Automated detection of buildings and roads in urban areas from VHR satellite images

Ali Özgün Ok*
Department of Geodesy and Photogrammetry, Nevşehir H.B.V. University, 50300, Nevşehir,

Abstract

In this paper, we present an unsupervised approach to detect regions belonging to buildings and roads in urban areas from very high resolution (VHR) satellite images. The proposed approach consists of three main stages. In the first stage, we extract information that is only related to building regions using shadow evidence and probabilistic fuzzy landscapes. First, the shadow areas cast by building objects are detected, and the directional spatial relationship between buildings and their shadows is modeled with the knowledge of the illumination direction. Thereafter, each shadow region is handled separately and the initial building regions are identified by iterative graph-cuts designed in two-label partitioning. The second stage of the framework automatically classifies the image into four classes: building, shadow, vegetation, and others. In this step, the previously labeled building regions as well as the shadow and vegetation areas are involved in a four-label graph optimization performed on the entire image domain to achieve the unsupervised classification result. The final stage aims to extend this classification to five classes, including the road class. For that purpose, we extract the regions that might belong to road segments and utilize that information in a final graph optimization. This final stage eventually characterizes the regions belonging to buildings and roads. Experiments performed on twelve test images selected from GeoEye-1 VHR datasets show that the presented approach has the ability to extract the regions belonging to buildings and roads in a single graph theory framework.

Keywords
Building Detection, Road Detection, Fuzzy Landscapes, Multi-label Graph Partitioning, VHR Satellite Imagery

Özet
Çok yüksek çözünürlüklü uydu görüntülerinden kentsel alanlardaki binaların ve yolların otomatik tespiti


Anahtar Sözcükler
Bina Tespiti, Yol Tespiti, Olasılık Haritaları, Çok Sınıflı Çizge-tabanlı Optimizasyon, Uydu Görüntüsü

* Corresponding Author:
E-mail: ozgunok@nevsehir.edu.tr © 2016 HKMO
1. Introduction

Amongst the large set of man-made objects available in very high resolution (VHR) satellite images, buildings and roads unquestionably form the two most important object classes of an urban area. This is mainly because most of the human population lives in urban and sub-urban environments, and extracting information belonging to these two classes in an automated manner could be very useful for a number of applications, e.g., urban area monitoring/detection, change detection, estimation of human population, transportation, and telecommunication. During the last four decades, a very large number of researchers has been involved with the detection of buildings and roads, and many research studies have been conducted. An extensive classification and summary of the previous work in the context of buildings can be found in excellent review papers published by Mayer (1999), Baltisvias (2004), Brenner (2005), and Haala and Kada (2010). In addition, the exceptional review papers of Baumgartner et al. (1997) and Mena (2003) also summarize and describe the literature conducted for road detection. The focus of this paper is on the design and development of a graph theory framework that enables the automatic detection of regions belonging to buildings and roads in urban areas from single VHR satellite images. Therefore, we briefly summarize the previous studies that aimed to automatically detect buildings and/or roads from single 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. In addition, a large number of proposed methods substantially benefit from the cast shadows of buildings (e.g., Lin and Nevatia, 1998; Katartzis and Sahli, 2008; Akgay 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; Unsalan and Boyer, 2005; Ingliad, 2007; Senaras et al., 2013). In addition, active contour (e.g., Karantzalos and Paragios, 2009), Markov Random Field (MRFs) (e.g., Katartzis and Sahli, 2008), graph-based (e.g., Akgay and Aksoy, 2010; Izadi and Saedi, 2012) and kernel-based (Sirmacek and Unsalan, 2011) approaches were also investigated. In recent papers, Ok et al. (2013) and Ok (2013) presented an efficient shadow-based approach to automatically detect buildings with arbitrary shapes in challenging environments.

The previous studies conducted on road detection from monocular optical images are also vast. In early works, rule-based systems and knowledge-based approaches were popular. Various studies integrated morphological processing to achieve a solution for the road extraction problem (e.g., Shi and Zhu, 2002; Guo et al., 2007). Classification strategies utilizing multi-band information have also been well-studied (e.g., Wiedemann and Hinz, 1999; Mena and Malpica, 2005). Approaches like support vector machines (e.g., Huang and Zhang, 2009), neural networks (e.g., Das et al., 2011), and MRFs (e.g., Katartzis et al., 2001) were also investigated. In a comparison study, Mayer et al. (2006) showed the possibility and extent of the extraction of a road centerline using six different approaches. In several recent works, Tournaire and Paparoditis (2009) developed an approach based on Marked Point Processes to extract road markings. Poullis and You (2010) presented a method based on the combination of perceptual grouping and graph-cuts to extract roads. Das et al. (2011) developed an approach based on salient features and a constraint satisfaction neural network to identify road networks. In a different work, Unsalan and Sirmacek (2012) detected the road networks using a probabilistic framework, and their approach was tailored for single-band datasets. In a rather recent work, a conditional random field based approach (Wegner et al., 2013) was proposed to accurately detect regions belonging to road segments from aerial images.

Considering the above prior works, researchers are motivated to obtain a single set of objects, either buildings or roads, and the approaches developed are specialized to distinguish only one of the objects. Despite the fact that the two classes have their own characteristics, both objects are in fact complementary to each other in several aspects: there is a road connection to most of the buildings, a road segment might be occluded by buildings, and so on (Hinz et al., 2001). Therefore, if the extraction of these two objects can be handled in a single framework, this might have a positive effect on the final quality of the detection of each individual class in an urban area. To the best of our knowledge, to date, there have only been a very limited number of studies that realize such an integrated behavior. Hinz and Baumgartner (2000) and Hinz et al. (2001) emphasized the importance of contextual information for road extraction. They proposed an approach based on both the global and local context of roads, including buildings. However, their approach utilized a DSM to detect shadow and building areas to facilitate the road extraction. Unsalan and Boyer (2005) detected houses and streets in a graph-related strategy. Their method extracted the street network using snakes that have unary and binary constraints, and the regions remaining after the detection of the street network were considered to be houses. However, their strategy is only valid for certain house and street formations observed in North America because of the assumptions involved during the detection. Aytekin et al. (2012) also proposed an approach to detect buildings and roads. Their approach was based on segmentation, and the regions belonging to buildings and roads were separated using very primitive morphological processing.

This paper presents a framework to detect regions belonging to buildings and roads from VHR satellite images in an automated manner (Figure 1). The basis of the framework relies on three main stages. Stage I aims to extract information that is only related to building regions. We benefit from the shadow information, and the directional spatial relationship between buildings and their shadows is modeled with fuzzy landscapes. We utilize a systematic procedure to eliminate the fuzzy landscapes that might belong to non-building objects. Thereafter, initial building regions are identified by iterative graph-cuts designed in a two-label (building and others) partitioning performed in region-of-interests (ROI) generated using shadow components. The goal of Stage II is to integrate the global evidence into the framework and to improve the extent of the classification in the entire image.
domain to four distinct classes (building, shadow, vegetation, and others). In this stage, the previously labeled building regions as well as the shadow, vegetation, and other areas are involved in a four-label graph optimization performed in the entire image domain. Stage III aims to eventually characterize the regions belonging to buildings and roads. We extract the regions that might belong to road segments from the other classes and utilize that information to initialize a final graph optimization performed with five different labels (building, shadow, vegetation, road, and others). This final step automatically divides the entire image into five classes where the buildings and roads are eventually identified.

The individual stages of the proposed framework will be described in the subsequent sections. Some of these stages are already well-described in Ok et al. (2013) and Ok (2013), and therefore, these stages are only somewhat revised to provide a complete overview of the methodology. These two previous papers are solely involved the detection of building objects, whereas the originality of this paper arises from the fact that it enables us to concentrate on the building and road classes in a multi-level graph partitioning framework.

The remainder of this paper is organized as follows. The proposed framework is presented in Section 2. The results are shown and discussed in Section 3. The concluding remarks and future directions are provided in Section 4.

2. Proposed Framework

2.1. Image and Metadata

The approach requires a single pan-sharped multi-spectral (B, G, R, and NIR) ortho-image. We utilize single images acquired from GeoEye-1 sensor, and all of the images are composed of one panchromatic (50 cm ground sampling distance - GSD) and four multi-spectral bands (R, G, B and NIR) (2 m GSD) with a radiometric resolution of 11 bits per band. We assume that metadata providing information about the solar angles (azimuth and elevation) of the image acquisition is attached to the image. We benefit from Smoothing Filter-based Intensity Modulation (SFIM) approach (Liu, 2000) available in an open-source image processing and remote sensing toolbox, Orfeo (OTB, 2012), to create a single pan-sharpened multi-spectral image.

2.2. Detection of Vegetation and Shadow Areas

Normalized Differential 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 Otsu’s method (Otsu, 1975) to compute a binary vegetation mask $M_v$ (Figure 2b). A recently proposed 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 color composite image (NIR, R, G). To detect shadow areas, as also utilized in the case of vegetation extraction, Otsu’s method is applied. Thereafter, the regions belonging to vegetated areas are subtracted to obtain a binary shadow mask, $M_s$. We perform a constrained region growing process on the detected shadow regions and apply new directional morphological processing (Ok, 2013) to filter out the shadow areas corresponding to relatively short objects to achieve a post-processed shadow mask $M_{sP}$ (Figure 2c).

2.3. The Generation and Pruning of Fuzzy Landscapes

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

$$\beta_\alpha (B) = (B^{\neg \alpha} \oplus v_{\alpha, \sigma, \kappa}) \cap B^c$$

In Eq. 1, $B^{\neg \alpha}$ 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 decrease while moving away from the shadow object and are bounded in a region defined by the object’s extents and the direction defined by angle $\alpha$ (Figure 3b). In Eq. 1, we use a line-based non-flat structuring element $v_{\alpha, \sigma, \kappa}$ (Figure 3a) generated by combining two different structuring elements with a pixel-wise multiplication: $v_{\alpha, \sigma, \kappa} = v_{\alpha, \sigma, \kappa} \ast v_{\sigma, \kappa}$ is an isotropic non-flat structuring element with kernel size $\kappa$, whereas the flat structuring element $v_{\alpha, \sigma, \kappa}$ is responsible for providing directional information, where $L$ denotes the line segment and $\alpha$ is the angle where the line is directed.

2.4. Stage I: Local Processing to Detect Initial Building Regions

In this stage, we consider the building detection task as a two-class partitioning problem where a given building region has...
to be separated from the background (building vs. others). 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. To solve the partitioning, we expand the GrabCut approach (Rother et al., 2004), in which an iterative binary-label graph-cut optimization (Boykov and Kolmogorov, 2004) is performed using the following energy formula:

$$E(f) = \sum_{n \in P} D_n(f_n) + \gamma_1 \sum_{(n,m) \in N} e^{\beta z_n z_m} I_{n \sim m} f_n$$

In Eq. 2, $D_n(f_n)$ favor the label preferences (foreground/background) for each pixel $n$ based on the observed pixel values $z$ and eventually depends on the results of mixture modeling. $N$ defines the set of neighboring pixel pairs computed in the 8-neighborhood, and $\beta$ and $\gamma_1$ are the constants that determine the degree of smoothness. To complete the partitioning and to estimate the final labels of all pixels in the image, a minimum-cut/max-flow algorithm is applied.

GrabCut is an originally semi-automated foreground/background partitioning algorithm. Given a group of pixels that is interactively labeled by the user, it partitions the pixels in an image using graph theory. In Ok et al. (2013), we adapted the GrabCut approach to an automated building detection framework. In that approach, the pixels corresponding to foreground/building ($TF$) and background/non-building ($TB$) classes are labeled automatically using the shadow regions and the generated fuzzy landscapes. We define the $TF$ region

$$\text{Figure 2: (a) GeoEye-1 pan-sharpened image (RGB), (b) vegetation mask (Mv), (c) post-processed shadow mask (Mps).}$$

$$\text{Figure 3: (a) Structuring element } V_{L,\alpha,\sigma,\kappa}. (b) The fuzzy landscapes generated using the shadow mask provided in Figure 2c, and (c) building regions detected after Stage I.}$$
in the vicinity of each shadow object that has extents that are outlined after applying a double thresholding \((\eta_1, \eta_2)\) to the membership values of the generated fuzzy landscape. To acquire a fully reliable \(T_f\) region, a refinement procedure that involves a single parameter, shrinking distance \((d)\), is also developed. For each shadow component, a bounding box that has an extent that is automatically determined after dilating the shadow region is generated to select the \(T_B\) and to define the ROI region in which the GrabCut partitioning is performed. Once a bounding box is selected, the pixels corresponding to background information within the selected bounding box are automatically determined: the shadow and vegetation regions as well as the regions outside the ROI region within the bounding box are labeled as \(T_B\).

### 2.5. Stage II: Global Processing to Detect Building Regions

After Stage I, the building regions are detected with relatively reduced completeness but with almost no over-detection (Figure 3c). The aim of this stage is to acquire building regions by investigating the global evidence collected for the building regions in the entire image space. The entire image can be divided into four distinct classes with the help of the pre-computed vegetation \((M_V)\) and shadow masks \((M_S)\) after detecting the building regions from the first stage. First, we assign unique labels for the regions belonging to each class: building, vegetation and shadow. Thereafter, the remaining regions that do not correspond to any of these three classes are assigned to a fourth class, others (Figure 4a).

In this stage, we follow the approach proposed in Ok (2013). In that approach, we proposed a single-step four-label graph-cut optimization. Given a set of pixels \(z = \{z_1, z_2, \ldots, z_N\}\) and a set of class labels \(L \in \{1, \ldots, l\}\) where \(l = 4\), our aim is to find the optimal mapping from data \(z\) to class labels \(L\). Each pixel has an initially assigned value \(\alpha_n\) corresponding to each class label \(L\), where \(\alpha_n \in \{1,2,3,4\}\). We follow the Gibbs energy function in Eq. 2 and initialize a GMM with \(K\) components for each of the four classes. We also follow the same expression for the spatial smoothness term provided in (Rother et al., 2004), which states the smoothness priors in relation to the optimal mapping from data \(z\) to class labels \(L\).

To minimize energy for multi-label optimization using graph-cuts, a special graph (Boykov et al., 2001) that depends on the smoothness term and the number of labels \(L\) is constructed. For the optimization, an effective approach is the \(\alpha\)-expansion move algorithm. Here, we briefly describe the \(\alpha\)-expansion approach, and for further information see Boykov et al. (2001). As a formal definition, let \(f\) and \(g\) be two different mappings from data \(z\) to class labels \(L\) and let \(\alpha\) be a specific class label. A mapping \(g\) is defined as an \(\alpha\)-expansion move from mapping \(\bar{f}\) to mapping \(g\). The approach performs cycles for mapping every class label \(L\) in a certain order that is fixed or random and aims to find a mapping from data \(z\) to class labels \(L\) that is a local minimum of the energy with respect to the expansion moves performed. The approach is guaranteed to terminate in finite number of cycles and finishes when no \(\alpha\)-expansion move with lower energy for any class label \(L\) exists.

Despite the fact that the four-label optimization identifies and correctly labels most of the building regions, several non-building regions may have been incorrectly labeled as buildings in the final result due to the spectral similarities involved between some building and non-building areas. Therefore, to solve this problem, we also proposed a new shadow verification approach (Ok, 2013). During the verification, we extract the regions belonging to class building and confirm these regions with the previously generated probabilistic landscape. For the regions that could not be confirmed, we exploit shadow information that may reveal information after the four-label optimization, and the regions of the vicinity of each shadow object that has extents that are outlined after applying a double thresholding \((\eta_1, \eta_2)\) to the membership values of the generated fuzzy landscape. To acquire a fully reliable \(T_f\) region, a refinement procedure that involves a single parameter, shrinking distance \((d)\), is also developed. For each shadow component, a bounding box that has an extent that is automatically determined after dilating the shadow region is generated to select the \(T_B\) and to define the ROI region in which the GrabCut partitioning is performed. Once a bounding box is selected, the pixels corresponding to background information within the selected bounding box are automatically determined: the shadow and vegetation regions as well as the regions outside the ROI region within the bounding box are labeled as \(T_B\).

![Figure 4](image.png)

*Figure 4: (a) The input of Stage II, (b) verified building regions, and (c) the output of Stage II. White, green, black, and grey colors indicate the regions for the classes building, vegetation, shadow and others, respectively.*
rejection are further tested for new shadow evidence.

Thus, our approach has the ability to recover building regions (Figure 4b) that have shadow regions that are missed in the initial shadow mask generated (Figure 2c). Finally, building regions that could not be validated with shadow evidence are joined to the class others (Figure 4c).

2.6. Stage III: Simultaneous Detection of Buildings and Roads

After Stage II, we expect the class others to involve any object other than buildings, vegetation, and shadows. Thus, if we extract only the regions belonging to the class others after Stage II (Figure 5a), this class principally covers a number of subclasses, such as roads, open lands, vehicles, parking lots and so on, in an urban area. However, note that, despite the quality of the four-label optimization achieved after Stage II, some of the building regions might still be incorrectly labelled as the class others (e.g. due to occlusions on shadow regions) or vice versa (e.g. due to bridges). For that reason, in this stage, our objective is to separate the regions that are likely to belong to roads using the information gathered from Stage II. However, as seen in Figure 5a, the regions belonging to the class others can be quite complex; therefore, simple post-processing generally does not work well. Therefore, this final stage aims to expand the previous classification to five classes in which the roads are involved as a separate class. We automatically extract representative regions that are likely to belong to road segments from the class others and utilize that information to initialize five-label graph optimization in the entire image domain. For that purpose, initially, some large patches and other features attached to the road segments must be removed. In a recent work, Das et al. (2011) proposed an efficient strategy to separate road regions from non-road regions. In short, they utilized region part segmentation (RPS) to separate the road regions from attached irrelevant non-road regions, and thereafter, a medial axis transform (MAT) based approach is employed to filter and verify road hypotheses. In this section, we follow their RPS approach to remove irrelevant objects (e.g., parking lots) and modify the MAT approach to collect the regions that are most likely to belong to road regions. We also follow their reasonable hypothesis for MAT processing: the widths of roads do not vary abruptly. After extracting each reasonable sub-component from MAT-based processing, we defined a buffer (2 pixels) around each component to initialize the road information required for the final graph optimization (Figure 5b). We follow the multi-label optimization framework introduced in Stage II, only this time with five classes instead of four: building, road, vegetation, shadow, and others. At the last step, we extract the regions labeled as buildings and roads from the optimization result (Figure 5c) and perform the shadow verification (cf. Stage II, only for the class building) to achieve the final results.

3. Results and Discussion

The assessments of the proposed approach are performed over twelve GeoEye-1 test images (50 cm GSD) that differ according to their urban area and building characteristics as well as their illumination and acquisition conditions. Reference data corresponding to buildings and roads are drawn by an experienced operator.

In this study, we use three well-known quality measures (Precision, Recall, and F₁-score) to evaluate the pixel-level performance of the proposed approach:

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

\[
F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}
\]

In Eqs. 3-5, TP are true positives, FP are false positives, and FN are false negatives. The operator \(| . |\) denotes the number of pixels assigned to each distinct category, and the F₁-score combines Precision and Recall into a single number. The object-level performance of the proposed approach for buildings can also be evaluated with the measures given in Eqs. 3-5. We label an output building object as TP if it has at least a 60% pixel overlap ratio (Aksoy et al., 2012; Ok et al., 2013; Ok, 2013) with a building object in the reference data. However, we label an output object as FP if the output object of the proposed approach does not coincide with any of the building objects in the reference data, and we label an output object as FN if the output object corresponds to a reference object with a limited amount of overlap (< 60%). Note that this object-level assessment might not produce meaningful results for roads because there may be a single whole connected road object in a test image. Therefore, we do not provide object-level results of road detection. The proposed framework requires no training data collection, thus, the results can be computed once the parameter values are determined. All of the parameters required for Stages I and II are already investigated comprehensively in Ok (2013). Therefore, in this study, we considered the parameters required for Stage III and fixed the values of the parameters to values suggested in the study conducted by Das et al. (2011).

We visualize the results in Figure 6, and according to the results presented, the proposed framework seems to be effective for both the detection of buildings and roads. The building regions are very well identified despite the complex characteristics of buildings in the test images, e.g., roof color and texture, shape, size and orientation. The road segments are also recognized for most of the cases on the condition that the segments are not occluded. The numerical results in Tables 1 and 2 favor these facts. For buildings, we achieved overall pixel-level precision and recall as 81.3% and 89.7%, respectively. The computed pixel-level F₁-score for twelve test images is around 85%. In addition, the overall object-level precision and recall are 96.9% and 84.9%, respectively. The computed object-level F₁-score for the twelve test images is 90.5%. For roads, the overall pixel-level precision and recall are computed as 55.1% and 68.8%, respectively. This corresponds to an overall pixel-level F₁-score of approximately 61.2%. If the complexities of the test images and the involved imaging conditions are taken into consideration, we believe that this is a promising performance for the detection of buildings and roads.
Figure 5: (a) The regions belonging to the class others, (b) input for the class road, and (c) the output of Stage III. White, red, green, black, and grey colors indicate the regions for the classes building, road, vegetation, shadow and others, respectively.
Automated Detection of Buildings and Roads in Urban Areas from VHR Satellite Images

According to the results presented in Figure 6, the proposed approach gives a strong impression that the method is highly robust for buildings and that the regions detected are quite convincing and representative. As seen, most of the building regions are extracted successfully without having a strict limitation influenced by the well-known complex characteristics of buildings, e.g., roof color and texture, shape, size and orientation. It is also evident that the approach distinctly separates building regions from other areas, except for a few cases. The lowest precision (25.9%) is obtained for test image #11. That image is acquired in winter season with a very low solar elevation angle (21.54°). Thus, the area is almost completely covered by snow and the low solar elevation angle causes severe shading effects on building rooftops (especially for buildings having gable roof styles) which makes the detection problematic. Not surprisingly, the lowest recall ratio also belongs to that image (#11). Besides, in some cases, our approach may over-detect some building boundaries, and typical examples for this problem are visible in test images #9, #4, and #12. This is due to two specific reasons. First, some parts of the building boundaries may

Table 1. Building detection results of the proposed approach.

<table>
<thead>
<tr>
<th>Test Image</th>
<th>Performance (%)</th>
<th>Object-level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>#1</td>
<td>74.0</td>
<td>81.8</td>
</tr>
<tr>
<td>#2</td>
<td>77.4</td>
<td>82.2</td>
</tr>
<tr>
<td>#3</td>
<td>76.9</td>
<td>89.3</td>
</tr>
<tr>
<td>#4</td>
<td>67.1</td>
<td>92.9</td>
</tr>
<tr>
<td>#5</td>
<td>92.8</td>
<td>96.5</td>
</tr>
<tr>
<td>#6</td>
<td>84.8</td>
<td>87.7</td>
</tr>
<tr>
<td>#7</td>
<td>87.1</td>
<td>90.4</td>
</tr>
<tr>
<td>#8</td>
<td>89.9</td>
<td>91.5</td>
</tr>
<tr>
<td>#9</td>
<td>65.2</td>
<td>97.3</td>
</tr>
<tr>
<td>#10</td>
<td>83.0</td>
<td>92.1</td>
</tr>
<tr>
<td>#11</td>
<td>25.9</td>
<td>69.3</td>
</tr>
<tr>
<td>#12</td>
<td>74.9</td>
<td>89.2</td>
</tr>
<tr>
<td>Total</td>
<td>81.3</td>
<td>89.7</td>
</tr>
</tbody>
</table>

Table 2. Road detection results of the proposed approach.

<table>
<thead>
<tr>
<th>Test Image</th>
<th>Performance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
</tr>
<tr>
<td>#1</td>
<td>41.6</td>
</tr>
<tr>
<td>#2</td>
<td>62.8</td>
</tr>
<tr>
<td>#3</td>
<td>84.3</td>
</tr>
<tr>
<td>#4</td>
<td>50.0</td>
</tr>
<tr>
<td>#5</td>
<td>72.6</td>
</tr>
<tr>
<td>#6</td>
<td>58.7</td>
</tr>
<tr>
<td>#7</td>
<td>56.9</td>
</tr>
<tr>
<td>#8</td>
<td>39.8</td>
</tr>
<tr>
<td>#9</td>
<td>35.8</td>
</tr>
<tr>
<td>#10</td>
<td>53.3</td>
</tr>
<tr>
<td>#11</td>
<td>14.2</td>
</tr>
<tr>
<td>#12</td>
<td>38.9</td>
</tr>
<tr>
<td>Total</td>
<td>55.1</td>
</tr>
</tbody>
</table>
have very smooth transition between their surroundings. The test image #9 that is acquired over a desert area is a very good example of this fact. Second, incorrect shadow detection may occur for buildings having gable roof, e.g. a large over-detection observed in the upper-center of the image #4. In that part of the image, unfortunately, a part of the gable roof is wrongly detected as a shadow region. As a result, the automatically selected $T_r$ region in the first stage contains a number of pixels corresponding to the background near the building, and therefore, an over-detection emerges in the final result. Our approach recovers buildings from single evidence, e.g., shadows. Therefore, if a large non-shadow region that has a spectral reflectance that is very close to the shadow regions exists, the approach may produce false positive regions, e.g., the two regions shown in the upper-left corner of test image #3. In addition, the cast shadows of two specific man-made objects, a building and a bridge, cannot be separated. Therefore, large bridges primarily used for vehicular traffic might also be incorrectly labeled as buildings. Such problematic cases are visible in test image #6. In addition to these problems, the results prove that the approach presented is generic for different roof colors, textures and types and has the ability to extract arbitrarily shaped buildings in complex environments. Various building shapes (rectangular (e.g. #1), #9), L-shaped (e.g. #5), hexagonal (e.g. #6), elliptic (e.g. #8), arc-style (e.g. #4, #10), contiguous (e.g. #3, #12), and complex (#7) are visible in the test images, and most of them are successfully handled. Not surprisingly, the approach works best for the cases where the cast shadows of buildings are clearly visible and not occluded by any other object (#5, #8).

The results of road regions shown in Figure 6 reveal that the approach proposed has the ability to extract road segments provided that the segments are not occluded by any other object, e.g., buildings and trees. In addition, a major difficulty arises due to the cast shadows of these objects (#4 and #7). In dense urban areas, the performance of road detection particularly depends on the solar angles of the image acquisition. Thus, in specific illumination conditions (e.g., acute sun elevation angles), it might not be possible to extract any of the road segments of a dense urban area. Nevertheless, for the presented test images (#4 and #7), our approach recovered many visible road segments. It is obvious that proposed approach will work best for the cases where we have specific distinct road type, e.g. asphalt (#3 and #5). However, if we cannot collect enough information related to specific road types (e.g. the white colored road region visible in #12), it might not be possible to correctly label these regions. More importantly, if the color of the road is somewhat similar to the color of the open lands (e.g. #11), unintended false positive regions appears. One other important point to be emphasized is road detection is the smoothness assumption enforced during global partitioning. Although the level of smoothness can be easily controlled during the global processing performed in Stage III, it is not always possible to handle all of the thin linear structures that might belong to road segments because of the smoothing involved, and this fact might negatively affect the recall ratios computed for the road segments. Nevertheless, despite the difficulties mentioned, we must highlight the fact that the proposed approach utilizes only a very basic assumption for road detection (that is the widths of roads do not vary abruptly) and that no prior information is employed for the extraction of buildings and roads. Thus, with this fact in mind, we believe that the proposed framework has a unique behavior to extract buildings and roads and provides fairly satisfactory results in complex environments.

4. Conclusions

In this paper, a novel graph-based approach is presented to automatically detect regions belonging to buildings and roads from a single VHR multispectral satellite image. Assessments performed on twelve test images selected from GeoEye-1 images reveal that the approach has the ability to detect buildings and roads in a single graph theoretic framework. The results also prove that the proposed approach can recover substantial amount of information related to regions of buildings and roads without any prior knowledge.

In the future, we will focus more on improving Stage III, where the automated information is extracted for the class road. In an urban area, one of our major tasks is to separate large bridges from buildings; therefore, we plan to expand the method such that a logical separation between the buildings and bridges could be achieved. In several recent works, higher-order cliques in a random field could be an interesting way to represent regions belonging to roads. Thus, this information might also be an interesting topic for further research. Finally, the simplification of the outlines of the building regions and the road network is also a required task that we will pursue in the future.

Acknowledgements

The author would like to thank HA VELSAN A.Ş. for providing the test images used in this study. The author is also grateful to anonymous reviewers for their helpful comments.

References


Mayer H., (1999), Automatic object extraction from aerial imagery-A survey focusing on buildings. Computer Vision and Image Understanding, 74(2), 138-149.


Toumaire O., Paparoditis N., (2009), A geometric stochastic approach based on marked point processes for road mark detection from high resolution aerial images. ISPRS Journal of Photogrammetry and Remote Sensing, 64(6), 621-631.


Wegner J.D., Montoya J., Schindler K., (2013), A higher-order CRF model for road network extraction. In: IEEE Conference on Computer Vision and Pattern Recognition, Portland, USA.