Spatiotemporal inferences for use in building detection using series of very-high-resolution space-borne stereo images

Rongjun Qin[a], Jiaojiao Tian[b] & Peter Reinartz[b]

[a] Future Cities Laboratory, Singapore-ETH Centre, 138602
Singapore

[b] German Aerospace Center (DLR), Remote Sensing Technology
Institute (IMF), 82234 Wessling, Germany
Published online: 13 Jul 2015.

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis
Taylor & Francis Group

Spatiotemporal inferences for use in building detection using series of very-high-resolution space-borne stereo images

Rongjun Qin[a]*, Jiaojiao Tian[b], and Peter Reinartz[b]

[a]*Future Cities Laboratory, Singapore-ETH Centre, 138602 Singapore;* [b]*German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF), 82234 Wessling, Germany*

Automatic building detection from very-high-resolution (VHR) satellite images is a difficult task. The detection accuracy is usually limited by spectral ambiguities and the uncertainties of the available height information. Feature extraction and training sampling collection for supervised methods are other sources of uncertainty. Most widely used VHR sensors have shorter revisit cycles (IKONOS/GeoEye-1/2, 3 days; WorldView 1/2, 1.1 days) due to large off-nadir viewing angles and hence are able to perform consistent acquisition of mono or stereo images. In this article, we investigate the possibility of using high-temporal stereo VHR images to enhance remote-sensing image interpretation under the context of building detection. Digital surface models, which contain the height information, are generated for each date using semi-global matching. Pre-classification is performed combining the height and spectral information to obtain an initial building probability map. With a reference land cover map available for one date, the training samples of the other dates are automatically derived using a rule-based validating procedure. A spatiotemporal inference filter is developed considering the spectral, spatial, and temporal aspects to enhance the building probability maps. This aims at homogenizing the building probability values of spectrally similar pixels in the spatial domain and geometrically similar pixels in the temporal domain, while being robust to the silhouette of the images and geometric discrepancies of the multitemporal data. The effectiveness and robustness of the proposed method are evaluated by performing three experiments on six stereo pairs of the same region over a time period of five years (2006–2011). The area under curve (AUC) of the receiver operating characteristic and kappa statistic (κ) are employed to assess the results. These experiments show that spatiotemporal inference filtering largely improves the accuracy of the building probability map (average AUC = 0.95) while facilitating building extraction in snow-covered images. The resulting building probability maps can be further used for other applications (e.g. building footprint updating).

1. Introduction

1.1. Background

Data interpretation of remote-sensing data is a major task for earth observation, and the accurate identification and localization of buildings is essential for building analysis, planning, and urban growth monitoring. The development of very-high-resolution (VHR) optical images has paved a path to study building properties on a large scale (Qin, Gong, and Fan 2010; Qin et al. 2013). In particular, space-borne platforms carrying these sensors usually have a short revisit time (e.g. IKONOS, 3.5 days; WorldView, 1.1 days), giving them the potential to capture remote-sensing image time series.

*Corresponding author. Email: rqin@studuent.ethz.ch

Moreover, advanced stereo matching algorithms (Gehrke et al. 2010; Gruen 2012; Hirschmüller 2008) have driven the incremental availability of digital surface models (DSMs) at relatively low cost, providing additional information for data interpretation, and are particularly useful for the identification of buildings (Qin and Fang 2014).

Current building detection methods mainly work with data sets of one date, while algorithms are designed under the scenarios that (1) only a single multispectral image is available, (2) only the DSM is available, or (3) both image and DSM are available. Though efforts have been made to improve building feature delineation, detection strategies, and learning methods (Huang and Zhang 2011; Meng, Wang, and Currit 2009; Ok, Senaras, and Yuksel 2013; Sirmacek and Unsalan 2011), algorithmic limitations have been reached owning to the existence of uncertainties in DSMs and spectral ambiguities between buildings and other impervious objects such as roads and ground. Artefacts occur for data captured under different acquisitions: humid surfaces (e.g. after rain) can lead to specular reflections and the snow coverage of a scene will significantly reduce the spectral information of the images. Moreover, atmospheric and other unpredictable conditions affect the images, sometimes substantially reducing the quality of the matched DSMs (Hirschmüller and Scharstein 2009). However, current building detection algorithms are not well enough advanced to interpret such complicated variations, which restricts their optimal performance to certain conditions.

If time-intensive multitemporal data are available, their joint use via probability inference may tackle such 'case-specific' limitations for current building detection algorithms. Research on analysing low- to medium-resolution image sequences is mainly aimed at pairwise change detection (Coppin et al. 2004) and visual analysis for independently interpreted results, such as vegetation and urban growth (Kastens and Legates 2002). Correlations between different dates were analysed among the interpreted results, whereas multitemporal images were not used for mutual support in a joint interpretation. To the authors' best knowledge, to date only a few studies have addressed the probability inferences of VHR images with high temporal resolution for joint data interpretation. In this article, we therefore propose a method to perform spatiotemporal inference on building probability maps, adopting a three-dimensional (3D) bilateral filter to study the possibility of enhancing data interpretation accuracy with time series of VHR stereo images, including the derived DSMs.

1.2.   Related works

This study investigates the performance improvement in regard to building detection using multitemporal VHR stereo images, which has rarely been done before. Our study is closely related to the general topic of building detection and multitemporal analysis, and therefore it is necessary to introduce the state-of-the-art techniques applied in these areas.

1.2.1.   Building detection method

Building detection methods have previously been intensively studied, mainly focusing on single images, DSMs, or combined ortho-images and DSMs. With the availability of advanced stereo matching algorithms, the current trend is to use both ortho-images and DSMs for building detection. Due to the availability of training procedures, these methods can generally classified as either supervised or unsupervised.

As buildings are usually seen as one type of off-terrain object, height information is a major indicator in assessing their probability. The normalized DSM (nDSM) is commonly

used to derive off-terrain objects with height thresholds, and can be generated either by subtracting a given digital terrain model (DTM) from the derived DSM or from the DSM alone. For most unsupervised methods, spectral information is usually used for detecting vegetation and sharpening the boundary of buildings. Chen et al. (2012) proposed a step-wise method combining multispectral ortho-imagery and nDSM: the initial building segments were obtained by truncating the nDSM and normalized difference vegetation index (NDVI) with given thresholds, and the final building masks were generated with rule-based consideration of the region size and relational constraints between buildings and trees. Grigillo, Kosmatin Fras, and Petrovič (2011) generated the initial building masks in the same manner, but eliminated tree masks using the homogeneity feature (Zhang 1999) and NDVI. Qin and Fang (2014) proposed a hierarchical method to derive building segments using morphological operations on DSM and NDVI, and they adopted graph cut optimization to refine building boundaries using multispectral images. (Lu, Trinder, and Kubik 2006) proposed using the Dempster–Shafer (Shafer 1976) algorithm for fusing building probability values extracted from multispectral imagery and DSM. Tian and Reinartz (2013) computed a building probability map based on random forests (RF) classification, and they adopted panchromatic images to get sharper building boundaries.

The advantage of unsupervised methods is their rapid computation and flexibility to allow intuitive implementation of prior knowledge for building mask refinement. However, threshold selection and parameter tuning may be based on a case-by-case fashion. The supervised method can deal with this problem more efficiently, as prior information is derived from the data per se. Methods involving binary classification (building and non-building) or land-cover classification have been intensively studied (Dópido et al. 2013; Lee, Shan, and Bethel 2003; Meng et al. 2012; Tuia et al. 2010; Turlapaty et al. 2012), with most of attention drawn to improvement in feature extraction and design of the classifier (Dópido et al. 2013; Qin 2014b). The resulting building class was commonly used as the final output (Lu, Trinder, and Kubik 2006) or initial building masks for further refinement (Rottensteiner et al. 2005). The recent trend in classification involves the development of spatial features to improve their separability (Qin 2014b; Zhang et al. 2006). Classification accuracy can be further improved by incorporating the available height information (Huang, Zhang, and Gong 2011). Turker and San (2010) adopted the support vector machine (SVM) classifier (Wang 2005) to classify pan-sharpened images for building detection. Lee, Shan, and Bethel (2003) employed the iterative self-organizing data analysis (ISODATA) techniques algorithm to classify IKONOS multispectral images, and then approximated the building class using shape elements. For classification methods such as SVM and RF (Breiman 2001), the confidence in classification is usually provided for each class, which could be further used to refine building detection.

1.2.2. Multitemporal image analysis

One of the major tasks in multitemporal analysis is to evaluate temporal evolution and changes in the ground scene over time. Bitemporal data have usually been studied for change detection (Akca et al. 2010; Qin 2014a; Tian, Cui, and Reinartz 2014), and multitemporal time series images commonly used for studying urban or vegetation growth at a coarse level (Coppin et al. 2004; Lu et al. 2004). Kastens and Legates (2002) used low-spatial resolution time series as a means of assessing vegetation changes and high-lighted sensitive areas for vegetation degradation. In their method, images in the time

series are used independently to compute statistics such as circular variance, which are computed for the data analysis. Petitjean, Inglada, and Gançarski (2012) proposed a dynamic time warping (DTW) on high-temporal- but low-spatial-resolution images to tackle the problem of irregular data sampling and pairwise comparison of time-series sequences. Herold, Goldstein, and Clarke (2003) used a set of old aerial photographs and some IKONOS images to model urban growth and change, under the scenario of modelling and studying the spatial constraints on urban growth.

One prominent inference model is Markov random field analysis (MRF) (Blake et al., 2011), which posts similarity constraints on spatially and temporally neighbouring pixels/ objects to propagate their probability. The MRF inference model has been widely used for change detection applications (Crispell, Mundy, and Taubin 2012; Qin and Gruen 2014; Schindler and Dellaert 2010). Taneja, Ballan, and Pollefeys (2011) proposed a voxel-based method to infer change probability through a MRF framework, which implicitly adopted a multi-matching model for sensing geometric discrepancies among data from different dates, and a similar method was described by Schindler and Dellaert (2010). However, the MRF inference model is substantially a global optimization method, which incurs the problem of high computation load for pixel-wise calculation. Due to the local nature of the ground scene, it is more efficient to use non-global methods for inference.

1.3. *Proposed strategy*

Despite previous work on low-to-medium resolution multitemporal analysis, particularly under the context of change detection, little work has been done on the enhancement of image interpretation using VHR stereo data. In regard to improving the interpretation accuracy for each date, most existing methods focus on interpreting urban/vegetation changes from multitemporal data with no investigation on how these could contribute to each other. Therefore we aim to close this gap: we first adopt the RF classification method (Breiman 2001) to derive the probability maps for data from each date, with the training samples generated from a reference map of only one date. Then the resulting building probability map of each date is updated using a 3D bilateral filter considering the spectral, spatial, and temporal information.

2.   Data preprocessing

To perform pixel-/object-wise processing of multitemporal stereo data, a key step is to generate well co-registered DSM and ortho-images for all the dates. In this study we used the 'Catena' system at DLR (German Aerospace Center) (Krauß et al. 2013), which implicitly adopts a multi-image block adjustment for bias correction (Fraser and Hanley 2003) of rational polynomial coefficients and semi-global matching (SGM) (Hirschmüller 2005) for DSM generation. Bias correction was done using a large number of multi-matching tie points to ensure accurate geometric alignment. SGM adopts a multi-path dynamic programming to minimize cost function:

$$E(\mathbf{D}) = \sum_{p} C(p, \mathbf{D}_p) + \sum_{q \in \mathrm{N}_p} P_1 T\big[|\mathbf{D}_p - \mathbf{D}_q| = 1\big] + \sum_{q \in \mathrm{N}_p} P_2 T\big[|\mathbf{D}_p - \mathbf{D}_q| > 1\big], \quad (1)$$

where $\mathbf{D}$ is the disparity map (a matrix) where the value of each pixel on this map grid corresponds to the parallax in the epipolar images. $\mathbf{D}_p$ denotes the values of pixel p in map

grid **D**, and N_p is a set of pixels neighbouring point p that follows an eight/four connectivity rule. The first of these terms denotes the matching cost of **D**, which is usually computed using census or mutual information cost (Hirschmüller [2005]). The second and third terms are the smoothing terms for disparity jumps of neighbouring pixels N_p at one and more than one pixel, with P_1 and P_2 being the penalty. $T[\cdot]$ is a Boolean function that equals 1 when the expression holds true, and 0 otherwise. In our experiment, $P_1 = 300, P_2 = 1000$ were set empirically. The reader will find a detailed description of the algorithm and implementation in d'Angelo and Reinartz ([2011]) and Hirschmüller ([2005]).

3. Methodology

In this study the building probability maps are first generated using supervised classification with the multispectral ortho-image and DSM. The RF classifier is adopted to perform the classification due to its low computational complexity and capability of handling large volumes of data. For each test sample, it also provides a probability estimation of belonging to a particular class (Breiman [2001]), which is descriptive for delineating building pixels numerically. With a land-cover reference map (hereafter referred as the reference map) available for one date of the time series, it is possible to derive the training samples for the other dates automatically. Our method is divided into three steps: (1) training sample generation; (2) feature extraction and classification; and (3) spatiotemporal inference using a 3D bilateral filter. The first two steps are used to generate the building probability map, where we present a step-wise automatic method to select training samples for each date. Our main contribution lies in the third step, where a 3D bilateral filter is proposed to locally infer building probability values from the data of all dates.

3.1. Training sample generation

In most cases the land-cover reference map is usually available in the geo-database, where the training samples for supervised classification are derived and with which the classification accuracy is assessed. However, the reference map may not be available for all dates. Our method requires reference data for only one date to derive training samples. Indeed, the reference map does not have to cover all the objects in the research region, but must include typical ground objects for each class. Therefore, we propose to derive coarse reference maps from the available data for training sample generation. As our main focus in this study is buildings, we categorize the scene into the five following classes: 'building', 'ground', 'road', 'tree', and 'shadow'. By checking the consistency of DSMs and ortho-images, it is possible to verify existing and newly detected objects using change detection techniques. It should be noted that this procedure is different from a normal change detection process, as the verified and new objects do not need to be inclusive providing the generated coarse reference map is representative. Our proposed reference map generation procedure is shown in Figure 1.

As shown in Figure 1, the procedure of automatic reference generation is built on a coarse change detection procedure, and thus effective change indicators are very important. As shown in Figure 2, spectral variation due to seasonal differences can be very significant and employing spectral differences as a change indicator may cause much unnecessary error. Therefore, we use height difference as a robust measure and, being more specific, the difference in nDSM, since that is inherently robust to small height shifts caused by data
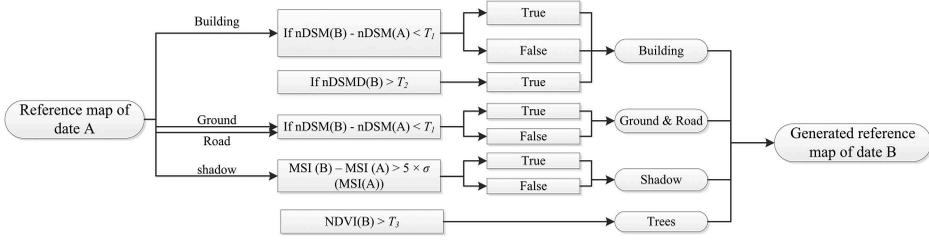
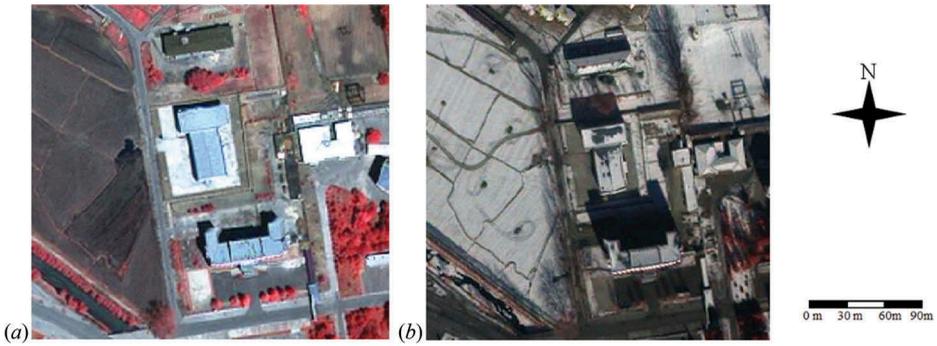Figure 1.   Flow chart for reference map generation.

Figure 2.   Examples of spectral changes across different seasons. Pan-sharpened IKONOS multi-spectral images (centre coordinate: 125°44′ E, 39°48′ N) of part of North Korea in summer (*a*) and in winter (*b*).

co-registration. Building change usually induces a height difference, with the ground object being its dual class. Therefore, buildings with height change will be eliminated from the reference map, as well as ground and roads. The nDSM in Figure 1 denotes an nDSM truncated by NDVI, serving as a good indicator for buildings. One of the key characteristics of shadow is its local luminance variance. We employ the morphological shadow index (MSI) as proposed by Huang and Zhang (2012) and Huang, Zhang, and Zhu (2013), which adopts grey-level top-hat reconstruction from the inverse of the brightness image. As the MSI is linearly correlated to shadow, only a very small number of samples are needed for training. Since top-hat reconstruction eliminates a zero-order shift between MSIs for different dates, the new shadow pixels are extracted using a 5σ threshold, where σ is the standard deviation of the MSI of the reference shadow segments. Segments whose MSI is different to the shadow reference map by less than 5σ are kept as shadow candidates. Trees are jointly determined by nDSM and NDVI with a certain threshold. In our experiment, in the reference map generation procedure (Figure 1) we take $T_1 = 1.5, T_2 = 3$, and $T_3 = 0.2$ as empirical values, these being determined by the DSM co-registration quality of multispectral information. Training samples are randomly selected from the derived reference map for each date, and only 200 pixels per class are used for training.

3.2.   Feature extraction and classification

Earlier works have demonstrated that the combined use of spectral and height information can significantly improve data interpretation accuracy (Huang, Zhang, and Gong 2011).

Height information is particularly useful in interpreting building properties. The RF classifier is adopted in our experiment, due to its low computational cost and high classification accuracy (Breiman 2001). Like other popular classifiers, RF provides a posteriori probability of the classification result belonging to a particular class, and the class label is determined by selecting the one with maximal probability. As our intention is to derive building probability values, we adopt pixel-wise classification rather than object-based strategy, which avoids pixels being diffused by erroneous segments.

The following features are used for the classification task:

(1) Principal component analysis (PCA) transformation of multispectral bands
(2) Differential morphological profile (DMP) of the panchromatic images
(3) Morphological top hat by reconstruction (MTHR) of the DSM

PCA (Jolliffe 2005) is widely used for dimension reduction of high-dimensional features. Due to its nature of maximizing variance in each dimension, it usually provides higher separability of different classes than the direct use of spectral information (Zhang et al. 2006). MTHR is regarded as an efficient indicator for off-terrain objects (Qin 2014c; Qin and Fang 2014), and exhibits a good capability to separate spectrally similar classes such as buildings and impervious ground. The MTHR of a DSM $\mathbf{J}$ (with its dimension being $m_{\mathbf{J}} \times n_{\mathbf{J}}$) can be computed as

$$\mathbf{T}_{\mathbf{J}}^{\mathbf{e}} = \mathbf{J} - \mathbf{B}_{\mathbf{J},\varepsilon(\mathbf{J},\mathbf{e})}, \qquad (2)$$

where $\mathbf{e}$ is a structural element (with dimensions $m_{\mathbf{e}} \times m_{\mathbf{e}}$), with ε being the grey-level erosion operator:

$$\varepsilon(\mathbf{J},\mathbf{e})(i,j) = \min\{\mathbf{J}(p-a,q-b)|, \quad \mathbf{e}(a,b) = 1, 0 < a,b < m_{\mathbf{e}} - 1, 0 < p,q < m_{\mathbf{J}}, n_{\mathbf{J}}\}. \qquad (3)$$

$\mathbf{B}_{\mathbf{J},\mathbf{I}}$ is the grey-level morphology reconstruction of $\mathbf{J}$ from $\mathbf{I}$, and in this case $\mathbf{I}$ is $\varepsilon(\mathbf{J},\mathbf{e})$. For more details on the use of grey-level morphological operation, the reader may refer to Vincent (1993).

DMP has been shown to be a valid spatial feature in improving classification accuracy (Benediktsson, Pesaresi, and Amason 2003). It adopts geodesic opening and closing operations at different scales to build DMPs in order to represent the image structural information, denoted as

$$\mathbf{DMP}_{\mathbf{J},i} = \mathbf{B}_{\mathbf{J},\varepsilon(\mathbf{J},\mathbf{e}_i)} - \mathbf{B}_{\mathbf{J},\varepsilon(\mathbf{J},\mathbf{e}_{i-1})}, \qquad (4)$$

where $\mathbf{e}_i, i = 1, 2, \ldots, n$, are the structural elements with different dimensions and $\mathbf{DMP}_{\mathbf{J},i}, i = 1, 2, \ldots, n$, is the DMP feature sequence of a raster grid $\mathbf{J}$, computed using the differentiations of grey-level morphological reconstruction with different structure elements $\mathbf{e}_i$. In our experiment, we use a disk-shaped structural element of varying radius to build the DMP sequence. The radius ranges from 3 to 30 pixels (with an interval of 3 pixels) to delineate ground objects.

We employ a simple vector-stacking method to fuse these features. To render each component of the feature vector numerically equivalent for computation, the values of each dimension are normalized to the range [0, 1]. RF is used to perform the training and

classification, with 500 trees initialized for the decision tree construction and testing, and the output of this approach is a probability map for each class.

3.3. Spatiotemporal inference using a 3D bilateral filter

The generated building probability maps vary with the data from different dates, due to their spectral and height uncertainties. Therefore, for the building probability map of each date, artefacts can frequently occur. For example, ground is likely to be identified as buildings under snowfall, and the edge of the buildings might be noisy due to their highly similar spectrum to the ground. The matching algorithm may fail for certain small buildings owing to poor spectral quality, resulting in their incorrect identification. According to Pacifici, Longbotham, and Emery (2014), the spectral response of the same object from different acquisitions can vary considerably, leading to different separability between urban classes. Our idea is to leverage the uncertainties by fusing building probability maps across all temporal acquisitions using a simple and fast method, while avoiding the fusion process diffused by data with large height discrepancies in the temporal direction.

The bilateral filter is regarded as an edge-aware adaptive kernel filter that performs spatial filtering, weighting each pixel according to its spatial and spectral proximity to the centric pixel for filtering (Tomasi and Manduchi 1998). It filters an image raster $\mathbf{I}$ (dimension $m_{\mathbf{I}} \times n_{\mathbf{I}}$) as follows:

$$\mathbf{k_I}(x,y) = \sum_{i,j} -e^{\frac{||\mathbf{I}(x,y)-\mathbf{I}(i,j)||^2}{2\sigma_1^2} + \frac{||[x,y]-[i,j]||^2}{2\sigma_2^2}} \mathbf{I}(i,j), 0 \leq i,j < m_{\mathbf{I}}, n_{\mathbf{I}}, \tag{5}$$

where $\mathbf{k_I}$ is the filtered raster. σ_1 and σ_2 are the spectral and spatial bandwidths, respectively, controlling the sensitivities of the spectral and spatial dissimilarities between the centric and surrounding pixels. The spectral difference $(\mathbf{I}(x,y) - \mathbf{I}(i,j))$ is usually computed as the Euclidean distance of the transformed colour space (e.g. CIELA; Joblove and Greenberg 1978) or PCA). The filtering process assigns a large weight to spatially close and spectrally similar pixels for filtering, while assigning a very small weight for pixels of varying spectral value.

Considering the spatial correlation among pixels, a smoothness constraint can be posted on neighbouring pixels having similar spectral responses, aiming promote the homogeneity of building probability values for locally similar pixels. The height information is robust in the temporal direction, so the homogeneity of the building probability values can be correlated to the height similarity in the temporal direction (Tian et al. 2013). Based on these considerations, we further develop the 2D bilateral filter to a 3D bilateral filter that implements the aforementioned constraints, and update the building probability map of each date:

$$P_f(x,y,t) = \frac{1}{\sum w(x,y,t)} \sum_{m=x-l}^{m=x+l} \sum_{n=y-l}^{n=y+l} \sum_{k=1}^{h} w(m,n,k)P(m,n,t), \tag{6}$$

where $P(m,n,t)$ is the raw building probability map at time t, with $P_f(x,y,t)$ being the filtered results; l is the window length, which is usually the value of the spatial bandwidth; h is the number of temporal data sets; and $w(m,n,k)$ is the 3D adaptive kernel used to compute the aggregated weight in the spatial and temporal directions:

$$w(m, n, k) = w_1(m, n, t)w_2(m, n, k, t), \tag{7}$$

where

$$w_1(m, n, t) = e^{\frac{-L(\mathbf{I}(x,y,t))-L(\mathbf{I}(m,n,t))^2}{2\sigma_1^2} + \frac{-||[x,y]-[m,n]||^2}{2\sigma_2^2}}$$

and

$$w_2(m, n, k, t) = e^{\frac{-||\mathbf{H}_t(m,n)-\mathbf{H}_k(m,n)||^2}{2\sigma_3^2}}. \tag{8}$$

$w_1(\cdot)$ is the normal bilateral filter that considers the spatial smoothness; $L()$ represents the colour transformation from RGB to CIELAB colour space; $\mathbf{H}_k$ represents the DSM at time k; $w_2(\cdot)$ is an extended part that constrains the height difference in the temporal direction, with σ_3 being the temporal bandwidth; and $w_2(\cdot)$ assigns large weights for DSMs on dates of similar height while assigning a low value to those with large height differences. It should be noted $w_2(\cdot)$ is constrained by weight $w_1(\cdot)$ in the spatial domain. The kernel value will be still small if the spectral value is different from that of the centric one, even with a similar height in temporal direction. This means that only spectrally similar pixels in the spatial domain and pixels of similar height in the temporal domain will be used to contribute to the building probability values of the centric pixels.

The bandwidth values of the 3D bilateral filter are usually determined empirically through trial-and-error approaches. These values are linearly related to the visual effect of the final results: a high σ_1 usually results in more smooth but less informative results, while a low value leads to sharper but noisier results. A high σ_2 considers more pixels among the spatial pixels, which also leads to more blurry boundaries; a low σ_2 leads to sharper boundaries but contributes less to the noise removal. σ_3 should be carefully determined according to the co-registration quality between DSMs: a very low value results in zero contribution from the other temporal DSMs, and a high value might blur areas where height changes occur.

The advantage of the 3D bilateral filter is its computational efficiency in comparison with global methods such as MRF. MRF optimizes the probability of buildings by considering all pixels in the spatial and temporal domains, which carries a high computational load. In fact, the correlations between building pixels only exist locally and pixels far apart make little contribution to each other, since building objects do not span a large area and are independent from each other. Moreover, the proposed 3D bilateral filter is easy to implement and requires only polynomial computational time. In addition, it does not place any constraints on the temporal order of the input data, which has the potential for parallel processing of large volumes of multitemporal data.

4. Experiment

4.1. Data description

To evaluate the performance of the proposed method, data sets of the same regions from six different dates are used. We selected an industrial area near Dong-an, North Korea, as the test site due to data availability. Table 1 shows detailed information on the available satellite stereo imagery. As listed, four data sets were captured in winter, including one

Table 1. Data description.

Data set	Satellite	Acquisition date	Resolution (m)		Approximate snow cover
			PAN	MS	
1	IKONOS	23 February 2006	1	4	4%
2	GeoEye-1	20 December 2009	0.5	2	3%
3	IKONOS	12 January 2010	1	4	85%
4	IKONOS	13 May 2010	1	4	No
5	IKONOS	7 January 2011	1	4	4%
6	IKONOS	2 May 2011	1	4	No

Note: PAN, panchromatic; MS, multispectral.

featuring very heavy snow coverage. The other two data sets were captured in summer. Among these six data sets, five are IKONOS stereo imagery with 1 m pixel size in the panchromatic band and 4 m pixel size in the multispectral bands, and one is from GeoEye-1 with 0.5 m pixel size (panchromatic) and 2 m pixel size (multispectral). The widely used Gram–Schmidt pan-sharpening method implemented in ENVI software (ENVI 2014) is applied to all multispectral channels for further interpretation. Since the proposed method is pixel-based, all images are resampled to 1 m pixel size for pixel-wise alignment.

In the present study, two sub-regions of this area are selected for the experiment with two data sets of 1230 × 1303 and 1264 × 650 pixels, respectively. The reference building footprints are manually extracted for all six temporal data sets. Since the selected regions are located in a rapidly developing district, within the previous six years many new buildings had been built. Therefore, for presentation convenience, we take the final data set (no. 6) as the main example since it contains the most buildings.

The multispectral images and reference building footprints of the two selected regions, which describe the building characters and distribution, are shown in Figure 3. The selected regions contain mainly industrial buildings and a few residential houses, varying in size and roofing materials. Special attention should be paid to very dense and small buildings (e.g. red circles in Figure 3(a)), since these might represent potentially challenging cases for building detection where the DSM generation algorithm may not produce height jumps due to the limited resolution.

As described in Section 2, the DSMs of all dates are generated using the same method, with block adjustment among all data sets. This approach theoretically guarantees sub-pixel accuracy in planimetry, with height accuracy ranging from 1 to 2 m depending on convergence angles between the respective viewing directions. The generated DSMs of data set no. 6 are displayed in Figure 4. Although the quality check of the absolute accuracy of the DSMs is important, for our particular purpose it is more crucial to evaluate their relative difference. As shown in Figure 5, the profiles of the six DSMs along the black line in Figure 4(b) are plotted. The red line represents the height values of the DSM in the year 2006, before these buildings had been built. All other five DSMs clearly show the cross-sections of four buildings approximately 12 m in height.

4.2. Experimental analysis

The aim of our experimental analysis is to investigate the effectiveness and robustness of the proposed spatiotemporal inference method. For this purpose, three experiments
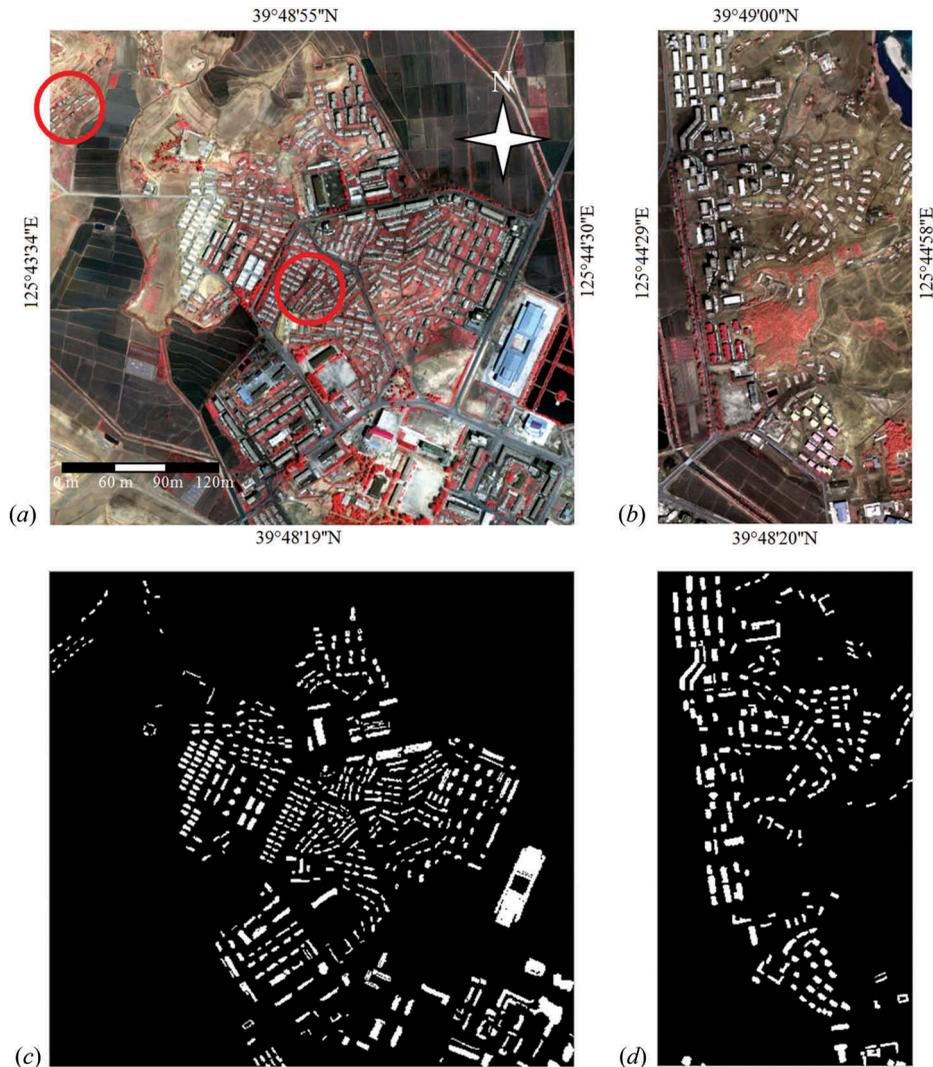
Figure 3. Ortho-rectified images for test regions 1 (*a*) and 2 (*b*), and the corresponding reference maps (*c*) and (*d*); red circles denote small and dense houses.

were designed and carried out based on the data sets: (1) comparative analysis on probability maps; (2) case analysis – performance of images with weak spectral information; and (3) accuracy of analysis regarding building detection. In the first experiment, we evaluate the differences between the raw probability map (generated from the classification) and those inferred using the area under the curve (AUC) values of the receiver operating characteristic (ROC) curve (Hand and Till 2001). This provides qualitative and quantitative results on the performance of our method. In the second experiment, we analyse the special case of 12 January 2010: besides heavy shading effects, almost the whole test region was covered in snow, resulting in very low spectral variability. Therefore, this experiment examines the performance of the proposed method on multi-seasonal images. The third experiment is an
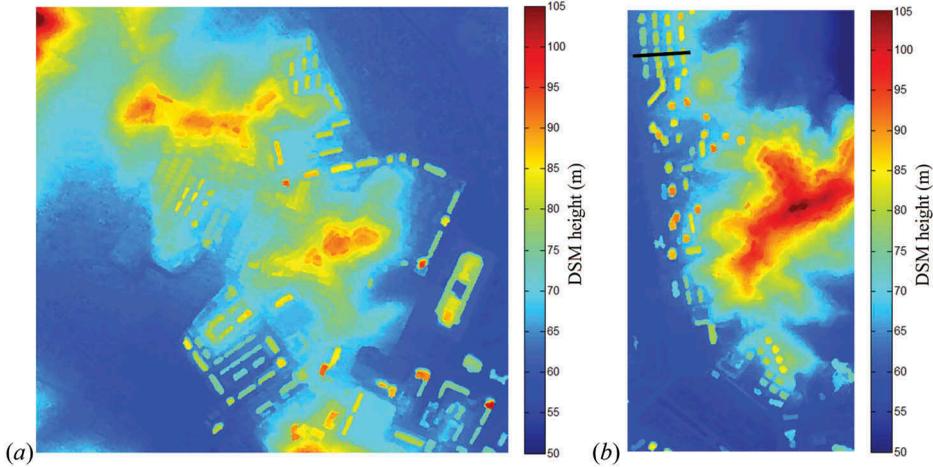
Figure 4.   Generated DSMs for data set no. 6 in test regions 1 (*a*) and 2 (*b*).
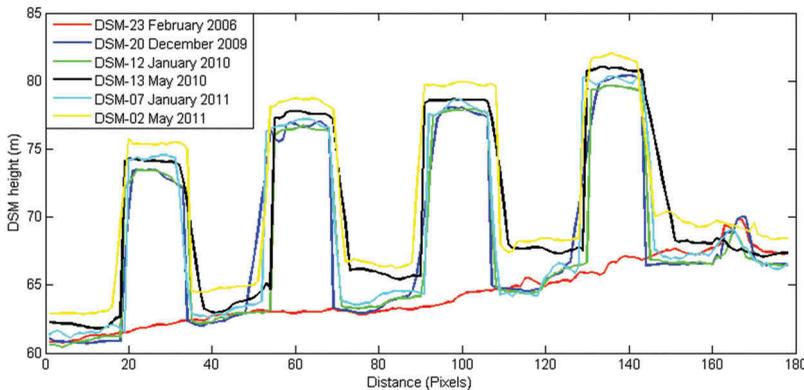

Figure 5.   DSM height value comparison along the black line in Figure 4(*b*).

application-based analysis, aiming to prove that the improved probability maps can provide a building footprint of good quality.

4.2.1. Experiment I

The proposed spatiotemporal inference filtering approach fuses raw building probability maps from different dates. To provide a descriptive analysis on the incremental effect of the proposed approach, we evaluate the accuracy improvement in building probability maps by increasing the number of raw probability maps incorporated in the inference process.

Figures 6 and 7 demonstrate the visual comparison of resulting probability maps of data set no. 6. In each sub-figure, four probability maps are displayed. Some small buildings in the raw probability map of test region 1 (Figure 6(*a*)) are not well high-lighted. After spatial inference (with $w(m, n, t) = w_1(m, n, t)$ in Equation (7), which is a standard bilateral filter) the 'salt and pepper' effect is effectively removed, but some roads
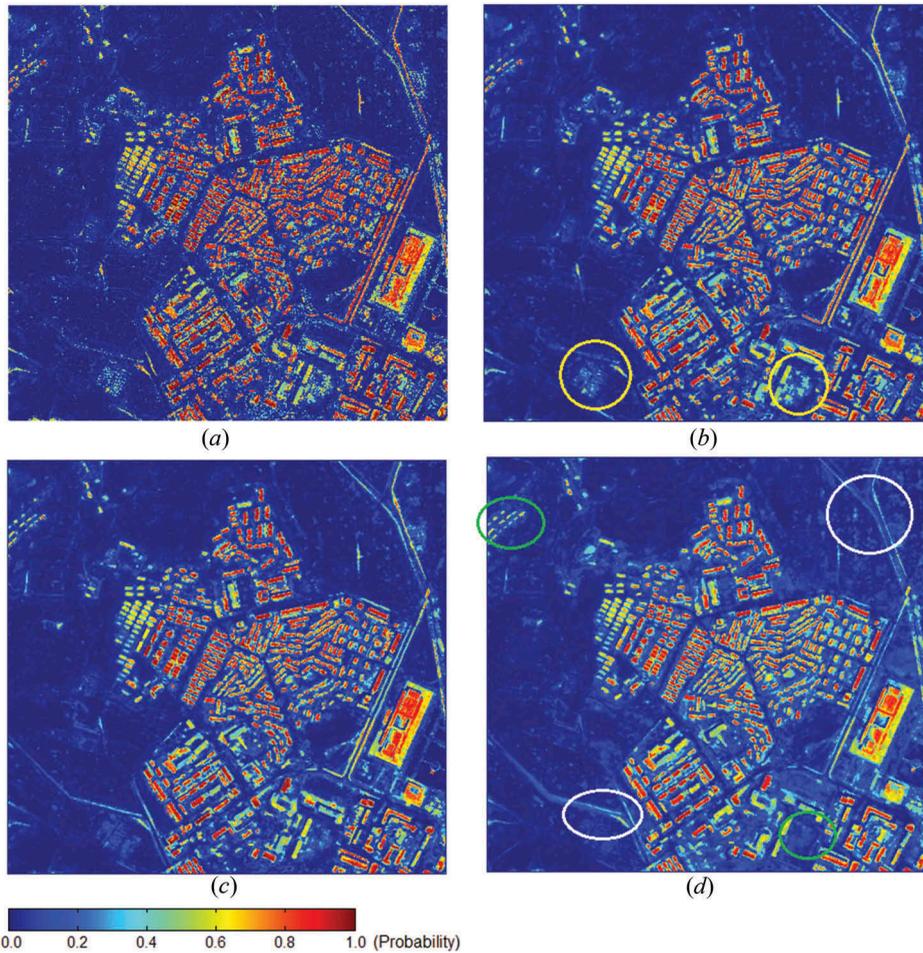
Figure 6. Comparison of building probability maps in test region 1. The significance of the circled areas is explained in the main text. (*a*) Raw probability map. (*b*) Probability map computed using only spatial inference; (*c*) probability map obtained using spatiotemporal inference on one reference data set; (*d*) probability map obtained using spatiotemporal inference with five reference data sets.

are still given a high probability of being buildings, as shown in Figure 6(*b*). As denoted by the yellow circle, some hilly ground is assigned high building probability values, but this situation is significantly improved after the spatiotemporal inference is applied (shown in Figure 6(*c*)), and this is further enhanced after incorporating all six data sets in the spatiotemporal inference (Figure 6(*d*)). In the region marked by white circles, roads and hilly ground are assigned lower building probability values while the actual buildings still show higher probability values. Another region, marked by a green circle, clearly shows the improvement in the probably map. Figure 7 shows the results obtained in the second test region, in which the probability map achieves a similar level of improvement and which can be observed clearly in the region marked with yellow circles.

The same experiment is carried out for all six data sets of the two test regions. The building probability maps using 1one to five temporal reference data sets (named as Prob_2, Prob_3, Prob_4, Prob_5, and Prob_6, respectively) are all calculated for a more
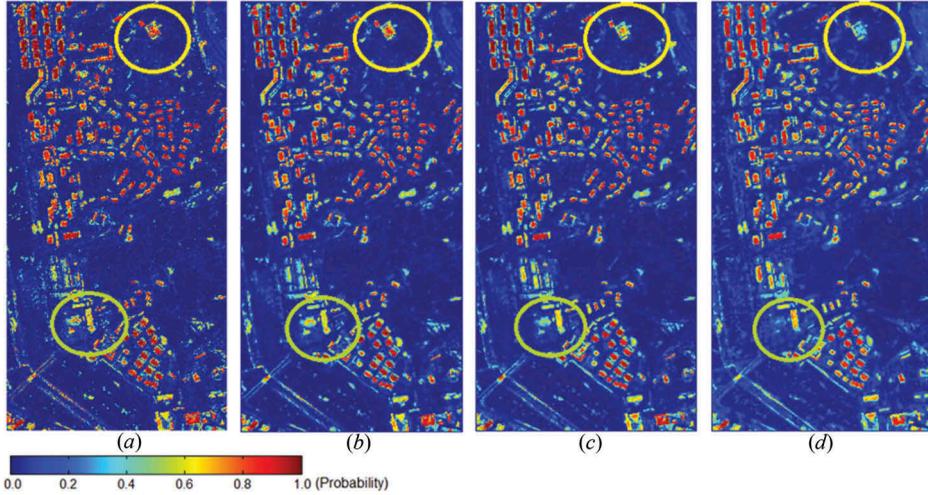
Figure 7.   Comparison of building probability maps in the second test region. The significance of the circled area is explained in the main text. (*a*) Raw probability map; (*b*) probability map computed only for spatial inference; (*c*) probability map using spatiotemporal inference on one reference data set; (*d*) probability map using spatiotemporal inference with five reference data sets.

comprehensive comparison. Together with the raw probability map (Prob_0) and that using only spatial inference (Prob_1), each data set has seven building probability maps for evaluation. The performance of the spatiotemporal inference filtering is evaluated through ROC curves (Hand and Till 2001) by comparing these probability maps to the reference building footprint. The ROC curves of test region 1 are shown in Figure 8. All six temporal data sets are involved in the evaluation procedure. For a numerical comparison, the AUC values are reported in Table 2. A larger AUC indicates a higher accuracy.

From Table 2, we note that the AUCs of the raw probability map (Prob_0) vary widely, this being influenced by the quality of the training data and seasonal disturbances. However, after spatiotemporal inference, 11 recorded an AUC over 0.95 (Prob_6), which is satisfactory for most building detection applications. A significant improvement in accuracy can be found between Prob_1 and Prob_2, which proves the necessity of temporal inference.

4.2.2. *Experiment II*

It is interesting to study the enhancement of our proposed method for some special scenarios, as in the case of high snow coverage. As shown in Figure 9(*a*), the true colour image of test region 1 captured on 12 January 2010 has very low spectral variability, with notable errors in the raw probability map as shown in Figure 9(*c*). Such errors can simply be linked to other similar cases (e.g. low spectral information for panchromatic images, high-specular reflecting impervious surfaces after rain).

A sub-region for further analysis is denoted by the red rectangle in Figure 9(*a*). The lack of texture may affect the quality of the DSM. As can be seen in Figure 9(*b*), the separability of the low buildings in the hilly area (denoted by a white ellipse) is not clearly observable. Moreover, the height of some buildings, as denoted by a yellow circle, may fail to record accurately. Since the training data used in this study are automatically
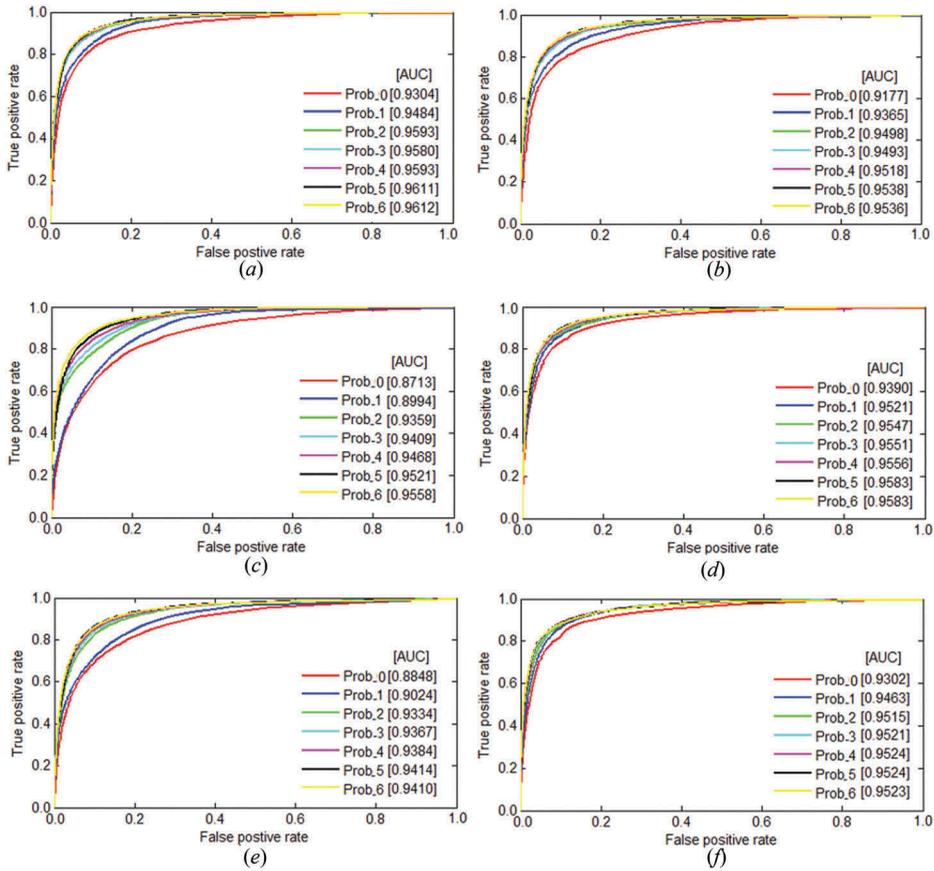
Figure 8. ROC comparisons of test region 1 for (a) 23 February 2006, (b) 20 December 2009, (c) 12 January 2010, (d) 13 May 2010, (e) 7 January 2011, and (f) 2 May 2011.

Table 2. Performance results for the two test regions.

|	Dates	Prob_0	Prob_1	Prob_2	Prob_3	Prob_4	Prob_5	Prob_6
Test region 1	23 February 2006	0.9304	0.9484	0.9593	0.9579	0.9593	0.9611	0.9612
	20 December 2009	0.9177	0.9365	0.9497	0.9493	0.9518	0.9538	0.9536
	12 January 2010	0.8713	0.8994	0.9359	0.9409	0.9468	0.9521	0.9558
	13 May 2010	0.9390	0.9521	0.9546	0.9551	0.9556	0.9583	0.9583
	7 January 2011	0.8848	0.9024	0.9334	0.9367	0.9384	0.9414	0.9410
	2 May 2011	0.9302	0.9463	0.9515	0.9521	0.9524	0.9524	0.9523
Test region 2	23 February 2006	0.9537	0.9681	0.9740	0.9726	0.9729	0.9736	0.9754
	20 December 2009	0.8989	0.9309	0.9515	0.9489	0.9542	0.9567	0.9581
	12 January 2010	0.8837	0.9235	0.9549	0.9553	0.9574	0.9588	0.9626
	13 May 2010	0.9567	0.9661	0.9723	0.9698	0.9698	0.9708	0.9722
	7 January 2011	0.8667	0.9026	0.9337	0.9421	0.9432	0.9462	0.9474
	2 May 2011	0.9175	0.9424	0.9479	0.9465	0.9469	0.9492	0.9504
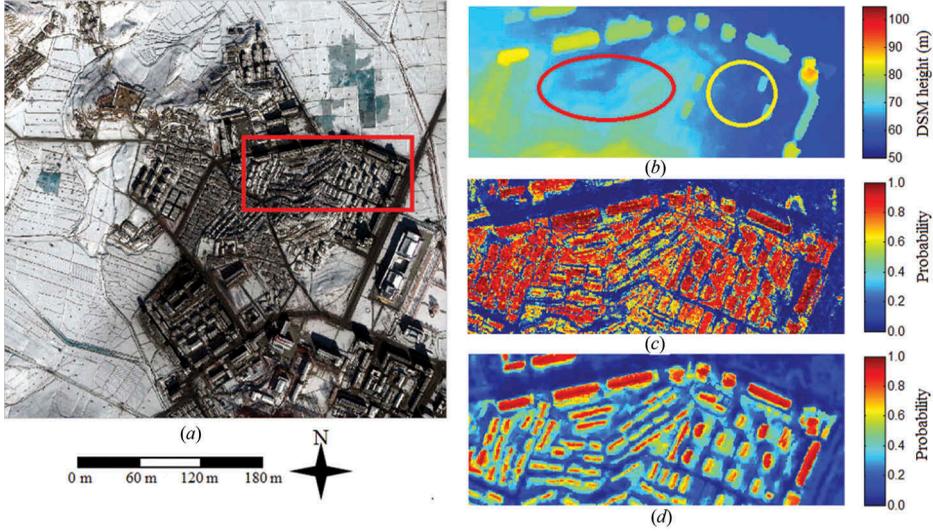
Figure 9.   Building probability map for a snow-covered data set: (*a*) a data set in the first test region with heavy snow cover; the area bounded by the red rectangle indicates an enlarged region, shown in (*b*)–(*d*); (*b*) DSM of the selected region; (*c*) original building probability map of the selected region; (*d*) final probability map of the selected region.

generated by reference to an outdated land-cover map, inaccurate building height will directly reduce the quality of the training data and further affect classification accuracy. For instance, if these buildings with lower height values are included in the training data, many snow-covered ground areas would be wrongly classified as buildings.

In addition to DSM quality, the high percentage of snow cover is another main reason for the less accurate results shown in Figure 9(*c*). The original spectral and texture of land-cover characters are hidden by the snow, severing the relationship between spectral features and land cover. After applying the proposed spatiotemporal inference, the quality of the building probability map is largely improved. As shown in Figure 9(*d*), the detected buildings are well separated from each other since these are assigned significantly higher values than ground.

4.2.3. *Experiment III*

We perform additional experiments to further analyse the robustness of the proposed method. As one important aspect of building detection is to obtain a precisely updated building footprint, probability maps of higher quality are capable of providing greater accuracy. Therefore, in this experiment a series of thresholds are manually selected to truncate these probability maps. The obtained building footprint is compared to the reference building footprint, and the similarity between them is measured using κ. In this step, three probability maps are selected for each data set for comparison, including the original map, one with spatial inference, and the final spatiotemporal inference probability map with all the data sets.

For each probability map, 18 building footprints are generated by assigning a series of thresholds T ($T = 0.1, 0.15, 0.2, \ldots, 0.95$). The evaluation results of these building footprints are reported in Figures 10 and 11, respectively. In all 12 data sets (two test
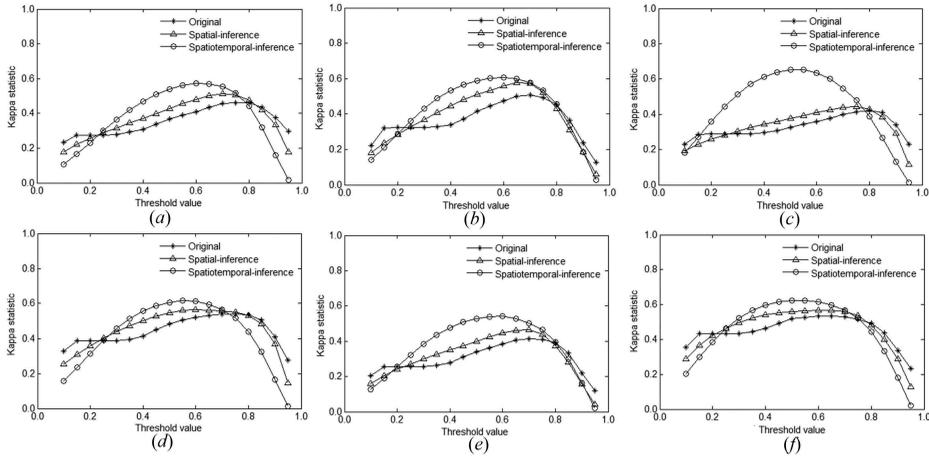
Figure 10. Plot of κ for the first test region of the data set captured on (*a*) 23 February 2006, (*b*) 20 December 2009, (*c*) 12 January 2010, (*d*) 13 May 2010, (*e*) 7 January 2011, and (*f*) 2 May 2011.

regions with six data sets), the probability maps obtained by spatiotemporal inference recorded the highest κ for the generated building footprint with finely tuned thresholds. Spatial inference also has a positive contribution to the accuracy of the obtained building footprint. The results confirm the effectiveness of the proposed spatiotemporal inference method.

Moreover, it may be noted that there is a distinct improvement between the probability map of spatial inference over that of the spatiotemporal inference in Figure 10(*c*), which was obtained for the snow-covered data set. This further proves that the proposed method has the potential to recover data interpretation on data sets of poor spectral quality.

Table 3 presents the commission (false-positive) and omission errors (false-negative) of these two experiments when κ reaches its maximum value. It will be seen that commission errors are minimal when spatiotemporal inference is adopted: on average,
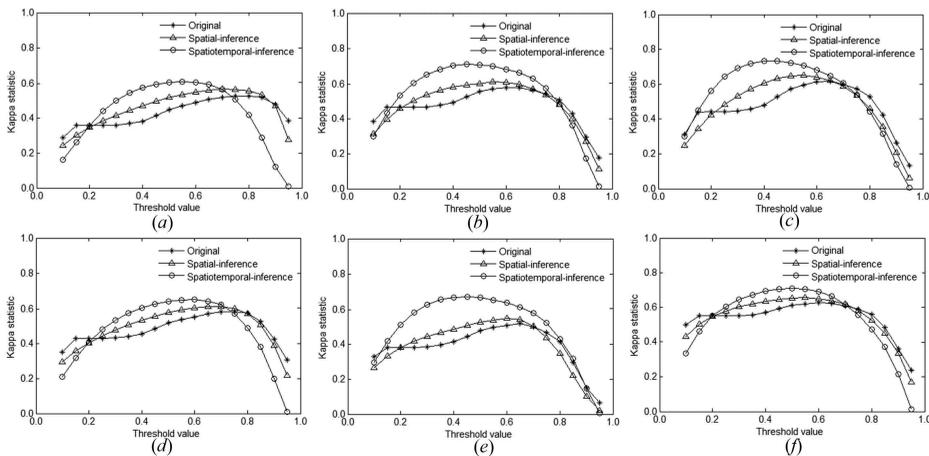

Figure 11. Plot of κ for the second test region of the data set captured on (*a*) 23 February 2006, (*b*) 20 December 2009, (*c*) 12 January 2010, (*d*) 13 May 2010, (*e*) 7 January 2011, (*f*) 2 May 2011.

Table 3.　Commission and omission errors for the two test regions.

		Test region 1			Test region 2		
	Dates	O	S	ST	O	S	ST
Commission error	23 February 2006	0.5867	0.4954	0.3911	0.5355	0.4537	0.3189
(false-positive rate)	20 December 2009	0.5752	0.4771	0.1714	0.3158	0.1837	0.1268
	12 January 2010	0.5012	0.4332	0.3322	0.3857	0.2955	0.2003
	13 May 2010	0.6106	0.5015	0.4256	0.4485	0.3079	0.1975
	7 January 2011	0.4627	0.4112	0.3226	0.4442	0.3807	0.3254
	2 May 2011	0.5176	0.3965	0.3429	0.4440	0.3278	0.2076
Omission error	23 February 2006	0.0174	0.0191	0.0219	0.0109	0.0125	0.0206
(false-negative rate)	20 December 2009	0.0369	0.0431	0.0519	0.0304	0.0355	0.0347
	12 January 2010	0.0244	0.0266	0.0294	0.0244	0.0284	0.0285
	13 May 2010	0.0248	0.0263	0.0263	0.0353	0.0403	0.0367
	7 January 2011	0.0242	0.0262	0.0325	0.0167	0.0185	0.0230
	2 May 2011	0.0237	0.0243	0.0266	0.0258	0.0313	0.0285

Note: O, original building probability; S, spatial inferred building probability; ST, spatiotemporal inferred probability.

30–40% of commission errors are reduced by comparing the original and spatially inferred building probability maps. Only a very small increase in false-negatives is observed. This is expected, as the spatiotemporal inference process is aimed at reducing uncertainty that could eliminate the pixels of some very small buildings.

5. Conclusion

In this study we proposed a 3D bilateral filter to investigate the possibility of enhancing building detection accuracy with multitemporal stereo images. First, DSMs are generated from stereo imagery using the SGM algorithm. Based on a given reference map of one date, coarse reference maps are generated through a rule-based method and then the training samples are derived randomly for classification with a RF classifier. The building probability map of each date is enhanced via a fast spatiotemporal 3D bilateral filtering which considers both spatial and temporal correlation of the multitemporal ortho-images and DSMs.

The proposed approach was successfully validated using six pairs of stereo satellite images captured over an industrial area over a time span of five years (2006–2011). The average AUC value of the resulting building probability maps is better than 0.95, which is a promising result for high-accuracy building detection. A prominent AUC improvement is obtained for data sets with very low texture information (e.g. data with a high snow cover rate). In this data set, the accuracy of the probability map of the two test regions improved from 0.8713 to 0.9558 and 0.8837 to 0.9626, respectively. The resulting building probability map is further tested with the reference building footprint, which demonstrates its effectiveness in providing useful input for city management.

It should be noted that the supervised method for raw building probability map generation might not be optimal, but this component can always be substituted with the most advanced algorithms to produce better results. The selection of the bandwidth parameter σ_3 is dependent on the DSM and the co-registration quality, which still needs to be optimized for robustness. Moreover, this study presents the possibility of enhancing

building probability maps with multitemporal stereo images and their derived DSMs. It is straightforward to extend this method to other urban classes such as trees, ground, and roads. In addition, our 3D spatiotemporal bilateral filter provides a probabilistic output and it would also be interesting to investigate further how this process might facilitate deterministic methods such as class labelling. Therefore, our future work will analyse the performance of the proposed method for other urban classes and focus on adaptive parameter selection for the bandwidth, as well as develop the spatiotemporal inference for deterministic problems.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

References

Akca, D., M. Freeman, I. Sargent, and A. Gruen. 2010. "Quality Assessment of 3D Building Data." *The Photogrammetric Record* 25 (132): 339–355. doi:10.1111/phor.2010.25.issue-132.

Benediktsson, J. A., M. Pesaresi, and K. Amason. 2003. "Classification and Feature Extraction for Remote Sensing Images from Urban Areas Based on Morphological Transformations." *IEEE Transactions on Geoscience and Remote Sensing* 41 (9): 1940–1949. doi:10.1109/TGRS.2003.814625.

Blake, A., P. Kohli, and C. Rother. 2011. *Markov Random Fields for Vision and Image Processing*, 463. Cambridge, MA: The MIT Press.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

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. doi:10.1080/01431161.2012.659355.

Coppin, P., I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin. 2004. "Digital Change Detection Methods in Ecosystem Monitoring: A Review." *International Journal of Remote Sensing* 25 (9): 1565–1596. doi:10.1080/0143116031000101675.

Crispell, D., J. Mundy, and G. Taubin. 2012. "A Variable-Resolution Probabilistic Three-Dimensional Model for Change Detection." *IEEE Transactions on Geoscience and Remote Sensing* 50 (2): 489–500. doi:10.1109/TGRS.2011.2158439.

d'Angelo, P., and P. Reinartz. 2011. "Semiglobal Matching Results on the ISPRS Stereo Matching Benchmark." ISPRS Hannover Workshop 2011: High-Resolution Earth Imaging for Geospatial Information, 79–84. Hannover, Germany, 14–17 June.

Dópido, I., J. Li, P. R. Marpu, A. Plaza, J. Bioucas Dias, and J. A. Benediktsson. 2013. "Semisupervised Self-Learning for Hyperspectral Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 51 (7): 4032–4044. doi:10.1109/TGRS.2012.2228275.

ENVI, 2014. Gram-Schmidt Pan Sharpening. Accessed December 10, 2014. http://www.exelisvis.com/docs/GramSchmidtSpectralSharpening.html.

Fraser, C. S., and H. B. Hanley. 2003. "Bias Compensation in Rational Functions for IKONOS Satellite Imagery." *Photogrammetric Engineering & Remote Sensing* 69 (1): 53–57. doi:10.14358/PERS.69.1.53.

Gehrke, S., K. Morin, M. Downey, N. Boehrer, and T. Fuchs. 2010. "Semi-Global Matching: An Alternative to LIDAR for DSM Generation." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Calgary, AB* 38 (B1): 6.

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. doi:10.15292/geodetski-vestnik.2011.01.028-045.

Gruen, A. 2012. "Development and Status of Image Matching in Photogrammetry." *The Photogrammetric Record* 27 (137): 36–57. doi:10.1111/phor.2012.27.issue-137.

Hand, D. J., and R. J. Till. 2001. "A Simple Generalisation of the Area under the ROC Curve for Multiple Class Classification Problems." *Machine Learning* 45 (2): 171–186. doi:10.1023/A:1010920819831.

Herold, M., N. C. Goldstein, and K. C. Clarke. 2003. "The Spatiotemporal Form of Urban Growth: Measurement, Analysis and Modeling." *Remote Sensing of Environment* 86 (3): 286–302. doi:10.1016/S0034-4257(03)00075-0.

Hirschmüller, 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*, 807–814.

Hirschmüller, H. 2008. "Stereo Processing by Semiglobal Matching and Mutual Information." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (2): 328–341. doi:10.1109/TPAMI.2007.1166.

Hirschmüller, 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. doi:10.1109/TPAMI.2008.221.

Huang, X., and L. Zhang. 2011. "A Multidirectional and Multiscale Morphological Index for Automatic Building Extraction from Multispectral Geoeye-1 Imagery." *Photogrammetric Engineering & Remote Sensing* 77 (7): 721–732. doi:10.14358/PERS.77.7.721.

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. doi:10.1109/JSTARS.2011.2168195.

Huang, X., L. Zhang, and W. Gong. 2011. "Information Fusion of Aerial Images and LIDAR Data in Urban Areas: Vector-Stacking, Re-Classification and Post-Processing Approaches." *International Journal of Remote Sensing* 32 (1): 69–84. doi:10.1080/01431160903439882.

Huang, X., L. Zhang, and T. Zhu. 2013. "Building Change Detection From Multitemporal High-Resolution Remotely Sensed Images Based on a Morphological Building Index." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* PP (99): 1–11.

Joblove, G. H., and D. Greenberg. 1978. "Color Spaces for Computer Graphics." Proceedings of ACM SIGGRAPH 1978, 20–25, Atlanta, GA, August 23–25.

Jolliffe, I. 2005. "Principal Component Analysis." *Encyclopedia of Statistics in Behavioral Science* 3 (1): 1580–1584.

Kastens, J. H., and D. R. Legates. 2002. "Time Series Remote Sensing of Landscape-Vegetation Interactions in the Southern Great Plains." *Photogrammetric Engineering & Remote Sensing* 68 (10): 1021–1030.

Krauß, T., P. d'Angelo, M. Schneider, and V. Gstaiger. 2013. "The Fully Automatic Optical Processing System CATENA at DLR." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 177–181. doi:10.5194/isprsarchives-XL-1-W1-177-2013.

Lee, D. S., J. Shan, and J. S. Bethel. 2003. "Class-Guided Building Extraction from Ikonos Imagery." *Photogrammetric Engineering & Remote Sensing* 69 (2): 143–150. doi:10.14358/PERS.69.2.143.

Lu, D., P. Mausel, E. Brondizio, and E. Moran. 2004. "Change Detection Techniques." *International Journal of Remote Sensing* 25 (12): 2365–2401. doi:10.1080/0143116031000139863.

Lu, Y. H., J. C. Trinder, and K. Kubik. 2006. "Automatic Building Detection Using the Dempster-Shafer Algorithm." *Photogrammetric Engineering & Remote Sensing* 72 (4): 395–403. doi:10.14358/PERS.72.4.395.

Meng, X., N. Currit, W. Le, and X. Yang. 2012. "Detect Residential Buildings from Lidar and Aerial Photographs through Object-Oriented Land-Use Classification." *Photogrammetric Engineering and Remote Sensing* 78 (1): 35–44. doi:10.14358/PERS.78.1.35.

Meng, X., L. Wang, and N. Currit. 2009. "Morphology-Based Building Detection from Airborne Lidar Data." *Photogrammetric Engineering & Remote Sensing* 75 (4): 437–442. doi:10.14358/PERS.75.4.437.

Ok, A. O., C. Senaras, and B. Yuksel. 2013. "Automated Detection of Arbitrarily Shaped Buildings in Complex Environments from Monocular VHR Optical Satellite Imagery." *IEEE Transactions on Geoscience and Remote Sensing* 51 (3): 1701–1717. doi:10.1109/TGRS.2012.2207123.

Pacifici, F., N. Longbotham, and W. J. Emery. 2014. "The Importance of Physical Quantities for the Analysis of Multitemporal and Multiangular Optical Very High Spatial Resolution Images." *IEEE Transactions on Geoscience and Remote Sensing* 52 (10): 6241–6256. doi:10.1109/TGRS.2013.2295819.

Petitjean, F., J. Inglada, and P. Gançarski. 2012. "Satellite Image Time Series Analysis under Time Warping." *IEEE Transactions on Geoscience and Remote Sensing* 50 (8): 3081–3095. doi:10.1109/TGRS.2011.2179050.

Qin, R. 2014a. "Change Detection on LOD 2 Building Models with Very High Resolution Spaceborne Stereo Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 96: 179–192. doi:10.1016/j.isprsjprs.2014.07.007.

Qin, R. 2014b. "A Mean Shift Vector-Based Shape Feature for Classification of High Spatial Resolution Remotely Sensed Imagery." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. Advance online publication. doi:10.1109/JSTARS.2014.2357832.

Qin, R. 2014c. "An Object-Based Hierarchical Method for Change Detection Using Unmanned Aerial Vehicle Images." *Remote Sensing* 6 (9): 7911–7932. doi:10.3390/rs6097911.

Qin, R., and W. Fang. 2014. "A Hierarchical Building Detection Method for Very High Resolution Remotely Sensed Images Combined with DSM Using Graph Cut Optimization." *Photogrammetry Engineering and Remote Sensing* 80 (8): 37–48. doi:10.14358/PERS.80.9.000.

Qin, R., J. Gong, and C. Fan. 2010. "Multi-Frame Image Super-Resolution Based on Knife-Edges." In *IEEE International Conference on Signal Processing (ICSP)*, 972–975. Beijing, China. 24–28, October.

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 & Remote Sensing* 79 (8): 717–730. doi:10.14358/PERS.79.8.717.

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 (2014): 23–35. doi:10.1016/j.isprsjprs.2014.01.006.

Rottensteiner, F., J. Trinder, S. Clode, and K. Kubik. 2005. "Using the Dempster–Shafer Method for the Fusion of Lidar Data and Multi-Spectral Images for Building Detection." *Information Fusion* 6 (4): 283–300. doi:10.1016/j.inffus.2004.06.004.

Schindler, G., and F. Dellaert. 2010. "Probabilistic Temporal Inference on Reconstructed 3d Scenes." In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 1410–1417.

Shafer, G. 1976. *A Mathematical Theory of Evidence*, 314. Princeton, NJ: Princeton University Press.

Sirmacek, B., and C. Unsalan. 2011. "A Probabilistic Framework to Detect Buildings in Aerial and Satellite Images." *IEEE Transactions on Geoscience and Remote Sensing* 49 (1): 211–221. doi:10.1109/TGRS.2010.2053713.

Taneja, A., L. Ballan, and M. Pollefeys. 2011. "Image Based Detection of Geometric Changes in Urban Environments." In *Proceedings of IEEE International Conference on Computer Vision*, 2336–2343. Barcelona, Spain, November 6–13.

Tian, J., S. Cui, and P. Reinartz. 2014. "Building Change Detection Based on Satellite Stereo Imagery and Digital Surface Models." *IEEE Transactions on Geoscience and Remote Sensing* 52 (1): 406–417. doi:10.1109/TGRS.2013.2240692.

Tian, J., and P. Reinartz. 2013. "Fusion of Multi-Spectral Bands and DSM from WorldView-2 Stereo Imagery for Building Extraction." In *2013 Joint Urban Remote Sensing Event (JURSE)*, Sao Paulo, April 21–23, 135–138.

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. doi:10.1016/j.isprsjprs.2013.02.017.

Tomasi, C., and R. Manduchi. 1998. "Bilateral Filtering for Gray and Color Images." In *International Conference on Computer Vision*, 839–846.

Tuia, D., F. Ratle, A. Pozdnoukhov, and G. Camps-Valls. 2010. "Multisource Composite Kernels for Urban-Image Classification." *IEEE Geoscience and Remote Sensing Letters* 7 (1): 88–92. doi:10.1109/LGRS.2009.2015341.

Turker, M., and K. San. 2010. "Building Detection from Pan-Sharpened Ikonos Imagery through Support Vector Machines Classification." In *ISPRS Technical Commission VIII Symposium, Networking the World with Remote Sensing International, ISPRS Archives*, 841–846.

Turlapaty, A., B. Gokaraju, Q. Du, N. H. Younan, and J. V. Aanstoos. 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. doi:10.1109/JSTARS.2011.2179792.

Vincent, L. 1993. "Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms." *IEEE Transactions on Image Processing* 2 (2): 176–201. doi:10.1109/83.217222.

Wang, L. 2005. *Support Vector Machines: Theory and Applications*, 431. Berlin: Springer.

Zhang, L., X. Huang, B. Huang, and P. Li. 2006. "A Pixel Shape Index Coupled with Spectral Information for Classification of High Spatial Resolution Remotely Sensed Imagery." *IEEE Transactions on Geoscience and Remote Sensing* 44 (10): 2950–2961. doi:10.1109/TGRS.2006.876704.

Zhang, Y. 1999. "Optimisation of Building Detection in Satellite Images by Combining Multispectral Classification and Texture Filtering." *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (1): 50–60. doi:10.1016/S0924-2716(98)00027-6.