

# Change detection on LOD 2 building models with very high resolution spaceborne stereo imagery

Rongjun Qin <sup>a\*</sup>

<sup>a</sup> Singapore-ETH Center, Future Cities Laboratory, ETH, Zurich. 1 CREATE Way, #06-01 CREATE Tower, Singapore 138602.

\*Corresponding author at: Singapore-ETH Center, Future Cities Laboratory, 1 CREATE Way, #06-01 CREATE Tower, Singapore 138602. Tel: +65 8389 5276. Email: [rjin@student.ethz.ch](mailto:rjin@student.ethz.ch) (Mr. Qin)

**ABSTRACT:** Due to the fast development of the urban environment, the need for efficient maintenance and updating of 3D building models is ever increasing. Change detection is an essential step to spot the changed area for data (map/3D models) updating and urban monitoring. Traditional methods based on 2D images are no longer suitable for change detection in building scale, owing to the increased spectral variability of the building roofs and larger perspective distortion of the very high resolution (VHR) imagery. Change detection in 3D is increasingly being investigated using airborne laser scanning data or matched Digital Surface Models (DSM), but rare study has been conducted regarding to change detection on 3D city models with VHR images, which is more informative but meanwhile more complicated. This is due to the fact that the 3D models are abstracted geometric representation of the urban reality, while the VHR images record everything. In this paper, a novel method is proposed to detect changes directly on LOD (Level of Detail) 2 building models with VHR spaceborne stereo images from a different date, with particular focus on addressing the special characteristics of the 3D models. In the first step, the 3D building models are projected onto a raster grid, encoded with building object, terrain object, and planar faces. The DSM is extracted from the stereo imagery by hierarchical semi-global matching (SGM). In the second step, a multi-channel change indicator is extracted between the 3D models and stereo images, considering the inherent geometric consistency (IGC), height difference, and texture similarity for each planar face. Each channel of the indicator is then clustered with the Self-organizing Map (SOM), with “change”, “non-change” and “uncertain change” status labeled through a voting strategy. The “uncertain changes” are then determined with a Markov Random Field (MRF) analysis considering the geometric relationship between faces. In the third step, buildings are extracted combining the multispectral images and the DSM by morphological operators, and the new buildings are determined by excluding the verified unchanged buildings from the second step. Both the synthetic experiment with Worldview-2 stereo imagery and the real experiment with IKONOS stereo imagery are carried out to demonstrate the effectiveness of the proposed method. It is shown that the proposed method can be applied as an effective way to monitoring the building changes, as well as updating 3D models from one epoch to the other.

**Keywords:** Stereo imagery; Semi-global matching; Building change detection; Self-organizing map; Markov random field.

## 1. Introduction

According to OGC (Open Geospatial Consortium) ([Gröger et al., 2007](#)), 3D city models are divided into different level of details (LOD) for visualization and data analysis. LOD 2 models contain the roof structures of individual houses, which are gradually incorporated into the contemporary geographic information systems (GIS) for a variety of urban applications ([Swisstopo, 2014](#); [Zurich, 2014](#)). However, due to the time-consuming process of 3D city modeling ([Gruen and Wang, 1998](#); [Rottensteiner et al., 2012](#)), how to keep these models up-to-date is of high interest. In the current process, updating of the 3D database and monitoring changes still largely rely on manual inspection ([Kamini et al., 2006](#); [Knudsen and Olsen, 2003](#)), which is both cost-ineffective and time-inefficient. In addition to the application of model updating, the 3D change in general is of much interest for volumetric analysis ([Martha et al., 2010](#)), urban growth monitoring ([Tian et al., 2013](#)), building code compliance (detection of illegal buildings), and historical studies of city evolution.

Change detection methods with 2D low-to-medium resolution spaceborne images were intensively studied in the past, mainly for land cover change applications, ([Collins and Woodcock, 1996](#); [Fung, 1990](#); [Jun et al., 2013](#); [Rogan et al., 2002](#); [Song et al., 2001](#)), and the used sensors were not limited to optical systems, but also radar systems ([Bazi et al., 2005](#); [Inglada and Mercier, 2007](#); [Moser and Serpico, 2009](#)). These methods assumed affine transformations between images from two dates, since the height variation was usually not significant enough to cause perspective distortion in low resolution images. Only a few 2D approaches considered change detection in the building scale ([Bouziani et al., 2010](#); [Doxani et al., 2010](#); [Pacifici et al., 2007](#); [Walter, 2004](#)) with very high resolution (VHR) images, yet these methods were limited to images with elevation angles close to 90 degrees for reliable co-registration, and high variability of texture pattern made them more sensitive to co-registration errors.

Despite the high spatial resolution of current spaceborne sensors (e.g. IKONOS with 1 meter, Worldview with 0.5 meters, and Quickbird with 0.6 meters GSD (Ground Sampling Distance)) as well as the aerial images ([Qin et al., 2010](#)), the significant impacts of perspective distortion for image-based co-registration ([Qin et al., 2013](#)), illumination, moisture, and atmosphere of different dates on the spectral responses hinder approaches with spectral information alone. Therefore, more and more researchers tend to incorporate the height information for change detection and consider the change in 3D.

There are a number of methods concerning the use of Digital Surface Models (DSM), which can be either generated from VHR images or airborne laser scanning (ALS) data. With the advanced development of automatic matching algorithm ([Hirschmuller, 2008](#); [Zhang and Gruen, 2006](#)), DSM generated from spaceborne or aerial VHR images get more preferences, simply because they are less expensive and widely available. [Gong et al. \(2000\)](#) subtracted two DSMs generated from aerial images for detection volumetric changes of vegetation, and similar approaches can be found in ([Martha et al., 2010](#); [Murakami et al., 1999](#); [Waser et al., 2008](#)). [Chaabouni-Chouayakh and Reinartz \(2011\)](#) argued that simple subtraction of DSM produced many false positives due to insufficient quality of DSM and co-registration errors. After thresholding the DSM difference, they post-classified the resulting change mask into “building change” and “non-building change”, with various shape features used in a Support Vector Machine (SVM) classifier. Similarly, [Jung \(2004\)](#) first compared two DSMs generated from aerial

images, and then the change mask was further refined by classifying the corresponding aerial images. However, the insufficient quality of DSM and tall vegetation were the main causes of false positives.

As object-based analysis shows more potential to change detection ([Hussain et al., 2013](#)), [Tian, J et al. \(2013\)](#) proposed a region-based method with DSMs generated from Cartosat-1 stereo pairs (2.5 m resolution), and the segmentation of panchromatic orthoimages were used as a support into a weighted change vector analysis (CVA) framework ([Johnson and Kasischke, 1998](#)). Meanwhile, using IKONOS and WorldView-2 stereo pairs, [Tian et al. \(2013\)](#) adopted Dempster-Shafer fusion method ([Shafer, 1976](#)) to determine the changed area based on the texture difference of the orthoimages and DSMs. Nevertheless, the methods highly relied on the quality of the generated DSM, which might vary with the intersection angles, and the scene complexity.

There were only a few work related to change detection directly on 3D models. [Akca et al. \(2010\)](#) performed change detection for 3D building models under a quality control context, where building models were compared with ALS point clouds after a 3D co-registration process ([Akca, 2007](#)). [Knudsen and Olsen \(2003\)](#) projected the 3D building models on 2D aerial views, and then performed a binary classification (with “building” and “non-building” class) on the image view. However, this approach assumed that new buildings retained the same spectral as old buildings, which was not commonly applicable. Moreover, false alarms occurred on roads, whose spectral response was similar to building roofs (high inter-class spectral similarity).

The 3D models are reliable in terms of the geometric positioning of their vertices. When they are applied for change detection, there are a number of issues concerning their characteristics:

1. Intuitively, the 3D models can be converted to a raster DSM, and the problem can be addressed by the DSM-based approach. Whilst one should note that 3D models are somehow abstractions of the real geometry, where the operators decide the type of roof objects to be modeled, e.g. large central air-condition compressor, water tanks, overhanging railways are not modeled sometimes. These objects exist in VHR stereo images, which are likely to be identified as changes by the DSM-based approaches.
2. The presence of building objects in the 3D models provides a perfect segmentation of building and terrains. Beyond the scope of segmentation, 3D models also provide the information of the planar faces, which means the change detection can be performed in face level.
3. The geometric connectivity between faces or objects sometimes poses strong semantic information, and the neighboring faces may share the same change status (change/non-change) due to their geometric similarity or spectral similarity (if textures of the models are available).

Little work addressing these issues has been found in the literature. This paper focuses on detecting changes from the LOD 2 models (including buildings and terrain, hereafter they are referred to 3D models) with VHR spaceborne stereo images, and proposes an object-based approach to address these issues. In the first step (pre-processing), the 3D models are resampled as a regular raster grid, encoded with object and face index, and DSM are generated from the VHR stereo images with a hierarchical semi-global matching (SGM). The second step addresses the aforementioned issues. A robust multi-

channel indicator is proposed concerning the inherent geometric consistency (IGC), the height difference (HEIDIF) and the texture difference (TEXDIF). They are fused robustly with Self-organizing Map (SOM) to obtain the changed faces of the buildings, and uncertain results are determined via a Markov Random Field (MRF) analysis concerning the spatial connectivity of the faces. In the third step, buildings are extracted combining the multispectral orthoimage and DSM using a morphology-based hierarchical method. New buildings are obtained by excluding the verified buildings in the second step. The workflow is shown in Fig. 1.

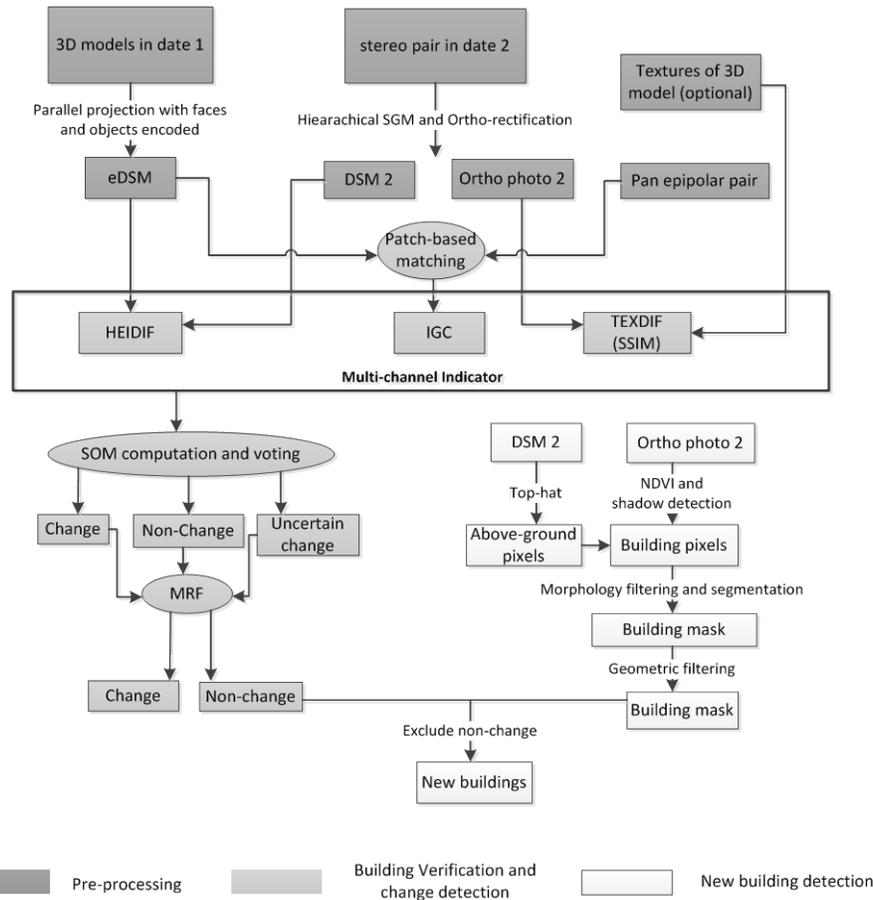


Fig. 1. The workflow of the proposed method; each step will be introduced in detail in the subsequent sections.

The remainder of the paper is organized as follows: Section 2 describes the pre-processing of the 3D models and the stereo images. Section 3 describes the building verification process, where the contributions of this paper lie in. It presents the computation of change indicators, the fusion mechanism, and MRF analysis for building verification. In section 4, a morphology-based building extraction method is introduced for obtaining the new buildings. Section 5 presents a synthetic experiment on WorldView-2 dataset, and a real data experiment on IKONOS images. The tunable parameters are then analyzed into detail. Section 6 draws the conclusion by discussing the pros and cons of the proposed method, as well as future improvements.

## 2. Pre-Processing

### 2.1. Resampling 3D models into encoded DSM

The 3D models used in this paper are composed of buildings and terrain, which generated purely from the stereo pairs. Firstly, the stereo pairs were geo-referenced with ground control points (GCP). Then the buildings were generated semi-automatically with CC-modeler software ([Gruen and Wang, 1998](#)), and the terrain were generated with manually measured features (points, break lines). The measurements were conducted stereoscopically, with a measurement accuracy around 1-2 pixels in image space. The intersection angles of the stereo pairs are around 30 degrees, thus the measurement accuracy is 1-2 pixel size (0.5-1 meters for WorldView-2, 1-2 meters for IKONOS) in the horizontal direction and 1.5-3 pixel size (0.75-1.5 meters for WorldView-2, 1.5-3 meters for IKONOS) in the vertical direction.

For computational purpose, the 3D models are resampled into a raster representation. The cell size is set as the same GSD as stereo images (e.g. IKONOS, 1 meter). Each cell is encoded with the following information during the resample process: 1) the individual building ID (identity) and the terrain ID (only one ID for terrain); 2) the planar face ID. After the resampling process, there might be some missing pixels on the building edges due to numerical reasons. An inverse distance (IVD) interpolation is adopted to fill up the height value of these missing pixels, and a nearest neighborhood approach is used to assign their object ID and face ID. For notational convenience, this encoded DSM is referred as eDSM hereafter.

What's more, the digital terrain model (DTM) can be also obtained by applying the same procedure to the terrain object.

### 2.2 DSM generation with hierarchical semi-global matching

Modern spaceborne images use Rational Polynomial Functions (RPF) sensor models for geometric modeling, such as Worldview (1/2), IKONOS, and QuickBird, etc. In order to get a high geometric accuracy in sub-pixel level, the first order bias correction ([Fraser and Hanley, 2003](#)) is performed on the stereo pairs using four GCPs. Thus a quasi epipolar stereo pair is generated ([Wang et al., 2011](#)), and a hierarchical SGM ([Rothermel et al., 2012](#)) is implemented for computing the disparity map of the epipolar images. Point clouds are then generated through forward intersection, and resampled into a raster grid. IVD is used to fill the small regions (whose radius are less than 20 meters) eliminated by left-right consistency check procedure. There are still some small unmatched areas (less than 400 square meters), due to the occlusion of the high-rise buildings, and these areas will not be considered for subsequent computation.

The 16-bit panchromatic band of the epipolar images are used for matching. The classic SGM algorithm adopts a multi-path dynamic programming strategy to minimize the following cost function ([Hirschmuller, 2005](#)):

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

where the first term denotes the sum of matching costs for the disparity  $D$ , and the second and third term control the smoothness of the disparity map.  $N_p$  denotes the neighboring pixels of the pixel  $p$ .  $T[\cdot]$  is a logical function which equals 1 if the argument is true, and 0 otherwise.  $P_1$  is the penalty value for the disparity jump of 1 pixel, which usually happens at slant surfaces.  $P_2$  is the penalty value for disparity jump of more than 1 pixel, which possibly occurs at surface discontinuities.  $P_2$  is usually larger than  $P_1$ . Census cost ([Zabih and Woodfill, 1994](#)) is used for computing the matching cost, which is said to be robust to radiometric difference ([Hirschmuller and Scharstein, 2009](#)), and a  $7 \times 9$  window for Census cost is employed to make full use of the 64-bit integer for computation. Moreover, the parameter tuning is insensitive for Census cost ([Hirschmuller and Scharstein, 2009](#)) since only the local ranking of grey values is used. The value of the Census cost only depends on the window size (for a  $7 \times 9$  window, the range of the cost is [1, 63]). In this paper,  $P_1 = 28$ ,  $P_2 = 50$ , and 8 paths are used for matching.

The hierarchical SGM ([Rothermel et al., 2012](#)) adopts a coarse-to-fine strategy. The images are down-sampled as pyramids, and the matching is performed in each pyramid layer. The disparity range for the finer layer is dynamically determined by the coarser layer, and the matching of the coarsest layer searches all possible disparity values. A big advantage of the hierarchical SGM is its computation and memory efficiency, as only a small disparity range need to be searched. Moreover, the matching is more stable in coarser layer, because large parallaxes created by high-rise buildings become smaller, with less occluded area affecting the global energy defined in equation (1). Thus the hierarchical strategy enables the algorithm to retain the stability from the coarser layer to the finer one.

However, this strategy may omit small buildings for the satellite images in the coarser layer. Therefore, only three pyramid layers are used for matching, where the GSD of the coarsest layer is small enough to keep small buildings (for original image with 1 meter GSD, the GSD of the coarsest layer is 4 meters). Furthermore, to save the memory cost for disparity computation, the searching range is limited to [-250, 250] for the coarsest layer (i.e. [-1000, 1000] for the original epipolar pair), which is large enough for most of the spaceborne stereo pairs.

### 3. Building verification and change detection

A major task of detecting changes on the 3D models is to verify if the same buildings in date 1 exist in date 2, thus the change can be found in non-existing buildings. Therefore, the geometric and textural differences need to be computed to indicate whether the change occurs. Due to the uncertainties of height and texture differences, a multi-channel change indicator is proposed concerning the IGC, height difference (explicit geometric consistency), and texture similarity, and then they are fused into SOM for robust change indication on each face. Concerning the connectivity of faces and objects, MRF is then adopted to propagate change status with neighboring faces.

#### 3.1 Multi-channel change indicators

##### 3.1.1 Inherent geometric consistency

Methods have been proposed to deal with the unreliability of the height comparison. Following a shift-based least squares adjustment, [Tian et al. \(2013\)](#) took the minimal height difference over a

window as robust change indicators, and then they adopted post-filtering techniques to eliminate errors induced by the uncertainties. As the 3D models are reliable in terms of their positioning accuracy, their IGC with the stereo images is considered as a robust measure: by projecting each image of the stereo pair on the building roof faces via bias-corrected RPF, the projected image patches from the two images should be the same if the geometry of the building in both dates is consistent (Fig. 2(a)). A similar idea has been used in the street level with laser scanning point clouds and has achieved satisfactory results (Qin and Gruen, 2014), whereas it required complete geometry of the objects in the scene (e.g. mobile laser scanning point clouds). As discussed in the first section, the 3D building models are the abstractions of the geometry, where facilities or complex man-made structures may not be modeled. Therefore the two image patches projected on the roof may have perspective distortions (Fig. 2(b)), which may fail similarity metrics of image patches that are based on pixel-wise comparison (e.g. Normalized Cross-correlation (NCC), direct subtraction). Moreover, for demolished buildings, the two projected image patches might be similar due to the homogeneity of surrounding area, which might produce false negatives for histogram-based methods (e.g. image patches shown in Fig. 2(c)).

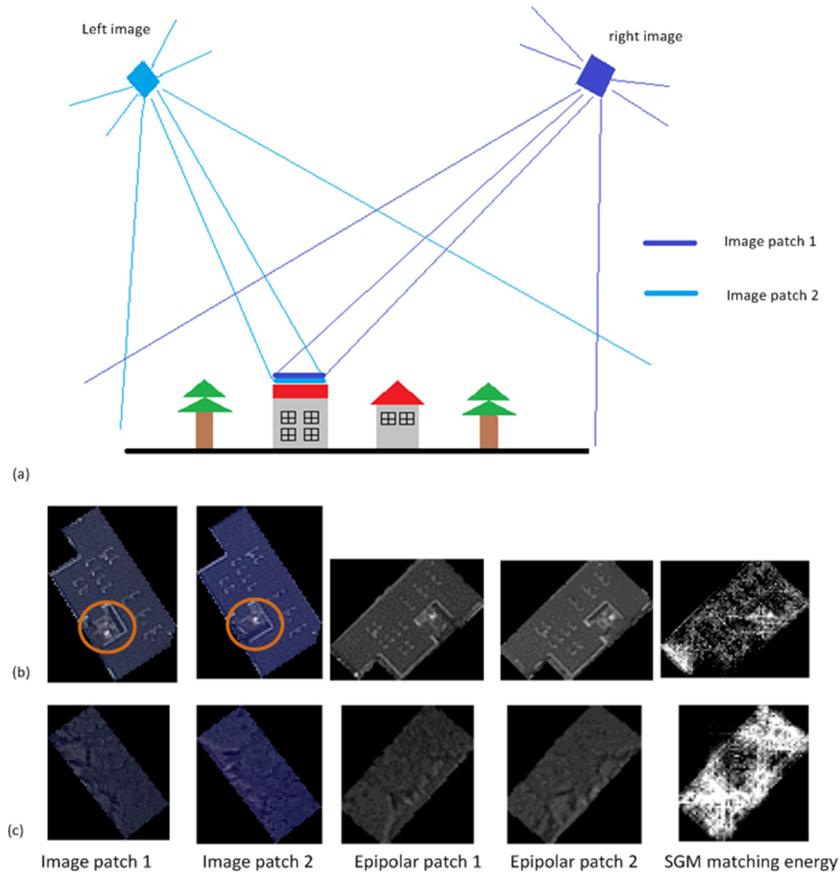


Fig. 2 Inherent geometric consistency and examples. (a) Graphic illustration of the IGC. (b) An example shows that the geometry in date 1 is not correctly modeled, resulting in perspective distortion (marked in circle), and it shows low SGM energy in the patch. (c) An example of demolished building in date 2, and the two textures are projected to surrounding areas with similar spectral. The SGM energy is able to detect the inconsistency of these two patches.

Therefore, a patch based matching strategy (PBMIGC) is proposed for IGC computation, and the idea is to measure the similarity of these two image patches under a matching framework: After a face of a building is projected on each image of the stereo pairs, the corresponding image patches are mapped onto the epipolar images as the epipolar patches (EPPA), and then these two EPPAs are matched by dense matching algorithms (Fig. 2(b, c)). The matching confidence is used to measure the similarity of the two patches, which can be defined as the mean of pixel-wise matching scores (e.g. correlation coefficient of NCC matching). In this work, the classic SGM is adopted for the EPPA matching, and the same parameters of SGM (type of matching cost,  $P_1, P_2$ ) defined in section 2 is adopted. The mean value of the SGM matching energy is computed as the PBMIGC metric. The matching energy of every pixel is denoted as the minimum of the aggregated cost (substantially  $E(D)$  defined in equation (1)), which is calculated by propagating the matching cost (Census cost) with multi-path dynamic programming. The readers may refer details of Census cost to [\(Zabih and Woodfill, 1994\)](#) and details of aggregated cost computation to [\(Hirschmuller, 2005\)](#). The PBMIGC of all the objects are normalized into [0, 1] for subsequent computation. The PBMIGC computation procedure is shown in Fig. 3.

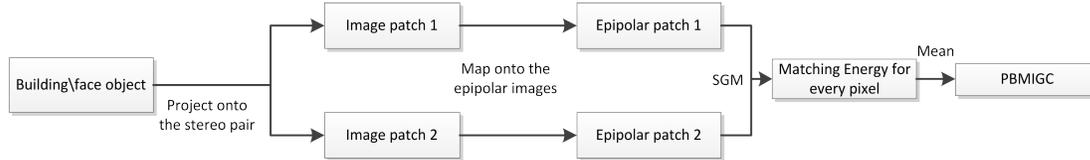


Fig.3 PBMIGC computation.

Matching the EPPA pair is more stable than matching the whole image, for the reason that: 1) each EPPA pair corresponds to a planar face in the 3D model, which implies the disparity range is usually small and continuous; 2) there is very little occlusion in the EPPA pairs, which is one of the main causes of instability of global matching methods [\(Gruen, 2012\)](#). SGM is known to have linear computation time of  $O(ImgHei \times ImgWid \times DispRange)$ , while the disparity range can be set to a small value (e.g. 5-10 pixels) for the EPPA pair, thus the computation is very fast for the patch-wise matching. The PBMIGC metric measures the overall energy of patch-wise matching. It is robust to perspective distortion induced by incomplete modeling of the geometry (Fig. 2(b)), and is able to detect high inconsistency of similar patches (Fig. 2(c)). PBMIGC is a geometric consistency metric between a 3D model and a stereo pair, thus it does not require textures from the 3D models in date 1, and only images patches from the stereo pairs are used to test the geometric difference.

### 3.1.2 Height difference computation

IGC provides one channel of information, and this channel normally produces omission errors, which mainly occur in areas with very high textural similarity. Therefore, it is also necessary to consider the HEIDIF as another channel of the change indicator, which usually produces commission errors (false positive) due to the high uncertainty of the matching algorithm. Before calculating the height difference, an essential step is to co-register two datasets, which is used for eliminating the systematic errors that might affect the HEIDIF computation. The systematic errors are mainly from the geo-referencing, where the GCP locations in the image space are manually measured (with 1-2 pixels measurement error).

Therefore, a least squares 3D surface matching (LS3D) (Gruen and Akca, 2005) is employed to co-register eDSM in date 1 and the matched DSM in date 2. The LS3D method minimizes the squared sum of the Euclidean differences between the two datasets with a 7-parameter similarity transformation ( $X, Y, Z, \omega, \varphi, \kappa$ , scale, and scale are fixed as 1 in this paper). It has an outlier removal procedure that adaptively removes gross errors that are larger than  $K\sigma$ , which is done in each iteration of the minimization.  $\sigma$  is the standard deviation of each iteration and  $K$  is a scalar that controls the final threshold for gross errors. In this paper,  $K$  is set as 5 to suppress errors that come from the changed buildings, modeling outliers and matching errors. An initial  $\sigma = 4$  (meter) is given for the first iteration, which is estimated from the measurement uncertainty ( $2 \times 2$  pixels geo-referencing uncertainty for the two dates, with 1 meter pixel size for IKONOS data).

A  $\sigma_0$  of 1.519 m is obtained by co-registering the IKONOS data (introduced later in section 5.2), with 20.4% gross errors (change buildings, matching errors and modeling outliers) eliminated.

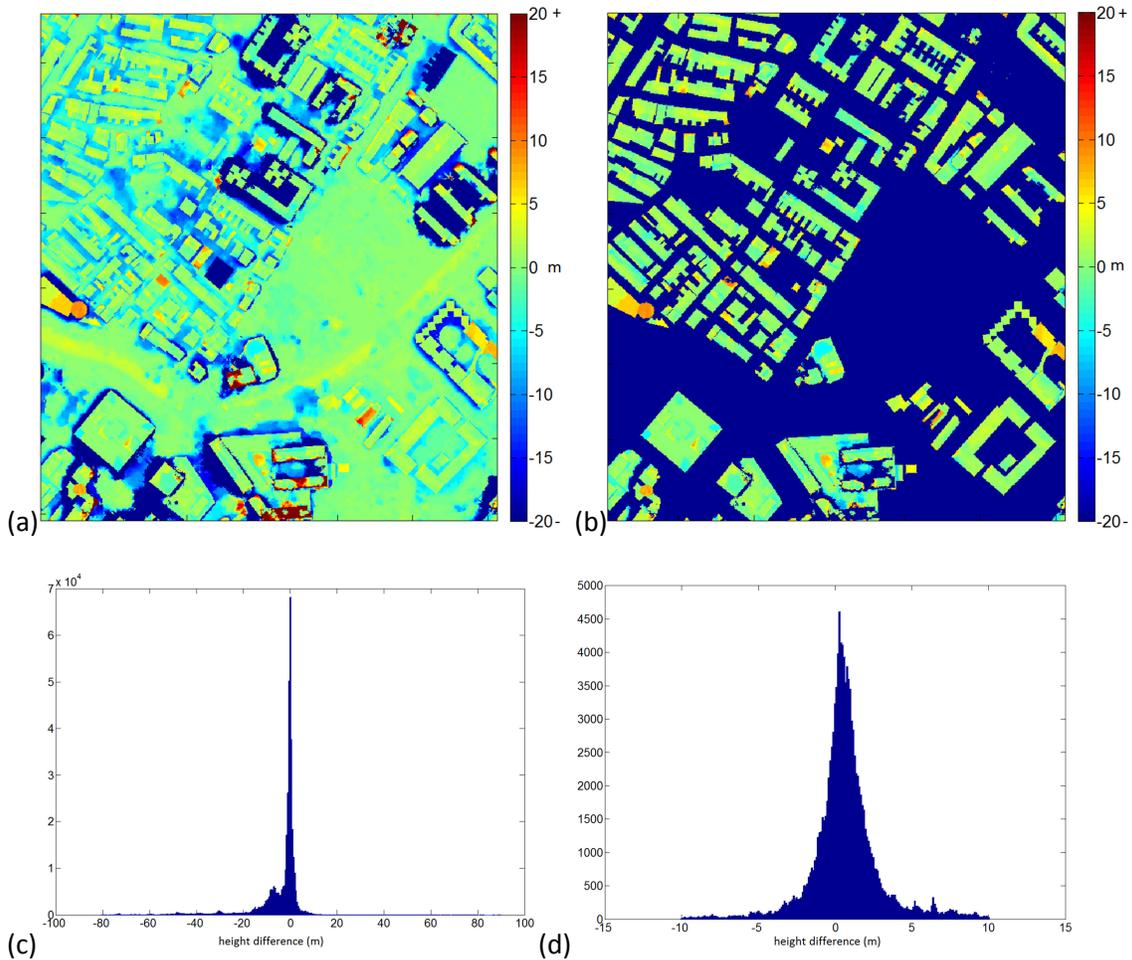


Fig. 4 Accuracy assessment after co-registration (IKONOS data), eDSM – DSM: (a) height differences (RMSE: 11.5 m) (b) height differences only with building objects (RMSE: 2.31 m); (c) the histogram of (a); (d) the histogram of (b).

Fig. 4 shows a small area of height differences, where changed buildings are not considered. Since eDSM is derived from the accurate 3D models, the quality of the matched DSM can then be assessed by

analyzing the height deviation to the eDSM after co-registration in the unchanged part. It can be seen from Fig. 4(a) that most of the errors occur near the boundary of the building, and this is a common problem of the state-of-the-art matching algorithms which impose the continuity constraint on the DSM. These errors lead a high RMSE (Root Mean Square Error) of 11.5 m. Fig. 4(b) shows that the HEIDIF on the buildings is much smaller (with RMSE of 2.31 m), which means the HEIDIF on buildings has only a small number of outliers considering the  $\sigma_0$  of co-registration (1.519 m), thus it can be used as an indicator for building change.

The HEIDIF of each face can be calculated by taking the mean of the absolute height difference on the face. In Fig. 4(d), the histogram of the height differences on the building shows that it approximately follows a normal distribution, while there are still some outliers which may affect the computation of the HEIDIF for each object. A common strategy is to remove points that are larger than the standard deviation when computing the mean value, while sometimes the distribution on each face may not follow the normal distribution. Therefore, a more robust way is used to calculate the HEIDIF of each face: first construct the histogram of the absolute height difference of the face, and then compute the mean value of pixels that took more than 10% of all the total pixels (as shown in Fig. 5, the mean is computed only on the green columns) in the histogram.

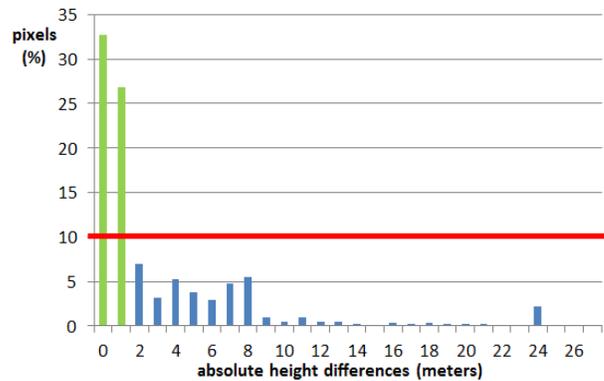


Fig. 5. Histogram of the absolute height difference for a face. Only pixels that take more than 10% are considered (in green).

For normalization purpose, the computed HEIDIF for each face is capped at 20 meters, namely HEIDIF values that are larger than 20 meters will be assigned as 20 meters when being normalized to [0, 1].

### 3.1.3. Texture similarity

If the textures of the 3D models are available, the similarity of texture between different dates could serve as another channel. Owing to the high temporal influences of light, atmosphere and humidity conditions, it is usually difficult to find robust measurement for the similarity. Therefore, although the texture plays an important role for manual interpretation of change, its temporally varying and unpredictable nature prevent it to be robust change information for automated means. Direct subtraction of the texture will result in many false positives if there is significant appearance difference. Instead of pixel-wise similarity measurement, statistical metrics are more preferable to reduce false positives ([Inglada and Mercier, 2007](#)), but they produces false negatives due to the inter-class similarity

([Huang and Zhang, 2008](#)). Structure Similarity (SSIM) ([Wang et al., 2004](#)) is known as a flexible metric which combines both pixel-wise metrics and statistics of local image patches: it considers the luminance  $l(x, y)$ , contrast  $c(x, y)$ , and structure  $s(x, y)$  in a local window with a weighting scheme, which is adjustable due to different imaging conditions. The SSIM is defined as:

$$SSIM(x, y) = [l(x, y)]^a \cdot [c(x, y)]^b \cdot [s(x, y)]^c \quad (2)$$

where

$$l(x, y) = \frac{2\mu_x\mu_y + K_1}{\mu_x^2 + \mu_y^2 + K_1}, c(x, y) = \frac{2\sigma_x\sigma_y + K_2}{\sigma_x^2 + \sigma_y^2 + K_1}, s(x, y) = \frac{\sigma_{xy} + K_3}{\sigma_x\sigma_y + K_1} \quad (3)$$

where  $a, b, c$  are the weighting values that control the significance of each component.  $K_1, K_2$  and  $K_3$  are prior parameters that control the sensitivity of each component, and are usually assigned as very small numbers (e.g.  $10^{-6}$ ) if no prior information is given, which is used to avoid being divided by zeros.  $\mu$  is the mean value of the window,  $\sigma$  is the variance, and  $\sigma_{xy}$  is the covariance of two local windows. The structure element  $s(x, y)$  is similar to NCC, which measures the coherence between the local windows. Since the local contrast is the more robust to luminance variations and co-registration error, a larger contribution is assigned to  $c(x, y)$  (the weight  $b$  should be smaller, as  $l(x, y)$ ,  $c(x, y)$  and  $s(x, y)$  are smaller than 1) by suppressing the other two ( $a = 2, b = 1, c = 2$ ). The relationship among  $a, b$  and  $c$  are important, while their actual value control the final value of  $SSIM$ , which is less important since the texture similarity will be normalized in order to be fused with other channels of the change indicators.

In this work, SSIMs are computed by comparing the texture of each face of the 3D models in date 1 with the two orthoimages generated by each image of the stereo pair in date 2. Then the mean value of the two SSIMs is taken as the texture similarity. The same procedure applies to each multispectral band. It is common that the texture of the 3D models is 8-bit RGB images, thus the multispectral 16-bit images need to be properly stretched accordingly to compute the SSIM. The negative value of the SSIM is seen as the texture difference, which is incorporated into the multi-channel indicator, and values in each channel are normalized into  $[0, 1]$ , which represents the probability of changes.

It should be noted that the SSIM texture similarity measurement is different from the PBMIGC metric: the SSIM measures that appearance difference, which needs textures from both dates, while the PBMIGC measures the geometric difference, and only image patches from the stereo pairs in date 2 is needed.

### 3.2 Fusion of multi-channel indicator with self-organizing map

The self-organizing map (SOM) is known as a neural network that has special properties to create spatially organized “internal representation” of the multi-dimensional input features ([Kohonen, 1998](#); [Moosavi and Qin, 2012](#)). The basic element of SOM is a neuron, which has the same dimension of the input feature, and neurons are spatially related to each other in a pre-defined 2D pattern (rectangle or hexagon lattice). Similar to k-means clustering ([Hartigan and Wong, 1979](#)), the training phase of SOM is an iterative process. K-means clustering iteratively computes the centroids (similar to neuron) of the clusters, where the values of the centroids are calculated independently. Differently, each neuron of

SOM is spatially related, and the training process enables the neurons to interact with each other: for each training sample, the best matching unit (BMU) is obtained by finding the neuron whose distance to the training sample is minimal, and then the value of BMU and its neighboring neurons are updated simultaneously. The detailed algorithm is introduced in (Kohonen, 1998).

The spatially related lattice of neurons presents the input features (in this case, it is the multi-channel change indicator) in a smooth manner, which is easier to find natural clusters robustly (Moosavi and Qin, 2012). Blunders can be absorbed by neighboring neurons during the training process. While in k-means clustering, such blunders might form an independent cluster. Moreover, the prior information can be employed to indicate the significance of each channel by weighting them during the training process (Vesanto et al., 1999). In this work, the contributions of the different channels are weighted as:  $w_{IGC} = 0.5$  for PBMIGC,  $w_{HD} = 0.4$  for height difference (HEIDIF), and  $w_{TD} = 0.1$  for texture difference (TEXDIF).

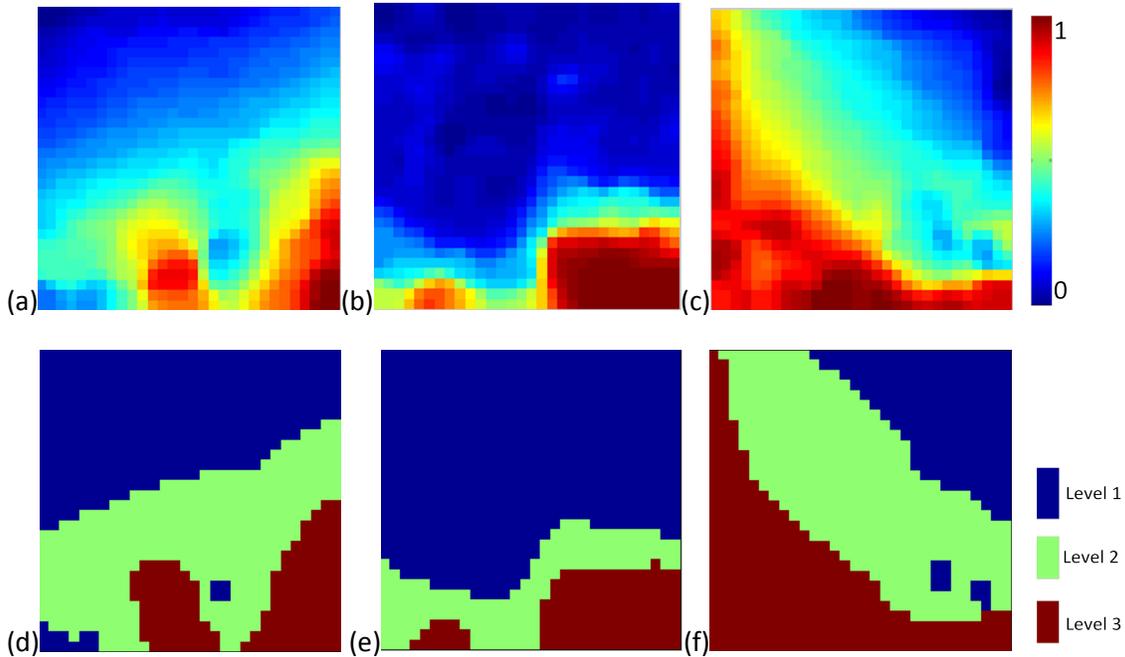


Fig. 6. SOM output for each channel with 30 X 30 neurons. (a) PBMIGC; (b) HEIDIF; (c) TEXDIF; (d),(e),(f) k-means clustering of (a),(b) and (c), respectively. Each channel is clustered into 3 levels of change. Level 1 is non-change, level 2 is uncertain change, and level 3 is change.

The training of SOM produces lattices of neurons for each channel of the change indicator (Fig. 6). In Fig. 6(a-c), each map represents the neuron's value in each channel, and the grids themselves are not important but the coherence between different channels is meaningful: high coherence means these channels support each other, otherwise contradictory. A higher coherence between PBMIGC and HEIDIF is observed (in Fig. 6(a-b)), while TEXDIF is deviated from the other two owing to its high uncertainties. It can be seen that the confliction exists among these channels, that is, one channel has high probability of change while another shows low probability. Therefore, a voting strategy is adopted concerning the contribution of different channels. A k-means method is firstly used to cluster the neurons of each channel into 3 levels, where level 1 and level 3 represent non-change and change, with level 2 in-

between (uncertain change). These levels are then summed up as votes, and 9 votes mean an absolute change and 3 votes mean an absolute non-change. A traffic-light rating system is used to differentiate these votes for operational purpose: neurons less than 5 votes represent non-change, and those that have more than 7 votes are changes; votes from 5-7 indicate that the decision should be made upon the operator.

After the neurons are determined by their votes, the multi-channel indicator of each building faces can be projected on the SOM through a KNN (K nearest neighborhood) method ([Silva and Del-Moral-Hernandez, 2011](#)) (in this work K = 4), and the most frequent status is chosen among the K neurons. For the uncertain neurons, their probability of being a change can be computed as the mean value of all the channels as  $\frac{1}{n} \sum_{i=1}^n C_i$ , where  $n$  is the number of channels and  $C_i$  is the neuron value of the channel  $i$ .

### 3.3 Change determination based on Markov random field analysis

To achieve fully automated performance, the change statuses of the uncertain faces need to be determined. It should be noted that geometrically (such as similar height, position, shape) or texturally similar faces are likely to have the same change status, and such similar faces can be either within the same building or from buildings that are very close to each other (less than 10 meters). In this work, such spatial relationship is applied under MRF analysis ([Blake et al., 2011](#)), where faces with uncertain change status are determined by neighborhood propagation. Let us consider change status of the faces  $X = \{X_1, \dots, X_n\}$  as a random vector, and  $n$  indicates the number of faces. Each random variable is either 0 or 1, representing “non-change” and “change”.  $P(X)$  denotes the *a priori* probability distribution. Based on the connectivity relationship, it forms a hidden Markov model (HMM):

$$P(X_i | X_1, \dots, X_{i-1}, X_{i+1}, X_n) = P(X_i | N(X_i)) \quad (4)$$

where  $N(X_i)$  is defined as the neighboring faces of  $X_i$ , subjected to a graph representation. Let  $Z = \{Z_1, \dots, Z_m\}$  be a set of observations (i.e. the multi-channel indicator), and based on the Bayesian framework, the *a posteriori* probability can be written as:

$$P(X|Z) \propto P(Z|X)P(X) \quad (5)$$

It is common to express this *a posteriori* probability as a sum of energies ([Blake et al., 2011](#)) to generalize the terms of observation and connectivity constraints of MRF:

$$P(X|Z) = \frac{1}{H(z, \omega)} \exp(-E(X, Z, \omega)) \quad (6)$$

where

$$E(X, Z, \omega) = \Phi(X, \omega) + \Psi(X, Z) \quad (7)$$

where  $\omega$  is an abstraction of the MRF model parameter and  $H(z, \omega)$  is attached for normalization purpose. The MAP (maximum a posterior) problem turns to minimizing  $E(X, Z, \omega)$ :

$$\hat{X} = \operatorname{argmin}_X [E(X, Z, \omega)] \quad (8)$$

$\Psi(X, Z)$  models the energy observations, which can be determined from the section 3.2:

$$\Psi(X, Z) = \begin{cases} 0, & \text{determined as change in section 3.2 (with 3 – 4 votes)} \\ 1, & \text{determined as non – change in section 3.2 (with 8 – 9 votes)} \\ 1 - \frac{1}{n} \sum_{i=1}^n C_i, & \text{otherwise (with 5 – 7 votes)} \end{cases} \quad (9)$$

where  $\{C_i\}$  are neuron values of each channel when the multi-channel indicator is projected onto the SOM.  $\Phi(X, \omega)$  models the inherent connectivity of  $\{X_i\}$ , and connectivity relationship  $N(X_i)$  is defined under a given radius  $r$  in 3D for the geometric center  $G(X_i)$  (computed as the mean of 3D coordinates in a face) of every face  $X_i$  and the dependency of faces to the building objects:

$$N(X_i) = \{X_j \mid |G(X_i) - G(X_j)| < r \text{ or } X_j \in B(X_i)\} \quad (10)$$

where  $B(X_i)$  denotes the building object that  $X_i$  belongs to (the building objects are encoded as building index in the eDSM), and  $r = 20$  (meters) in the experiments. Thus  $\Phi(X, \omega)$  can be computed as:

$$\Phi(X, \omega) = \sum_{X_i} \sum_{X_j \in N(X_i)} \omega_{ij} |X_i - X_j| \quad (11)$$

where  $\omega_{ij}$  measures the similarity between two faces. Concerning the high texture variability, and that sometimes textures are not available for 3D models, thus  $\omega_{ij}$  are measured solely on the geometric similarities in a truncated form:

$$\omega_{ij} = T[HvD_{ij} < 10] \left(1 - \frac{|HvD_{ij}|}{10}\right) \times T[AreaDif_{ij} < 100] \left(1 - \frac{|AreaDif_{ij}|}{100}\right) \quad (12)$$

where  $HvD_{ij}$  denotes the difference of height variation of face  $X_i$  and  $X_j$ . The height variance is computed as the variance of the height value of pixels in a face, which describes the inclination of a face.  $AreaDif_{ij}$  denotes the area difference of two faces.  $T[\cdot]$  is a logical function as used in equation (1), which truncates  $HvD_{ij}$  at 10 meters and  $AreaDif_{ij}$  at 100 m<sup>2</sup>, meaning that if either  $HvD_{ij}$  or  $AreaDif_{ij}$  is larger than the given threshold, their connectivity will be lost.

Noting that equation (11) is an *Ising* model ([Winkler, 2003](#)), and it is well proven that problem (8) can be solve by graph cuts (GC) ([Vicente et al., 2008](#)) within polynomial time, where  $\Psi(X, Z)$  and  $\Phi(X, \omega)$  are data term and smooth term, respectively. By give a weight  $\beta$  for the smooth term, GC solves the following problem via a max\min cut algorithms ([Boykov and Kolmogorov, 2004](#)):

$$\hat{X} = \operatorname{argmin}_X [\Psi(X, Z) + \beta \Phi(X, \omega)] \quad (13)$$

The output of GC gives binary labeling for each face over the MRF, where 0 indicates “non-change”, and 1 indicates “change”. In the experiments, the weight  $\beta$  for the smooth term is set to be 0.2.

#### 4. New building detection

The building verification detects changes of the buildings in date 1, and the new buildings exist in date 2 should also be detected. A simple subtraction of eDSM and DSM will result in many false alarms, since not everything is modeled in eDSM. A building object exists in both dates but not modeled in date 1 will result in large height difference. Therefore, the new buildings are extracted by firstly detecting buildings

in date 2, and then excluding the verified unchanged buildings. There has been a large number of building detection methods for the combined use of multispectral information, such as rule-based approaches ([Chen et al., 2012](#); [Grigillo et al., 2011](#)), classification-based approaches ([Lu et al., 2006](#); [Rottensteiner et al., 2007](#); [Turlapaty et al., 2012](#)).

In this work, building detection is closely related to the workflow, but not the focus for change detection with 3D models. Therefore, a simple and effective procedure is proposed to detect building patches using the generated orthoimage and DSM:

- 1) Blob detection based on top-hat reconstruction;
- 2) NDVI (Normalized Difference Vegetation Index) for tree elimination and morphology shadow index (MSI) for shadow patch elimination;
- 3) Noise and irregular structure filtering.

The following subsections briefly introduce the above steps.

#### 4.1 Blob detection

The first step is to detect above-ground objects, and it can be done straightforwardly when DTM is available. Thus the normalized DSM (nDSM) can be computed by subtracting DTM from DSM. The DTM from the 3D models is available for date 1, and it is a valid assumption that the DTM does not change between date 1 and date 2. But to provide a general solution, the above-ground objects are computed only with DSM in date 2, by using the top-hat reconstruction ([Vincent, 1993](#)).

Top-hat reconstruction originates from morphological reconstruction of grey-level images, where a mask image  $J$  is reconstructed as  $B_{JI}$  from a marker image  $I$ , and this is done by finding the peaks of the mask which are marked by the marker  $I$  (the reader may refer to ([Vincent, 1993](#)) for detail.). The top-hat reconstruction is computed by subtracting  $B_{JI}$  from  $J$ , where  $I$  is less than  $J$  pixel-wisely.  $I$  is usually computed by morphology erosion of  $J$  with structural element  $e$ . Thus the top-hat reconstruction  $T_J^e$  can be written as:

$$T_J^e = J - B_{J\varepsilon(J,e)} \quad (14)$$

where  $\varepsilon(J, e)$  is the grey-level morphological erosion, defined as:

$$\varepsilon(J, e)(i, j) = \min\{J(a, b) \mid e(a - i, b - j) = 1\} \quad (15)$$

By applying the top-hat reconstructions on the DSM, one can effectively extract the above-ground pixels (Fig. 7(c)). In this work, a disk-shaped element  $e$  is used, the radius of which is defined as the largest building radius. This radius can be approximated via the 3D models in date 1 (in the experiments, 100 m is used as the radius). As mentioned in equation (14) and (15), grey level morphology erosion based on the shape element  $e$  is first applied to the DSM by taking the minimum of DSM for every pixel, denoted as  $I$ . Then a morphology reconstruction  $B_{JI}$  is carried out on  $I$  by finding the maximal possible value  $g$  on the DSM, so that for two logical images  $b_1 = \{\text{DSM} > g\}$  and  $b_2 = \{I > g\}$ ,  $b_1$  can be marked by  $b_2$ . Finally the top-hat can be obtained by subtracting  $B_{JI}$  from the DSM. After this, the initial

above-ground objects are binarized by truncating the top-hat pixels that is higher than 2 meters. Given the  $\sigma_0$  of 1.519 meters of the co-registration, this threshold is an empirical value which prevents low buildings being omitted, and the wrongly detected blobs due to matching errors will be eliminated by subsequent refining procedure, as such blobs usually have irregular shapes.

#### 4.2 NVDI and MSI extraction

The orthoimage is generated based on the matched DSM, and NDVI is extracted based on the multispectral information, where an empirical value (0.2) is used to truncate them as vegetation pixels. As a strong characteristic of shadow is its strong variance in luminance, MSI ([Huang and Zhang, 2012](#)) is computed by performing the top-hat reconstruction on the negative panchromatic orthoimage. Dark blobs are then detected as shadows by truncating the top-hat reconstruction. Both vegetation and shadow pixels are eliminated from the above-ground mask computed in section 4.1.

#### 4.3. Noise and irregular structure filtering

The pixel-wise mask produces many isolated pixels, which is problematic for object-based analysis. A binary morphology opening and closing operation is adopted using a disk-shaped structural element with 2-pixel radius ([Meng et al., 2009](#)). Then, the binary mask is segmented by a fast segmentation method with 4-neighborhood connectivity ([Davies, 2004](#)), where each segment is identified as an object.

The above-ground objects in an urban scenario are not necessarily building objects, some are suspending railways, bridges, etc. and these objects might not be included in the 3D building models. Therefore, such objects are filtered out based on their geometric attributes: elongation and expansion, and they are computed as:

$$\text{elongation} = \frac{\text{Marjor axis length}}{\text{Minor axis length}}$$

$$\text{expansion} = \text{sqrt}[(\text{Marjor axis length})^2 + (\text{Minor axis length})^2] \quad (16)$$

where the major axis and minor axis length refer to that of the fitted ellipse of the segments, and elongation reveals the thinness of the segments. Expansion shows to the lateral extension of the segments – how long does the building cover through. Segments whose elongation and expansion are larger than given thresholds will be eliminated, and the thresholds are computed by taking the largest elongation and expansion values of the 3D models in date 1. Fig. 7(e-f) shows an example of the geometric and morphological filtering, where the isolated pixels are filtered, and threadlike structures (suspending rail ways) are eliminated by geometric filtering.

In a final step, the filtered mask of each objects intersect with the verified unchanged buildings described in section 3.2. Segments being covered more than 70% are seen as consistent with the unchanged building in date 1, which will be eliminated from the new building mask. An example of new building detection is shown in Fig. 7, and the new building mask can be used to mark the DSM, thus to perform the volumetric analysis on the new buildings. Moreover, the intersection between the new building and the changed buildings of the 3D models (described in section 3) can differentiate the types

of change between demolition and rebuilding: changed buildings overlapped with detected new buildings are reconstructed buildings, while the others are demolished buildings.

This building detection procedure is parameter free, as most of threshold values are estimated from the 3D building models.

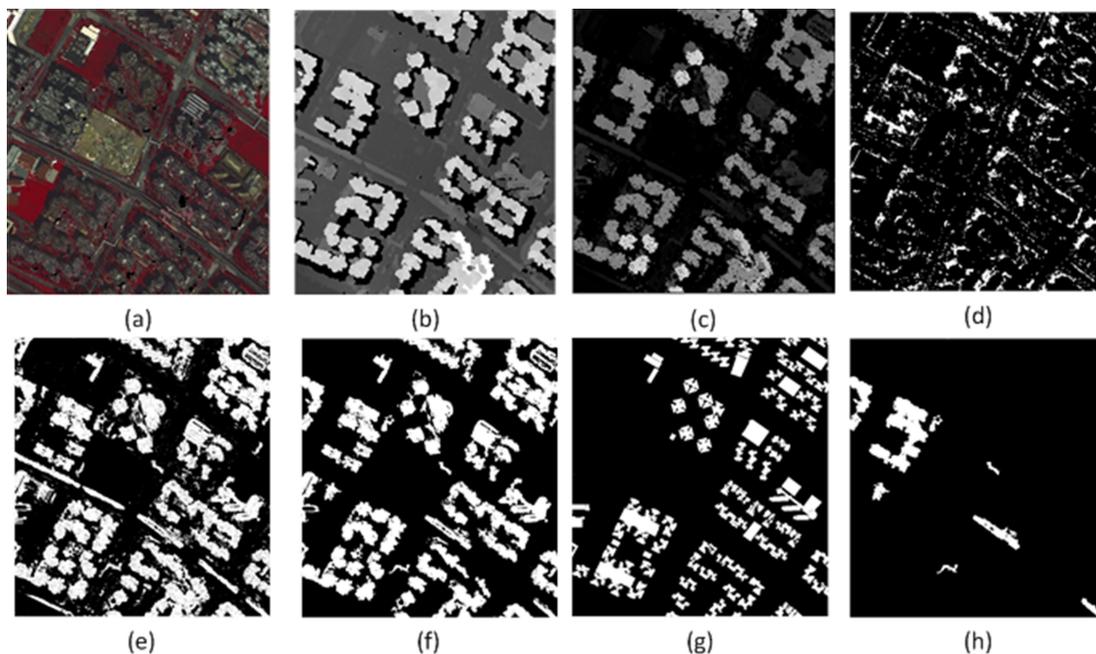


Fig. 7. New building detection procedure. (a) False-color multispectral orthoimage; (b) DSM; (c) Top-hat reconstruction of the DSM; (d) detected shadow with MSI; (e) Building mask truncated with NDVI and MSI; (f) building mask after morphology filtering and geometric filtering; (g) verified non-change building from date 1; (h) new building by excluding the non-change buildings.

## 5. Experiment and Analysis

Both the synthetic and real data experiments are performed to validate the methodology. Worldview-2 stereo pair is used for the synthetic experiment, where only one date is available, and this is under the scenario of quality control on the 3D models. IKONOS stereo pairs are adopted to conduct the real data experiment, and this is under the application of model updating, building change assessment. The reference data of change is manually marked by careful inspection, both pixel-wise and object-wise evaluation are carried out by measuring the TPR (true positive rate), FPR (false positive rate), FNR (false negative rate), correctness, Kappa Coefficient (KC), which are computed as:

$$\begin{aligned}
 TPR &= TP/GP, FPR = FP/PD, FNR = FN/GP \\
 \text{Correctness} &= TP/(TP + FP) \\
 KC &= \frac{(TP+TN)/Npix-M}{1-M}, M = \frac{(TP+FP) \times (TP+FN) + (FN+TN) \times (FP+TN)}{Npix \times Npix} \quad (17)
 \end{aligned}$$

where  $Npix$  denotes the number of pixels in the area of interest (AOI).  $TP, FP, FN, TN$  denote true positives, false positives, false negatives, true negatives, respectively.  $GP, PD$  represent ground truth

positives and detected positives. *TPR* describes the completeness of the detection, *FPR* describes the commission errors, and *FNR* describes the omission errors.

In the general cases of change detection, the 3D models can be available in an earlier date or a later date, thus the models can be updated either to a newer situation or to a historical status. For notional convenience, the date when 3D models are available is referred as date 1, and the one with stereo images is referred as date 2, regardless of their chronological order.

### 5.1. Synthetic experiment

The purpose of the synthetic experiment was to test the proposed method in a controlled situation, where the misregistration between the models and the stereo images only depended on the manual measurement of the model. The Worldview-2 in-track stereo pair with 4-band multispectral channel was used, and the images were stretched to 8-bit depth by the data provider. 4 GCPs were used for geo-referencing. The data covers newly developed area in the city of Singapore, with 12.68 km<sup>2</sup>, where there are still some buildings under construction. 982 building were measured semi-automatically using CC-modeler ([Gruen and Wang, 1998](#)) software, while the under-constructed buildings were not modeled.

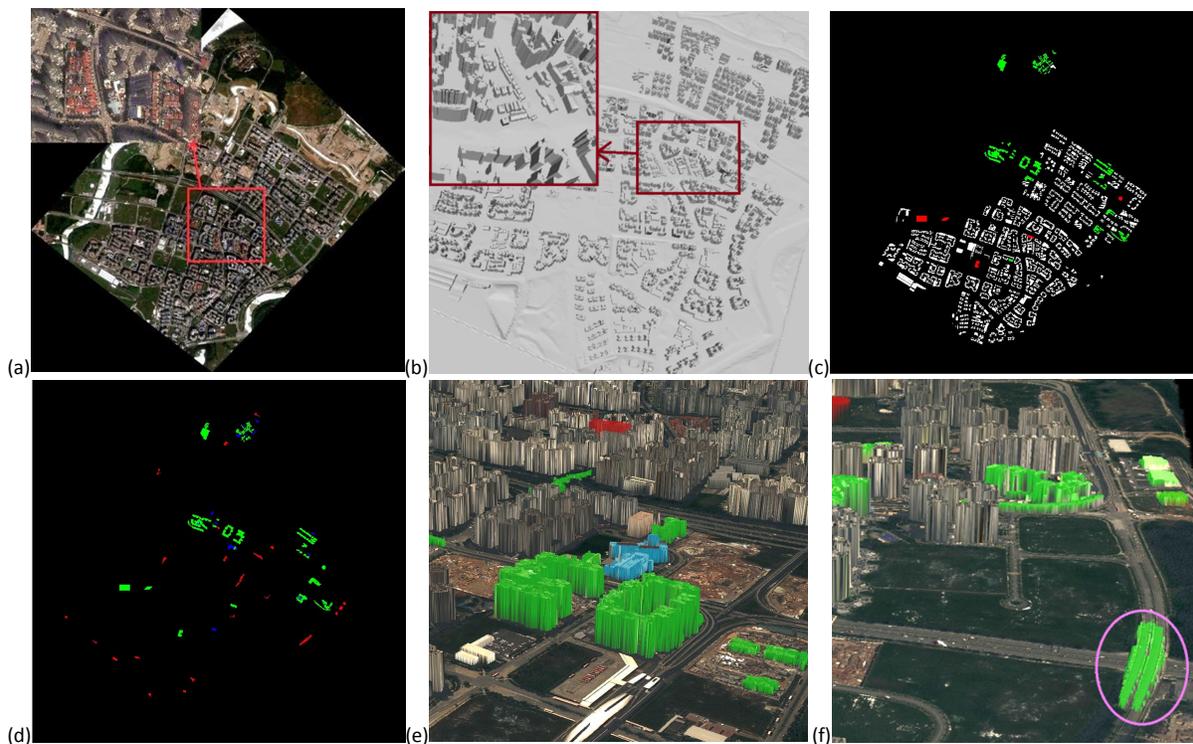


Fig. 8 Results of the synthetic experiment. (a) Left image of the stereo pair; (b) measured 3D building models; (c) reference data, red color indicates changed buildings detected from the 3D models, and green color represents new buildings detected in date 2; (d) detection results, green: TP, red: FP, blue: FN; (e) 3D visualization of detected changes: red: demolition from date 1 to date 2; green: new building from date 1 to date 2; blue: rebuilding from date 1 to date 2; (f) an example of a bridge that is wrongly identified as a new building (marked in circle).

This experiment aimed to detect the existing under-constructed buildings, which had been finished to a large extent but not modeled. Moreover, some artificial buildings were added, some existing buildings were deleted, and the height of some buildings were intentionally changed in the 3D models to test the

propose method. Therefore, the modified 3D models were referred as from date 1, and the stereo images were from date 2.

Table 1. Change detection results of the synthetic experiment

(pixel-based)	<i>TPR</i>	<i>FPR</i>	<i>FNR</i>	correctness	<i>KC</i>	(object-based)			
						<b>NO</b>	<b>NCOGT</b>	<b>NCOCD</b>	<b>Rate</b>
Building verification	0.819	0.147	0.180	0.852	0.889	5034	6	5	0.833
New building detection	0.959	0.200	0.093	0.799	0.946	5067	33	22	0.667
Overall	0.934	0.192	0.109	0.807	0.936	5067	39	27	0.692

(**NO**: number of objects (faces or new building segments); **NCOGT**: number of changed objects (ground truth); **NCOCD**: number of changed objects that are correctly detected)

The main challenge of this experiment is the highly building-like bridges, and segments of overhanging railways, which are likely to be detected as buildings. The change detection results are shown in Fig. 8. It can be seen that most of the manually added buildings are correctly detected (Fig. 8(c-d)), while the statistics in Table. 1 shows that some of the new buildings are not detected in the object-based evaluation, which is mainly caused by their small size. The correctness of the building verification is higher than the new building detection, while the building detection produces more false alarms. For the statistics of building verification,  $N_{pix}$  is the number of buildings pixels in the eDSM, and for the new building detection and overall change detection,  $N_{pix}$  is all the possible pixels within the AOI, excluding the occluded pixels from the image matching.

## 5.2. Real dataset experiment

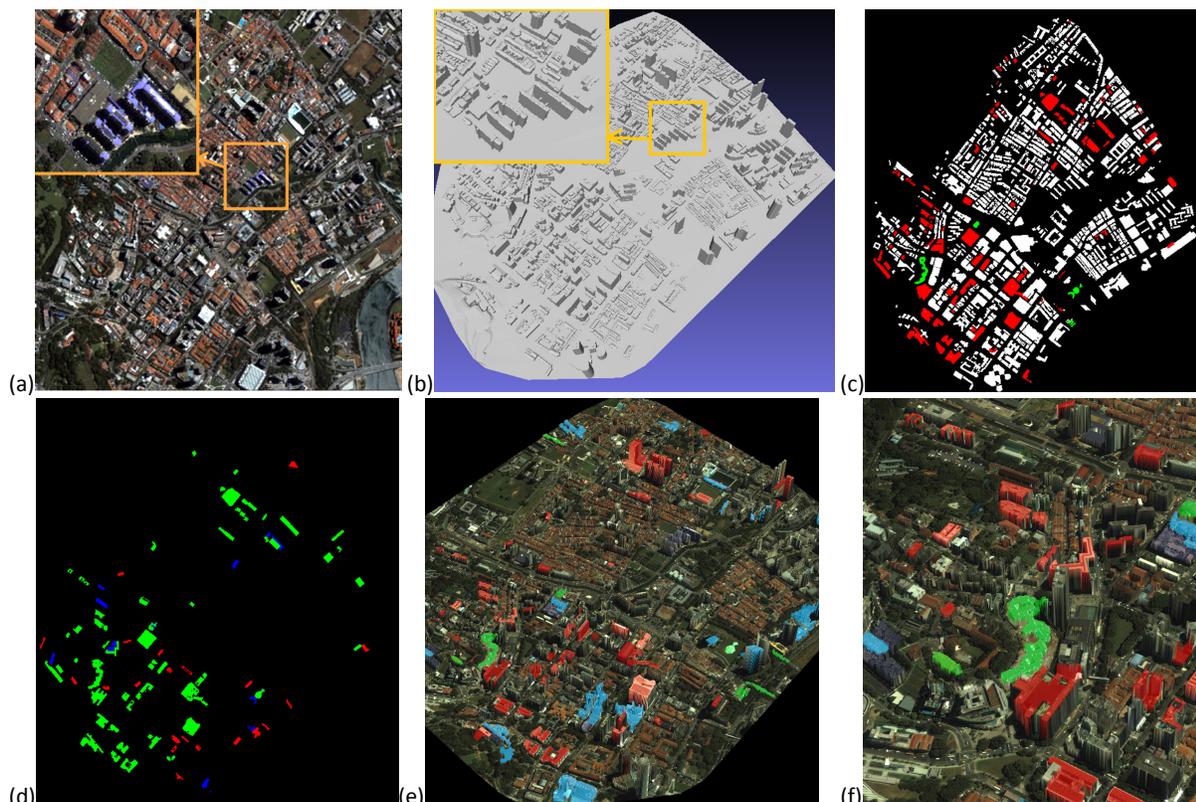


Fig. 9. Real data experiment with IKONOS data. (a) One of the stereo pairs in 2010;(b) building models; (c) reference data, red color indicates the detected changes from date 1, and green color indicates the buildings that existing in date 2 (not in date 1); (d) detection results. Green: TP, red: FP, blue: FN; (e) 3D visualization of detected changes: red: demolition from date 1 to date 2; green: building in date 2 but not in date 1; blue: rebuilding from date 1 to date 2; (f) enlarged view of (e).

The study area locates in Rochor area in Singapore, covering 4.2 km<sup>2</sup>, with varying building types, which results in high differences on building shapes, scales, and height. 927 buildings were modeled by professional operators in a surveying company using the IKONOS pairs (with 1 meter GSD) in 2010, and the IKONOS pairs in 2002 were used to update the model to an earlier stage for historical analysis. Both of the two stereo pairs were bias-corrected using 4 GCPs. The challenge of this dataset is its high building density (ca. 217.19 building objects/ km<sup>2</sup>) and different size of buildings, and some of them only take tens of pixels when projecting back to the IKONOS images, which easily cause false positives by noise induced by the DSM and co-registration. As discussed before, year 2010 was denoted as date 1, and year 2002 was denoted as date 2, which fitted to the scenario of updating model to a historical situation.

The detection results are shown in Fig. 9. Most of the changes are correctly detected, with a few false positives (commission errors) and false negatives (omission errors). To take out the possible disturbance of the noise due to the small segment size, only buildings that are larger than 200 m<sup>2</sup> are evaluated in Table 2:

Table 2. Change detection results of the real dataset experiments

(pixel-based)	<i>TPR</i>	<i>FPR</i>	<i>FNR</i>	correctness	<i>KC</i>	(object-based)			
						<b>NO</b>	<b>NCOGT</b>	<b>NCOCD</b>	<b>Rate</b>
Building verification	0.901	0.105	0.098	0.894	0.936	3849	95	78	0.821
New building detection	1	0.32	0	0.678	0.813	3852	3	3	1
Overall	0.927	0.128	0.093	0.871	0.946	3852	98	81	0.826

(**NO**: number of objects (faces or new building segments); **NCOGT**: number of changed objects (ground truth); **NCOCD**: number of changed objects that are correctly detected)

Table 2 reveals high detection accuracy in terms of *TPR* and *KC*, which means most of the changes are detected. There are still some errors for the small changed building segments. The commission errors mainly occur to small isolated buildings, and omission errors happen to small buildings that are very closed to large buildings. This is mainly due to the inaccuracy of the matching algorithm on small buildings: the continuity constraint omits the small buildings while creates extension from the borders of the large buildings, thus the omitted small building segments cause false positives and the extended building segments lead to false negatives. The change indicators become less reliable due to the lack of textural content of small building patches and the increased noise level, which also produce false positives.

### 5.3. Parameter analysis

There are several tunable parameters in the proposed method, and some of them are trivial and self-explanatory, such as the threshold on the histogram for computing the height differences described in section 3.1.2, and height threshold in the top-hat reconstruction of DSM for computing the initial building mask mentioned in section 4.1. These parameters can be seen as constants, as these parameters are data independent. The number of neurons for the SOM is usually problem-related, in our case, the number of neurons (30 × 30) is an empirical value according to repetitive tests, as it is

able to show clear clusters and at the same time requires acceptable computation time. This parameter (number of neurons for the SOM) is not sensitive to different data and can be fixed for the proposed method.

The most important parameters are the data-related prior parameters which controls the fusion and MRF propagation, whose sensitivities are crucial to the transferability of the proposed method: 1) weight of each channels  $w_{IGC}$ ,  $w_{HD}$  and  $w_{TD}$  when fused into SOM; 2) threshold of voting determining the “change”, “non-change”, “uncertain” in section 3.2; 3) the smooth term  $\beta$  in equation (13). Based on the real data experiment, Fig. 8 shows the influence of different parameters.

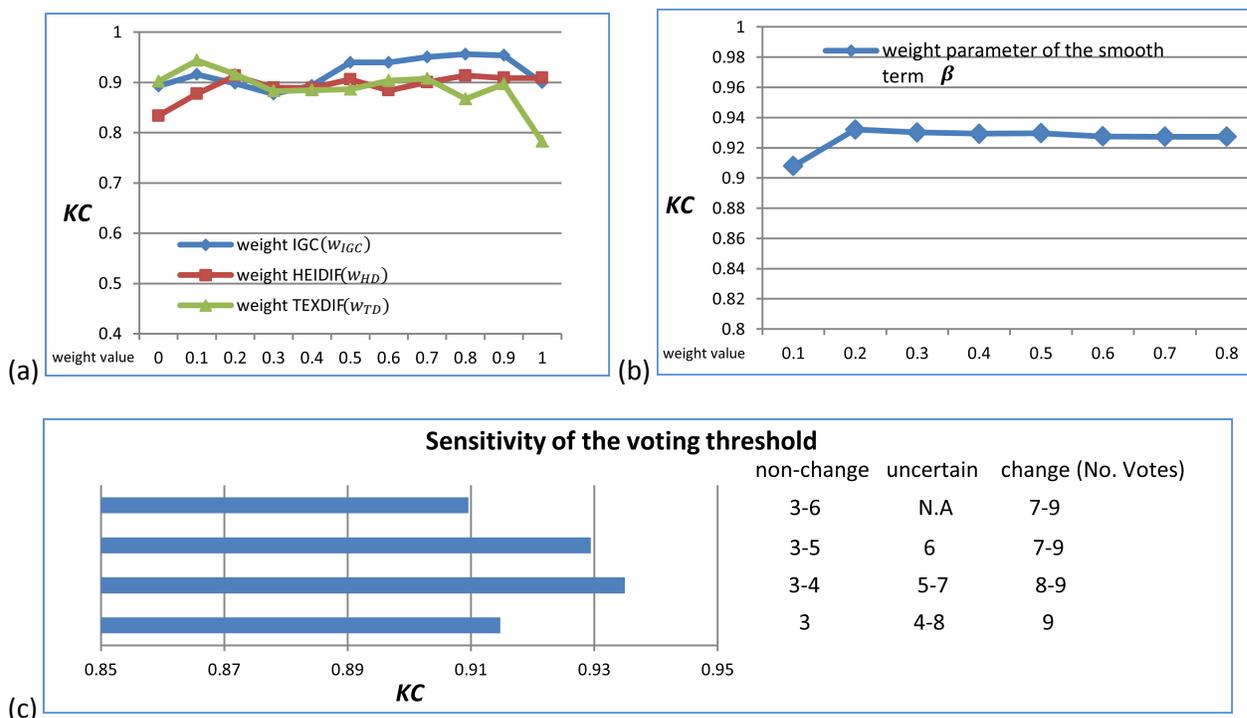


Fig. 10. Sensitivity of different parameters: (a) the sensitivity of  $w_{IGC}$ ,  $w_{HD}$  and  $w_{TD}$ ; (b) the sensitivity of  $\beta$  in equation (13); (c) the sensitivity of different voting threshold configuration.

Fig. 10(a) shows the sensitivity of  $w_{IGC}$ ,  $w_{HD}$  and  $w_{TD}$  with regard to final  $KC$ . Noting that the sum of these parameters is 1, thus by adjusting one of them from 0 to 1, the other two parameters share the rest amount (e.g.  $w_{IGC} = 0.2$ , then  $w_{HD} = w_{TD} = \frac{1-0.2}{2} = 0.4$ ). When one of the weights is assigned as 1, the contribution of other channels will not be considered. Fig. 10(a) demonstrates that the highest  $KC$  is obtained when  $w_{IGC} = 0.8$ , whilst the lowest  $KC$  is observed when considering the TEXTDIF alone. When considering the IGC or HEIDIF alone (when  $w_{HD} = 1$  or  $w_{TD} = 1$ ), this method achieves approximately the same  $KC$ . The  $KC$  is relatively stable when  $w_{IGC}$  vary from 0.5 to 0.9, which shows the robustness of tuning this parameter. As the highest  $KC$  is observed when  $w_{IGC} = 0.8$ , it reveals that the optimal performance should combine all the channels, and the IGC channel is dominant in this experiment.

In equation (13), a stronger constraint on neighborhood connectivity is posted when the smooth term  $\beta$  is larger. Fig. 10(b) shows that the highest  $KC$  is observed when  $\beta = 0.2$ , and the  $KC$  slightly decreases as  $\beta$  increases from 0.2, which is relatively stable for parameter tuning.

Fig. 10(c) shows the relationship between the  $KC$  and the voting threshold as described in section 3.2. As shown on the right-hand side of Fig. 10(c), the voting threshold configuration is based on the bandwidth (the range of the voting threshold) of the “uncertain” category (ascending from the first to the fourth row). The highest  $KC$  is observed when the bandwidth of “uncertain” is 3 (5-7 votes), and decreases when the bandwidth is too large (4-8 votes). This implies the MRF propagation procedure can effectively improve the change detection results than a simple cut-off threshold, but the bandwidth of the “uncertain” faces for MRF propagation should not be too large, as faces with higher probability of “non-change” or “change” might be inappropriately affected by the neighborhood constraints.

## 6. Conclusion

In this paper, a robust changed detection method is developed to detect changes directly from LOD 2 3D models in face-level with VHR stereo images. The proposed method first computes a multi-channel indicator for each face of the 3D building models concerning the IGC, HEIDIF and TEXDIF, and then these channels are fused with SOM for robust analysis. A voting strategy is used on the SOM to classify the faces as “change”, “non-change” and “uncertain change”, and these uncertain changes are determined through a MRF analysis considering the neighborhood connectivity of the building faces. New buildings are detected by combining the DSM and multispectral orthoimage. The proposed method has addressed and utilized the characteristics 3D building models, and the experiments demonstrate that it has archived satisfactory results.

In general, the contribution of this paper lies in the following aspects:

1. A new framework for change detection fusing multi-channel indicator is developed for change detection on 3D models. The proposed method treats each face of the 3D model as a change unit, and MRF analysis is adopted to interpret the spatial and contextual connectivity of the building faces.
2. A new IGC metric PBMIGC is introduced to robustly deal with uncertainties and the incompleteness characteristics of the 3D models.
3. Both synthetic and real data experiments with large amount of buildings in complex urban scenarios are conducted to illustrate the effectiveness of the proposed method, and the sensitivity of parameters is carefully studied to demonstrate the transferability of this method.

The proposed method is robust for building verification, however the synthetic experiment still shows false detection of non-building structures, which mainly attribute to the building detection method. Most of the change detection errors occur to small buildings, due to the inaccuracy of the matcher, the lack of textural content and the increased impact of noise on small building patches. Thus better post-filter method for building detection and matching algorithm on small building patches should be explored. Moreover, the matched DSM in date 1 can also be used to eliminate objects that are not modeled. The future work will focus on the development of new building detection method under the

framework of change detection, where the characteristics and features of the buildings can be learned from the existing 3D models. The development of matching algorithm that utilizes scale-adaptive strategies for improving the accuracy on small buildings is also planned. Furthermore, different types of data and multi-channel indicator and their reliability will be analyzed with the proposed framework.

## Acknowledgement

This work was established at the Singapore-ETH Centre for Global Environmental Sustainability (SEC), co-funded by the Singapore National Research Foundation (NRF) and ETH Zurich. I would like to thank the Digital Globe for providing the Worldview and IKONOS images, Dr. Heiko Hirschmueller for sharing technique details of Census cost implementation, Mr. Tangwu Du for combining some open source codes for some basic RPC processing, Mr. Vahid Moosavi for the helpful discussion of SOM, Dr. Devrim Akca for providing the LS3D software. I would like to thank the anonymous reviewers for their valuable comments and suggestions. I would like to express my particular thanks to my advisors Prof. Armin Gruen for his valuable advice and Prof. Gerhard Schmitt for his support to this work.

## References

- Akca, D., 2007. Matching of 3D surfaces and their intensities. *ISPRS Journal of Photogrammetry and Remote Sensing* 62 (2), 112-121.
- Akca, D., M. Freeman, I. Sargent and A. Gruen, 2010. Quality assessment of 3D building data. *The Photogrammetric Record* 25 (132), 339-355.
- Bazi, Y., L. Bruzzone and F. Melgani, 2005. An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. *IEEE Transactions on Geoscience and Remote Sensing* 43 (4), 874-887.
- Blake, A., P. Kohli and C. Rother, 2011. *Markov random fields for vision and image processing*. The MIT Press.
- Bouziani, M., K. Goïta and D.-C. He, 2010. Automatic change detection of buildings in urban environment from very high spatial resolution images using existing geodatabase and prior knowledge. *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (1), 143-153.
- Boykov, Y. and V. Kolmogorov, 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (9), 1124-1137.
- Chaabouni-Chouayakh, H. and P. Reinartz, 2011. Towards automatic 3D change detection inside urban areas by combining height and shape information. *Photogrammetrie-Fernerkundung-Geoinformation* 2011 (4), 205-217.
- Chen, L., S. Zhao, W. Han and Y. Li, 2012. Building detection in an urban area using lidar data and QuickBird imagery. *International Journal of Remote Sensing* 33 (16), 5135-5148.
- Collins, J. B. and C. E. Woodcock, 1996. An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data. *Remote Sensing of Environment* 56 (1), 66-77.
- Davies, E. R., 2004. *Machine Vision: Theory, Algorithms, Practicalities*. Morgan Kaufmann. 934 p.
- Doxani, G., K. Karantza and M. Tsakiri-Strati, 2010. Automatic change detection in urban areas under a scale-space, object-oriented classification framework. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII (Part4/C7)*, 7 p. (on CDROM).
- Fraser, C. S. and H. B. Hanley, 2003. Bias compensation in rational functions for IKONOS satellite imagery. *Photogrammetric engineering and remote sensing* 69 (1), 53-58.
- Fung, T., 1990. An assessment of TM imagery for land-cover change detection. *IEEE Transactions on Geoscience and Remote Sensing* 28 (4), 681-684.
- Gong, P., Greg.S. Biging and R. Standiford, 2000. Technical Note: Use of Digital Surface Model for Hardwood Rangeland Monitoring. *Journal of Range Management* 53 (6), 622-626.
- Grigillo, D., M. Kosmatin Fras and D. Petrovič, 2011. Automatic extraction and building change detection from digital surface model and multispectral orthophoto. *Geodetski Vestnik* 55 (1), 28-45.

Gröger, G., T. Kolbe and A. Czerwinski, 2007. Candidate OpenGIS® CityGML Implementation Specification (City Geography Markup Language). Open Geospatial Consortium Inc, OGC

Gruen, A., 2012. Development and status of image matching in photogrammetry. *The Photogrammetric Record* 27 (137), 36-57.

Gruen, A. and D. Akca, 2005. Least squares 3D surface and curve matching. *ISPRS Journal of Photogrammetry and Remote Sensing* 59 (3), 151-174.

Gruen, A. and X. Wang, 1998. CC-Modeler: a topology generator for 3-D city models. *ISPRS Journal of Photogrammetry and Remote Sensing* 53 (5), 286-295.

Hartigan, J. A. and M. A. Wong, 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28 (1), 100-108.

Hirschmuller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 807-814.

Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (2), 328-341.

Hirschmuller, H. and D. Scharstein, 2009. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (9), 1582-1599.

Huang, X. and L. Zhang, 2008. An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* 46 (12), 4173-4185.

Huang, X. and L. Zhang, 2012. Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (1), 161-172.

Hussain, M., D. Chen, A. Cheng, H. Wei and D. Stanley, 2013. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing* 80 (2013), 91-106.

Inglada, J. and G. Mercier, 2007. A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis. *IEEE Transactions on Geoscience and Remote Sensing* 45 (5), 1432-1445.

Johnson, R. D. and E. Kasischke, 1998. Change vector analysis: a technique for the multispectral monitoring of land cover and condition. *International Journal of Remote Sensing* 19 (3), 411-426.

Jun, C., L. Miao, X. Chen, J. Chen and L. Chen, 2013. A spectral gradient difference based approach for land cover change detection. *ISPRS Journal of Photogrammetry and Remote Sensing* 85 (2013), 1-12.

Jung, F., 2004. Detecting building changes from multitemporal aerial stereopairs. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (3), 187-201.

Kamini, J., S. C. Jayanthi and V. Raghavswamy, 2006. Spatio-temporal analysis of land use in urban mumbai-using multi-sensor satellite data and gis techniques. *Journal of the Indian Society of Remote Sensing* 34 (4), 385-396.

Knudsen, T. and B. P. Olsen, 2003. Automated change detection for updates of digital map databases. *Photogrammetric Engineering and Remote Sensing* 69 (11), 1289-1296.

Kohonen, T., 1998. The self-organizing map. *Neurocomputing* 21 (1), 1-6.

Lu, Y. H., J. C. Trinder and K. Kubik, 2006. Automatic building detection using the Dempster-Shafer algorithm. *Photogrammetric Engineering and Remote Sensing* 72 (4), 395-403.

Martha, T. R., N. Kerle, V. Jetten, C. J. van Westen and K. V. Kumar, 2010. Landslide volumetric analysis using Cartosat-1-derived DEMs. *IEEE Geoscience and Remote Sensing Letters* 7 (3), 582-586.

Meng, X., L. Wang and N. Currit, 2009. Morphology-based building detection from airborne LiDAR data. *Photogrammetric Engineering and Remote Sensing* 75 (4), 437-442.

Moosavi, S. V. and R. Qin, 2012. A New Automated Hierarchical Clustering Algorithm Based on Emergent Self Organizing Maps. In: *16th International Conference on Information Visualisation (IV)* pp. 264-269.

Moser, G. and S. B. Serpico, 2009. Unsupervised change detection from multichannel SAR data by Markovian data fusion. *IEEE Transactions on Geoscience and Remote Sensing* 47 (7), 2114-2128.

Murakami, H., K. Nakagawa, H. Hasegawa, T. Shibata and E. Iwanami, 1999. Change detection of buildings using an airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (2), 148-152.

Pacifici, F., F. Del Frate, C. Solimini and W. J. Emery, 2007. An innovative neural-net method to detect temporal changes in high-resolution optical satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing* 45 (9), 2940-2952.

Qin, R., J. Gong and C. Fan, 2010. Multi-frame Image super-resolution based on knife-edges. In: *2010 IEEE 10th International Conference on Signal Processing (ICSP)*, pp. 972-975.

Qin, R., J. Gong, H. Li and X. Huang, 2013. A Coarse Elevation Map-based Registration Method for Super-resolution of Three-line Scanner Images. *Photogrammetric Engineering and Remote Sensing* 79 (8), 717-730.

Qin, R. and A. Gruen, 2014. 3D change detection at street level using mobile laser scanning point clouds and terrestrial images. *ISPRS Journal of Photogrammetry and Remote Sensing* 90 23-35.

Rogan, J., J. Franklin and D. A. Roberts, 2002. A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery. *Remote Sensing of Environment* 80 (1), 143-156.

Rothermel, M., K. Wenzel, D. Fritsch and N. Haala, 2012. SURE: Photogrammetric Surface Reconstruction from Imagery. In: *Proceedings LC3D Workshop, Berlin*, pp.

- Rottensteiner, F., G. Sohn, J. Jung, M. Gerke, C. Baillard, S. Benitez and U. Breitkopf, 2012. The ISPRS benchmark on urban object classification and 3D building reconstruction. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* I-3 293-298.
- Rottensteiner, F., J. Trinder, S. Clode and K. Kubik, 2007. Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis. *ISPRS Journal of Photogrammetry and Remote Sensing* 62 (2), 135-149.
- Shafer, G., 1976. *A Mathematical Theory of Evidence*. Princeton University Press.
- Silva, L. A. and E. Del-Moral-Hernandez, 2011. A SOM combined with KNN for classification task. In: *The 2011 International Joint Conference on Neural Networks* pp. 2368-2373.
- Song, C., C. E. Woodcock, K. C. Seto, M. P. Lenney and S. A. Macomber, 2001. Classification and change detection using Landsat TM data: when and how to correct atmospheric effects? *Remote Sensing of Environment* 75 (2), 230-244.
- Swisstopo, 2014. Swiss Simple 3D City Models, <http://www.swisstopo.admin.ch/internet/swisstopo/en/home/products/landscape/swissBUILDINGS3D.html>. (Accessed on January 21, 2014)
- Tian, J., S. Cui and P. Reinartz, 2013. Building Change Detection Based on Satellite Stereo Imagery and Digital Surface Models. *IEEE Transaction on Geoscience and Remote Sensing* 52 (1), 406 - 417.
- Tian, J., P. Reinartz, P. d'Angelo and M. Ehlers, 2013. Region-based automatic building and forest change detection on Cartosat-1 stereo imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 79 226-239.
- Turlapaty, A., B. Gokaraju, Q. Du, N. H. Younan and J. V. Aansoos, 2012. A hybrid approach for building extraction from spaceborne multi-angular optical imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (1), 89-100.
- Vesanto, J., J. Himberg, E. Alhoniemi and J. Parhankangas, 1999. Self-organizing map in Matlab: the SOM toolbox. In: *Proceedings of the Matlab DSP conference*, pp. 16-17.
- Vicente, S., V. Kolmogorov and C. Rother, 2008. Graph cut based image segmentation with connectivity priors. In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, Anchorage, Alaska, 24-26 June, pp. 1-8.
- Vincent, L., 1993. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE Transactions on Image Processing* 2 (2), 176-201.
- Walter, V., 2004. Object-based classification of remote sensing data for change detection. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (3), 225-238.
- Wang, M., F. Hu and J. Li, 2011. Epipolar resampling of linear pushbroom satellite imagery by a new epipolarity model. *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (3), 347-355.
- Wang, Z., A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, 2004. Image quality assessment: From error visibility to structural similarity. *Image Processing, IEEE Transactions on* 13 (4), 600-612.
- Waser, L., E. Baltsavias, K. Ecker, H. Eisenbeiss, E. Feldmeyer-Christe, C. Ginzler, M. Küchler and L. Zhang, 2008. Assessing changes of forest area and shrub encroachment in a mire ecosystem using digital surface models and CIR aerial images. *Remote Sensing of Environment* 112 (5), 1956-1968.
- Winkler, G., 2003. *Image analysis, random fields and Markov chain Monte Carlo methods: a mathematical introduction*. Springer Verlag.
- Zabih, R. and J. Woodfill, 1994. Non-parametric local transforms for computing visual correspondence. In: *Proceedings of European Conference on Computer Vision*, Stockholm, Sweden, May, pp. 151-158.
- Zhang, L. and A. Gruen, 2006. Multi-image matching for DSM generation from IKONOS imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 60 (3), 195-211.
- Zurich, 2014. 3D City Models of Zurich, [http://www.stadt-zuerich.ch/ted/de/index/geoz/3d\\_stadtmodell.html#datenbestellung](http://www.stadt-zuerich.ch/ted/de/index/geoz/3d_stadtmodell.html#datenbestellung). (Accessed on January 21, 2014)