International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,
Volume XL-1/W1, ISPRS Hannover Workshop 2013, 21 – 24 May 2013, Hannover, Germany

EXPLOITING SHADOW EVIDENCE AND ITERATIVE GRAPH-CUTS FOR EFFICIENT DETECTION OF BUILDINGS IN COMPLEX ENVIRONMENTS

A. O. Ok [a, *], C. Senaras [b, c], B. Yuksel [d]

[a] Department of Civil Engineering, Faculty of Engineering, Mersin University, 33343, Mersin, Turkey - oozgun@mersin.edu.tr
[b] HAVELSAN A.S., Ankara 06520, Turkey
[c] Informatics Institute, Middle East Technical University, Ankara 06531, Turkey - csenaras@havelsan.com.tr
[d] Department of Computer Engineering, Middle East Technical University, Ankara 06531, Turkey - e1449826@ceng.metu.edu.tr

Commission III, WG III/4

KEY WORDS: Building Detection, Shadow Evidence, Iterative Graph-cuts, Fuzzy Landscapes, Optical Imagery

ABSTRACT:

This paper presents an automated approach for efficient detection of building regions in complex environments. We investigate the shadow evidence to focus on building regions, and the shadow areas are detected by recently developed false colour shadow detector. The directional spatial relationship between buildings and their shadows in image space is modelled with the prior knowledge of illumination direction. To do that, an approach based on fuzzy landscapes is presented. Once all landscapes are collected, a pruning process is applied to eliminate the landscapes that may occur due to non-building objects. Thereafter, we benefit from a graph-theoretic approach to accurately detect building regions. We consider the building detection task as a binary partitioning problem where a building region has to be accurately separated from its background. To solve the two-class partitioning, an iterative binary graph-cut optimization is performed. In this paper, we redesign the input requirements of the iterative partitioning from the previously detected landscape regions, so that the approach gains an efficient fully automated behaviour for the detection of buildings. Experiments performed on 10 test images selected from QuickBird (0.6 m) and Geoeye-1 (0.5 m) high resolution datasets showed that the presented approach accurately localizes and detects buildings with arbitrary shapes and sizes in complex environments. The tests also reveal that even under challenging environmental and illumination conditions (e.g. low solar elevation angles, snow cover) reasonable building detection performances could be achieved by the proposed approach.

1. INTRODUCTION

Space-borne imaging is a standard way of acquiring information about the objects on the Earth surface. Today, the information obtained is rather diverse and high-quality due to the advanced capabilities of satellite imaging such as the availability of sub-meter resolution optical sensors, broadened spectral sensitivity, and increased data availability. Thus, satellite images are one of the most important data input source to be utilized for the purpose of object detection.

It is a fact that most of the human population lives in urban and sub-urban environments. Therefore, the detection of man-made features from satellite images is of great practical interest for a number of applications such as urban monitoring, change detection, estimation of human population etc. In an early work, Huertas and Nevatia (1988) emphasized the importance of the automation for the detection, and they also stated the major task: the extraction and description of man-made objects, such as buildings. Up to now from their early paper, various researchers belonging to different scientific communities involved for the same task, and accordingly, a significant number of research studies have been published. Since this paper is devoted to the automated detection of buildings from a single optical image, we very briefly summarize the previous studies aimed to automatically detect buildings from monocular optical images.

The pioneering studies for the automated detection of buildings were in the context of single imagery, in which the low-level features were grouped to form building hypotheses (e.g. Huertas and Nevatia, 1988; Irvin and Mckeown, 1989). Besides, a large number of methods proposed substantially benefit from the cast shadows of buildings (e.g. Huertas and Nevatia, 1988; Irvin and Mckeown, 1989; McGlone and Shufelt, 1994; Lin and Nevatia, 1998; Peng and Liu, 2005; Katartzis and Sahli, 2008; Akçay and Aksoy, 2010). Further studies devoted to single imagery utilized the advantages of multi-spectral evidence, and attempted to solve the detection problem in a classification framework (e.g. Benediktsson et al., 2003; Lee et al., 2003; Shackelford and Davis, 2003; Ünsalan and Boyer, 2005; Inglada, 2007; Senaras et al., 2013; Sümer and Turker, 2013). Besides, approaches like active contours (e.g. Peng and Liu, 2005; Cao and Yang, 2007; Karantzalos and Paragios, 2009; Ahmadi et al., 2010), Markov Random Fields (MRFs) (e.g. Krishnamachari and Chellappa, 1996; Katartzis and Sahli, 2008), graph-based (e.g. Kim and Muller, 1999; Sirmacek and Unsalan, 2009; Izadi and Saeedi, 2012) and kernel-based (Sirmacek and Unsalan, 2011) approaches were also investigated.

In this paper, we present an automated approach for the detection of building regions from single optical satellite imagery. To focus on building regions, we exploit the cast shadows of buildings, and the shadow areas are detected by recently proposed false colour shadow detector (Teke et al., 2011). The directional spatial relationship between buildings and their shadows in image space is modelled with the prior knowledge of illumination direction. To do that, an approach based on fuzzy landscapes is presented. Once all landscapes are collected, a pruning process is applied to eliminate the landscapes that may occur due to non-building objects.

Thereafter, we benefit from a graph-theoretic approach to accurately detect building regions. In this paper, we consider the building detection task as a binary partitioning problem where a building region has to be accurately separated from its background. One of our insights is that such a problem can be formulated as a two-class labelling problem (building/non-building) in which a building class in an image corresponds only to the pixels that belong to building regions, whereas a non-building class may involve pixels that do not belong to any of building areas (e.g., vegetation, shadow, and roads). To solve the two-class partitioning, an iterative binary graph-cut optimization (Rother et al., 2004) is carried out. This optimization is performed in region-of-interests (ROIs) generated automatically for each building region, and assigning the input requirements of the iterative partitioning in an automated manner turns the framework into a fully unsupervised approach for the detection of buildings.

The individual stages of our approach will be described in the subsequent section. Some of these stages are already well-described in Ok et al. (2013), and therefore, these stages are only revised here. Besides, this paper extends our previous work from two aspects. First, we aim to improve the pruning step before the detection of building regions. Because water bodies appear dark both in visible and NIR spectrum, the shadow detector utilized detects water bodies as shadow. To mitigate this problem, we extend the pruning step in which we investigate the length of each shadow component in the direction of illumination by enforcing a pre-defined maximum height threshold for buildings. In this way, we eliminate the landscapes generated from large water bodies before the detection of building regions. Second, we improve the way used to generate ROIs. In our previous work, the bounding box of each ROI was extracted automatically after dilating the shadow regions. However, we realized that this might cause large ROI regions particularly where the cast shadows of multiple building objects are observed as a single shadow region. To avoid this problem, in this paper, we generate ROIs from the foreground information extracted from the shadow regions, thereby allowing us to better focus on building regions and their close neighbourhood.

The remainder of this paper is organized as follows. The approach is presented in Section 2. The results of the approach are given and discussed in Section 3. The concluding remarks are provided in Section 4.

2. BUILDING DETECTION

2.1 Image and Metadata

The approach requires pan-sharped multi-spectral (B, G, R, and NIR) ortho-images. We assume that the metadata files providing information about the solar angles (azimuth and elevation) of the image acquisition are also attached to the images. By definition, the solar azimuth angle (A) in an ortho-rectified image space is the angle computed from north in a clockwise direction, whereas the solar elevation angle (ϕ) is the angle between the direction of the geometric centre of the sun and the horizon.

2.2 The Detection of Vegetation and Shadow Regions

Normalized Difference Vegetation Index (NDVI) is utilized to detect vegetated areas. The index is designed to enhance the image parts where healthy vegetation is observed; larger values produced by the index in image space most likely indicate the vegetation cover. We use the automatic histogram thresholding based on the Otsu's method (Otsu, 1975) to compute a binary vegetation mask, M_V (Fig. 1b, e). A new index is utilized to detect shadow areas (Teke et al., 2011). The index depends on a ratio computed with the saturation and intensity components of the Hue-Saturation-Intensity (HSI) space, and the basis of the HSI space is a false colour composite image (NIR, R, G). To detect the shadow areas, as also utilized in the case of vegetation extraction, Otsu's method is applied. Thereafter, the regions belonging to the vegetation cover are subtracted to obtain a binary shadow mask, M_S (Fig. 1c, f).

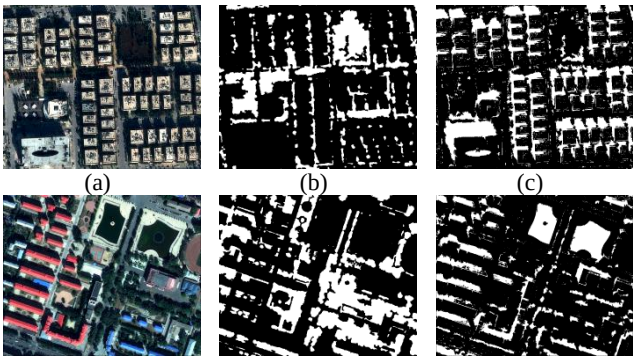


Figure 1. (a, d) Geoeye-1 pan-sharped images (RGB), the (b, e) vegetation masks (M_V), and (c, f) shadow masks (M_S).

2.3 The Generation and Pruning of Fuzzy Landscapes

Given a shadow object B (e.g. each 8-connected component in M_S) and a non-flat line-based structuring element $v_{L,\alpha,\sigma,\kappa}$, the landscape $\beta_\alpha(B)$ around the shadow object along the given direction α can be defined as a fuzzy set of membership values in image space (Ok et al., 2013):

$$\beta_\alpha(B) = \left(B^{per} \oplus v_{L,\alpha,\sigma,\kappa}\right) \cap B^c \,. \tag{1}$$

In Eq. 1, B^{per} represents the perimeter pixels of the shadow object B, B^C is the complement of the shadow object B, and the operators $\oplus$ and $\cap$ denote the morphological dilation and a fuzzy intersection, respectively. The landscape membership values are defined in the range of 0 and 1, and the membership values of the landscapes generated using Eq. 1 decrease while moving away from the shadow object, and bounded in a region defined by the object's extents and the direction defined by angle α. In Eq. 1, we use a line-based non-flat structuring element $v_{L,\alpha,\sigma,\kappa}$ generated by combining two different structuring elements with a pixel-wise multiplication (*):

$$v_{L,\alpha,\sigma,\kappa} = v_{L,\kappa,\alpha} * v_{\sigma,\kappa} \,. \tag{2}$$

In Eq. 2, $v_{\sigma,\kappa}$ is an isotropic non-flat structuring element with kernel size κ, and the decrease rate of the membership values within the element is controlled by a single parameter σ

$$v_{\sigma,\kappa}(x) = e^{-\left(\frac{\|\overrightarrow{ox}\|}{\sigma}\right)} \max\left\{0, 1 - \frac{2\|\overrightarrow{ox}\|}{\kappa}\right\} \,, \tag{3}$$

where $\|\overrightarrow{ox}\|$ is the Euclidean distance of a point x to the centre of the structuring element. On the other hand, the flat structuring element $v_{L,\kappa,\alpha}$ is responsible to provide directional information $D(L)$ where L denotes the line segment and α is the angle where the line is directed

$$v_{L,\kappa,\alpha}(x) = round\left(1 - \frac{\theta_\alpha(x,o)}{\pi}\right) D(L) \,, \tag{4}$$

where the *round*(.) operator maps the computed membership values to the nearest integer and $\theta_\alpha(x,o)$ denotes the angle

differences computed between the unit vector along the direction α and the vector from kernel centre point (o) to any point x on the kernel. In this paper, we utilized the parameter combination κ = 40 m and σ = 100 which successfully characterizes the neighbourhood region of a building region.

During the pruning step, we investigate the vegetation evidence within the directional neighbourhood of the shadow regions. At the end of this step, we remove the landscapes that are generated from the cast shadows of vegetation canopies. To do that, we define a search region in the immediate vicinity of each shadow object by applying two thresholds (T_{low} = 0.7, T_{high} =0.9) to the membership values of the fuzzy landscapes generated. Once the region is defined, we search for vegetation evidence within the defined region using the vegetation mask, M_V, and reject a fuzzy landscape region generated from a cast shadow if there is substantial evidence of vegetation ($\geq$ 0.7) within the search region (Fig. 2).

We assess the height difference of the objects compared to the terrain height to separate the landscapes of building and other non-building objects. Based on the assumption that the surfaces on which shadows fall are flat, it is possible to investigate the length of the shadow objects in the direction of illumination to enforce a pre-defined height threshold value. To do that, for a given solar elevation angle (ϕ) and height threshold (T_H), we compute the shadow length (L_H) that should be cast by a building: $L_H = T_H / \tan(\phi)$. Thereafter, we generate a directional flat structuring element whose length is equal to L_H in the direction of illumination. Since the perimeter pixels of the shadow objects are already computed (B^{per}), for each shadow object, we use a directional flat structuring element to search the number of perimeter pixels that satisfies the length L_H. In this study, we apply two height thresholds two limit the height of building regions. The lower threshold T_H^1 = 3 m eliminates the fuzzy landscapes arise due to short non-building objects such as cars, garden walls etc., and if none of the perimeter pixels of a shadow object is found to be satisfying L_H^1, the generated fuzzy landscape is rejected. The upper threshold T_H^2 = 50 m discards the fuzzy landscapes generated from large dark regions such as water bodies which are incorrectly identified as shadow region by the shadow detector (Fig. 2f). To do that, we eliminate the landscapes if at least one of the perimeter pixels of a shadow object satisfies L_H^2.

2.4 Detection of Building Regions using Iterative Graph-cuts

In this paper, we consider the building detection task as a two-class partitioning problem where a given building region has to be separated from its background accurately (building/non-building). Therefore, the class *building* in an image corresponds only to the pixels that belong to building regions, whereas the class *non-building* may involve pixels that do not belong to any of building areas (e.g. vegetation, shadow, roads etc.). To solve the partitioning, we utilized the GrabCut approach (Rother et al., 2004) in which an iterative binary-label graph-cut optimization is performed.

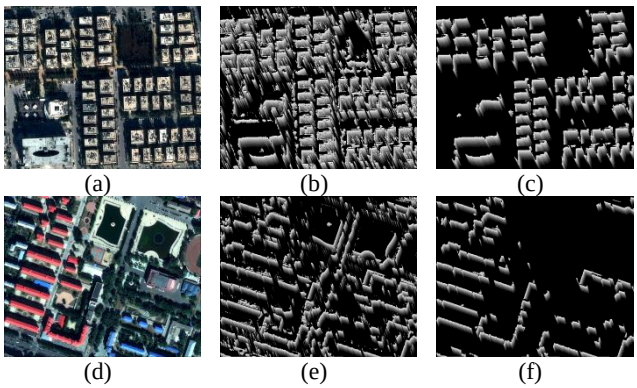


Figure 2. (a, d) Geoeye-1 pan-sharped images (RGB). The fuzzy landscapes generated using the shadow masks provided in Fig. 1c, f are illustrated in (b, e), respectively. The fuzzy landscapes after applying the pruning step are shown in (c, h).

GrabCut is originally semi-automated foreground/background partitioning algorithm. Given a group of pixels interactively labelled by the user, it partitions the pixels in an image using a graph-theoretic approach. Given a set of image pixels $\mathbf{z} = (z_1, z_2, \dots, z_N)$ in an image space, each pixel has an initial labelling from a trimap $T = \{T_B, T_F, T_U\}$, where T_B and T_F represent the background and foreground label information provided by the user respectively, and T_U denotes the unlabelled pixels. In addition, each pixel has an initially assigned value $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_N)$ corresponding to background or foreground where $\alpha_n \in \{0,1\}$ and the underline operator indicates the parameters to be estimated/solved. At the first stage of the algorithm, two GMMs with K components for the foreground (K_F) and the background classes (K_B) are constructed from the pixels manually labelled by the user. Let us define $\mathbf{k} = \{k_1, k_2, \dots, k_N\}$ with $k_n \in \{1, \dots, K\}$ as the vector representing the mixture components for each pixel. Then, the Gibbs energy function for the partitioning can be written as

$$E(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{z}) = U(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{z}) + V(\underline{\alpha}, \mathbf{z}) \quad (5)$$

where $\underline{\theta}$ denotes the probability density function to be obtained by mixture modeling for each pixel. In Equ. 5, $U(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{z})$ denotes the fit of the background/foreground mixture models to the data $\mathbf{z}$ considering α values, and defined as

$$U(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{z}) = -\sum_n D(\alpha_n, k_n, \underline{\theta}, z_n) \quad (6)$$

where $D(\alpha_n, k_n, \underline{\theta}, z_n)$ favor the label preferences for each pixel z_n based on the observed pixel values. On the other hand, $V(\underline{\alpha}, \mathbf{z})$ is the boundary smoothness and is written as

$$V(\underline{\alpha}, \mathbf{z}) = \gamma_1 \sum_{(m,n)\in C} [\alpha_n \neq \alpha_m] e^{-\beta \|z_m - z_n\|^2} \quad (7)$$

where the term $[\alpha_n \neq \alpha_m]$ can be considered as an indicator function getting a binary value 1 if $\alpha_n \neq \alpha_m$, and 0 if $\alpha_n = \alpha_m$, C is the set of neighboring pixel pairs computed in 8-neighborhood, β and γ_1 are the constants determining the degree of smoothness. The smoothness term β is computed automatically after evaluating all the pixels in an image, and the other smoothness term γ_1 is fixed to a constant value (that is 50) after investigating a set of images. To complete the partitioning and to estimate the final labels of all pixels in the image, a minimum-cut/max-flow algorithm is utilized. Thus, the whole framework of the GrabCut partitioning algorithm can be summarized as a two-step process (Rother et al., 2004):

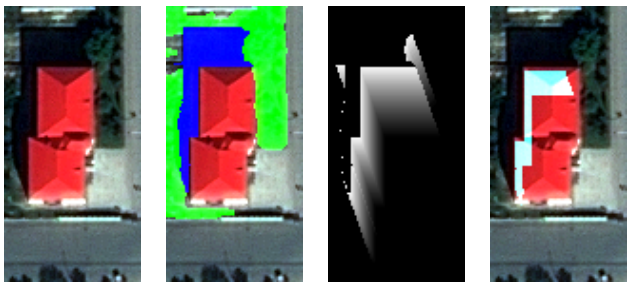
Initialization:
(i) Initialize T_B, T_F, and T_U from the user.
(ii) Set $\alpha_n = 0$ for $n \in T_B$, and $\alpha_n = 1$ for $n \in \overline{T_B}$, complement of the background.
(iii) Initialize mixture models for T_B and T_F.

Iterative minimization:
(iv) Assign GMM components for each n in T_U , are assigned.
(v) Extract GMM parameters from data z.
(vi) Solve the optimization using min-cut/max-flow
(vii) Repeat steps *(iv)-(vi)* until convergence.

As can be seen from (*i*), the initialization of the iterative partitioning requires user interaction. The pixels corresponding to foreground (T_F) and background (T_B) classes must be labelled by the user, and after that, the rest of the pixels in an image is partitioned. In this part, we integrate the iterative partitioning approach to an automated building detection framework. We term T_F to the image pixels that are most likely to belong to building areas. On the other hand, T_B of an image corresponds to the pixels of non-building areas. We present a shadow component-wise approach to focus on the local neighbourhood of the buildings to define T_F. It is a basic common fact of all images is that the shadows cast by building objects are located next to their boundaries (Fig. 3a). Thus, T_F can be extracted automatically from the directional neighbourhood of each shadow component with the previously generated fuzzy landscapes. To do that, we define the T_F region in the vicinity of each shadow object whose extents are outlined after applying a double thresholding (η_1 = 0.9, η_2 = 0.4) to the membership values of the fuzzy landscape generated (Fig. 3d). To acquire a fully reliable T_F region, a refinement procedure that involves a single parameter, shrinking distance (d = 2 m), is also performed (Ok et al., 2013).

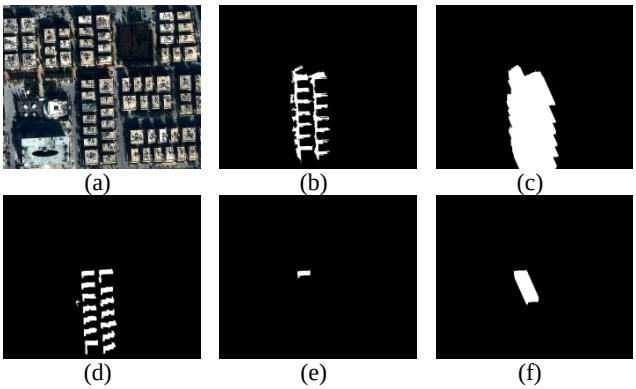
In this study, we present a region-of-interest (ROI) based iterative partitioning. In Ok et al. (2013), we performed the iterative partitioning locally for each shadow component in a bounding box covering only a specific ROI region whose extents were extracted automatically after dilating the shadow region. The dilation was performed with a flat line kernel defined in the opposite direction of illumination, and since the ROI must include all parts of a building to be detected, the size of the building in the direction of illumination was taken into account. During the generation of the ROIs, the size was controlled by a single dilation distance parameter, ROI size (= 50 m), which was also defined in the opposite direction of illumination. The bounding boxes generated by dilating the shadow components works well for most of the cases; however for certain conditions (e.g. acute solar elevation angles, dense environments etc.), it might cause large ROI regions to be produced for multiple building objects (Fig. 4c). To avoid this problem, as original to this work, we generate ROIs from the foreground information T_F (Fig. 4d). Since the generated T_F regions are separated for such cases, this provides an opportunity to define the ROIs in a separate manner (Fig. 4f). Thus, this strategy allows us to better focus on individual building regions and their close neighbourhood independent from the shadow component utilized.

Once the bounding box of a specific ROI is determined, we automatically set up the pixels corresponding to background information (T_B) within the selected bounding box. To do that,



(a) (b) (c) (d)

Figure 3. (a) Geoeye-1 image (RGB), (b) the detected shadow (blue) and vegetation (green) masks. (c) Fuzzy landscape generated from the shadow object with the proposed line-based non-flat structuring element, (d) the final foreground pixels (T_F) overlaid with the original image.



(a) (b) (c) (d) (e) (f)

Figure 4. (a) Geoeye-1 image (RGB), (b) a single shadow component detected, and (c) the large ROI region generated. (d) The foreground information T_F (without refinement) generated from the shadow component in (b). (e) One of the T_F regions and (f) the ROI formed for that region after dilation.

we search for the shadow and vegetation evidences within the bounding box and we label all those areas as T_B. In addition, we also label the regions outside the ROI region within the bounding box as T_B since we only aim to detect buildings within the ROI region for a given foreground information.

Finally, we remove the small-sized artefacts that may occur after the detection stage. To do that, a threshold (T_{area} = 30 m^2) is employed to define the minimum area enclosed by a single building region.

3. RESULTS AND DISCUSSION

The test data involve images acquired from two different satellites (QuickBird and Geoeye-1) which are capable of providing sub-meter resolution imagery, and all images are composed of four multi-spectral bands (R, G, B and NIR) with a radiometric resolution of 11 bits per band. The assessments of the proposed approach are performed over 10 test images which differ from their urban area and building characteristics as well as from their illumination and acquisition conditions. The first three test images (#1-3) belong to a QuickBird image, whereas the rest (#4-10) is selected from different Geoeye-1 images. The solar elevation angles tested range between 21.54° and 78.12° and the images were acquired with off-nadir angles of at most ≈ 18 degrees. To assess the quality of our results, they are compared to reference data. The precision, recall and F_1-score (Aksoy et al., 2012; Ok et al., 2013) performance measures are determined both on a per-pixel and per-object level. For the object based evaluation, a building region is considered to be a true positive if 60% of its area is covered by a building region in the reference.

We visualize the detection results in Fig 5, and according to the results presented, the developed approach seems to be robust and the regions detected are found to be satisfactory. The building regions are well detected despite the complex characteristics of buildings in the test images, e.g. roof colour and texture, shape, size and orientation. The numerical results in Table 1 favour these facts. Considering the per-pixel evaluation, overall mean ratios of precision and recall are computed as 79.1% and 85.5%, respectively. The computed pixel-based F_1-score for all test images is around 82%. In view of the per-object evaluation, overall mean ratios of precision and recall are computed as 92.8% and 79.9%, respectively. This corresponds to an overall object-based F_1-score of approximately 86%. If the complexities of the test images and the involved imaging conditions are jointly taken into consideration, we believe that this is a promising building detection performance.

Figure 5. (first column) Test dataset (#1-10), (second column) the results of per-pixel evaluation, and (third column) the results of per-object evaluation. Green, red and blue colours represent true-positive, false-positive and false-negative, respectively.

Table 1. Performance results of the proposed approach.

ID	Performance (%)					
	Per-Pixel Level			Per-Object Level		
	Precision	*Recall*	F_1-*score*	*Precision*	*Recall*	F_1-*score*
#1	89.6	95.6	92.5	100	93.9	96.8
#2	72.5	89.6	80.2	97.2	85.4	90.0
#3	64.8	95.8	77.3	78.6	91.7	84.6
#4	82.7	90.2	86.3	95.8	86.3	90.8
#5	74.2	76.4	75.3	98.6	75.5	85.5
#6	79.6	90.1	84.5	100	76.9	87.0
#7	78.4	81.6	80.0	77.8	76.1	76.9
#8	78.3	83.1	80.6	97.3	83.7	90.0
#9	87.0	87.7	87.3	76.4	79.7	78.0
#10	40.6	63.5	49.6	73.7	60.9	66.7
Total	79.1	85.5	82.2	92.8	79.9	85.9

The lowest precision and recall ratios for both per-pixel and per-object assessment are obtained for test image #10. Actually, this is not surprising since that image is acquired in winter season with a very low solar elevation angle (21.54°). Thus, the region is covered by snow. This fact and the fact that the low solar elevation angle causes severe shading effects on building rooftops (especially for buildings having gable roof styles with specific orientation) limit the detection. Besides, it is rather difficult to detect shadow areas in a snow covered image because the cast shadows of buildings fall over a bright colour may significantly reduce the saturation component of the shadow region. As a result, the effectiveness and the performance of the index used to detect shadow areas reduce dramatically, which also have a major influence on the final performance of the proposed approach. The second lowest precision performance of per-pixel evaluation is achieved for test image #3 and the main reason is the two large bridges used for vehicular traffic on the upper-right corner of the image. The height threshold T_H^1 works well to eliminate the landscapes generated from non-building objects since the shadows of these objects generally have height differences less than 3m compared to terrain height. However, in certain cases such as large bridges, the height of non-building objects exceeds the given threshold. As a result, it is not possible to avoid such cases and some parts of the road segments might be labelled as building regions. Besides, our approach may over-detect some building boundaries. This is due to two specific reasons. First, some parts of the building boundaries may have very smooth transition between their surroundings. Second, a building may involve several roof parts that are identical to their surroundings although the main colour of the rooftop is distinguishable from its background. Nevertheless, we think that most of the over-detections can be corrected with further high-level processing.

The results show that the approach presented is generic for different roof colours, textures and types, and has the ability to detect arbitrarily shaped buildings in complex environments. According to the results provided in Table 1, the highest F_1-scores are achieved for test image #1 where the buildings are formed in a single-detached style. Besides, for most of the test images, our approach provides quite satisfactory object-based ratios. Apparently, this is due to the reason that our approach labels a region as building only if a valid shadow region is detected. Therefore, we can conclude that the presented approach for building detection is robust from an object-based point-of-view. Besides the mentioned advantages, the proposed approach is also time-efficient. The images are processed on a PC with a CPU Intel i5 2.6GHz and 4GB RAM, and the processing requires less than 30 seconds for each image on average.

4. CONCLUSIONS

In this paper, a novel approach is presented to detect building regions from a single high resolution multispectral image. First, vegetation and shadow areas are extracted with the help of the multi-spectral information widely accessible to the most of the high resolution satellite images. The spatial relationship between buildings and their cast shadows is modelled by means of a fuzzy landscape approach and a pruning process is applied to eliminate the landscapes belonging to non-building objects. The final building regions are detected by iterative graph partitioning. In this study, the input requirements of the iterative partitioning are extracted automatically so that the framework turns out to be an efficient approach for the detection of buildings. Assessments performed on 10 test images selected from QuickBird and Geoeye-1 images reveal that the approach accurately localizes and detects buildings with arbitrary shapes, sizes, colours in complex environments. The tests also reveal that even under challenging environmental and illumination conditions, reasonable building detection performances could be achieved by the proposed approach.

In the near future, we will focus to reduce the limitations of the proposed approach. A major task is to separate large bridges from buildings; therefore, we plan to develop and integrate a different method that is particularly designed for road and/or bridge detection. In this way, the road segments that are erroneously labelled due to large bridges can be identified and eliminated. As a different work, we plan to extend the graph-cut optimization in a multi-label manner, and this improvement will further improve the results of the presented approach. An additional post-processing step that involves the simplification of the outlines of the detected building regions is also a required task and we will pursue in the near future.

ACKNOWLEDGEMENTS

This work was supported by HAVELSAN A.Ş.

REFERENCES

Ahmadi, S., Zoej, M.J.V., Ebadi, H., Moghaddam, H.A., Mohammadzadeh, A., 2010. Automatic urban building boundary extraction from high resolution aerial images using an innovative model of active contours. International Journal of Applied Earth Observation and Geoinformation, 12(3), pp. 150-157.

Akçay, H.G., Aksoy, S., 2010. Building detection using directional spatial constraints. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 1932–1935.

Aksoy, S., Yalniz, I.Z., Tasdemir, K., 2012. Automatic detection and segmentation of orchards using very high resolution imagery. IEEE Transactions on Geoscience and Remote Sensing, 50(8), pp. 3117-3131.

Benediktsson, J.A., Pesaresi, M., Arnason, K., 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), pp. 1940-1949.

Cao, G., Yang, X., 2007. Man-made object detection in aerial images using multi-stage level set evolution. International Journal of Remote Sensing, 28(8), pp. 1747-1757.

Huertas, A., Nevatia, R., 1988. Detecting buildings in aerial images. Computer Vision, Graphics, and Image Processing, 41(2), pp. 131-152.

Inglada, J., 2007. Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. ISPRS Journal of Photogrammetry and Remote Sensing, 62(3), pp. 236-248.

Irvin, R.B., Mckeown, D.M., 1989. Methods for exploiting the relationship between buildings and their shadows in aerial imagery. IEEE Transactions on Systems, Man, and Cybernetics, 19(6), pp. 1564-1575.

Izadi, M., Saeedi, P., 2012. Three-Dimensional polygonal building model estimation from single satellite images. IEEE Transactions on Geoscience and Remote Sensing, 50(6), pp. 2254-2272.

Karantzalos, K., Paragios, N., 2009. Recognition-driven two-dimensional competing priors toward automatic and accurate building detection. IEEE Transactions on Geoscience and Remote Sensing, 47(1), pp. 133-144.

Katartzis, A., Sahli, H., 2008. A stochastic framework for the identification of building rooftops using a single remote sensing image. IEEE Transactions on Geoscience and Remote Sensing, 46(1), pp. 259-271.

Kim, T.J., Muller, J.P., 1999. Development of a graph-based approach for building detection. Image and Vision Computing, 17(1), pp. 3-14.

Krishnamachari, S., Chellappa, R., 1996. Delineating buildings by grouping lines with MRFs. IEEE Transactions on Image Processing, 5(1), pp. 164-168.

Lee, D.S., Shan, J., Bethel, J.S., 2003. Class-guided building extraction from Ikonos imagery. Photogrammetric Engineering and Remote Sensing, 69(2), pp. 143-150.

Lin, C., Nevatia, R., 1998. Building detection and description from a single intensity image. Computer Vision and Image Understanding, 72(2), pp. 101-121.

McGlone, J.C., Shufelt, J.A., 1994. Projective and object space geometry for monocular building extraction. In: Proc. of Computer Vision and Pattern Recognition, pp. 54-61.

Ok, A.O., Senaras, C., Yuksel, B., 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), pp. 1701-1717.

Otsu, N., 1975. A threshold selection method from gray-level histograms. Automatica, 11, pp. 285-296.

Peng, J., Liu, Y.C., 2005. Model and context-driven building extraction in dense urban aerial images. International Journal of Remote Sensing, 26(7), pp. 1289-1307.

Rother, C., Kolmogorov, V., Blake, A., 2004. Grabcut: interactive foreground extraction using iterated graph cuts, ACM Transactions on Graphics, 23(3), pp. 309-314.

Senaras, C., Özay M., Vural, F. Y., 2013. Building detection with decision fusion. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 6(3), in-press.

Shackelford, A.K., Davis, C.H., 2003. A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas. IEEE Transactions on Geoscience and Remote Sensing, 41(10), pp. 2354-2363.

Sirmacek, B., Unsalan, C., 2009. Urban-area and building detection using SIFT keypoints and graph theory. IEEE Transactions on Geoscience and Remote Sensing, 47(4), pp. 1156-1167.

Sirmacek, B., Unsalan, C., 2011. A Probabilistic Framework to Detect Buildings in Aerial and Satellite Images. IEEE Transactions on Geoscience and Remote Sensing, 49(1), pp. 211-221.

Sümer, E., Turker, M., 2013. An adaptive fuzzy-genetic algorithm approach for building detection using high-resolution satellite images. Computers, Environment and Urban Systems (in-press).

Teke, M., Başeski, E., Ok, A.Ö., Yüksel, B., Şenaras, Ç., 2011. Multi-spectral false color shadow detection, In: Stilla, U., Rottensteiner, F., Mayer, H., Jutzi, B., Butenuth, M. (Eds.), Photogrammetric Image Analysis. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 109-119.

Ünsalan, C., Boyer, K.L., 2005. A system to detect houses and residential street networks in multispectral satellite images. Computer Vision and Image Understanding, 98(3), pp. 423-461.