International Journal of Applied Earth Observation and Geoinformation 44 (2016) 217–225

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Region-based urban road extraction from VHR satellite images using Binary Partition Tree

Mengmeng Li[a,*], Alfred Stein[a], Wietske Bijker[a], Qingming Zhan[b]

[a] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500AE Enschede, The Netherlands
[b] School of Urban Design and Research Centre for Digital City, Wuhan University, 430072 Wuhan, China

ARTICLE INFO

Article history:
Received 22 May 2015
Received in revised form 25 August 2015
Accepted 22 September 2015

Keywords:
Binary Partition Tree
Fuzzy aggregation
Morphological profiles
Road extraction
Very high resolution images

ABSTRACT

This paper provides a hierarchical method for urban road extraction. It consists of (1) obtaining the road region of interest from a VHR image, (2) hierarchically representing this road region of interest in a Binary Partition Tree (BPT), and extracting the roads based on this BPT at hierarchical levels. Besides using two existing geometrical features (i.e. compactness and elongation), we define two other structural features based on orientation histograms and morphological profiles to guide the region merging of BPT. The morphological profiles are constructed using a series of path openings, which facilitate modeling linear or curved structures. The proposed method was applied to two types of VHR images with different urban settings, corresponding to a Pléiades-B image of Wuhan, China, and a Quickbird image of Enschede, the Netherlands. Experimental results show that the proposed method was able to group adjacent small segments that have high spectral heterogeneity and low road-like geometrical properties to form more meaningful roads sections, and performed superior to the existing methods. Furthermore, we compared the proposed method with two other existing methods in the literature. We conclude that the proposed method can provide an effective means for extracting roads over densely populated urban areas from VHR satellite images.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

A subject of considerable interest for information extraction from remote sensing images is the extraction of roads. It is essential for many applications, such as transportation management, geodata updating, and land cover and land use analysis. Road extraction from remote sensing images can be traced back to Bajcsy et al. (1976), and has been applied to a large variety of sensors covering a wide range of spatial resolutions (Mena, 2003). Very High Resolution (VHR) satellite images have a spatial resolution below 1 m, from which roads exhibit as image regions with more detailed spatial information compared with low or mid-low resolution satellite images. In particular, the growing availability of VHR images, e.g. QuickBird, GeoEye, WorldView and Pléiades images, has caused an increase in the number of studies on road extraction from VHR images. Nonetheless, these studies fail to obtain an ideal result, particularly in complex urban areas where road obstacles, e.g. from vehicles, trees and shadows are commonly found, thus affecting the accuracy of road extraction.

Road extraction methods are found regularly in remote sensing literature. According to Mena (2003), road extraction methods are categorized as road tracking (Vosselman and Knech, 1997; Hu et al., 2007; Movaghati et al., 2010; Tang et al., 2014), active contour models (snakes) (Ravanbakhsh et al., 2008; Butenuth and Heipke, 2012), mathematical morphology (Zhang et al., 1999; Soille and Pesaresi, 2002; Valero et al., 2010), segmentation and classification (Song and Civco, 2004; Bouziani et al., 2010; Shi et al., 2014), and knowledge-based methods (Baltsavias, 2004). Road extraction from images, similar to, e.g. blood vessel detection from medical images, is also a well-known issue in computer vision (Chai et al., 2013). Chai et al. (2013) distinguished three types of approach: pixel-based (Stoica et al., 2004; Mnih and Hinton, 2010; Montoya-Zegarra et al., 2014), line-based (Lacoste et al., 2005), and graph-based (Gerke et al., 2004; Turetken et al., 2012, 2013; Ünsalan and Sirmacek, 2012; Wegner et al., 2013). A road extraction system often involves several different types of techniques. There is no ideal road extraction for all types of land structures, in rural, suburban and urban areas. Besides using the spectral information of roads, many studies have stressed the

* Corresponding author.
E-mail addresses: m.li@utwente.nl (M. Li), a.stein@utwente.nl (A. Stein), w.bijker@utwente.nl (W. Bijker), qmzhan@whu.edu.cn (Q. Zhan).

http://dx.doi.org/10.1016/j.jag.2015.09.005
0303-2434/© 2015 Elsevier B.V. All rights reserved.

importance of using the geometrical, contextual, and structural knowledge for road extraction (Baltsavias, 2004).

In a VHR image, roads appear as image regions, and hence a region-based approach could be of interest. For example, Grote et al. (2012) presented a region-based method for suburban road extraction using high resolution color infrared images as well as a digital surface model (DSM). The authors employed a set of radiometric and geometric features to group small road regions to form meaningful road-like regions. This method, however, involves a rather large number of parameters and was not tested on just a single VHR satellite image, but also involved availability of a DSM.

This paper aims to provide a novel region-based method for urban road extraction from VHR images. The proposed method uses a Binary Partition Tree (BPT) (Salembier and Garrido, 2000; Valero et al., 2013) to hierarchically represent road regions, and relies on the road region of interest (RoadROI) preliminarily obtained from an image. The RoadROI highlights the road information, whereas it reduces the effects of vegetation, shadow and buildings on roads. In particular, by eliminating buildings via shadows, the risk of confusing between buildings and roads because of similar spectral characteristics is reduced. The construction of a BPT uses two widely used geometrical features, i.e. compactness and elongation indices, and two newly defined structural features based on orientation histogram and morphological profiles. An automatic strategy to obtain the RoadROI from an image is provided.

The remainder of this paper is organized as follows. Section 2 and Section 3 describes the study area and used data, and illustrates the proposed method for road extraction, respectively. Section 4 presents experimental results and related analysis, followed by discussion in Section 5 and the conclusions of this research in Section 6.

2. Study area and data

The proposed method was investigated using two different VHR images over urban areas, namely study areas in Wuhan and Enschede (see Fig. 1). For study area A, the road extraction is part of a large study on land use changes of Wuhan, China, being a dynamic area. A piece of a Pléiades image was acquired on 11 July 2013. This image was acquired with an azimuth angle of 140.8°, four multispectral bands of 2 m spatial resolution, a panchromatic band of 0.5 m, and has a size of 1599 × 2019 pixels. This 0.5 m panchromatic band was resampled from the 0.7 m raw panchromatic band by the image provider using spline resampling. The multispectral bands were fused with the panchromatic band using the Gram-Schmidt pan-sharpening method (as same as study area B). Study area B is a more stable and established urban area. We used a Quickbird image acquired on 21 September 2006. This image has an azimuth angle of 172°, four multispectral bands of 2.4 m spatial resolution, a panchromatic band of 0.6 m, and a size of 928 × 1089 pixels. These two study areas show that roads have a similar spectral reflectance as buildings, parking lots, squares and other built-up areas, making road extraction intrinsically difficult. Furthermore, the vehicles on roads and the trees surrounding roads give rise to obstacles. Both of these two study areas therefore make the road extraction from VHR images challenging. The reference road map of study area A was produced based on the official land use map by manually digitizing a part of small roads. The reference map of study area B was produced from the public Open Street Map.

3. Methods

3.1. Obtain the road region of interest

The workflow of the proposed road extraction is presented in Fig. 2. It is a region-based method, which starts from an initial segmentation, i.e. an over-segmentation case where a single object is partitioned into several subdivisions. Vegetation is removed first, followed by building shadows and buildings. In this way, the road region of interest (RoadROI) is obtained. This RoadROI is used to construct a BPT for hierarchically representing road-like regions. In doing so, we make use of geometrical and structural information that is used for extracting roads based on fuzzy logic.

Let I be an image, and $P = \{R_i\}$ be an initial segmentation with $i = 1, 2, \cdots, M$ regions R_i. Initial segmentation is obtained by any existing segmentation methods. Here we use the Fractal Net Evolution Approach (FNEA) (Benz et al., 2004) implemented in eCognition Software by giving a small scale parameter. For each region R_i, we first smooth the within-pixel values by using the mean value of R_i with respect to a spectral band. In other words, each pixel in region R_i is replaced by the mean of the pixel values in that region. The smoothing process aims to reduce image noise with respect to vegetation, shadow, and building regions obtained by pixel-based methods. The smoothed image is therefore used for obtaining the RoadROI. We remove the vegetation regions based on the Normalized Difference Vegetation Index (NDVI) using Otsu's threshold method (Otsu, 1979). Next, shadow regions are automatically detected by using the method proposed by Teke et al. (2011), where a Hue-Saturation-Intensity (HSI) color space is converted from a Near-Infrared-Red-Green false color space to formulate an index $(Saturation - Intensity)/(Saturation + Intensity)$, by which the Otsu's threshold is applied for shadow detection. Dark objects,

(a) Study area A (b) Study area B

Fig. 1. (a) The Pléiades image (with true color composite) of Wuhan, China (study area A). (b) The Quickbird image (with true color composite) of Enschede, the Netherlands (study area B).

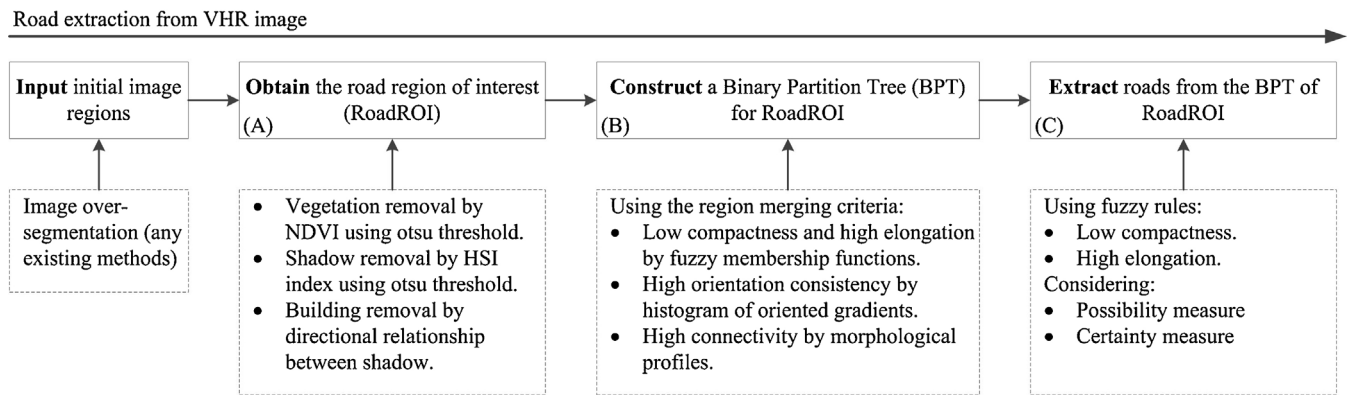

Fig. 2. Workflow of the proposed road extraction method.

however, might be mis-detected as shadows if we directly use this method for the entire image. To improve the detection performance, we subset the images into small patches for improving the bimodality of the image histogram for Otsu thresholding (Adeline et al., 2013).

Moreover, as buildings have similar spectral and textural properties as roads, we make use of the directional relationship between buildings and their shadows to exclude the buildings from the image. The directional relationship is modeled by a fuzzy landscape of building shadows. The detailed description and computation of a fuzzy landscape can be found in Li et al. (2015). A reference (shadow) region R_{ref} and a direction with angle β is considered. Then for each image pixel $x \in I$, the fuzzy landscape $F_\beta(R_{ref})(x)$ specifies a membership function to the reference region R_{ref}. The $F_\beta(R_{ref})(x)$ ranges from 0 to 1, and specifies the degree of satisfying the directional relationship of $x \in I$ to R_{ref}. This fuzzy landscape is further smoothed at the region-level by averaging the within pixel values of an image region R_i. Based upon the smoothed fuzzy landscape F, the building regions are automatically detected if their directional relationship is larger than 0.5. By using this method, other shadows, i.e. mainly tree shadows, are removed from the set of that can then be used for building detection. Removal of tree shadows is further handled by using directional relationships between trees (i.e. vegetation) and their shadows. To do so, we first obtain the vegetation boundaries adjacent to shadows, then construct another fuzzy landscape that has the opposite direction angle $180^\circ - \beta$ where the pixel value specifies the degree that a shadow pixel is casted by a tree. Similarly, tree shadow regions can be detected if their directional relationship is larger than 0.5. The rest of image regions that are not vegetation, buildings and their shadows will be used for RoadROI BPT construction.

3.2. Construction of a BPT for RoadROI

A BPT provides a tree-structure for a region-based representation of an image, and has been used for image processing, segmentation and objects extraction (Salembier and Garrido, 2000; Sigurdsson et al., 2014; Valero et al., 2015). Let T be a BPT, and v be a node of T, corresponding to an initial image region. As shown in Fig. 3, the top node R11 is the root of T with two children nodes R10 and R6. The node R10 is referred as a branch node and the R6 is a leaf node as it has no children nodes.

3.2.1. Region model

The construction of a BPT involves two key components: the region model and region merging. The region model specifies how regions are represented and how to compute the union of two adjacent regions (Kurtz et al., 2012). Two shape features, namely compactness and elongation indices, and two newly defined structural features, namely the orientation histogram and connectivity, are used to define the region model with respect to an image region in RoadROI. The compactness cm and elongation eln indices (Bouziani et al., 2010) are defined for each region R_i as

$$cm_i = 2 \cdot \sqrt{\pi \cdot a_i}/p_i \tag{1}$$

$$eln_i = a_i/(len_i^2) \tag{2}$$

where a_i, p_i and len_i are the area, perimeter and length (i.e. the major axis of an polygon) of R_i, respectively.

Inspired by the Histogram of Oriented Gradients (HoG) (Dalal and Triggs, 2005), for a region R_i, we calculate an orientation histogram with k bins, labeled HoG_i, to represent the orientation distribution of R_i. Differing from the original HoG in Dalal and Triggs (2005), for a region R_i, the local neighborhood of computing its HoG_i is determined by the size of the bounding box of R_i and a given cell size w_{cell}. Specifically, the bounding box of R_i is buffered if either the width or the length of this bounding box is less than w_{cell}. By doing so, the minimum local neighborhood of a HoG_i is guaranteed. The rest of the computation of a HoG_i is the same as the orginial HoG computation.

Morphological profiles (MP) have been widely used to model spatial information of the images (Ghamisi et al., 2015). An MP is created by applying a series of traditional openings and closings with a structuring element (SE) of increasing size to an image. A major shortage of using traditional openings or closings for building an MP, however, is the difficulty in determining a proper SE for

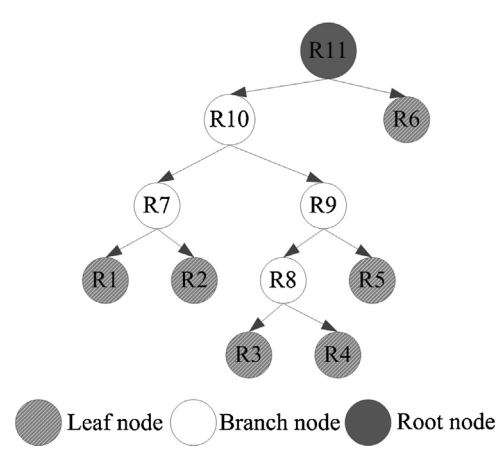

Fig. 3. Tree-structure representation of a BPT sample with six initial regions.

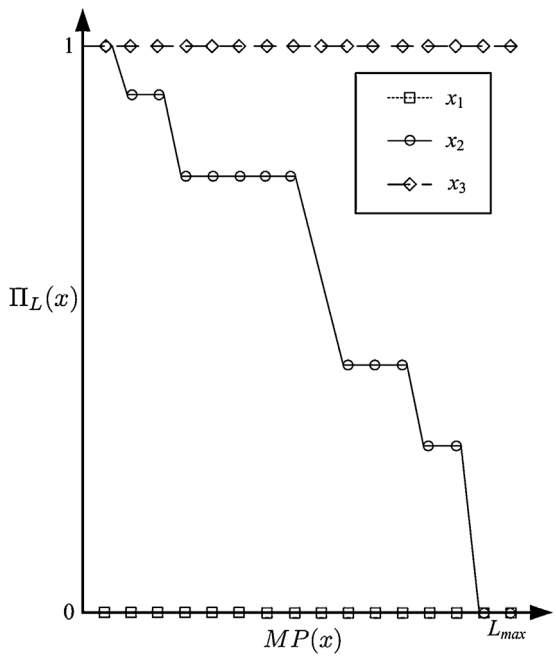

Fig. 4. Three MPs of pixel x_1, x_2, x_3 created by using path openings with the length L ranging from 0 to L_{max} pixels. Specifically, x_1, x_2, x_3 refer to not a road pixel, probably a road pixel, and a road pixel in a RoadROI.

conducting the opening or closing operation. Furthermore, traditional openings and closings fail to extract curved structures, for instance, a curved road in an image. To overcome this shortage, Heijmans et al. (2005) proposed two variants, called path openings and closings, to facilitate extracting pixel structures that are locally oriented but not perfectly straight, from binary and gray-level images. It removes the pixels that are not connected in an adjacency graph under a given length and orientation. The theoretical foundations of path openings and closings are detailed given in Heijmans et al. (2005). The L-tuple $(a_1, a_2, \cdots, a_L)$ is called a *δ-path of length L* if $a_k \mapsto a_{k+1}$, or equivalently, if $a_{k+1} \in \delta(\{a_k\})$, for $k = 1, 2, \cdots, L-1$. The $\mapsto$ refers to a directed adjacency relation, meaning that there is an edge going from a_k to $a_k + 1$. Let the set of all *δ-paths of length L* be Π_L, and let the set of *δ-paths of length L* contained in a subset X of an image I be $\Pi_L(X)$, $X \subset I$. The operator $\alpha_L(X)$ called the path opening is defined as the union of all *δ-paths of length L* contained in X. We apply a series of path openings with a length of increasing size to form an MP on the pixels of the RoadROI. The morphological profile $MP(x)$ of a pixel x is defined as

$$MP(x) = \{\Pi_L(x)\}, L \in \{1, \cdots, L_{\max}\}, \tag{3}$$

where L_{max} indicates the longest length of the used path openings. We exemplify the characteristics of the MPs with respect to different types of pixels, namely non-road, possible road and road pixels (Fig. 4).

Based upon the MP of RoadROI, we define a connectivity feature to measure to which degree an image region is linear or curved. The connectivity feature con_i of R_i is defined as

$$con_i = \frac{1}{n_i}\sum_{p=1}^{n_i}\sum_{L=0}^{L_{\max}} w_L \cdot \Pi_L(x_p), x_p \in R_i, \tag{4}$$

where w_L refers to the weight of a path opening with length L being considered in con_i. For a simple linear case, the w_L equals to $\frac{1}{L_{\max}}$. In addition, weight values calculated from a Gaussian distribution can also be used for w_L.

3.2.2. Region merging

The region merging criterion determine the order of pair-wise regions being merged in a BPT creation. The two geometrical features of the union of adjacent regions are quantified via a Z-shape (or spline-based) fuzzy membership function,

$$f(x; a, b) = \begin{cases} 1, & x \le a \\ 1 - 2 \cdot \left(\frac{x-a}{b-a}\right)^2, & a \le x \le \frac{a+b}{2} \\ 2 \cdot \left(\frac{x-a}{b-a}\right)^2, & \frac{a+b}{2} \le x \le b \\ 0, & x \ge b \end{cases} \tag{5}$$

$$\mu_{1,ij} = f(cm_{ij}, a_1, b_1), \tag{6}$$

$$\mu_{2,ij} = f(eln_{ij}, a_2, b_2), \tag{7}$$

where $\mu_{1,ij}$ and $\mu_{2,ij}$ refer to the membership values of cm_{ij} or eln_{ij} respectively. Theoretically, the values of cm_{ij} and eln_{ij} have the range of (0, 1). Here the variable a_1 and a_2 are equal to 0 for both cases.

Next, we use histogram intersection to quantify the orientation similarity between two adjacent regions. Let HoG_i, HoG_j and HoG_{ij} be the orientation histograms of the objects R_i, R_j, and their union R_{ij}, respectively. The orientation similarity $\mu_{3,ij}$ of two adjacent regions is defined as,

$$\mu_{3,ij} = \max(f_{bin_i}(R_i) - f_{bin_i}(R_j), f_{bin_j}(R_j) - f_{bin_j}(R_i)) \tag{8}$$

where $f_{bin_i}(R_i)$ (resp. $f_{bin_j}(R_j)$) is the maximum normalized frequency (equal to 1) of the HoG_i (resp. HoG_j). Here, we normalize a histogram by dividing it using the largest value of this histogram. The subscript bin_i (resp. bin_j) indicates a bin location that has the maximum normalized frequency of the HoG_i (resp. HoG_j). The $f_{bin_i}(R_j)$ refers to the normalized frequency of HoG_j at the bin location bin_i. Likewise, the $f_{bin_j}(R_i)$ is the normalized frequency of the HoG_i at the bin location bin_j.

The MP is used to measure the connectivity of two adjacent regions. Let con_i and con_j be the connectivity features of R_i and R_j. The connectivity of two adjacent regions $\mu_{4,ij}$ is defined as

$$\mu_{4,ij} = (con_i \cdot a_i + con_j \cdot a_j)/(a_i + a_j) \tag{9}$$

where a_i and a_j are the areas of R_i and R_j, respectively.

In summary, we define region merging criteria as

$$O_{ij} = \alpha \cdot \sqrt{\mu_{1,ij} \cdot \mu_{2,ij}} + (1-\alpha) \cdot \sqrt{\mu_{3,ij} \cdot \mu_{4,ij}} + \epsilon, \tag{10}$$

$$\epsilon = \left| wd_{ij} - \max(wd_i, wd_j) \right| / wd_{ij}, \tag{11}$$

where $\sqrt{\mu_{1,ij} \cdot \mu_{2,ij}}$ and $\sqrt{\mu_{3,ij} \cdot \mu_{4,ij}}$ are two terms representing the geometrical and structural features, respectively, and the parameter α weights these two terms. In this paper, we take $\alpha = 0.5$. The parameter ϵ aims at constraining the width of the two adjacent merged regions, where wd_i, wd_j and wd_{ij} are the width of the regions R_i, R_j and R_{ij} respectively. A high value of O_{ij} gives high priority to the two adjacent objects R_i and R_j to be merged. Based upon the defined region model and region merging criteria, the BPT of a RoadROI is created, referring to the Algorithm A.1 in Appendix A.

3.3. Road extraction

Based upon the constructed RoadROI BPT, road extraction is performed. The initial image regions are merged to form more road-like regions represented by BPT nodes. Given an image region R_v represented by a BPT node v, the membership degree of R_v being a road region is quantified by two fuzzy membership values $\mu_{1,v}$ and $\mu_{2,v}$. We then apply the method of Sebari and He (2013), to

detect road regions from a BPT based on these two fuzzy rules. Furthermore, the adopted method also quantifies the uncertainty of extracted road regions. Specifically, a possibility measure $\prod(R_v)$ represents the possibility that an image region R_v belongs to road,

$$\prod(R_v) = \max(\mu_{1,v}, \mu_{2,v}) \tag{12}$$

$$\prod(\bar{R}_v) = \max(1 - \mu_{1,v}, 1 - \mu_{2,v}) \tag{13}$$

where $\prod(\bar{R}_v)$ denotes the possibility that the region R_v belongs to non-road. In addition, a necessity measure $N(R_v)$ is defined to represent the certainty that an image region R_v is road,

$$N(R_v) = 1 - \prod(\bar{R}_v) \tag{14}$$

$$N(\bar{R}_v) = 1 - \prod(R_v) \tag{15}$$

where $N(\bar{R}_v)$ denotes the possibility that the region R_v belongs to non-road. The image region R_v is therefore detected as a road class if $\prod(R_v) > \prod(\bar{R}_v)$ and $N(R_v) > N(\bar{R}_v)$.

3.4. Quality assessment

We compare the proposed method with two other road extraction methods found in the literature. We first use the Orfeo ToolBox (Inglada and Christophe, 2009), labeled as OTB method, to extract roads from both Pléiades and Quickbird study images. The Orfeo ToolBox is an open source C++ library for remote sensing images processing, which is developed by the French space agency to promote the use of Pléiades images. Hence, this toolbox can be well used to compare the results between the proposed and existing methods, particularly using Pléiades images. In addition, we compare our extracted roads with the method from Valero et al. (2010), labeled the S. V. method, that investigated the use of MP for road extraction. Because the proposed method also uses MP, we consider it a proper option for the comparison purpose. For implementation of the OTB and S.V. methods, we set the corresponding parameters as follows; For OTB extraction, an input pixel of roads is used for each study area, see Fig. 6; other parameters are kept the same as suggested in the user guide of Orfeo Toolbox. For S.V. extraction, the *MGL* parameter is obtained by using a path opening operation of length 400 pixels; the L_{min} equals to 400 pixels; and the *T* parameter is equal to 50 which is the same as used in the literature (Valero et al., 2010).

We evaluate the extracted roads using three accuracy measures, namely completeness, correctness, and quality, used by Clode et al. (2007),

$$completeness = \frac{TP}{TP + FN}, \tag{16}$$

$$correctness = \frac{TP}{TP + FP}, \tag{17}$$

$$quality = \frac{TP}{TP + FP + FN}, \tag{18}$$

where *TP* indicates the true positive and is the number of pixels existing in both reference and extracted roads; *FN* indicates the false negative and is the number of pixels existing in the reference roads, but not in the extracted roads; *FP* indicates the false positives and is the number of pixels existing in the extracted roads, but not in the reference roads.

4. Experiments

4.1. Obtain the road region of interest

Fig. 5 shows the results of obtaining the RoadROIs at both study areas A and B. Fig. 5a1 and b1 gives the detected vegetation and shadows by Otsu's threshold. To erase the tree shadows, we constructed a fuzzy landscape to model the directional relationship between shadows and their adjacent trees (i.e. vegetation in this work). Table 1 gives the size of the used fuzzy structuring element, as well as other parameter settings associated with the step of obtaining the RoadROI. The rest of the detected shadows, corresponding to building shadows, were used to detect buildings. Fig. 5a2 and b2 displays the detected building shadows. The

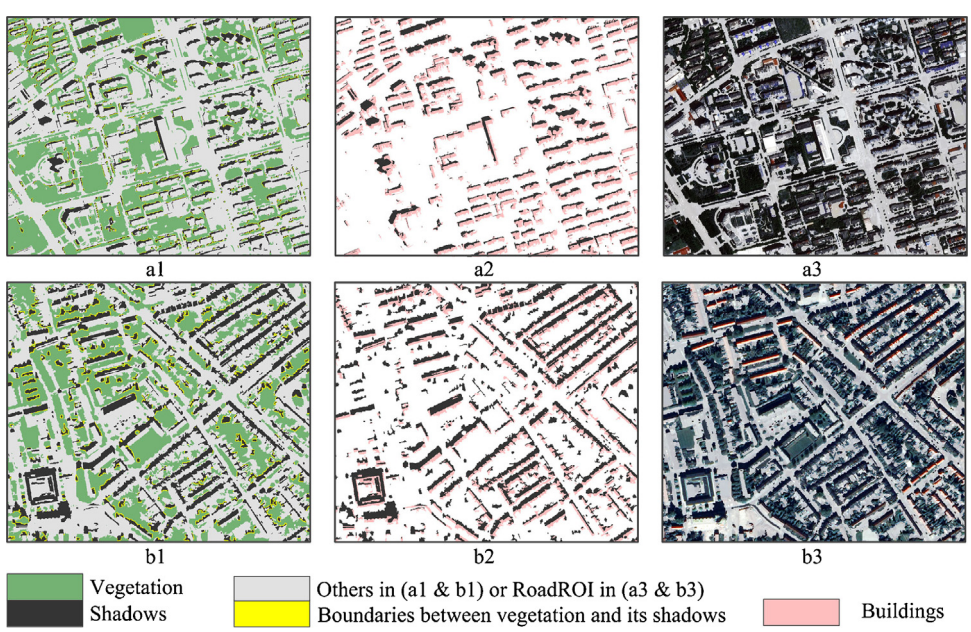

Fig. 5. The obtained RoadROIs of both study areas A and B, corresponding to the top and bottom panels, respectively. (a1 and b1) are the detected vegetation and shadows by Otsu's threshold. (a2 and b2) are the detected buildings by directional relationship between buildings and their shadows. (a3 and b3) are the obtained RoadROIs overlapping with the corresponding true color VHR images. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

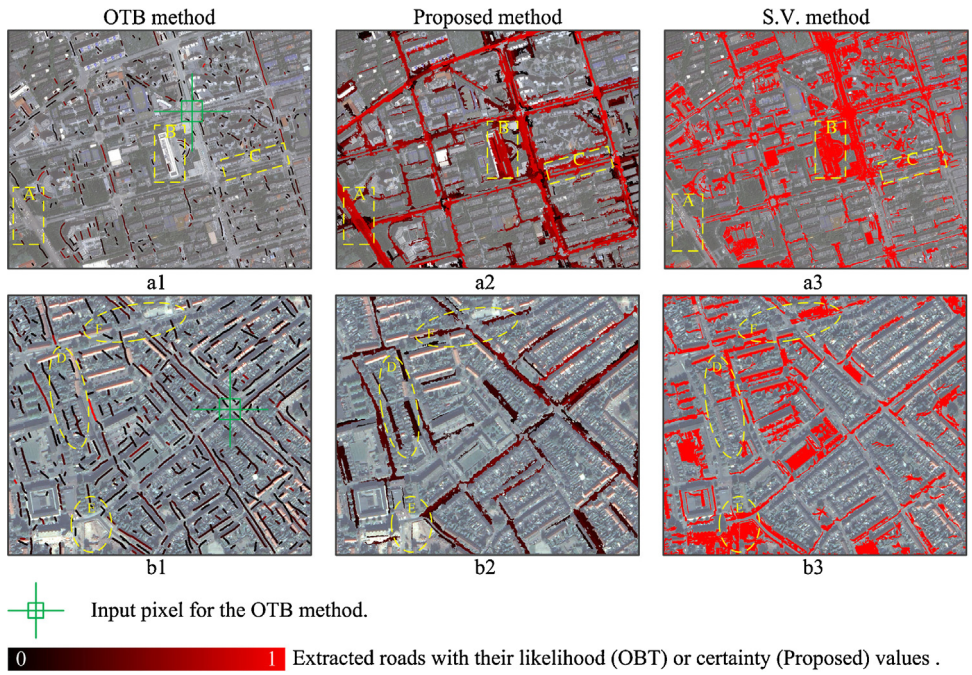

Fig. 6. Road extraction by the proposed and other two existing methods. (a1 and b1), (a2 and b2), and (a3 and b3) are the results produced from the OTB, proposed, and S.V. methods. The proposed and OTB methods also provide the extracted results with associated uncertainty measures, corresponding to the certainty and likelihood values respectively. (For interpretation of the references to color in the text, the reader is referred to the web version of the article.)

RoadROIs of study areas A and B were obtained by masking the detected vegetation, building shadows, and buildings (see Fig. 5a3 and b3).

4.2. Construction of a BPT for RoadROI

The construction of a RoadROI BPT involves in calculating two shape features, namely compactness and elongation, as well as two structural features, namely orientation histogram and connectivity. Table 1 lists the parameter settings of this BPT construction. To define the parameters for the Z-shape membership functions in Eqs. (6) and (7), we used the compactness and elongation information about roads in the literature, i.e. the compactness of most roads was lower than 0.3 and the elongation was lower than 0.2. We set the parameters, for instance, $b_1 = 0.6$ to make sure that the possibility of a region with compactness less than 0.3 is larger than 0.5. To calculate an orientation histogram of an object, we set the minimum neighborhood w_{cell} equal to 50 pixels, and chose the number of bins k equal to 18. To calculate the connectivity feature, we set the maximum length L_{max} of path opening equal to 400 pixels.

Table 1
Parameter settings associated with the steps of road extraction. Internal parameters refer to those described in this paper. External parameters refer to those associated with existing algorithms but were not in detail described in this paper.

Parameters setting	Description
Obtain RoadROI	
Scale = 40	(**External**) Scale parameter in eCognition software for creating an image segmentation. Other parameters were set as default.
$\varphi = 140.8^\circ, 172^\circ$	(**External**) Azimuth angle of a VHR image, used for calculating the direction parameter β (Li et al., 2015).
Construct RoadROI BPT	
$b_1 = 0.6$; $b_2 = 0.4$	(**Internal**) Parameters for defining Z-shape membership functions.
$w_{cell} = 50$; $k = 18$	(**Internal**) Parameters for calculating a HoG_i.
$L_{max} = 400$	(**Internal**) Parameter for calculating a MP.
$\alpha = 0.5$	(**Internal**) Parameter for defining the region merging criteria.

4.3. Road extraction

Fig. 6 gives the extracted results at study areas A and B by the proposed and the OTB and S.V. methods. The proposed method gave the best extraction as compared to OTB and S. V. for both study areas A and B. Fig. 6a1 and b1 shows that the roads extracted by OTB exhibit relatively low likelihood at both study areas A and B. For the roads with heavy obstacles like vehicles and vegetation, the OTB gave the worst extraction performance followed by S.V. for these two study areas, see yellow boxes A and C in Fig. 6a1–a3 and yellow ellipses F in Fig. 6. The proposed method was able to merge those small regions caused by the appearance of vehicles to form more meaningful road-like regions. Fig. 6a3 and a4 shows that a large part of buildings and other built-up areas, particularly those close to roads, were extracted as roads, see yellow box B in Fig. 6a3 and yellow ellipse E in Fig. 6b3, because those adjacent buildings and other built-up areas had similar spectral reflectance as roads. By using the proposed method, some adjacent buildings were removed in the first step. Because built-up regions generally had a low elongation and a high compactness, these regions were less likely to be merged as more meaningful road-like segments by the proposed method. As shown in the yellow box B in Fig. 6a2 and yellow ellipse D in Fig. 6b2, however, several building were mis-extracted as roads. In fact, these buildings were fully surrounded by trees, and no building shadows were detected. Hence, these buildings failed to be detected by shadow information at the stage of obtaining the RoadROI. In addition, these buildings were similar to roads, which have a high elongation and low compactness. In addition to road extractions at study areas A and B, the reader is referred to the supplementary data for road extraction on a large VHR image (see Fig. B.1 in Appendix B).

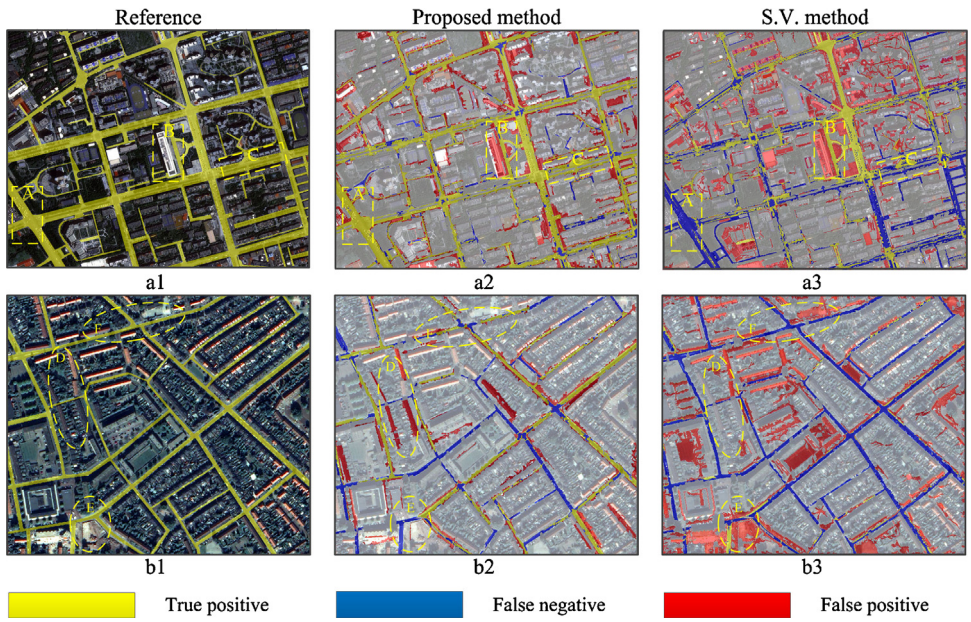

Fig. 7. Evaluation of the extracted roads, and compared with the S.V. method. (a1 and b1) are the reference maps at study areas A and B. (a2 and b2) and (a3 and b3) refer to the valuation of extracted results by the proposed and S.V. methods, in terms of true positive, false negative and false positive measures. (For interpretation of the references to color in the text, the reader is referred to the web version of the article.)

4.4. Quality assessment

Fig. 7 shows the evaluation of the extracted roads of the proposed method as compared with the S.V. method. We however did not evaluate the extracted roads by OTB method, because it extracts the roads as lines instead of regions. Further, as shown in Fig. 6, OTB produced the worst results from a visual inspection.

Fig. 7 shows that road extraction at study area A gave a better performance than at study area B. This figure also shows that the proposed method performed better than the S.V. method at both study areas A and B. For some regions with evident obstacles like vehicles and vegetation, see the yellow boxes A and C in Fig. 7a1–a3, the S.V. method yielded a high proportion of false negatives. The reason was that, for instance, the vehicles on roads increased the spectral heterogeneity which affected the representation of linear structures by path openings. In contrast, the proposed method was able to merge small regions to form larger, more meaningful regions so that the linear structures of roads were retained. For some regions adjacent to buildings and other built-up regions, see the yellow ellipse E in Fig. 7b1–b3, the S.V. method yielded a higher number of false positives as compared with the proposed method. The reason was that these areas had a similar spectral reflectance as roads. It is thus difficult to separate roads from their surroundings just relying on spectral characteristics. The reason was that during the first step of the proposed method, buildings were removed to reduce the effects. In addition, those built-up areas were initially segmented to regions with low elongation and high compactness properties which were constrained in the region merging process of a BPT creation. Nonetheless, some buildings were mis-extracted as roads, causing false positives in the evaluation, see yellow box B in Fig. 7a2 and yellow ellipse D in Fig. 7b2. The reason was that these buildings were not removed from the corresponding RoadROI, and had similar geometrical properties as roads, i.e. high elongation and low compactness. Furthermore, the final road extraction only considered these two geometrical properties.

Table 2 gives the quantitative evaluation of the roads extracted by the proposed and S. V. methods, in terms of completeness, correctness, and quality measures. This table denotes that the roads extracted by the proposed method had a higher accuracy than the S. V. method for both study areas A and B. The extracted roads by either the proposed or the S. V. method at study areas B had a lower accuracy than those at study area A, in terms of these three accuracy measures. Regarding these three accuracy measures, the completeness of the extracted roads was higher than correctness and quality measures. For study area B, although the proposed method produced a better extraction than the S. V. method, the extracted roads were still evaluated with low accuracy, which might be caused by the following reasons: (1) this study area covered a dense urban area, where the buildings were arranged with small spaces between them, and (2) a large part of small roads were fully blocked by vegetation and shadows, and the image regions of these small roads were not connected and could not be grouped to more meaningful roads.

Table 2
Evaluation of the extracted roads by the proposed and S.V methods using completeness, correctness, and quality measures. The A and B refer to the study areas A and B respectively.

	Proposed method		S.V. method	
	A	B	A	B
Completeness	0.8592	0.5766	0.4799	0.1283
Correctness	0.5875	0.4690	0.3758	0.0695
Quality	0.5359	0.3489	0.2670	0.0472

5. Discussion

This study investigated the use of BPT for region-based urban road extraction from VHR images. Our results confirm that BPT provides an effective representation for urban roads in VHR images, to obtain a set of road-like regions, which have high elongation and low compactness, for road extraction. Moreover, the BPT constructed in this work is application driven. This differs from a more common data driven BPT construction. To the best of our

knowledge, it is hard to construct an effective BPT for various applications. An application driven BPT is more suited for a specific extraction task like road extraction, because it uses a specific region model and a merging criterion. In addition, an application driven BPT can make use of expert's knowledge for BPT construction. Therefore, this paper uses fuzzy logic to model expert's knowledge. Many studies on road extraction from remote sensing images can be found in the literature, particularly the recent work by Grote et al. (2012), who also investigated region-based road extraction by grouping image segments into more meaningful road sections. The difference exists in several aspects, e.g. (1) the proposed method starts from an over-segmentation, which can be done by any existing segmentation methods, (2) DSM data is not used, and (3) less parameters are used.

From the examples in this study it is clear that buildings have an effect on road extraction, in particular those which are adjacent to roads or have similar shapes as roads. This finding was also confirmed by previous work (Mena, 2003; Grote et al., 2012). In this study, the proposed road extraction relies on a RoadROI, corresponding to image regions excluding vegetation, buildings and building shadows. The RoadROI places the emphasis on road information, whereas it first reduces the effect of vegetation, shadow and buildings on roads; it is obtained in an automated way. The definition of the RoadROI, however, could be dependent on study images. If an image shows many building shadows on roads, as in this study, we suggest the corresponding RoadROI to include building shadows, whereas otherwise, we suggest it to exclude building shadows. Future work could focus on analyzing the effect of building shadows on extracted roads, and could provide quantitative information to end-users whether a RoadROI should include building shadows or not. Moreover, because tree shadows affect building detection by means of shadow information (Li et al., 2015), they are removed from the set of detected shadows that can be used for building detection.

A Z-shape fuzzy membership function is used to quantify to which degree a pair of adjacent regions should be merged according to a geometrical feature. Other fuzzy membership functions (e.g. linear or Gaussian functions) (Zadeh, 1994) can also be used for this quantification. Experimentally, there is no evident difference between different membership functions in this task. The question is to assign proper parameters for the chosen membership function. We used the geometrical information of roads in the literature (Bouziani et al., 2010), i.e. the compactness of most roads was lower than 0.3 and the elongation was lower than 0.2, corresponding to $b_1 = 0.6$ and $b_2 = 0.4$. Our results again confirm this knowledge. We also defined two new structural features, namely the orientation histograms and connectivity. The computation of the orientation histogram HOG_i of region R_i needs to specify the size w_{cell} of the minimum local neighborhood and the number k of bins, where the w_{cell} mainly guarantees the minimum local neighborhood for the orientation histograms of small regions. The effect of w_{cell} and k on the computation of orientation histograms can be found in the literature (Dalal and Triggs, 2005). The computation of the connectivity con_i of region R_i needs to specify the longest length L_{max} of the used path openings. A smaller value of L_{max} may lead to a non-road region with a relatively higher value of connectivity, while a larger value of L_{max} may give a small-road region with a relatively lower value of connectivity. The region merging criterion in Eq. (10) involve the combination of the geometrical and structural terms, where $\alpha = 0.5$. A larger α gives more importance on the geometrical term, and vice versa. In the literature, other alternative methods can be used for this combination, for example, the decision fusion method by Fauvel et al. (2006). Our method, however, can also achieve comparable results by a relatively straightforward combination.

Road regions are hierarchically represented by a BPT, from which road regions are assigned as roads or non-roads using two fuzzy rules, referring to elongation and compactness properties. Although our results gave the best extraction compared with two other existing methods, more additional rules may improve the extraction performance. Moreover, new rules can also be easily integrated in the decision step, by means of possibility and necessity measures. The results of study area B showed that, for those small roads fully blocked by vegetation and building shadows, the corresponding image regions failed to be merged. Because, no adjacency relationships between two blocked regions were defined in the Region Adjacency Graph (RAG) (see the Algorithm A.1 in Appendix A). Future improvements can be conducted to deal with this problem, for example, if the distance of two blocked regions is smaller than a specific value, then these two regions can be seen as adjacent. In this paper, we compared the proposed method with two other existing road extraction methods, namely S.V., which is also region-based and also uses MP, and line-based OTB, which was especially developed for Pléiades images. The proposed method gave higher completeness, correctness and quality than the S.V. method. Visual inspection showed that the proposed method performed better than the OTB method. However, results could not be compared based on completeness, correctness and quality measures, because OTB produces roads as lines and not as regions. As can be seen from Appendix B, the proposed method also performs consistently well on a larger subset of the image. Although we compared our method to the best and closest alternatives for road extraction, further comparison with other existing road-extraction methods could be interesting and induce further improvements.

6. Conclusion

BPTs have a considerable potential for region-based representation of images. For remote sensing images, so far, the BPTs have been mainly used for image segmentation, based mainly on low-level spectral, textural and geometrical information. We investigated the use of BPTs for region-based urban road extraction from VHR images, integrating geometrical and structural information of roads. Based upon fuzzy logic, different types of information can be effectively fused for constructing a BPT, in which small regions are grouped to form larger and more meaningful road-like regions. It has been known that vegetation, shadows and buildings affect the road extraction from VHR images, particularly in urban environments. We investigated to use a set of automatic thresholding methods to preliminarily remove vegetation, buildings and building shadows to obtain a RoadROI, upon which the proposed road extraction was conducted. The experimental results showed that the proposed road extraction outperformed two other existing methods over the Wuhan and Enschede study areas using Pléiades and Quickbird VHR images respectively. We conclude that the proposed method can provide an alternative means for road extraction over dense urban areas from VHR images, including varying densities of roads, buildings adjacent to roads and partial covering of roads by vehicles and vegetation.

Acknowledgments

The authors sincerely thank the China Scholarship Council (CSC) and ITC Research Fund. For the financial support, the fourth author would like to thank the Chinese National Natural Science Fund Project under grant 41331175. The authors would also like to thank the anonymous reviewers for their valuable comments and suggestions to improve the manuscript.

Appendix A.

Algorithm A.1. BPT creation procedure

Initialize a image partition (i.e. the initial image segmentation that was used for preliminary SVM classification) for an image RoadROI. Suppose it has M objects (i.e. M leaves for BPT).

Generate the Region Adjacency Graph (RAG) for defining the neighboring relations between pair-wise objects. Calculate the dissimilarity $O(R_i, R_j)$ of each pair of adjacent objects R_i and R_j, and choose the pair of adjacent objects with lowest $O(R_i, R_j)$.

Merge the pair of adjacent objects R_i and R_j to create a new object, labeled as R_{M+1} (BPT nodes excluding leaves).

Update the RAG by removing old edges between R_i and R_j, and creating new edges linking R_{M+1}.

Return to Step 2. The stop criterion is that the RoadROI is formed into a single object or the lowest value of $O(R_i, R_j)$ reaches 1.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jag.2015.09.005.

References

Adeline, K., Chen, M., Briottet, X., Pang, S., Paparoditis, N., 2013. Shadow detection in very high spatial resolution aerial images: a comparative study. ISPRS J. Photogramm. Remote Sens. 80, 21–38.

Bajcsy, R., Tavakoli, M., 1976. Computer recognition of roads from satellite pictures. IEEE Trans. Syst. Man Cybern. SMC-6 (9), 623–637.

Baltsavias, E., 2004. Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. ISPRS J. Photogramm. Remote Sens. 58 (3–4), 129–151.

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS J. Photogramm. Remote Sens. 58 (3–4), 239–258.

Bouziani, M., Goita, K., He, D.-C., 2010. Rule-based classification of a very high resolution image in an urban environment using multispectral segmentation guided by cartographic data. IEEE Trans. Geosci. Remote Sens. 48 (8), 3198–3211.

Butenuth, M., Heipke, C., 2012. Network snakes: graph-based object delineation with active contour models. Mach. Vision Appl. 23 (1), 91–109.

Chai, D., Forstner, W., Lafarge, F., 2013. Recovering line-networks in images by junction-point processes. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Oregon, Portland, June 25–27, pp. 1894–1901.

Clode, S., Rottensteiner, F., Kootsookos, P., Zelniker, E., 2007. Detection and vectorization of roads from lidar data. Photogramm. Eng. Remote Sens. 73 (5), 517–535.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. I, pp. 886–893.

Fauvel, M., Chanussot, J., Benediktsson, J., 2006. Decision fusion for the classification of urban remote sensing images. IEEE Trans. Geosci. Remote Sens. 44 (10), 2828–2838.

Gerke, M., Butenuth, M., Heipke, C., Willrich, F., 2004. Graph-supported verification of road databases. ISPRS J. Photogramm. Remote Sens. 58 (3–4), 152–165.

Ghamisi, P., Dalla Mura, M., Benediktsson, J., 2015. A survey on spectral-spatial classification techniques based on attribute profiles. IEEE Trans. Geosci. Remote Sens. 53 (5), 2335–2353.

Grote, A., Heipke, C., Rottensteiner, F., 2012. Road network extraction in suburban areas. Photogramm. Rec. 27 (137), 8–28.

Heijmans, H., Buckley, M., Talbot, H., 2005. Path openings and closings. J. Math. Imaging Vis. 22 (3), 107–119.

Hu, J., Razdan, A., Femiani, J., Cui, M., Wonka, P., 2007. Road network extraction and intersection detection from aerial images by tracking road footprints. IEEE Trans. Geosci. Remote Sens. 45 (12), 4144–4157.

Inglada, J., Christophe, E., 2009. The orfeo toolbox remote sensing image processing software. In: International Geoscience and Remote Sensing Symposium (IGARSS), vol. 4, Cape Town, South Africa, July 12–17, pp. IV733–IV736.

Kurtz, C., Passat, N., Gançarski, P., Puissant, A., 2012. Extraction of complex patterns from multiresolution remote sensing images: a hierarchical top-down methodology. Pattern Recogn. 45 (2), 685–706.

Lacoste, C., Descombes, X., Zerubia, J., 2005. Point processes for unsupervised line network extraction in remote sensing. IEEE Trans. Pattern Anal. Mach. Intell. 27 (10), 1568–1579.

Li, M., Bijker, W., Stein, A., 2015. Use of Binary Partition Tree and energy minimization for object-based classification of urban land cover. ISPRS J. Photogramm. Remote Sens. 102, 48–61.

Mena, J., 2003. State of the art on automatic road extraction for GIS update: a novel classification. Pattern Recogn. Lett. 24 (16), 3037–3058.

Mnih, V., Hinton, G., 2010. Learning to detect roads in high-resolution aerial images. Lecture Notes in Computer Science 6316 LNCS (PART 6)., pp. 210–223.

Montoya-Zegarra, J., Wegner, J., Ladicky, L., Schindler, K., 2014. Mind the gap: modeling local and global context in (road) networks. In: Jiang, X., Hornegger, J., Koch, R. (Eds.), Pattern Recognition. Vol. 8753 of Lecture Notes in Computer Science. Springer International Publishing, pp. 212–223.

Movaghati, S., Moghaddamjoo, A., Tavakoli, A., 2010. Road extraction from satellite images using particle filtering and extended kalman filtering. IEEE Trans. Geosci. Remote Sens. 48 (7), 2807–2817.

Otsu, N., 1979. Threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. SMC-9 (1), 62–66.

Ravanbakhsh, M., Heipke, C., Pakzad, K., 2008. Road junction extraction from high-resolution aerial imagery. Photogramm. Rec. 23 (124), 405–423.

Salembier, P., Garrido, L., 2000. Binary Partition Tree as an efficient representation for image processing, segmentation, and information retrieval. IEEE Trans. Image Process. 9 (4), 561–576.

Sebari, I., He, D.-C., 2013. Automatic fuzzy object-based analysis of VHSR images for urban objects extraction. ISPRS J. Photogramm. Remote Sens. 79, 171–184.

Shi, W., Miao, Z., Wang, Q., Zhang, H., 2014. Spectral-spatial classification and shape features for urban road centerline extraction. IEEE Geosci. Remote Sens. Lett. 11 (4), 788–792.

Sigurdsson, E., Valero, S., Benediktsson, J., Chanussot, J., Talbot, H., Stefánsson, E., 2014. Automatic retinal vessel extraction based on directional mathematical morphology and fuzzy classification. Pattern Recogn. Lett. 47, 164–171.

Soille, P., Pesaresi, M., 2002. Advances in mathematical morphology applied to geoscience and remote sensing. IEEE Trans. Geosci. Remote Sens. 40 (9), 2042–2055.

Song, M., Civco, D., 2004. Road extraction using SVM and image segmentation. Photogramm. Eng. Remote Sens. 70 (12), 1365–1371.

Stoica, R., Descombes, X., Zerubia, J., 2004. A Gibbs point process for road extraction from remotely sensed images. Int. J. Comput. Vis. 57 (2), 121–136.

Tang, X., Song, S., Zhan, Y., 2014. A novel road extraction algorithm for high resolution remote sensing images. Appl. Math. Inform. Sci. 8 (3), 1435–1443.

Teke, M., Baeski, E., Ok, A., Yksel, B., Senaras, C., 2011. Multi-spectral false color shadow detection. Lecture Notes in Computer Science 6952 LNCS., pp. 109–119.

Turetken, E., Benmansour, F., Andres, B., Pfister, H., Fua, P., 2013. Reconstructing loopy curvilinear structures using integer programming. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Oregon, Portland, June 25–27, pp. 1822–1829.

Turetken, E., Benmansour, F., Fua, P., 2012. Automated reconstruction of tree structures using path classifiers and mixed integer programming. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, RI, USA, June 16–21, pp. 566–573.

Ünsalan, C., Sirmacek, B., 2012. Road network detection using probabilistic and graph theoretical methods. IEEE Trans. Geosci. Remote Sens. 50 (11, Pt 1), 4441–4453.

Valero, S., Chanussot, J., Benediktsson, J., Talbot, H., Waske, B., 2010. Advanced directional mathematical morphology for the detection of the road network in very high resolution remote sensing images. Pattern Recogn. Lett. 31 (10), 1120–1127.

Valero, S., Salembier, P., Chanussot, J., 2013. Hyperspectral image representation and processing with Binary Partition Trees. IEEE Trans. Image Process. 22 (4), 1430–1443.

Valero, S., Salembier, P., Chanussot, J., 2015. Object recognition in hyperspectral images using Binary Partition Tree representation. Pattern Recogn. Lett. 56, 45–51.

Vosselman, G., Knech, J.D., 1997. Road tracing by profile matching and kaiman filtering. In: Gruen, A., Kuebler, O., Agouris, P. (Eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images. Birkhäuser, pp. 265–274.

Wegner, J., Montoya-Zegarra, J., Schindler, K., 2013. A higher-order CRF model for road network extraction. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Oregon, Portland, June 25–27, pp. 1698–1705.

Zadeh, L.A., 1994. Fuzzy logic, neural networks, and soft computing. Commun. ACM 37 (3), 77–84.

Zhang, C., Murai, S., Baltsavias, E., 1999. Road network detection by mathematical morphology. In: ISPRS Workshop 3D Geospatial Data Production: Meeting Application Requirements, Paris, France, pp. 185–200.