AUTOMATED EXTRACTION OF BUILDINGS AND ROADS IN A GRAPH PARTITIONING FRAMEWORK

Ali Ozgun Ok

Department of Civil Engineering, Faculty of Engineering, Mersin University, 33343, Mersin, Turkey -
oozgun@mersin.edu.tr, ozguneo@gmail.com

Commission III, WG III/4

ABSTRACT:

This paper presents an original unsupervised framework to identify regions belonging to buildings and roads from monocular very high resolution (VHR) satellite images. The proposed framework consists of three main stages. In the first stage, we extract information only related to building regions using shadow evidence and probabilistic fuzzy landscapes. Firstly, the shadow areas cast by building objects are detected and the directional spatial relationship between buildings and their shadows is modelled with the knowledge of illumination direction. Thereafter, each shadow region is handled separately and initial building regions are identified by iterative graph-cuts designed in a 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 labelled building regions as well as the shadow and vegetation areas are involved in a four-label graph optimization performed in the entire image domain to achieve the unsupervised classification result. The final stage aims to extend this classification to five classes in which the class road is involved. 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 seven test images selected from GeoEye-1 VHR datasets show that the presented approach has ability to extract the regions belonging to buildings and roads in a single graph theory framework.

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 of the fact that 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 have been involved for 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), Baltsavias (2004), Brenner (2005), and Haala and Kada (2010). Besides, exceptional review papers of Baumgartner (1997) and Mena (2003) also summarizes and describes the literature conducted for road detection. The focus of this paper is on the design and development of a graph theory framework which enables us to automatically extract regions belonging to buildings and roads from monocular VHR satellite images. Therefore, in this part, we very briefly summarize the previous studies aimed to automatically detect buildings and/or roads 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. Besides, a large number of methods proposed substantially benefit from the cast shadows of buildings (e.g. Lin and Nevatia, 1998; 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; Ünsalan and Boyer, 2005; Inglada, 2007; Senaras et al., 2013). Besides, approaches like active contours (e.g. Karantzalos and Paragios, 2009), Markov Random Fields

(MRFs) (e.g. Katartzis and Sahli, 2008), graph-based (e.g. Akçay and Aksoy, 2010; Izadi and Saeedi, 2012) and kernel-based (Sirmacek and Ünsalan, 2011) approaches were also investigated. In rather 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 the 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 Zu, 2002; Guo et al., 2007). Classification strategies utilizing multi-band information are also 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 the extents of the extraction of road centreline using six different approaches. In several recent works, Tournaire and Paparoditis (2009) developed an approach based on Marked Point Processes to extracted 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 constraint satisfaction neural network to identify road networks. In a different work, Ünsalan 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, 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 work, the researchers are motivated for a single set of object, 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 of each

other in several aspects: there is a road connection to most of the buildings, a road segment might be occluded by buildings etc. (Hinz et al., 2001). Therefore, if the extraction of these two objects can be handled in a single framework, this might have a positive affect on the final quality of the detection of each individual class in an urban area. To the best of our knowledge, so far, there are only a very limited number of studies that realizes such an integrated behaviour. 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 global and local context of roads including buildings. However, their approach utilizes a DSM to detect shadow and building areas to facilitate the road extraction. Ünsalan and Boyer (2005) detected houses and streets in a graph-related strategy. Their method extracted the street network using the snakes having unary and binary constraints, and the regions remaining after the detection of the street network was considered as 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 an original framework to extract regions belonging to buildings and roads from VHR satellite images in an automated manner. The basis of the framework relies on three main stages. Stage I aims to extract information only related to building regions. We benefit from the shadow information, and directional spatial relationship between buildings and their shadows is modelled 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 labelled 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 class others 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 shortly revised in order to provide a complete overview of the methodology. Those two previous papers are fully motivated for the extraction of building objects solely, whereas the originality of this paper arises from the fact that it enables us to concentrate the classes building and road 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-sharpened multi-spectral (B, G, R, and NIR) ortho-image. We assume that metadata providing information about the solar angles (azimuth and elevation) of the image acquisition is attached to the 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 (Fig. 1b). 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 colour 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 detected shadow regions and apply a 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_{PS} (Fig. 1c).

2.3 The Generation and Pruning of Fuzzy Landscapes

Given a shadow object B (e.g. each 8-connected component in M_{PS}) 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:

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

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 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 $v_{L,\alpha,\sigma,\kappa}$ (Fig. 2a) generated by combining two different structuring elements with a pixel-wise multiplication: $v_{L,\alpha,\sigma,\kappa} = v_{L,\kappa,\alpha} * v_{\sigma,\kappa}$, where $v_{\sigma,\kappa}$ is an isotropic non-flat structuring element with kernel size κ, whereas the flat structuring element $v_{L,\kappa,\alpha}$ is responsible for
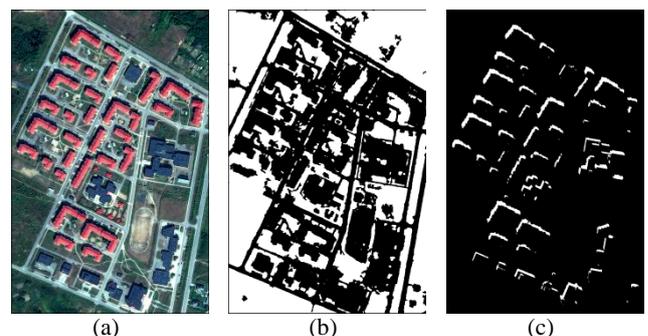

Figure 1. (a) GeoEye-1 pan-sharped image (RGB), (b) vegetation mask (M_V), (c) post-processed shadow mask (M_{PS}).

providing directional information where L denotes the line segment and α is the angle where the line is directed.

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.

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 utilize the GrabCut approach (Rother et al., 2004) in which an iterative binary-label graph-cut optimization (Boykov and Kolmogorov, 2004) is performed.

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 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 (T_F) and background/non-building (T_B) classes are labelled automatically using the shadow regions and the generated fuzzy landscapes. We define the T_F region in the vicinity of each shadow object whose extents are outlined after applying a double thresholding (η_1, η_2) to the membership values of the fuzzy landscape generated. 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 whose extent 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 labelled as T_B.

2.5 Stage II: Global Processing to Extract Building Regions

After Stage I, the building regions are detected with relatively reduced completeness but with almost no over-detection (Fig. 2c). 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_{PS}) 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* (Fig. 3a).

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 $\mathbf{z} = (z_1, z_2, \ldots, z_N)$, and a set of class labels $L \in \{1, \ldots l\}$ where $l = 4$, our aim is to find the optimal mapping from data $\mathbf{z}$ to class labels L. Each pixel has an initially assigned value (α_n) corresponding to each class labels L where $\alpha_n \in \{1, 2, 3, 4\}$. We follow the Gibbs energy function provided in (Rother et al., 2004), 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 $\mathbf{z}$ to class labels L.

In order 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 α-expansion move algorithm. Here, we briefly describe the α-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 $\mathbf{z}$ to class labels L, and let α be a specific class label. A mapping g is defined as an α-expansion move from mapping f if $g_n \neq \alpha$ implies $g_n = f_n$, where n denotes a pixel in data $\mathbf{z}$. With this definition, the set of pixels assigned to the label α has increased from mapping 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 $\mathbf{z}$ 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 there is no α-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 might still be incorrectly labelled 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 after the four-label optimization, and the regions rejected are further tested for new shadow
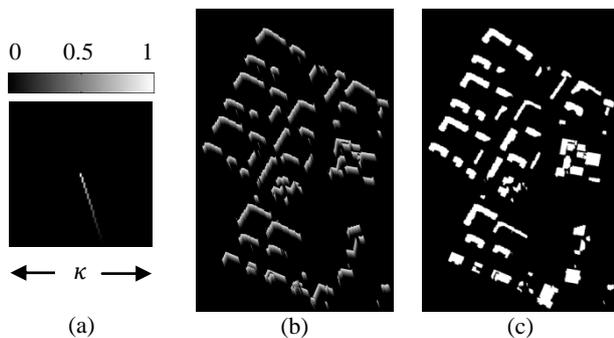

Figure 2. (a) Structuring element $v_{L,\alpha,\sigma,\kappa}$. (b) The fuzzy landscapes generated using the shadow mask provided in Fig. 1c, and (c) building regions detected after Stage I.
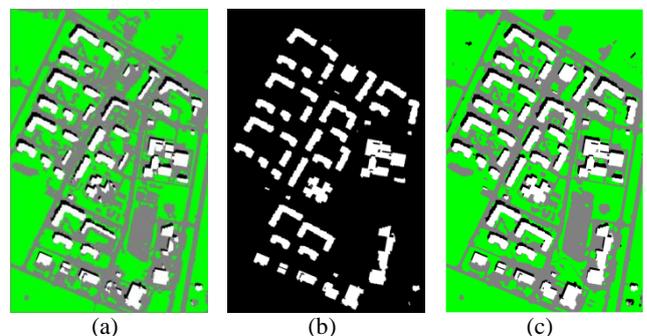

Figure 3. (a) The input of Stage II, (b) verified building regions, and (c) the output of Stage II. White, green, black, and grey colours indicate the regions for the classes building, vegetation, shadow and others, respectively.

evidence. Thus, our approach has ability to recover building regions (Fig. 3b) whose shadow regions are missed in the initial shadow mask generated (Fig. 1c). Finally, building regions which could not be validated with shadow evidence are joined to the class others (Fig. 3c).

2.6 Stage III: Simultaneous Extraction 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 class others after Stage II (Fig. 4a), this class principally covers a number of subclasses like roads, open lands, vehicles, parking lots etc. 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 class others (due to occlusions on shadow regions etc.) or vice versa (due to bridges etc.). 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 can be seen in Fig. 4a, the regions belonging to class others can be quite complex, therefore a simple post-processing generally does not work well. Therefore, this final stage aims to extend 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 utilize region part segmentation (RPS) to separate the road regions from attached irrelevant non-road regions and thereafter, medial axis transform (MAT) based approach is employed to filter and verify road hypotheses. In this part, 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 around each component to initialize the road information required for the final graph optimization (Fig. 4b). We follow the multi-label optimization framework introduced in Stage II, however, this time with five classes: *building*, *road*, *vegetation*, *shadow*, and *others*. At the last step, we extract the regions labelled as building and road from the optimization result (Fig. 4c), and perform the shadow verification (cf. Stage II, only for the class building) to achieve the final results.
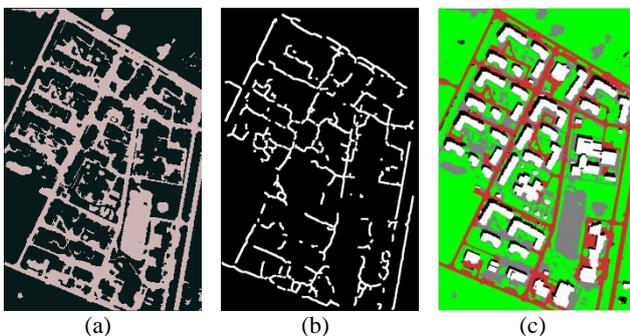

Figure 4. (a) The regions belonging to class others, (b) input for the class road, and (c) the output of Stage III. White, red, green, black, and grey colours indicate the regions for the classes building, road, vegetation, shadow and others, respectively.

3. RESULTS AND DISCUSSION

The test data include images acquired from GeoEye-1 sensor with 50 cm ground sampling distance, 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 seven test images which differ from their urban area and building characteristics as well as from their illumination and acquisition conditions. Reference data corresponding to buildings and roads are drawn by an experienced operator. Pixel-based precision, recall and F_1-score (Aksoy et al., 2012) performance measures are determined for each of the two classes to assess the quality of the results of the extraction. The proposed framework requires no training data collection, thus, the results can be computed once the parameter values are determined. All parameters required for Stages I and II are already investigated comprehensively in Ok (2013). Therefore, in this study, we considered and investigated the parameters required for stage III, and utilized a fixed parameter set to run the proposed framework for all test images.

We visualize the results in Fig 5, and according to the results presented, the proposed framework seems to be effective for both the extraction of buildings and roads. The building regions are very well identified despite the complex characteristics of buildings in the test images, e.g. roof colour 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 Table 1 favour these facts. For buildings, we achieved overall mean ratios of precision and recall as 83.0% and 89.4%, respectively. The computed pixel-based F_1-score for seven test images is around 86%. On the other hand, overall mean ratios of precision and recall are computed as 62.9% and 71.5%, respectively, for road segments. This corresponds to an overall object-based F_1-score of approximately 67%. 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 extraction of buildings and roads.

According to the results presented in Fig. 5, the proposed approach gives the strong impression that the method is highly robust for buildings, and the regions detected are quite convincing and representative. As can be 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 colour and texture, shape, size and orientation. It is also evident that the approach distinctively separates building regions from other areas except for a few cases. The lowest precision ratio (67.1%) is obtained for test image #4. This is due to large over-detection observed in the upper-centre of the image. In that part of the image, unfortunately a part of the gable roof is wrongly detected as a shadow region. As a result, automatically selected T_F region in the first stage contains a number of pixels corresponding to the background near the building, and therefore, an over-detection is emerged in the final result. Our approach recovers buildings from single evidence, shadows. Therefore, if a large non-shadow region whose spectral reflectance 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. Besides, 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 labelled as buildings. Such problematic cases are visible in test image #6. Besides these problems, the results prove that the

Figure 5. (first column) Test dataset (#1-7, RGB), (second column) the results of building regions, and (third column) the results of road regions. Green, red and blue colours represent true-positive, false-positive and false-negative, respectively.

approach presented is generic for different roof colours, textures and types, and has the ability to extract arbitrarily shaped buildings in complex environments. 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).

The results of road regions shown in Fig. 5 reveal that the approach proposed has ability to extract road segments provided that the segments are not occluded by any other object, e.g. buildings and trees. Besides, a major difficulty arises due to cast shadows of these objects (#4 and #7). In dense urban areas, the

Table 1. Test results of the proposed approach.

ID	Performance (%)					
	Buildings			Roads		
	Precision	*Recall*	*F₁-score*	*Precision*	*Recall*	*F₁-score*
#1	74.0	81.8	77.7	41.6	69.1	51.9
#2	77.4	82.2	79.7	62.8	52.8	57.4
#3	76.9	89.3	82.6	84.3	91.3	87.7
#4	67.1	92.9	77.9	50.0	50.8	50.4
#5	92.8	96.5	94.6	72.6	87.1	79.2
#6	84.8	87.7	86.2	58.7	77.7	66.9
#7	87.1	90.4	88.7	56.9	52.1	54.4
Total	83.0	89.4	86.1	62.9	71.5	66.9

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 most of the visible road segments. One other important point to be emphasized for road detection is the smoothness assumption enforced during the global partitioning. Although the level of smoothness can be easily controlled during global processing performed in Stage III, it is not always possible to handle all 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 strongly 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 no prior information is employed for the extraction of buildings and roads. Thus, having this fact in mind, we believe that the proposed framework has unique behaviour 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 extract regions belonging to buildings and roads from a single VHR multispectral satellite image. Assessments performed on seven test images selected from GeoEye-1 images reveal that the approach has ability to extract buildings and roads in a single graph theoretic framework. In the near future, we will focus more to improve 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 a rather recent work, Wegner (2013) showed that 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 and we will pursue in the future.

ACKNOWLEDGEMENTS

REFERENCES

Akçay, H.G., Aksoy, S., 2010. Building detection using directional spatial constraints. In: 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.

Aytekin, Ö., Erener, A., Ulusoy, İ., Düzgün, Ş., 2012. Unsupervised building detection in complex urban environments from multispectral satellite imagery, *International Journal of Remote Sensing*, 33(7), pp. 2152-2177.

Baltsavias, E.P., 2004. Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3-4), pp. 129-151.

Baumgartner, A., Steger, C., Mayer, H., Eckstein, W., 1997. *Multiresolution, Semantic Objects, and Context for Road Extraction*. Semantic Modeling for the Acquisition of Topographic Information From Images and Maps. Cambridge, Birkhauser-Verlag, pp. 140-156.

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.

Boykov, Y., Kolmogorov, V., 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9), pp. 1124-1137.

Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11), pp. 1222-1239.

Brenner, C., 2005. Building reconstruction from images and laser scanning. *International Journal of Applied Earth Observation and Geoinformation*, 6(3-4), pp. 187-198.

Das, S., Mirnalinee, T.T., Varghese, K., 2011. Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10), pp. 3906-3931.

Guo, D., Weeks, A., Klee, H., 2007. Robust approach for suburban road segmentation in high-resolution aerial images. *International Journal of Remote Sensing*, 28(2), pp. 307-318.

Haala, N., Kada, M., 2010. An update on automatic 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6), pp. 570-580.

Hinz, S., Baumgartner, A., 2000. Road Extraction in Urban Areas supported by Context Objects. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 33(B3), pp. 405-412.

Hinz, S., Baumgartner, A., Mayer, H., Wiedemann, C., Ebner, H., 2001. *Road Extraction Focussing on Urban Areas*. Automatic Extraction of Man-Made Objects from Aerial and Space Images (III). Balkema Publishers, Rotterdam, pp. 255-265.

Huang, X., Zhang, L., 2009. Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines. *International Journal of Remote Sensing*, 30(8), pp. 1977-1987.

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.

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.

Katartzis, A., Sahli, H., Pizurica, V., Cornelis, J. , 2001. A model-based approach to the automatic extraction of linear features from airborne images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(9), pp. 2073-2079.

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

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

Mayer, H., Hinz, S., Bacher, U., Baltsavias, E., 2006. A test of automatic road extraction approaches. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(3), pp. 209-214.

Mena, J.B., 2003. State of the art on automatic road extraction for GIS update: a novel classification. *Pattern Recognition Letters*, 24(16), pp. 3037-3058.

Mena, J.B., Malpica, J.A., 2005. An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery. *Pattern Recognition Letters*, 26(9), pp. 1201-1220.

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.

Ok, A.O., 2013. Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts. *ISPRS Journal of Photogrammetry and Remote Sensing*, (in-press).

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

Poullis, C., You, S., 2010. Delineation and geometric modeling of road networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(2), pp. 165-181.

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), pp. 1295-1304.

Shi, W., Zhu, C., 2002. The line segment match method for extracting road network from high-resolution satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 40(2), pp. 511-514.

Sirmacek, B., Ünsalan, 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.

Teke, M., Başeski, E., Ok, A.Ö., Yüksel, B., Şenaras, Ç., 2011. *Multi-Spectral False Color Shadow Detection*, Photogrammetric Image Analysis. Springer, Berlin, Heidelberg, pp. 109-119.

Tournaire, 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), pp. 621-631.

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.

Ünsalan, C., Sirmacek, B., 2012. Road network detection using probabilistic and graph theoretical methods. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11), pp. 4441-4453.

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.

Wiedemann, C., Hinz, S., 1999. Automatic extraction and evaluation of road networks from satellite imagery. In: *International Archives of Photogrammetry and Remote Sensing*, 32(3-2W5), pp. 95-100.