leads to little information [39]. Therefore, the half-correlated building segments contain the highest joint information to evaluate their change status. We adopt the entropy metric [39] to measure such information by adaptively weighting the contribution of the overlapping building segments to update the CI. Assuming the overlap measures $G(A, B)$ and $M(A, B)$ of the two segments as information sources, the joint information of their correlation can be defined as their joint entropy, being

$$W(A, B) = H(G(A, B))H(M(A, B)). \quad (9)$$

where $H(X) = -\sum p_i(X)\log(p_i(X))$ is the entropy measurement. $G(A, B)$ and $M(A, B)$ follow the Bernoulli distribution

$$p_1(G(A, B)) = G(A, B), p_0(G(A, B)) = 1 - G(A, B). \quad (10)$$

The distribution of $M(A, B)$ can be similarly defined as (10). As for the Bernoulli distributions, $H(X)$ reaches its maximum when $p_1(X) = p_0(X) = 0.5$. It is self-evident that $W(A, B)$ reaches its maximum when $G(A, B) = M(A, B) = 0.5$. This means a segment B that creates 50% mutual overlap with A ($M(A, B) = 0.5$, i.e., half correlated) should give its strongest support for updating CI. Based on this fact, we use a 2-D Gaussian function centered at $(0.5, 0.5)$ as the weight for CI updating

$$g(x, y) = e^{-\left[\frac{(x-0.5)^2 + (y-0.5)^2}{2\sigma^2}\right]} \quad (11)$$

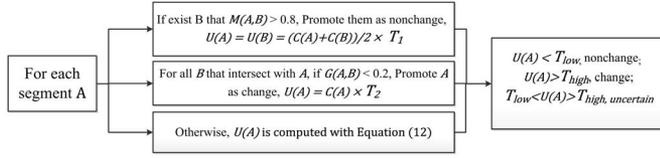


Fig. 6. Change determination process considering the overlapping.

where x and y are $G(A, B)$ and $M(A, B)$, respectively, and $\sigma = 0.5$. Let $C(A)$ be the initial CI of segment A computed in Section V-A, its updated values $U(A)$ are [(12), shown at the bottom of the page]

where $N(A)$ denotes the overlapped segments with A in the other date. To consider the influence by the size of the segment, the area of the segment is also incorporated when weighting different CIs.

The CI updating strategy is shown in Fig. 6. We define $T_1 (< 1)$ and $T_2 (> 1)$ as the relaxing factor to suppress the values of the CI for building objects that find highly overlapped segments ($M(A, B) > 0.8$) and to increase the CI of those do not find corresponding building objects ($G(A, B) < 0.2$). In our experiment, $T_1 = 0.8$, $T_2 = 1.2$. We define a dual threshold (T_{low}, T_{high}) after repetitive tests to classify each segment as “changed” ($U > T_{high}$), “nonchange” ($U < T_{low}$), and “uncertain” ($T_{low} < U < T_{high}$).

Building segments with a “change” status from the earlier date are demolished buildings, and those from the later dates are new buildings. Overlapping building segments with different statuses are promoted as “change” buildings if one of them has a “change” status. In our experiment, the thresholds are: $T_{low} = 0.4$, $T_{high} = 0.5$, which are selected by empirical tests.

The CI updating process can effectively suppress the false positives induced by the matching errors, as well as leveraging CIs of buildings with irregular overlaps with a weighting scheme. Fig. 7 shows the CIs of the building segments in a small area. It demonstrates that the updated CI is lower than that on the nonchange buildings (as can be seen in the green circle), while it is higher in the rebuilt area (within the red circle) due to the adaptive support from other buildings.

VI. EXPERIMENT AND RESULT ANALYSIS

The scanned aerial images are used to validate our method, since they are important resources to study the urban evolutions over the past times. They are provided by SWISSTOPO [40], together with orientation parameters, with an average GSD (ground sampling distance) of 0.38 m. The generated point clouds for each date are resampled to a regular grid DSM with 0.4 m cell size, and orthophotos are generated based on

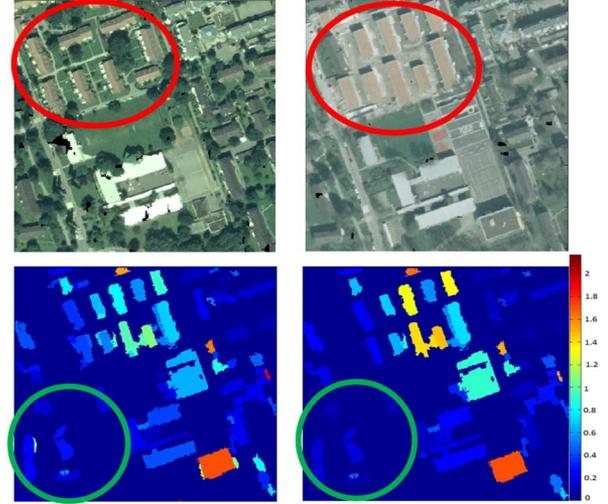


Fig. 7. Example of the computed CI. From left to right: image in the year 2002; image in the year 2007; initial CI [computed with equation (5)]; updated CI.



Fig. 8. Data used for experiment. First row: images; second row: DSMs. Left column: year 2002 (test area for experiment 1); middle column: year 2007 (test area for experiment 2.); right column: year 1994 (test area for experiment 3).

the DSM. Due to the relatively low quality of the film and noise produced during the scanning process, the images from different dates have different radiometric quality and sharpness, which limits the performance of algorithms using direct color and DSM comparisons. Moreover, the old scanned photos (before 2002) are in gray scale, which provides even less spectral information. Therefore, it is challenging to perform change detection analysis on such dataset. Three sites are selected for our experiments, which include typical urban scenarios such as residential area, mixtures of residential, and industrial area. Two of them are selected from the year 2002 (in summer) and 2007 (in winter), and one from the years 1994 and 2002,

$$U(A) = \frac{C(A) * \text{Area}(A) + \sum_{N(A)} g(G(A, B), M(A, B)) * \text{Area}(B) * C(B)}{\text{Area}(A) + \sum_{N(A)} g(G(A, B), M(A, B)) * \text{Area}(B)} \quad (12)$$

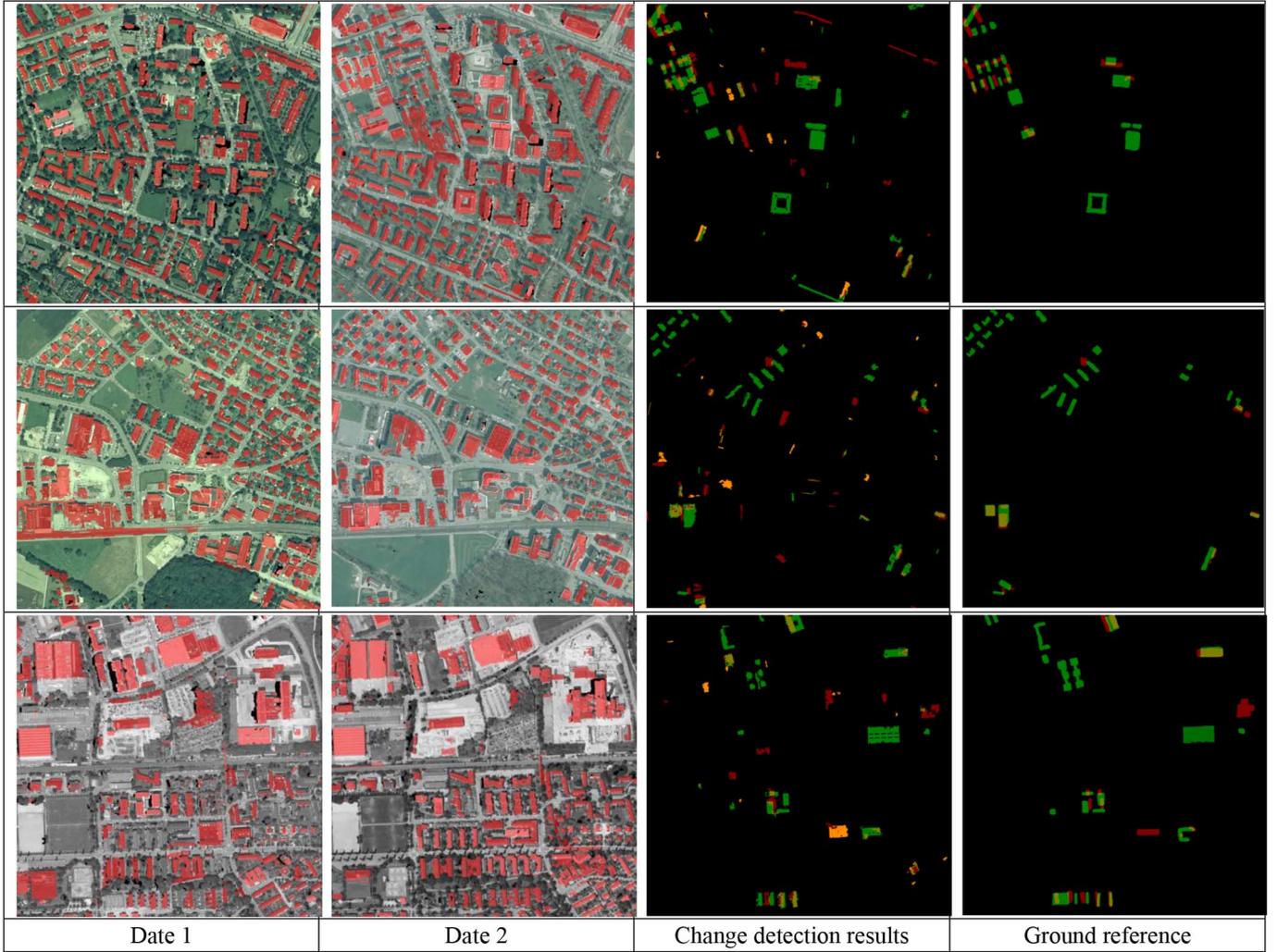


Fig. 9. Change detection results of the experiments. Rows 1–3: Experiments 1–3. Color representation of the third and fourth column: red: demolished; green: new; orange: uncertain changes.

with the stereo pair in 1994 being in gray scale. The reference data are manually sketched by careful inspection, and both pixel-based and object-based measures are carried out to assess the results of the proposed method. In our experiment, true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) are employed and the TP rate (TPR), FP rate (FPR), FN rate (FNR), overall accuracy (OA), and Kappa coefficient (KC) are calculated for evaluating the change detection accuracy [(13), shown at the bottom of the page]

where GP , PD , and N are the numbers of positives in the reference data, positives detected, and the total number of pixels/objects. TPR examines the completeness of detected changes, and FPR presents the wrongly detected changes. KC denotes the total agreement between the detected results and the reference change mask. Our proposed method uses

a dual threshold for determining the “change,” “nonchange,” and “uncertain changes,” and we only evaluate the “change,” and “nonchange” results in our experiment, since the uncertain changes should be identified by the operators.

Fig. 8 shows the examples of the data at these dates, which are also the test areas of the three experiments. It can be seen that there is a clear seasonal difference between the data in 2002 and 2007, where the leaves on the trees are sparser in the image of 2007, and this can be observed clearer in the first two images of Fig. 7. There is no near-infrared band for the color images, and the weak spectral information may not be sufficient to identify the vegetated area with the greenness index [9] of the color images.

To validate the proposed method, three simple methods are used for comparison: 1) DSM differencing (DSMDIF);

$$\begin{aligned}
 TPR &= \frac{TP}{GP}, FPR = \frac{FP}{PD}, FNR = \frac{FN}{GP}, OA = (TP + TN)/N \\
 KC &= \frac{OA - M}{1 - M}, M = \frac{(TP + FP) \times (TP + FN) + (FN + TN) \times (FP + TN)}{N \times N}
 \end{aligned} \tag{13}$$

TABLE I
STATISTICS OF THE TRAINING SAMPLES

		Experiment 1	Experiment 2	Experiment 3
Building	Date 1	183	224	219
	Date 2	231	207	208
Ground	Date 1	175	219	254
	Date 2	36	168	204
Vegetation	Date 1	217	201	246
	Date 2	25	135	200
Total segments	Date 1	22 902	17 763	15 314
	Date 2	11 358	11 802	14 265
Percentage (%)	Date 1	2.51	3.63	4.7
	Date 2	2.57	4.32	3.49

TABLE II
CLASSIFICATION RESULTS FOR DIFFERENT METHODS

		Experiment	Method	Building	Ground	Vegetation	OA
Date 1	1 (2002–2007)	DT-SVM	0.985	0.837	0.994	0.964	
		MSVM	0.975	0.992	0.996	0.959	
		RF	0.992	0.998	0.987	0.992	
	2 (2002–2007)	DT-SVM	0.979	0.988	0.955	0.974	
		MSVM	0.974	0.992	0.954	0.973	
		RF	0.979	0.846	0.952	0.941	
	3 (1994–2002)	DT-SVM	0.849	0.982	0.940	0.898	
		MSVM	0.839	0.983	0.942	0.893	
		RF	0.856	0.998	0.947	0.908	
Date 2	1 (2002–2007)	DT-SVM	0.938	0.910	0.985	0.935	
		MSVM	0.884	0.957	0.982	0.914	
		RF	0.942	0.998	0.783	0.943	
	2 (2002–2007)	DT-SVM	0.942	0.973	0.995	0.956	
		MSVM	0.886	0.997	0.982	0.941	
		RF	0.921	0.999	0.999	0.950	
	3 (1994–2002)	DT-SVM	0.743	0.993	0.999	0.825	
		MSVM	0.795	0.884	0.977	0.828	
		RF	0.883	0.992	0.994	0.917	

2) image differencing (IMGDIF); and (3) a region-based change detection method (REGB) [12] that combines the height and spectral/radiometric information. Thresholds are given to truncate the differences as change masks in these three methods. These methods usually deliver the change detection results for all land cover classes. In order to compare them with the proposed method, we only consider the area covered by building class from our classification results. The thresholds of these methods are tuned repetitively according the ground reference change mask, and only the results that produce the optimal *KCs* are used for the comparison.

A. Experiment

The experiment results are shown in Fig. 9. A first visual comparison between the third and fourth row of Fig. 8 indicates that most of the significant changes are detected, with only a few uncertain changes. This implies that the final CI has good separability between “change” and “nonchange.” The detected building segments contain some ground segments, which are usually determined as nonchange or uncertain change, as the sizes of the ground segments are usually large to tolerate small DSM errors. The image in the year 2002 is converted to gray

TABLE III
CHANGE DETECTION RESULTS

		Ex	Method	TPR	FPR	FNR	OA	KC	
Pixel-based	1 (2002–2007)	DT-SVM	0.848	0.418	0.152	0.979	0.680		
		MSVM	0.714	0.459	0.285	0.975	0.602		
		RF	0.718	0.461	0.281	0.975	0.603		
	2 (2002–2007)	DT-SVM	0.798	0.415	0.202	0.983	0.666		
		MSVM	0.753	0.373	0.246	0.984	0.671		
		RF	0.706	0.374	0.294	0.983	0.655		
	3 (1994–2002)	DT-SVM	0.791	0.364	0.208	0.978	0.694		
		MSVM	0.776	0.424	0.224	0.973	0.648		
		RF	0.805	0.369	0.194	0.978	0.696		
							NCSR D		
		Ex	Method	TPR	FPR	FNR	Date 1	Date 2	Date 3
Object-based	1 (2002–2007)	DT-SVM	0.941	0.767	0.058	15	22	20	
		MSVM	0.882	0.771	0.118				
		RF	0.933	0.750	0.067				
	2 (2002–2007)	DT-SVM	0.960	0.617	0.040	11	28	28	
		MSVM	0.963	0.526	0.030				
		RF	0.963	0.527	0.037				
	3 (1994–2002)	DT-SVM	0.909	0.463	0.091	16	21	22	
		MSVM	0.857	0.579	0.143				
		RF	0.952	0.523	0.048				

Ex, experiment; NCSR D: number of changed segments in the reference data.

scale in the third experiment to test our proposed method on data with poor spectral quality.

B. Result Analysis

Since the supervised classification is a crucial step for deriving the building segments, it is essential to evaluate the classification accuracy. We manually marked most of the segments with their class information on the test dataset, and less than 5% of them were randomly selected as training data, with the rest of the segments used for classification accuracy assessment. The statistics of training samples is shown in Table I. Besides the DT-SVM classification approach, we have tested the multi-class SVM [36] (MSVM) and the random forest (RF) approach [41] for classification. For the MSVM classification, we adopt the “one-against-all strategy” [42] with a “Gaussian radial basis function” as the kernel (with scaling factor set as 1). 500 trees are used for the RF training. All the features described in Section IV-A are used in a vector-stack fashion, with each component of the feature vector normalized to [0, 1] across all the segments.

Table II shows the classification results. It can be seen that there is no significant difference in terms of *OA*. The results of the experiments 1 and 2 have achieved higher classification accuracy, since color images contain more information. The RF has overall a better performance in identifying the building area, and the DT-SVM has higher classification accuracy in the vegetation class.

Table III shows the change detection results using different classification method. The object-based evaluation does not show significant differences, as the number of changed segments is small. The proposed method has achieved over 85% TPR and an average of 60% of FPR, meaning that most of the changed segments can be identified. Most of the false positives are small in size, occurring at the wrongly classified vegetation (trees) or places with matching errors, which can be further filtered according to their size and shapes. In the pixel-based

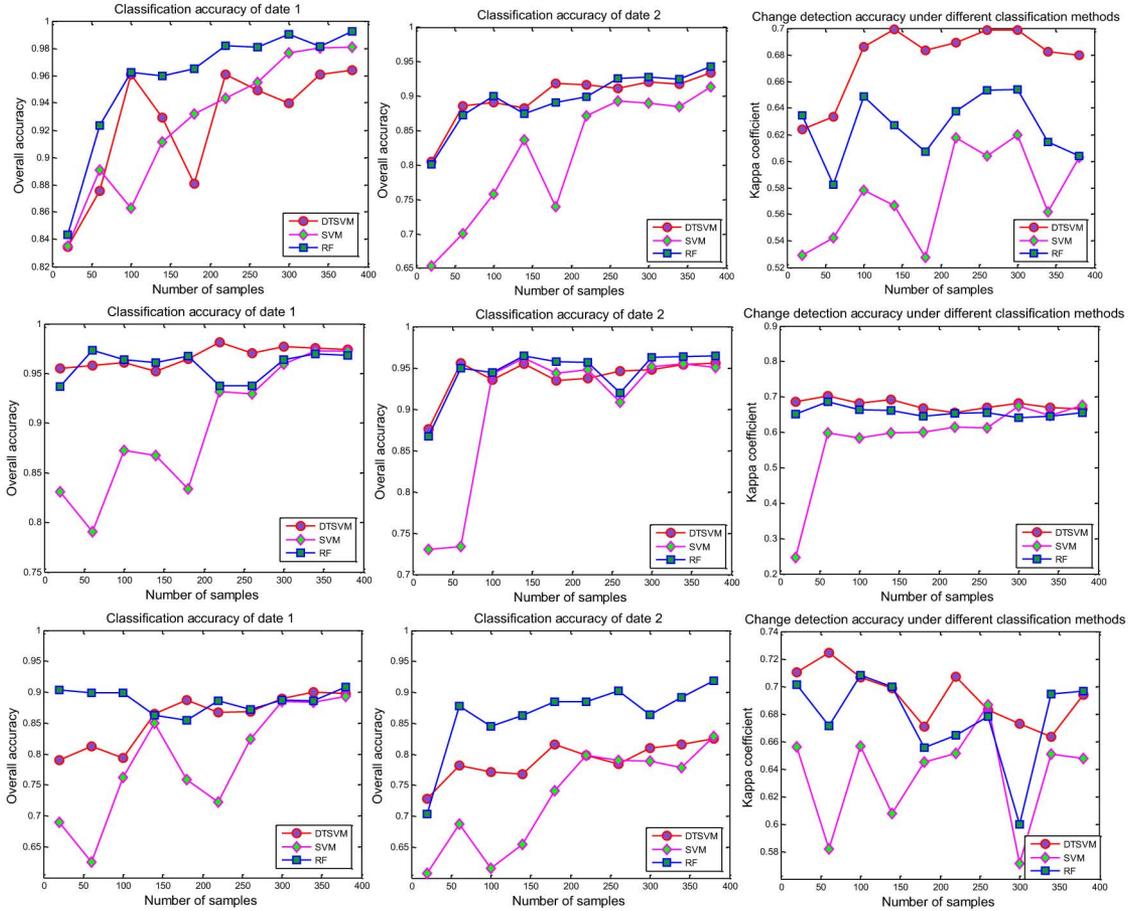


Fig. 10. Correlation between the number of training samples and the classification accuracy and change detection accuracy. Rows 1–3: Experiments 1–3. Columns 1–3: classification OA in date 1, classification OA in date 2 and change detection KC .

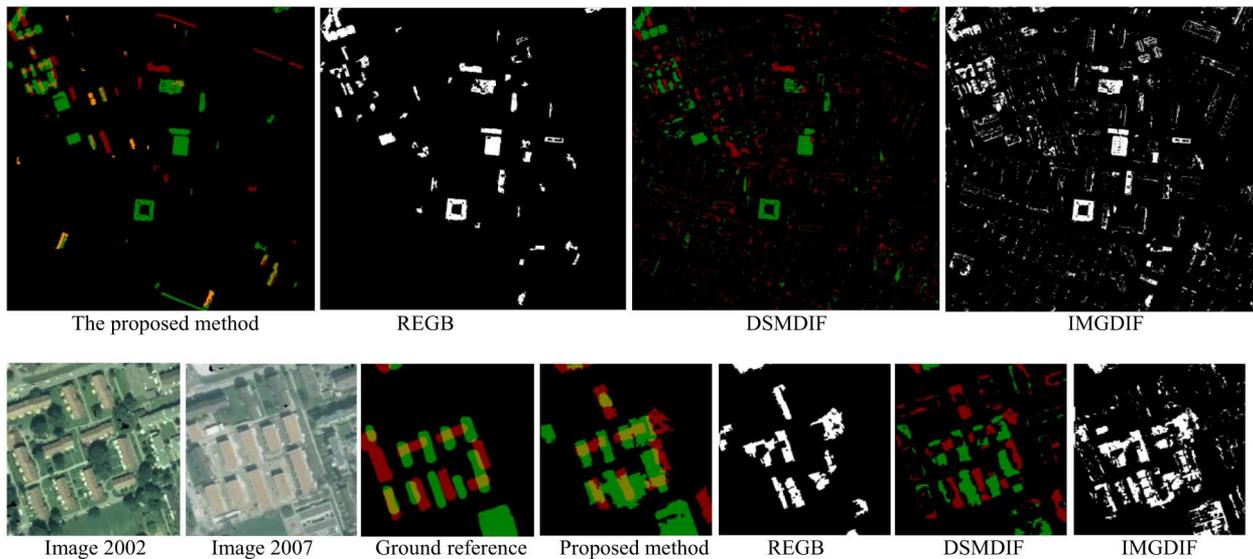


Fig. 11. Comparative study in test site 1 (red: demolished; green: new buildings). First row: the computed change mask; second row: an enlarged area with rebuilt buildings. The threshold of REGB (region-based method), DSMDIF (DSM differencing) and IMGDIF (image differencing) are optimized for the highest KC .

evaluation of the experiment 1, although the classification OA of DT-SVM is lower than that of the RF, it delivers higher KC in change detection, which is mainly due to its high TPR and low FPR . In experiment 3, the classification OAs of all the

three methods are similar, however, the KC of MSVM in the change detection results is lower than the other two, and this means that the classification is not necessarily correlated to the change detection results.

TABLE IV
A COMPARISON OF THE CHANGE DETECTION RESULTS

Method	Experiment 1			Experiment 2			Experiment 3		
	TPR	FPR	KC	TPR	FPR	KC	TPR	FPR	KC
Proposed	0.848	0.418	0.680	0.798	0.415	0.666	0.791	0.364	0.694
REGB	0.616	0.486	0.546	0.627	0.423	0.591	0.570	0.524	0.501
DSMDIF	0.677	0.685	0.407	0.544	0.342	0.586	0.639	0.443	0.580
IMGDIF	0.582	0.694	0.376	0.533	0.831	0.229	0.423	0.724	0.306

To study the performance of proposed method more comprehensively, we have performed the experiments for several times using different number of training samples, aiming to evaluate its robustness and understand how the training samples affects the change detection results. Fig. 10 demonstrates the relationship between the number of training samples and the classification, as well as change detection accuracy. It can be seen that in general, RF has higher performances in terms of classification, while the MSVM delivers the lowest classification OA . This is because each component of the feature vector contributes equivalently for the MSVM classification, which causes reduced classification accuracy due to the weak spectral information. Although RF has obtained the best classification performance for most of the experiments (first two columns of Fig. 10), the change detection results using DT-SVM has achieved the highest KC [as shown in Fig. 10 (the third column)]. As DT-SVM delineates the direct relationship between the THDSM and the above-terrain objects, it has higher chances to maintain the completeness of individual buildings. RF might detect more buildings in the classification, but they might be fragmental, which causes reduced accuracy in the change evaluation. Therefore, the high classification does not necessarily lead to high change detection accuracy. Within a certain range of classification accuracy, the completeness of the buildings is more important. Moreover, Fig. 10 (the third column) also shows that the change detection accuracy does not have constant improvement with the increasing number of training samples. It keeps stable for as long as the number of samples is beyond 100 in our case, where change detection with DT-SVM has obtained the highest KC .

Therefore, the proposed method using DT-SVM classification is more robust and is able to work with small numbers of training samples. The change detection results are mainly affected by the missing buildings with large sizes during the classification/building detection procedure, which leads to false negatives. Consequently, the change detection accuracy is not directly related to the classification accuracy of all the buildings, but largely dependent on the performance of the classification on the changed buildings with large sizes.

Fig. 11 shows the comparison between the proposed method and three methods including REGB (region based method), DSMDIF (DSM differencing), and IMGDIF (image differencing). An enlarged area (Fig. 11, second row) of the test site 1 shows a rebuilt region. It can be seen that the proposed method has extracted more complete regions than the other methods, in particular in the rebuilt area. Since both IMGDIF and REGB include the color information as direct change measure, while not specifying the building segments for each date, they are not

able to separate the changed object as “demolished” and “new.” The DSMDIF takes the negative height difference (DSM2–DSM1) as demolished region and the positive difference as the newly built regions, while it cannot delineate the change status of the overlapping regions (shown in Fig. 11, second row). Table IV shows the statistics of the comparison over the three test sites, where the TPR , TFR , and KC are used for the evaluation. As the DSMDIF and IMGDIF are pixel-based method, we only use the pixel-based evaluation for the comparison. It shows that among the tested methods, the proposed method shows the highest KC agreement to the ground reference. The IMGDIF shows the worst results, and the REGB performs better than DSMDIF. Although the parameters of the three methods (RGB, DSMDIF, and IMGDIF) are tuned to meet the best performance, it still shows lower accuracy than our proposed method.

C. Parameters

There are several tunable parameters in the proposed method, which are mainly from the change detection stage, since the parameters of feature vector for classification and segmentation (i.e., the radius for top hat by reconstruction, H_s , H_r , and β for segmentation) are fixed according to the resolution and quality of the data, as well as the segmentation. The thresholds for determining highly overlapped ($M(A, B) > 0.8$) and nonoverlapped ($G(A, B) > 0.2$) segments are trivial and should be adjusted according to the quality of segmentation. The basic height value m_b is closely related to the quality of DSMs and RMSE of their coregistration, and the weight ω is related to the quality of both spectral and height information. The relaxing parameters T_1 and T_2 are related to the quality of the initial CI. T_1 should be decreased and T_2 be increased when the initial indicators have lower quality. It is interesting to see how T_1 and T_2 and the adaptive process affects the value of the CI. To demonstrate the best achievable results of the proposed method with different relaxing values (T_1 and T_2), we compute the highest KC (hKC) by varying the thresholds (T_{low} and T_{high}) within the range of the CI for different T_1 and T_2 . For this particular purpose, we define $T_{low} = T_{high}$ for generating the change masks. The hKC for the updated CI are shown in Fig. 12. The hKC of initial CI is shown as a constant, since it does not change with T_1 and T_2 . It can be seen that the hKC is relatively robust with the value of T_1 and T_2 , and it reaches the peak when $T_1 = 0.7$, $T_2 = 1.3$ in our evaluation. A clear difference of hKC can be observed between the initial and updated CI. In all the parameter settings, the hKC of the updated CI is approximately 0.07 higher than the initial CI, which demonstrates the effectiveness of our adaptive CI updating approach. T_{low} and T_{high} controls the truncated threshold for change mask generation. A building segment with a CI value greater than 1 indicates strong evidences of being as a changed segment, and a value within $[0, 1]$ implies its uncertainty representing as the probability. Since 0.5 is the median value of the probability. T_{low} and T_{high} should be centered around this value, the larger difference between them, the more segments will be categorized as uncertain segments for operator’s decision.

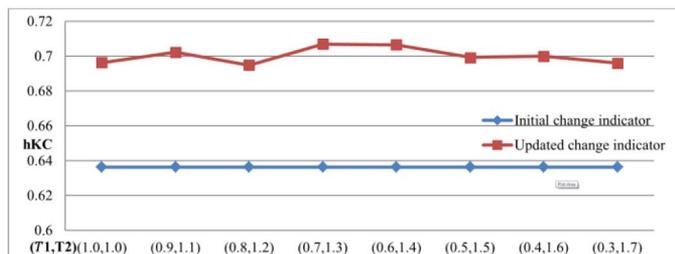


Fig. 12. hKC for the updated CI with different T_1 and T_2 .

VII. CONCLUSION

In this paper, we have proposed a novel framework for 3-D building change detection combining height, texture, and shape information. The basic idea is to first classify the images combining the height information to derive building segments, and then these building segments are further refined to compute the CI based on the height and texture differences. An adaptive strategy is used to update the CI based on segment overlapping. Our contributions of this paper are mainly three-fold: 1) the synergic mean-shift segmentation constrained by DSM; 2) a hierarchical decision tree analysis for classification; and 3) adaptive CI update based on the information measure with segment overlap.

The scanned aerial stereo images are used for experiments. The results have demonstrated that our proposed method has achieved high change detection accuracy even for images with poor spectral information (i.e., images with weak color or monochromatic images). The proposed workflow with the DT-SVM classification method is more robust toward the number of training samples, and it has achieved the highest change detection accuracy with only small number of training samples. The adaptive CI updating strategy incorporates the shape information of the building segments, which is more robust to the matching uncertainties and spectral/radiometric ambiguity. Three methods including 1) DSM differencing; 2) image differencing; and 3) region-based method are compared to our proposed method. The change detection results of the proposed method have achieved the best performance among these methods, and have demonstrated better capabilities of detecting changes in the rebuilt areas.

The proposed method adopts a preclassification strategy to derive building segments, and the change detection is tolerant toward misclassification of large ground segments to the buildings. Most of the false positives occur at tree segments that are misclassified as buildings, as well as small segments induced by matching errors. Part of these false positives can be eliminated by shape filters (e.g., size of the region, elongation, and convexity). The undetected buildings will not cause change detection errors in the nonchange area. However, those undetected buildings in the changed area will lead to large false negatives, regardless of their high classification accuracy. It raises an interesting observation that the higher classification accuracy does not necessarily lead to higher KC in our proposed change detection method, while it is more important that the building detection procedure can obtain high completeness in the changed area. Therefore, our future work will include the

height difference map as prior information of changes to correct the undetected buildings or incomplete building segments in the classification stage, thus to further improve the robustness of the proposed change detection method.

ACKNOWLEDGMENT

The authors would like to thank the SWISSTOPO for providing the scanned aerial images. They would also like to thank the anonymous reviewers for their precious comments and suggestions.

REFERENCES

- [1] J. Rogan, J. Franklin, and D. A. Roberts, "A comparison of methods for monitoring multitemporal vegetation change using thematic mapper imagery," *Remote Sens. Environ.*, vol. 80, pp. 143–156, 2002.
- [2] C. Song, C. E. Woodcock, K. C. Seto, M. P. Lenney, and S. A. Macomber, "Classification and change detection using Landsat TM data: When and how to correct atmospheric effects?," *Remote Sens. Environ.*, vol. 75, pp. 230–244, 2001.
- [3] R. Qin, J. Gong, H. Li, and X. Huang, "A coarse elevation map-based registration method for super-resolution of three-line scanner images," *Photogramm. Eng. Remote Sens.*, vol. 79, pp. 717–730, 2013.
- [4] R. Qin, J. Gong, and C. Fan, "Multi-frame image super-resolution based on knife-edges," in *Proc. IEEE Int. Conf. Signal Process. (ICSP)*, Beijing, China, 2010, pp. 972–975.
- [5] M. Bouziani, K. Goita, 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. Photogramm. Remote Sens.*, vol. 65, pp. 143–153, 2010.
- [6] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, pp. 328–341, 2008.
- [7] S. Gehrke, K. Morin, M. Downey, N. Boehrer, and T. Fuchs, "Semiglobal matching: An alternative to LIDAR for DSM generation," in *Proc. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Calgary, AB, Canada, 2010, vol. 38, (B1), p. 6.
- [8] A. Gruen, "Development and status of image matching in photogrammetry," *Photogramm. Rec.*, vol. 27, pp. 36–57, 2012.
- [9] R. Qin, "An object-based hierarchical method for change detection using unmanned aerial vehicle images," *Remote Sens.*, vol. 6, pp. 7911–7932, 2014.
- [10] R. Qin and A. Gruen, "3D change detection at street level using mobile laser scanning point clouds and terrestrial images," *ISPRS J. Photogramm. Remote Sens.*, vol. 90, pp. 23–35, 2014.
- [11] F. Rottensteiner, "Automated updating of building data bases from digital surface models and multi-spectral images: Potential and limitations," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 37, pp. 265–270, 2008.
- [12] J. Tian, P. Reinartz, P. d'Angelo, and M. Ehlers, "Region-based automatic building and forest change detection on Cartosat-1 stereo imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 79, pp. 226–239, 2013.
- [13] T. R. Martha, N. Kerle, V. Jetten, C. J. van Westen, and K. V. Kumar, "Landslide volumetric analysis using Cartosat-1-derived DEMs," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 582–586, Jul. 2010.
- [14] H. Chaabouni-Chouayakh and P. Reinartz, "Towards automatic 3D change detection inside urban areas by combining height and shape information," *Photogramm. Fernerkundung Geoinf.*, vol. 2011, pp. 205–217, 2011.
- [15] J. Shi, J. Wang, and Y. Xu, "Object-based change detection using georeferenced UAV images," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 38, pp. 177–182, 2011.
- [16] R. D. Johnson and E. Kasischke, "Change vector analysis: A technique for the multispectral monitoring of land cover and condition," *Int. J. Remote Sens.*, vol. 19, pp. 411–426, 1998.
- [17] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton University Press, 1976, vol. 1.
- [18] J. Tian, S. Cui, and P. Reinartz, "Building change detection based on satellite stereo imagery and digital surface models," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 406–417, Jan. 2014.

- [19] R. Qin, "Change detection on LOD 2 building models with very high resolution spaceborne stereo imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 96, pp. 179–192, 2014.
- [20] M. Rothmel, K. Wenzel, D. Fritsch, and N. Haala, "SURE: Photogrammetric surface reconstruction from imagery," in *Proc. LC3D Workshop*, Berlin, Germany, 2012.
- [21] A. Gruen and D. Akca, "Least squares 3D surface and curve matching," *ISPRS J. Photogramm. Remote Sens.*, vol. 59, pp. 151–174, 2005.
- [22] L. Waser *et al.*, "Assessing changes of forest area and shrub encroachment in a mire ecosystem using digital surface models and CIR aerial images," *Remote Sens. Environ.*, vol. 112, pp. 1956–1968, 2008.
- [23] X. Huang, L. Zhang, and W. Gong, "Information fusion of aerial images and LIDAR data in urban areas: Vector-stacking, re-classification and post-processing approaches," *Int. J. Remote Sens.*, vol. 32, pp. 69–84, 2011.
- [24] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. Int. Conf. Comput. Vis.* 1998, pp. 839–846.
- [25] H. Hirschmüller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2005, pp. 807–814.
- [26] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proc. Eur. Conf. Comput. Vis.*, Stockholm, Sweden, 1994, pp. 151–158.
- [27] H. Hirschmüller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1582–1599, Sep. 2009.
- [28] C. M. Christoudias, B. Georgescu, and P. Meer, "Synergism in low level vision," in *Proc. 16th Int. Conf. Proc. Pattern Recog.*, 2002, pp. 150–155.
- [29] D. Comanicu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
- [30] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.
- [31] R. Qin, "A mean shift vector-based shape feature for classification of high spatial resolution remotely sensed imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 99, pp. 1–12, Oct. 2015. doi: 10.1109/JSTARS.2014.2357832.
- [32] 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 Sens.*, vol. 5, no. 1, pp. 161–172, Feb. 2012.
- [33] R. Qin and W. Fang, "A hierarchical building detection method for very high resolution remotely sensed images combined with DSM using graph cut optimization," *Photogramm. Eng. Remote Sens.*, vol. 80, pp. 37–48, 2014.
- [34] L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 176–201, Apr. 1993.
- [35] G. H. Joblove and D. Greenberg, "Color spaces for computer graphics," in *Proc. ACM SIGGRAPH*, Atlanta, Georgia, 1978, pp. 20–25.
- [36] L. Wang, *Support Vector Machines: Theory and Applications*. New York, NY, USA: Springer, 2005, vol. 177.
- [37] E. R. Davies, *Machine Vision: Theory, Algorithms, Practicalities*, 4th ed. San Mateo, CA, USA: Morgan Kaufmann, 2004.
- [38] J. Tian, S. Cui, and P. Reinartz, "Building change detection based on satellite stereo imagery and digital surface models," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 406–417, Jan. 2014.
- [39] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 106, p. 620, 1957.
- [40] Swisstopo. (2014). *Aerial Photographs and Satellite Images* [Online]. Available: <http://www.swisstopo.admin.ch/internet/swisstopo/en/home/services/images.html>, accessed on Jul. 23, 2014.
- [41] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
- [42] J. Milgram, M. Cheriet, and R. Sabourin, "One against one" or "one against all": Which one is better for handwriting recognition with SVMs?," in *Proc. 10th Int. Workshop Front. Handwriting Recog.*, 2006.



Rongjun Qin (S'14) received the B.S. degree in computational mathematics from Wuhan University, Wuhan, China, in 2009, and the M.S. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIEMRS), Wuhan University, in 2011. He is currently pursuing the Ph.D. degree in photogrammetry and remote sensing at ETH Zurich, Zürich, Switzerland.

Since 2011, he has been a Researcher with Singapore ETH Centre, Future Cities Laboratory, Singapore, under the Simulation Platform Module. His research interest includes remote sensing image classification, UAV images processing, image dense matching, 3-D modeling, and change detection.

Mr. Qin serves as a Reviewer for several international journals including *ISPRS Journal of Photogrammetry and Remote Sensing*, *Photogrammetric Engineering and Remote Sensing*, *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING JOURNAL*. His rewards include the first prize of national mathematic modeling competition in China, and several scholarship awards.



Xin Huang (M'13–SM'14) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009

He is currently working as a Full Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIEMRS), Wuhan University. He has authored more than 50 peer-reviewed articles in the international journals. His research interests include hyperspectral data analysis, high-resolution image processing, pattern recognition, and urban remote

sensing applications.

Dr. Huang was recognized by the IEEE Geoscience and Remote Sensing Society (GRSS) as the Best Reviewer of IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2011. He was the winner of the IEEE GRSS 2014 Data Fusion Contest. Since 2014, he serves as an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He was the recipient of the Top-Ten Academic Star of Wuhan University in 2009; the Boeing Award for the Best Paper in Image Analysis and Interpretation from the American Society for Photogrammetry and Remote Sensing in 2010, the New Century Excellent Talents in University from the Ministry of Education of China in 2011, and the National Excellent Doctoral Dissertation Award of China in 2012.



Armin Gruen graduated in geodetic sciences and received the Doctorate degree in photogrammetry from TU Munich, Munich, Germany, in 1974.

He has been an Emeritus Professor at the Chair of Information Architecture, ETH Zurich, Switzerland. He is a Principle Investigator with the Singapore-ETH Center, Future Cities Laboratory, Singapore, since 2010. From 1981 to 1984, he was an Associate Professor with the Department of Geodetic Science and Surveying, Ohio State University, Columbus, OH, USA. From 1984 to 2009, he was a Professor and

Head of the Chair of Photogrammetry and Remote Sensing with the Institute of Geodesy and Photogrammetry, ETH Zurich. He has authored more than 500 articles and papers and edited and co-edited 21 books and conference proceedings. Currently, he is mainly involved in cultural heritage and 3-D/4-D city and environmental modeling projects.



Gerhard Schmitt received the Dipl.-Ing. and the Dr.-Ing. degree in architecture engineering from the Technical University of Munich (TUM), Munich, Germany, and a Master's degree in architecture from the University of California, Berkeley, CA, USA.

He is a Professor of Information Architecture with ETH Zurich, Zurich, Switzerland, and a Senior Vice President of ETH Global, Zurich, Switzerland. He is the Founding Director of the Singapore-ETH Center, Singapore, and leads its Simulation Platform. Prior to becoming a Professor with ETH Zurich in 1988, he

pursued Computer Aided Architectural Design (CAAD) research and teaching at Carnegie Mellon University, Pittsburgh, PA, USA, from 1984 to 1988. He was a Visiting Professor at the Harvard Graduate School of Design, Cambridge, MA, USA, at the Katholieke Universiteit Leuven, Leuven, Belgium, at the Technical University of Denmark, Lyngby, Denmark, and at the Technical University of Delft, Delft, The Netherlands. His research interests include urban management, urban planning, sustainable urbanism, urban ecology, 3-D urban models, simulation and visualization.

Prof. Schmitt initiated and conceptualized ETH Science City and received for this work the 2010 European Culture of Science Award.