

## Hierarchical graph-based segmentation for extracting road networks from high-resolution satellite images

Rasha Alshehhi <sup>\*</sup>, Prashanth Reddy Marpu

*Institute Center for Smart and Sustainable Systems, Masdar Institute of Science and Technology, Abu Dhabi, United Arab Emirates*



### ARTICLE INFO

#### Article history:

Received 31 July 2016

Received in revised form 9 February 2017

Accepted 13 February 2017

Available online 8 March 2017

#### Keywords:

Road networks

Graph-based segmentation

Gabor

Morphological filtering

Hierarchical merging and splitting

### ABSTRACT

Extraction of road networks in urban areas from remotely sensed imagery plays an important role in many urban applications (e.g. road navigation, geometric correction of urban remote sensing images, updating geographic information systems, etc.). It is normally difficult to accurately differentiate road from its background due to the complex geometry of the buildings and the acquisition geometry of the sensor. In this paper, we present a new method for extracting roads from high-resolution imagery based on hierarchical graph-based image segmentation. The proposed method consists of: 1. Extracting features (e.g., using Gabor and morphological filtering) to enhance the contrast between road and non-road pixels, 2. Graph-based segmentation consisting of (i) Constructing a graph representation of the image based on initial segmentation and (ii) Hierarchical merging and splitting of image segments based on color and shape features, and 3. Post-processing to remove irregularities in the extracted road segments. Experiments are conducted on three challenging datasets of high-resolution images to demonstrate the proposed method and compare with other similar approaches. The results demonstrate the validity and superior performance of the proposed method for road extraction in urban areas.

© 2017 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

## 1. Introduction

Accurate and up-to-date road information is essential for many urban applications, such as automated road navigation (Li et al., 2014), geometric correction of urban remote sensing images (Auclair-Fortier et al., 2000) and updating geographic information systems (GIS) (Mena, 2003; Bonnefon et al., 2002). Rapidly changing urban environments need frequent updates of road database. Roads can be defined as long narrow regions with various orientations, lengths, and widths. Usually, the width of a road is several pixels in high-resolution satellite images. The length of a road is usually longer than buildings and longer than or equal to a street block (Liu et al., 2015; Sujatha and Selvathi, 2015). Extracting road network from up-to-date satellite images is a challenging problem due to noise and occlusions that create non-homogeneous regions leading to inaccurate classification of road segments. For instance, complex background and contextual structures (e.g., trees, shadow and vehicles on the roads) usually appear in high-resolution images. Also, some road-like segments, which have identical spec-

tral/spatial properties such as railways and parking lots can be misclassified as roads.

Many approaches have been developed to extract roads from remotely sensed images for urban areas. Recently, there is a great interest in including spatial information e.g., morphological filtering (Liu et al., 2015; Sujatha and Selvathi, 2015; Valero et al., 2010; Maurya et al., 2011; Chaudhuri et al., 2012), shape and texture features (e.g., elongation, Gabor filtering, etc.) (Liu et al., 2015; Sujatha and Selvathi, 2015; Maurya et al., 2011; Zhou et al., 2010; Jin et al., 2012; Shi et al., 2014) along with various machine learning techniques (Maurya et al., 2011; Mokhtarzade and Zoj, 2007; Wegner et al., 2013; Mnih and Hinton, 2010).

Road extraction methodologies can be mainly classified based on two taxonomies. First, they can be divided into either road-area extraction or road-centerline extraction. Road-area extraction mainly depends on image classification and segmentation (Mnih and Hinton, 2010; Unsalan and Sirmacek, 2012; Cheng et al., 2014; Peng et al., 2011; Li et al., 2014). Road-centerline extraction methods concentrate on detecting road-skeletons (Liu et al., 2015; Sujatha and Selvathi, 2015; Shi et al., 2014; Miao et al., 2013; Miao et al., 2014; Cao and Sun, 2014; Hu et al., 2014; Shi et al., 2014; Sironi et al., 2014; Cheng et al., 2015). Second, extraction methods can be semi-automatic or automatic. In semi-automatic

\* Corresponding author.

E-mail addresses: [ralshehhi@masdar.ac.ae](mailto:ralshehhi@masdar.ac.ae) (R. Alshehhi), [pmarpu@masdar.ac.ae](mailto:pmarpu@masdar.ac.ae) (P.R. Marpu).

approaches (Liu et al., 2015; Chaudhuri et al., 2012), some prior information such as user input (e.g., seed points) or prior geographical information are required. In automatic approaches, no such prior information is required (Maurya et al., 2011; Mokhtarzade and Zoj, 2007; Wegner et al., 2013; Mnih and Hinton, 2010; Huang and Zhang, 2009).

Chaudhuri et al. (2012) proposed a semi-automatic method to extract road-centerline by applying directional Mathematical Morphology (MM), which is one of the commonly used methods to extract oriented, thin and line-like features. In Liu et al. (2015), the authors integrated MM (e.g., path opening/closing) approach by adapting volunteered geographic information (VGI), captured in the OpenStreetMap (OSM) database along with shape features (e.g., elongation and compactness) as prior knowledge to extract main road-networks from satellite images. Unsalan and Sirmacek (2012) suggested a method, that utilizes edge pixels to detect road centers by kernel-based density estimation. This process is followed by extracting road-shape segments based on a binary balloon algorithm and tracking road-segments based on a graph theoretic approach. Miao et al. (2013) proposed an automatic method, in which potential road segments are first obtained based on shape and spectral features, followed by multivariate adaptive regression splines modeling to extract road centerline. Hu et al. (2014) proposed a method where adaptive mean shift algorithm is first used to define road-center points. It is followed by voting to enhance salient linear features and then Hough transform to extract arc regions of the road-centerline. Most of these mentioned approaches integrated spectral and spatial features. However, selection of the most appropriate spatial features that can be used to extract different topologies (e.g., shape, orientation and scale) of roads in different urban images is a challenging problem. Some methods also consider contextual feature descriptors based on neighborhood, but the window-size of feature descriptors is defined subjectively.

Mokhtarzade and Zoj (2007) extracted roads from satellite images using artificial Neural Network (NN) classification based on texture features derived using Gray Level Co-occurrence Matrix (GLCM)(e.g., contrast, energy, entropy, and homogeneity). Huang and Zhang (2009) developed a methodology to extract road centerlines based on geometrical features of road-spectral variations by using Support Vector Machine (SVM) classifier. These methods are pixel-based methods where the complete contextual structure is not considered. Recently, probabilistic graphical models such as Markov Random Fields (MRF) (Perciano et al., 2011) and Conditional Random Fields (CRFs) (Wegner et al., 2013) are widely used for extracting roads based on contextual information. Furthermore, deep Convolution Neural Networks (CNNs) (Mnih, 2013; Shu, 2014; Saito and Aoki, 2015; Saito et al., 2016) have been achieving impressive state-of-the-art performance for extracting roads. In Mnih (2013), a supervised deep Neural Network (NN) architecture was proposed to extract roads more accurately from noisy data. Saito and Aoki (2015) and Saito et al. (2016) proposed a model averaging methods with various CNN parameters to produce more accurate roads. Although, the accuracy of these approaches is high, they need a large training database to have good road/non-road classification. Moreover, these approaches produce results where they have many discontinuous road-regions affecting the completeness of the extracted road network.

Urban classes (e.g., roads, building, parking lots, trees, etc.) typically appear at different scales in an image. The extraction method needs to take into account the scale to detect road and eliminate non-road pixels. Therefore, it is important to work with a multi-scale approach.

In this paper, we propose a graph-based segmentation to extract roads in urban areas. An input image is first segmented into superpixels to represent homogeneous regions. In order to elimi-

nate the background, multi-channel Gabor and morphological filters are used to derive features to produce the initial road-network. However, road-superpixels of the initial road-network maybe be discontinuous and can be affected by over and under-segmentation. Therefore, hierarchical merging and then splitting are carried out to improve the results. Merging is based on grouping smaller segments to produce an approximate road-network. Splitting is the process to eliminate non-road superpixels from the extracted road-network. The criteria for merging and splitting are based on similarity of spectral and spatial features of superpixels. A graph-theoretic approach is then used to model contextual structure between adjacent segments to produce the complete road-network. The main contributions of the proposed method are summarized as follows:

- Many road extraction methods are based on the pixel-based classification which produces heterogeneous results. Pixel-based approaches cannot discriminate between contextual structures. Therefore, we suggest using superpixels to incorporate spectral and spatial information based on contextual structure.
- We introduce a new hierarchical graph-based approach to extract road-network more efficiently and accurately based on two main steps: merging and splitting. Hierarchical merging is used to group possible road superpixels to produce an intermediate result based on contextual, spectral and spatial features. Splitting is used to eliminate artifacts from the intermediate output.
- Many recent methods fail to differentiate road from other road-like regions (e.g., parking lot, railway, narrow shadow). We integrate texture features to discriminate road-regions from other regions which have identical shape and spectral properties.
- To overcome the weakness of the existing methods while dealing with intersections and other discontinuous regions due to complex backgrounds, the shortest path algorithm is proposed to complete these unconnected road-regions.
- A simple regularization method is suggested to smooth the segmentation results.

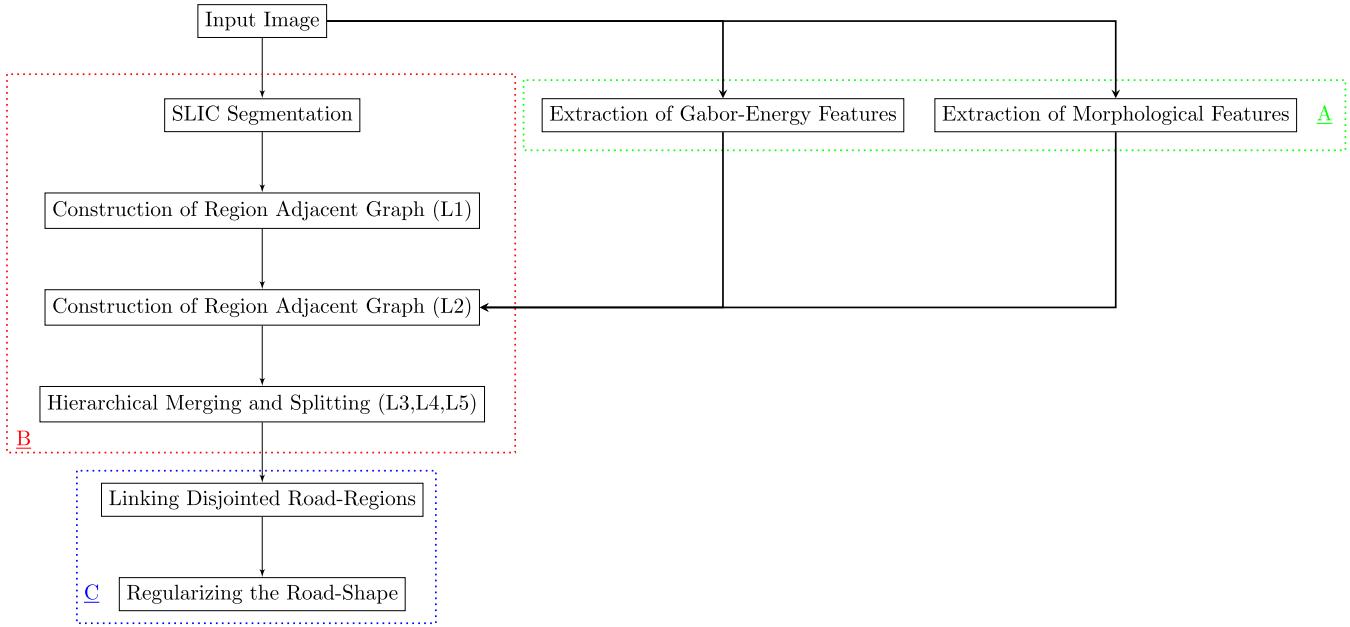
The remaining of this paper is organized as follows: we first describe the proposed method in Section 2, followed by a description of the experiments and the discussion on the final road-network including a comparison with other methods, Section 3. Finally, conclusions and future works are drawn in Section 4.

## 2. Methodology

The flowchart of the proposed method is depicted in Fig. 1. There are three main stages of the proposed method: (A) pre-processing, (B) graph-based segmentation and (C) post-processing. Pre-processing consists of two major steps to extract the most informative features based on: (i) Gabor filtering and (ii) morphological filtering. Graph-based segmentation is based on hierarchical merging or splitting. Post-processing consists of two steps: linking segments and removing small artifacts to obtain a complete road-network map.

### 2.1. Pre-processing

Pre-processing is used to filter the effect of background variations and produce more effective feature image which shows a high contrast between road and non-road pixels to facilitate the segmentation process. Here, we consider two well known filtering methods: (i) Gabor filtering to discriminate road texture from non-road texture and (ii) Morphological filtering to eliminate the background. Fig. 2 summarizes the pre-processing stage.



**Fig. 1.** Flowchart of the proposed method. It consists of (A) Pre-processing, (B) Graph-based segmentation, and (C) Post-processing.

### 2.1.1. Extraction of Gabor-energy feature

Gabor filtering is one of the well-known methods for texture analysis and edge detection (Jain and Farrokhnia, 1990; Idrissa and Achery, 2002; Jirik et al., 2011). It has the ability to perform multi-resolution decomposition with various spatial-frequencies and orientations (multi-channel Gabor filtering). A two-dimensional (2D) Gabor filter is obtained by modulating a sinusoidal plane wave of various frequencies and orientations with a 2D Gaussian (Idrissa and Achery, 2002; Daugman, 1985). Let  $G(x,y,f,\theta)$  be the function defining a Gabor filter centered at the origin  $(x,y)$  with  $f$  as spatial frequency and  $\theta$  as the orientation. The Gabor filter function is written as:

$$G(x,y,f,\theta) = \exp\left(-\frac{x'^2 + y'^2}{2\sigma^2}\right) \cos(2\pi f x'), \quad (1)$$

where

$$x' = x\cos(\theta) - y\sin(\theta), \quad (2)$$

$$y' = y\cos(\theta) - x\sin(\theta), \quad (3)$$

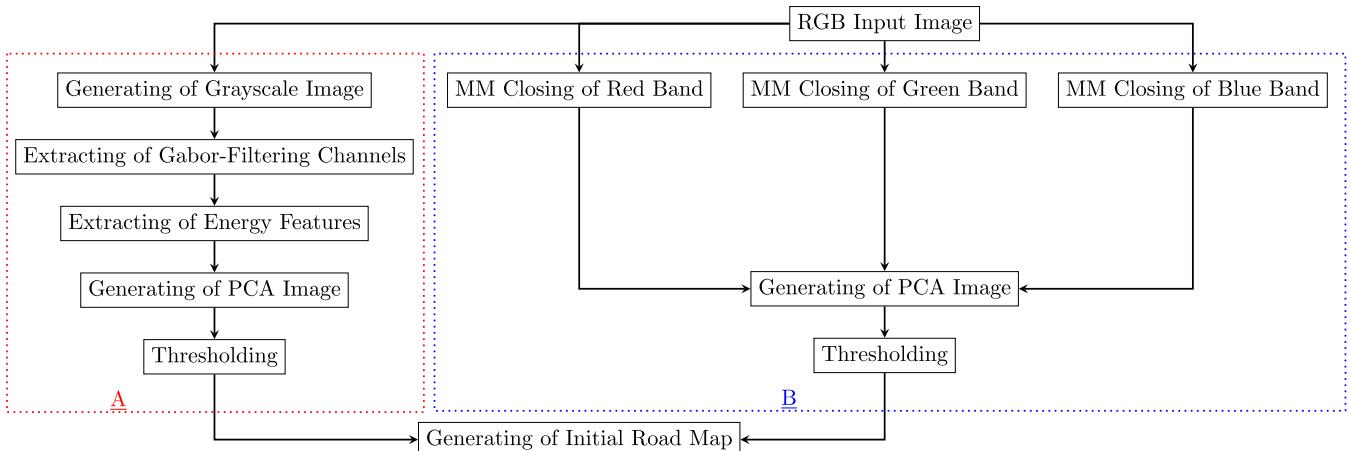
where  $\sigma$  is a standard deviation of the Gaussian kernel.

In order to localize road-texture boundaries, where road-texture repetition can be characterized by its spatial frequency and directionality by its orientation, we extract Gabor-energy features from the input image based on three steps, as follows:

- Generation of Gabor filter Bank: Gabor filtering is based on the selected orientations and frequencies.

The response of a Gabor filter to the grayscale image is obtained by 2D convolution operation. Let  $I(xa, ya)$  denote the grayscale image and  $I_G(xa, ya, f, \theta)$  denote the response of a Gabor filter with frequency  $f$  and orientation  $\theta$  to an image at point  $(xa, ya)$  in the image plane.  $I_G(\cdot)$  is obtained as:

$$I_G(xa, ya, f, \theta) = I(xa, ya) * G(x, y, f, \theta) \\ = \sum_x \sum_y I(xa - x, ya - y) G(x, y, f, \theta), \quad (4)$$



**Fig. 2.** Flowchart of the pre-processing stage. It includes: (A) Extraction of Gabor-Energy features and, (B) Extraction of morphological closing features.

- Extraction of Energy-Features: In order to identify road-texture features and extract the most useful information from the filtered image, non-linear sigmoidal transformation and then spatial smoothing are applied, as recommended in [Jain and Farrokhnia \(1990\)](#). The energy-feature is extracted using nonlinear sigmoid function ( $v$ ), as shown in Eq. (5). Spatial smoothing is applied above the transformed image to enhance results since spatial smoothing suppresses large variations in areas that belong to the same texture ([Jain and Farrokhnia, 1991](#)). It is applied by computing the average absolute deviation (AAD) from the mean in small overlapping windows, as shown in Eq. (6).

$$v(I_G) = \tanh(\alpha I_G) = \frac{1 - e^{-2\alpha I_G}}{1 + e^{-2\alpha I_G}}, \quad (5)$$

$$E_G(xb, yb) = \left( \frac{1}{M^2} \sum_{(xa, ya) \in W_{xbyb}} |v(I_G(xa, ya))| \right) - v(I_G), \quad (6)$$

where  $\alpha$  is a constant. The energy feature image  $E_G(xb, yb)$  corresponds to filtered image  $I_G$  and  $W_{xbyb}$  is a  $M \times M$  window centered at the pixels with coordinates  $(xb, yb)$ . The constant parameter  $\alpha$  of non-linear sigmoid function  $v(I_G)$  is assigned as 0.25, as recommended in [Jain and Farrokhnia \(1991\)](#).

- Principal Component Analysis (PCA) ([Jolliffe, 2002](#)) is applied on the stack of all Gabor-energy features to reduce the redundancy while preserving the contrast derived from the features, and the first principal component, ( $PC_{E_G}$ ) is used in the subsequent processing.

### 2.1.2. Extraction of morphological features

Mathematical morphology (MM) is a non-linear approach which uses the concepts of set, topology and geometry to analyze geometrical structures (e.g., shape and form). It examines the geometric structure of an image by probing with desired structuring elements (B) ([Soille, 2003; Soille, 2006](#)). In this work, morphological closing with an oriented linear B is applied to filter three image bands (Red, Green, and Blue). Morphological closing is used to fill structures that contain oriented linear Bs. Oriented linear is selected because roads are oriented linear structures. Morphological closing is defined as:

$$\phi(I) = \epsilon_B(\delta_B(I)), \quad (7)$$

where  $\delta_B(\cdot)$  is dilation process and  $\epsilon_B(\cdot)$  is erosion process. Finally, PCA is applied and the principal component ( $PC_{\phi(I)}$ ) is derived from Red, Green, and Blue closing images. The length of Bs is chosen based on the spatial resolution of the input image, by assuming an approximate road-width.

### 2.1.3. Extraction of initial road-network

Although oriented linear Bs are not flexible enough to detect linear and curved road-structure at the same time ([Valero et al., 2010](#)), simple linear closing can be combined with Gabor filtering to distinguish between road and non-road pixels. In our experiments, all pixels that satisfy Eq. (8) are labeled as initial road pixels ( $R(PC_E \cap PC_\phi)$ ).

$$R(PC_E \cap PC_\phi) \Rightarrow |PC_E| \geq (\delta(PC_E)) \cap PC_\phi \leq (\delta(PC_\phi)), \quad (8)$$

where  $PC_E$  denotes the principal component of Gabor-energy features and  $PC_\phi$  denotes the principal component derived from morphological filtering.  $\delta(PC_E)$  and  $\delta(PC_\phi)$  are corresponding standard-deviations. The initial road-network is determined by Eq. (8) according to the following observations:

- The brightness of road-pixels is low; thus, road-pixels should be negative in the corresponding principal component of the image obtained after morphological closing.
- The majority of road-pixels have similar values in all Gabor-energy channels. For instance, if Gabor-value of road-pixels is high/low, road-pixels should be positive/negative in the corresponding principal component. Therefore, absolute  $PC_E$  is considered.

## 2.2. Graph-based segmentation

Segmentation in this work consists of three stages: initial segmentation using simple linear iterative clustering (SLIC) algorithm ([Achanta et al., 2012](#)), Region Adjacency Graph (RAG) representation ([Tremeau and Colantoni, 2000](#)) and hierarchical merging and splitting of segments. SLIC is used to construct an initial segmentation based on spectral-spatial distance. RAG is the graph representation of an input image, where superpixels are vertices and they are connected by edges whose values denote the similarity or dissimilarity between adjacent superpixels ([Peng et al., 2011; Tremeau and Colantoni, 2000; Felzenszwalb and Huttenlocher, 2004](#)). Merging is used to hierarchically group adjacent segments (superpixels in the first step) based on merging functions, which are mainly based on similarity of road spectral and spatial features. The merged segments are also split into smaller segments to remove artifacts from road segments based on dissimilarity in road properties. The initial road-network  $R(PC_E \cap PC_\phi)$ , as described in Eq. (8), will be used as the first splitting function to eliminate road-background.

### 2.2.1. Simple Linear Iterative Clustering (SLIC)

SLIC ([Achanta et al., 2012](#)) is a clustering method to generate concise and neat superpixels based on color similarity and proximity in the image plane. The idea of SLIC is similar to k-mean clustering method ([Haykin, 2008](#)); where each pixel is iteratively clustered to the nearest k-centroid based on the smallest Euclidean distance between a pixel and all nearest k-centroids in color space, until there are no more changes in k-centroids. In SLIC segmentation, a new distance is used based on five-dimensional [labxy] space. [lab] is the pixel color vector in Lab color space where L defines the lightness of the color, and a and b define the color along a red/green and blue/yellow axis, respectively ([Hunter et al., 1948](#)). [xy] is the pixel position ([Li et al., 2014; Achanta et al., 2012](#)). [Lab] distance measures color similarity between (super)pixels, where smaller distance indicates more uniform color. [xy] distance measures proximity between (super)pixels, where smaller distance indicates more adjacent pixels. The distance between the pixels is calculated as:

$$D(i, j) = D_{lab}(i, j) + \frac{m}{S} D_{xy}(i, j), \quad (9)$$

$$D_{lab}(i, j) = \sqrt{(l_i - l_j)^2 + (a_i - a_j)^2 + (b_i - b_j)^2}, \quad (10)$$

$$D_{xy}(i, j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \quad (11)$$

where  $D_{lab}(i, j)$  is the Euclidean distance in spectral [lab] space.  $D_{xy}$  is the Euclidean distance in spatial [xy] space.  $D(i, j)$  is the sum of [lab] distance and normalized [xy] distance. For an image of  $N$  pixels and  $K$  superpixels, the size of each superpixel is about  $\frac{N}{K}$ . The distance between every two adjacent superpixel blocks is  $s = \sqrt{\frac{N}{K}}$ . The variable  $m$  controls the compactness of superpixels. When  $m$  is large, the weight of [xy] distance is higher. Consequently, spatial proximity is more important and the resulting superpixels are more compact. When  $m$  is small, equal weights are assigned to both [xy] and

[lab] distances. As a result, the importance of both [Lab] spectral similarity and [xy] spatial proximity are approximately the same, and superpixels have less regular size and shape (Achanta et al., 2012).

SLIC parameters are chosen according to the spatial resolution of the input image. The number of segments ( $K$ ) is chosen to be reasonably high to facilitate over-segmentation. In general, road-width is relatively small (e.g.,  $3 \leq w \leq 8$  m) compared with other urban man-made classes. For example, if we consider an image with spatial resolution of 0.5 m and image-size of  $1000 \times 1000$ , where  $3-8$  m corresponding to 6–16 pixels, the number of superpixels can be roughly between ( $K = \frac{1000 \times 1000}{6 \times 6} \approx 27000$ ) and ( $K = \frac{1000 \times 1000}{16 \times 16} \approx 4000$ ). To be able to facilitate over-segmentation, we can choose a value closer to the lower limit.

In order to determine the most optimal compactness parameter  $m$ ; we try to maximize the homogeneity within a superpixel. The difference between mean and median is often used to define heterogeneity within a segment. Eq. (12) is computed with different parameters of  $m$  in the range (Li et al., 2014; Soille, 2006):

$$D_m = \max(\mu(d_m) - \rho(d_m)), \quad (12)$$

where  $d_m$  is the [labxy] distance of each pixel from its associated superpixel center.  $\mu(d_m)$  and  $\rho(d_m)$  are mean and median functions, respectively.  $\mu(d_m) - \rho(d_m)$  is used to define heterogeneity within a segment. Therefore, the most optimal compactness parameter is obtained by minimizing the maximum heterogeneity within all superpixels in the entire image.

The advantages of using primitive SLIC superpixels are twofold: the segments carry more information in describing the spatial structure of the regions, and the number of primitive regions is much fewer than pixels in an image and thus help in largely speeding-up region merging processes. SLIC considers both spectral and spatial similarities between pixels. However, the result can include many over and under segmented regions. Subsequently, the input image is represented into Region Adjacency Graph (RAG) to facilitate merging and splitting of superpixels.

### 2.2.2. Region Adjacency Graph (RAG)

To introduce adjacent relationships between segments, we suggest using Region Adjacency Graph (RAG).  $RAG = (V, E)$  is an undirected graph, where  $v \subseteq V$  is a set of vertices corresponding to superpixels, and  $E$  is a set of edges connecting the pairs of neighboring vertices. In other words, there is an edge between two vertices if the vertices are adjacent. Each edge  $(v_i, v_j) \in E$  has a corresponding weight  $w(v_i, v_j)$  to measure the dissimilarity of the two vertices connected by that edge. Superpixels are initial regions which are hierarchically merged to obtain a complete road-region. They can also be eliminated from larger segments to remove artifacts using splitting criteria.

### 2.2.3. Hierarchical merging and splitting regions

RAG is a graph representation showing connectivity-view of adjacent SLIC superpixels, which take into account spectral similarity and spatial proximity; however, these features are not enough to completely separate road-region from background (e.g., road and railway segments have different texture features, road and parking lots have different shape properties such as asymmetry and elongation, etc.). On the other hand, urban environments have different structures (e.g., small houses, large buildings, narrow streets, wide highways, etc.); therefore, one scale is not enough to discriminate a target from the background. Also, the results of SLIC segmentation have many over/under segmented regions (e.g., road segment includes tree pixels, building pixels and shadow). To compensate this, we suggest a multi-level structure, where the number of segments is hierarchically reduced from

bottom-level to the next top-level so that segments are merged to create a complete road-network. There are two important measures to build multi-level structure: weighting and difference functions. The weighting function defines spectral and spatial relationship between adjacent vertices as in Eq. (13). The difference function between two neighboring regions  $R_1, R_2 \subseteq V$  is the minimum weight edge connecting them, as shown in Eq. (14).

$$w(v_i, v_j) = D(v_i, v_j), \quad (13)$$

$$Diff(R_1, R_2) = \min_{v_i \in R_1, v_j \in R_2, (v_i, v_j) \in E} w(v_i, v_j), \quad (14)$$

where  $w(v_i, v_j)$  is the weighting function between vertices  $v_i$  and  $v_j$ .  $D(v_i, v_j)$  is the Euclidean distance between vertices  $v_i$  and  $v_j$ . This distance depends on a set of spectral or spatial features of  $v_i$  and  $v_j$ , that discriminates road regions from the background. Let us assume  $a$  and  $b$  are features that differentiate road-regions from non-road regions, then we can define  $w(v_i, v_j) = \sqrt{(a_{v_i} - a_{v_j})^2 + (b_{v_i} - b_{v_j})^2}$ . The value of  $Diff(R_1, R_2)$  is between 0 and 1. It is close to 0 if adjacent regions are more homogeneous. Therefore,  $Diff(R_1, R_2)$  acts as a threshold function to merge adjacent regions.

Before discussing the hierarchical structure, we will introduce our assumptions on road/non-road regions, which translate into merging and splitting principles in the multi-level structure.

- Brightness (B): Road regions have low variations in gray levels (Wang et al., 2011; Sheeren et al., 2006; Jabari and Zhang, 2013; Yu et al., 2016); therefore, lower standard-deviation of gray-level, the higher the possibility to be assigned to road class.
- Area (A): A road is generally a continuous feature with a relatively larger area compared to other man-made features (Liu et al., 2015). Therefore, segments with small area can be viewed as belonging to non-road class and can be eliminated. Area is defined as the total number of pixels in the segment (Yu et al., 2016; Akbari and Safari, 2013).
- Elongation and Asymmetry: Roads are narrow elongated regions (Liu et al., 2015; Jabari and Zhang, 2013). We use Elongation Index (EI) and Asymmetry Index (AI) as follows:

$$EI = \frac{I_{\text{major}}}{I_{\text{minor}}}, \quad (15)$$

where  $I_{\text{major}}$  is the length of the major axis of a region and  $I_{\text{minor}}$  is the length of the minor axis of a region (Liu et al., 2015; Yu et al., 2016). The EI for a square is 1 and for a rectangle, it is greater than 1. Thus, road segments have large value of EI (Liu et al., 2015).

$$AI(v_i, v_j) = \frac{2\sqrt{\frac{1}{4}(\sigma_x^2 + \sigma_y^2)^2 + (\sigma_{xy})^2 - \sigma_x^2 \sigma_y^2}}{\sigma_x^2 + \sigma_y^2}, \quad (16)$$

where  $\sigma_x^2$  is variance of  $x$  of a region and  $\sigma_y^2$  is variance of  $y$  of a region, where  $(x, y)$  corresponds to the pixel-coordinate. Road regions have large value of AI.

- Compactness: The width of road is relatively small. Therefore, its compactness is low (Liu et al., 2015; Wang et al., 2011; Jabari and Zhang, 2013; Akbari and Safari, 2013; Cao et al., 2015) and compactness is expressed by the Compactness Index (CI), as follows:

$$CI = \frac{2\sqrt{A\pi}}{P}, \quad (17)$$

where  $A$  is the area of the region and  $P$  is a perimeter of the region (Yu et al., 2016).

- Density: It describes the distribution of the pixels of a region in geographical space (Jabari and Zhang, 2013). The more the region is like a square, the higher is the value of density. Thus, road regions have lower density values; however, buildings have high value. Density Index (DI) is defined as follows:

$$DI = \frac{\sqrt{N}}{1 + \sqrt{\sigma_x^2 + \sigma_y^2}}, \quad (18)$$

where  $N$  is number of pixels in a region, and  $\sigma_x$  and  $\sigma_y$  are standard deviations of  $x$  and  $y$ , respectively.  $(x, y)$  corresponds to the pixel coordinate, identifying pixel-position with an image.

Based on the above assumptions, the adjacent regions are hierarchically merged to obtain an approximate road-map. Different combinations of spectral and spatial features are empirically tested to come up with general relation to extract road-map. The merging criteria  $M(R_1, R_2)$  between two regions  $R_1$  and  $R_2$  at different hierarchical-levels are based on different features, as follows:

$$M(R_1, R_2) \Rightarrow \min(\max(R_1), \max(R_2)) \leqslant Diff(R_1, R_2), \quad (19)$$

where

$$\max(R_1) = \max_{v_i, v_{ii} \in R_1}(w(v_i, v_{ii})), \quad (20)$$

$$\max(R_2) = \max_{v_j, v_{jj} \in R_2}(w(v_j, v_{jj})), \quad (21)$$

$Diff(R_1, R_2)$  is relatively large compared to internal differences within the individual regions. The merging function, which acts as a threshold function, shows that the difference  $Diff(R_1, R_2)$  between regions must be larger than the minimum internal difference  $\min(\max(R_1), \max(R_2))$ . The weighting function  $w$  is changed based on the selected feature(s), as described in L3 (Eq. (22)) and L4 (Eq. (23)), and consequently  $Diff(R_1, R_2)$ ,  $\min(\max(R_1), \max(R_2))$  and  $M(R_1, R_2)$  are changed. Fig. 3 illustrates the general concept of the proposed hierarchical representation including merging and splitting processes and it is summarized, as follows:

1. SLIC merging (L1): Pixels are clustered to build SLIC superpixels.
2. Gabor and Morphology Filtering (L2): SLIC superpixels from L1 which are not parts of the previous initial road-network  $R(PC_E \cap PC_\phi)$  (Eq. (8)) are eliminated. In Fig. 3, red nodes are removed in L2.

3. Spectral merging (L3): The adjacent superpixels are iteratively merged until convergence based on dissimilarity of standard-deviation of gray-level. Eq. (22) shows the distance function used for merging regions from L2.

$$w(v_i, v_j) = \sqrt{(B_{v_i} - B_{v_j})^2}, \quad (22)$$

However, the output of spectral-grouping is not robust enough to obtain a complete road-network. Therefore, merging based on spatial features could be used to improve the result.

4. Spatial merging (L4): As defined earlier, roads are elongated narrow regions with large area; therefore, we combine all road-features to construct the next-level. Distance function to group regions from L3 is presented as:

$$w(v_i, v_j) = \sqrt{(EI_{v_i} - EI_{v_j})^2 + (AI_{v_i} - AI_{v_j})^2 + (A_{v_i} - A_{v_j})^2}, \quad (23)$$

5. Spatial splitting (L5): There can be some narrow elongated regions, which are misclassified in L4 (e.g., buildings). Therefore, building needs further processing to differentiate from road-regions. Buildings have high Compactness Index (CI) and Density Index (DI). Therefore, superpixels not belonging to the road are eliminated from L2 based on two parameters: CI and DI, as follows:

$$CI_{L2} \geqslant CI_{L4} \cap DI_{L2} \geqslant DI_{L4}, \quad (24)$$

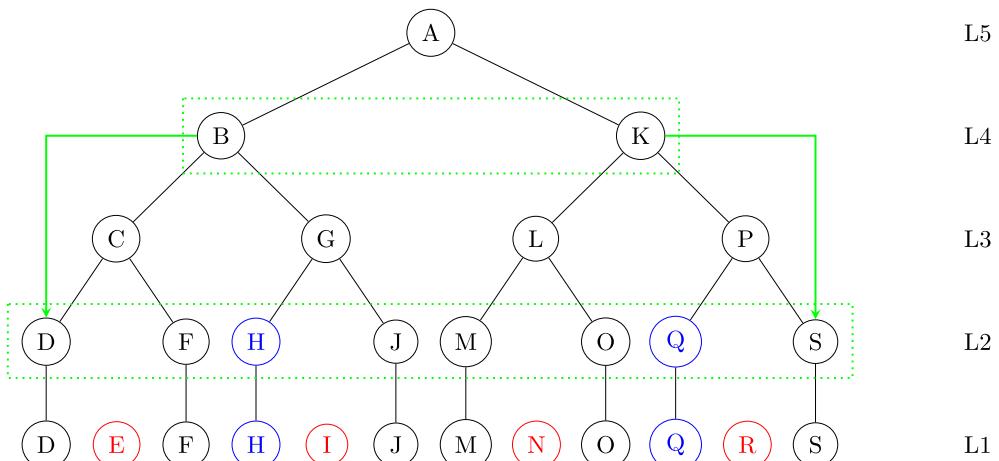
In Fig. 3, blue nodes are eliminated because they satisfy Eq. (24).

## 2.3. Post-processing

Because of road-occlusion by shadows, trees or other structures, there are discontinuous vertices in the graph-based map. Post-processing is used to merge the discontinuous vertices and to regularize the shape of road-segments within defined road-width.

### 2.3.1. Linking road segments

Dijkstra's algorithm (Russell and Norvig, 2003) is used to merge two discontinuous road superpixels from the previous intermediate road map (L5). To find the shortest path in L5, all regions from L1, which are eliminated are added. Let us assume a subgraph  $G$  of RAG with predefined window-size of the road map. If we have two road superpixels, belonging to different regions  $R_1$  and  $R_2$ ,  $R_1$  is



**Fig. 3.** Hierarchical graph-based representation. It shows merging and splitting processes in multi-level structure. Red and Blue nodes are eliminated nodes from L1 and L2 to construct L2 and L5, respectively. Green rectangles present spatial splitting process to construct L5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

considered as source vertex ( $S$ ),  $R_2$  is considered as a target vertex ( $T$ ), and weights between adjacent regions are based on Euclidean distance of brightness-mean between neighboring regions, as shown in Eq. (25).

$$w(v_i, v_j) = \sqrt{(\mu_{v_i} - \mu_{v_j})^2}, \quad (25)$$

The Dijkstra's shortest path is the path from source  $R_1$  to target  $R_2$  considering all vertices with the lowest total weight from source to target.

### 2.3.2. Regularization of the shape of the road-region

After iterative merging and subsequent splitting, the final width of the road region has to be regularized. Therefore, if the width of a road superpixel from  $L_1$ , is greater than the width of the parent road region, as shown in Eq. (26), road superpixel is eliminated.

$$W_{L1} > W_{L5}, \quad (26)$$

$$W = \frac{N}{\gamma}, \quad (27)$$

where  $N$  is number of pixels in a region, and  $\gamma$  is length-width ratio of a region. The ratio of length-width is estimated using the ratio of the eigenvalues ( $\lambda_1$  and  $\lambda_2$ ) of the covariance matrix,  $\gamma = \min \frac{\lambda_1}{\lambda_2}$  (Trimble-Definiens Joint Press Release, 2010).

## 3. Experimental analysis and discussion

To verify the effectiveness of the proposed method, extensive experiments on the road extraction from remote sensing images have been conducted on three datasets. The proposed method is also compared with other automatic methods. In this section, we will describe the experimental setup and discuss our experimental results.

### 3.1. Experimental setup

#### 3.1.1. Datasets description

- Massachusetts Road dataset: The dataset consists of 1171 images covering a large metropolitan area with urban, suburban, and rural regions. The size of all images is  $1500 \times 1500$  pixel with a spatial resolution of 1.2 m per pixel covering an area of  $500 \text{ km}^2$ . This dataset was built by Mnih and Hinton (2010). In this work, we report the results on 100 randomly selected images from this dataset.
- Zurich Road dataset: The dataset consists of 20 images acquired by the QuickBird satellite in 2002. The average image size is  $1000 \times 1150$  pixels with spatial resolution of 0.61 meter per pixel. It is composed of multispectral Very High-Resolution (VHR) images of 4 channels spanning near-infrared to visible spectrum (NIR-R-G-B). In this work, we concentrate on RGB channels. This dataset was built by Volpi and Ferrari (2015).
- Abu Dhabi Road dataset: The dataset consists of 210 images. The size of each image is  $1500 \times 1500$  pixel with spatial resolution of 0.5 m per pixel, composed of red, green and blue channels. In this work, we report the results on 20 randomly selected images from this dataset.

#### 3.1.2. Parameter setting

For Gabor filtering (Section 2.1.1), we use orientation separation angles ( $\theta$ ) of  $30^\circ$  as recommended in Clausi and Jernigan (2000), so  $\theta = \{0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ\}$ . We use frequency separation ( $f$ ) of 0.1 between 0.0 and 0.5 ( $0.0 \leq f \leq 0.5$ ), as recommended in Zhang and Tan (2002).

We assume that the road-width is relatively small ( $3 \leq w \leq 8$  m) in these datasets. The length of structuring elements of morphological closing process (described in Section 2.1.2), number of SLIC superpixels (described in Section 2.2.1) and search window-size to link road-segments (described in Section 2.3.1) are chosen based on the image-resolution. In Massachusetts data, the morphological closing ( $\phi_B(I)$ ) is performed by probing with the maximum of horizontal structuring elements  $B = 3 \times 7$  with directions  $\theta = [0, \frac{\pi}{4}, \frac{3\pi}{4}]$  and vertical structuring elements  $B = 7 \times 3$  with same directions. The number of superpixels is assigned as  $K \approx 36000$ , by assuming  $w = 3$ .

The spatial resolution of Zurich and Abu Dhabi data is higher, compared to Massachusetts data. Therefore, morphological closing ( $\phi(I)$ ) is performed by probing with larger structuring element  $B = 3 \times 13$  and  $B = 13 \times 3$ . The number of superpixels is calculated as  $K \approx 50000$ , assuming  $w = 5$ . For all datasets, we chose a search window-size of around 75 m.

The optimal compactness  $m$  which generates more regular SLIC superpixels is estimated as described in Section 2.2.1. For a test image from Massachusetts data, the corresponding plot is shown in Fig. 4, where we choose  $m = 25$ .

### 3.1.3. Evaluation metrics

The most common metrics for evaluating an extraction method are completeness (recall), correctness (precision) and quality (Sujatha and Selvathi, 2015; Maurya et al., 2011; Mnih and Hinton, 2010; Rajeswari et al., 2011). The completeness of a set of predictions is the fraction of true road pixels that are correctly extracted, while, the correctness is the fraction of predicted road pixels that are true roads. Quality is a measure of the goodness of the final result. It takes into account both the completeness and correctness.

$$\text{Completeness} = \frac{TP}{TP + FN}, \quad (28)$$

$$\text{Correctness} = \frac{TP}{TP + FP}, \quad (29)$$

$$\text{Quality} = \frac{TP}{TP + FP + FN}, \quad (30)$$

where TP is defined as the number of road pixels correctly extracted, FP is defined as the number of non-road pixels extracted

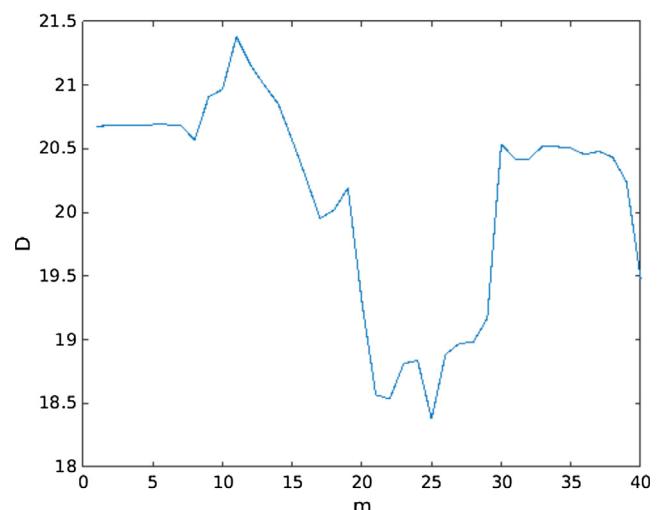


Fig. 4. [Labxy] Distance within superpixels (D) vs. Compactness (C).

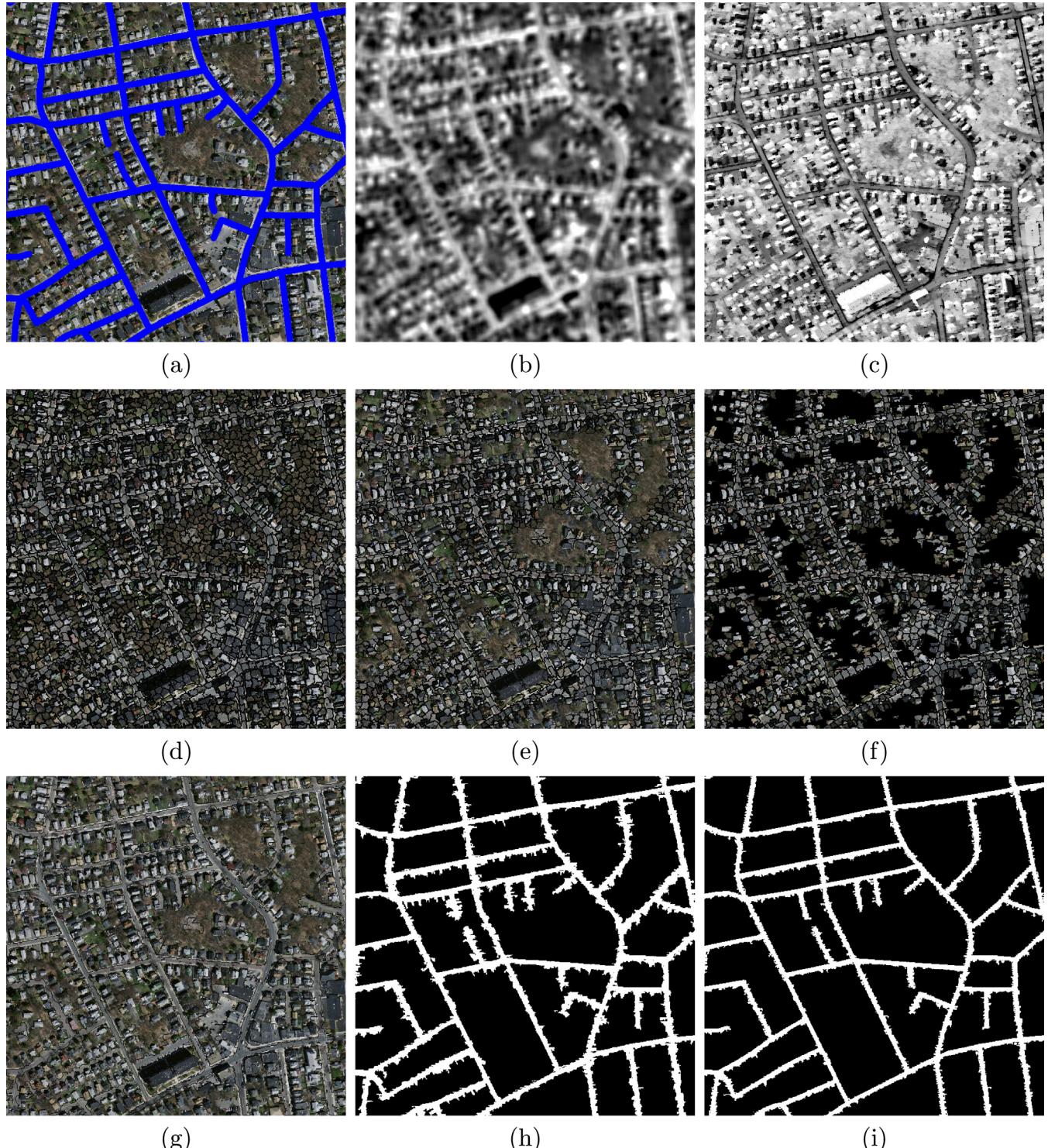
as roads, and FN is the number of road pixels extracted as non-road pixels.

### 3.1.4. Compared methods

To verify the performance, the proposed method is compared with two related methods: [Sujatha and Selvathi \(2015\)](#) and [Maurya et al. \(2011\)](#). We selected these methods because they are automatic methods which are based on segmentation tech-

niques and integrate spectral and spatial properties to connect road-segments, assuming that road-segments are elongated and narrow regions.

[Sujatha and Selvathi \(2015\)](#) presented a road extraction method based on adaptive threshold segmentation to build initial segmentation map, followed by connected components and opening/closing/thining morphological processes to remove unwanted non-road regions. This method is compared with seven other methods



**Fig. 5.** An example of all stages of road extraction method from Massachusetts data. 1. Pre-processing (b and c): (b) PCA of Gabor-filtering, and (c) PCA of morphological-filtering. 2. Hierarchical graph-based segmentation (d, e, f and g): (d) RAG after SLIC segmentation, (e) RAG after Gabor and morphological filtering, (f) RAG after Gabor and morphology without background, (g) RAG after merging and splitting. 3. Post-processing (h and i): (h) After connecting, and (i) After regularizing.

(Maurya et al., 2011; Shi et al., 2014; Miao et al., 2014; Mena and Malpica, 2005; Jin and Davis, 2005; Hu et al., 2007; Huang et al., 2012) and had the best performance.

The method of Maurya et al. (2011) uses unsupervised k-means clustering segmentation; however, spectral clusters might have other non-road parts like buildings and parking lots. To improve the effectiveness of extracted road-regions, morphological operations are used to remove non-road parts based on the assumptions that road-region is an elongated area that has the largest connected component.

### 3.2. Results

#### 3.2.1. Massachusetts data

The results of all steps of the proposed method in one of Massachusetts images are given in Fig. 5.

A test image with the corresponding ground truth is given in Fig. 5a. In order to enhance the contrast between the target and its background, Gabor filters and morphological filters are applied. Gabor-energy feature is able to discriminate between different texture-classes. For example, in most cases, road-regions have higher energy than non-road regions (Fig. 5b). On the other hand, road-regions have lower intensity in morphological closing image (Fig. 5c).

The input image is segmented using SLIC technique; which is based on distance in color and position. The segmented image is presented in Fig. 5d. Gabor and morphological operators are used for further processing to remove unwanted road regions. As shown in Fig. 5e, most of the non-road segments are removed by fusing Gabor-energy and morphological features. Fig. 5f presents the final output of graph-based-segmentation after integrating color and shape features as weighting functions. The result has some discontinuous road-regions due to artifacts on roads (e.g., tree, shadow, vehicle, etc.). These regions are connected using the shortest path between road-regions and its nearest road-regions (Fig. 5g). The result is further processed to smooth the boundaries (Fig. 5h).

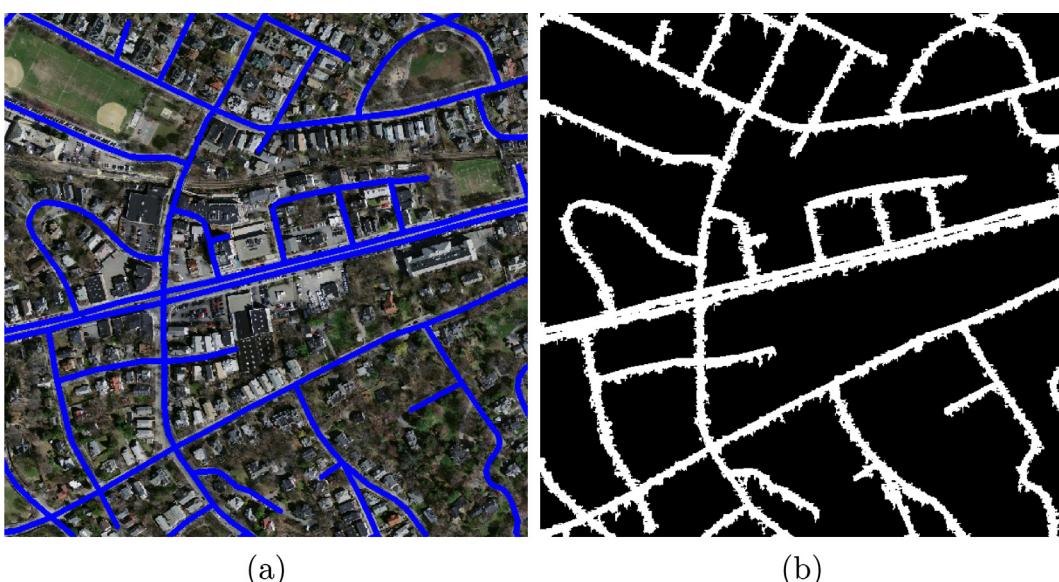
Fig. 6 shows an example of typical false negative detection, which tends to be caused by parallel roads. Our method is able to deal with shadows and occlusions caused by large objects such as buildings and with wider roads. However, in some cases it is unable to differentiate two road-lines. One possible solution of

dealing with such problem is reducing segment-size. However, reducing segment-size will lead to inaccurate region properties. Another possible approach is integrating predefined features (e.g., road-width) to separate between two road-lines. The resulting road-network is compared with extracted road-network implemented in Maurya et al. (2011) and Sujatha and Selvathi (2015). Some examples are shown in Figs. 7 and 8. In Fig. 7, it can be seen that most of the unwanted regions such as roofs, shadows, and bare soil have been removed by applying all methods; however, the result by the proposed method is visually superior to those produced by other compared methods. This is because both the compared methods (Maurya et al., 2011; Sujatha and Selvathi, 2015) initially depend on spectral dissimilarity, either by segmentation thresholding (Sujatha and Selvathi, 2015) (Fig. 7b) or unsupervised learning (Maurya et al., 2011) (Fig. 7c), and consequently they could not differentiate between roads and parking-areas or areas around buildings, since spectral properties are identical. The proposed method constructs the graph-based architecture based on over-weighting spatial proximity (distance between adjacent regions) than spectral similarity (distance in lab space) (Fig. 7d).

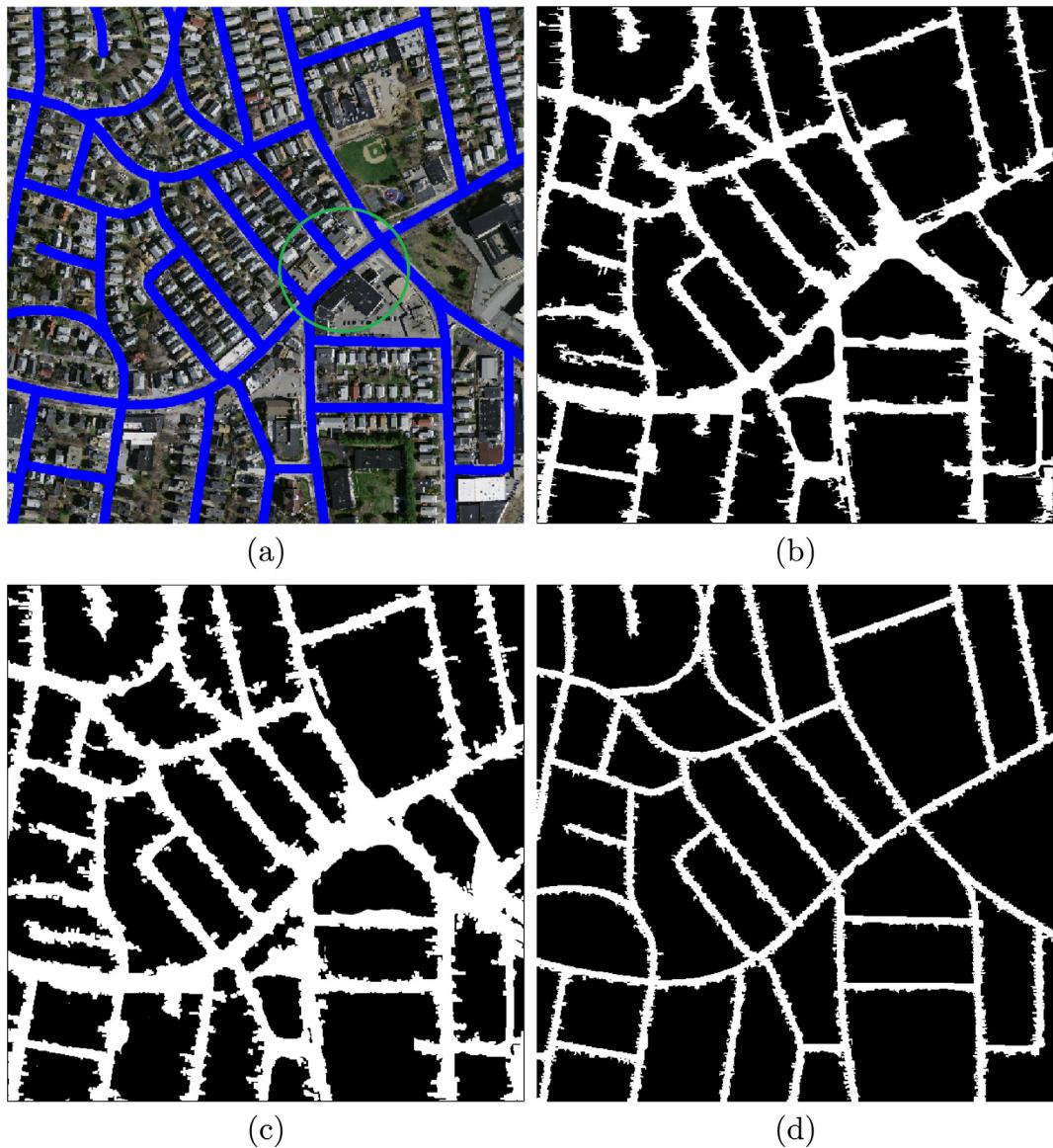
Fig. 8 demonstrates that the method of Sujatha and Selvathi (2015) and Maurya et al. (2011) achieve good performance in detecting most of the road regions; while, both of them are inefficient in differentiating road-regions from railway regions (Fig. 8b and c). Our method is able to discriminate between road and railway regions by applying Gabor filtering with different orientations and frequencies to differentiate road-texture (Fig. 8d).

It is also interesting to note that in both compared methods, the region around intersections with no salient elongated narrow features are predicted as road-regions. This is because many intersections in the dataset are so close to each other. This implies the positive effect of merging based on asymmetry and elongation and then splitting based on compactness and density to eliminate superpixels from adjacent regions in hierarchical multilevels.

Table 1 presents a comparison between the proposed method, Sujatha and Selvathi (2015) and Maurya et al. (2011) approaches in Massachusetts data according to three criteria: completeness, correctness, and quality. It presents an average of completeness, correctness, and quality for the results of applying the proposed method in 100 different images. It shows that the proposed



**Fig. 6.** An example of false negative from Massachusetts data: (a) test image overlapped by ground-truth, and (b) the final segmented image (Quality = 90.0%).



**Fig. 7.** 1st example of the comparison in Massachusetts road data. (a) Test image with ground-truth, (b) The result obtained by [Sujatha and Selvathi \(2015\)](#) (Quality = 69.4%), (c) The result obtained by [Maurya et al. \(2011\)](#) (Quality = 80.5%), and (d) The result obtained by the proposed method (Quality = 90.1%). Green-circle presents some regions around intersections, which are misclassified by Sujatha and Selvathi and Maurya et al. approaches. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

method achieves the best quality among other methods. This is more apparent when looking at a correctness of 91.0%. This demonstrates the importance of the initial combination of spectral, texture and shape properties to merge road-regions removing false alarms because of parking lots, buildings, trees, shadows and other classes. It is also interesting to note that completeness is 92.5%. This, in turn, implies a success for road-network extraction.

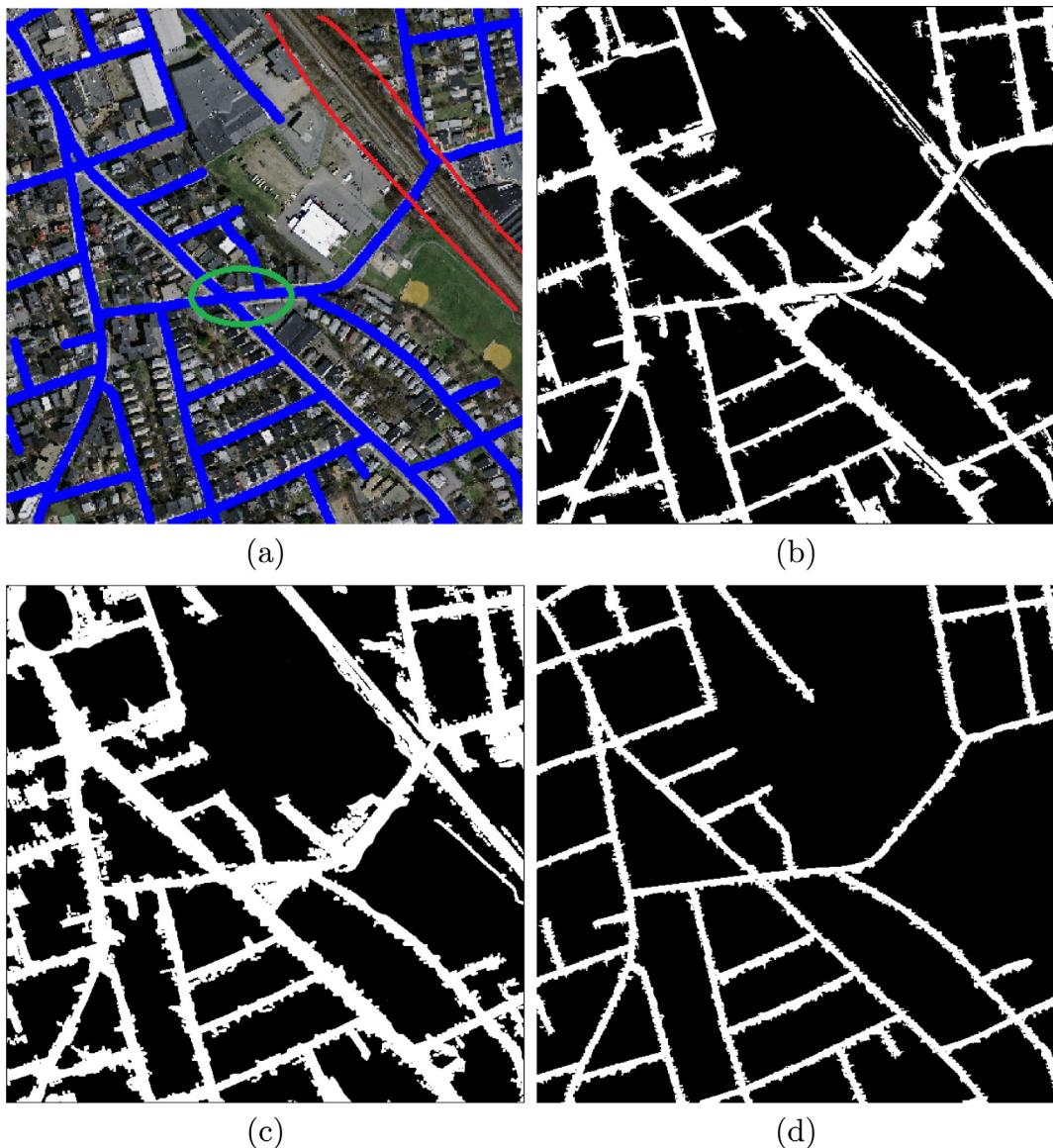
### 3.2.2. Zurich data

All stages of applying the proposed method in one of the Zurich images are shown in Fig. 9. In contrast with previous data, Gabor-energy values are lower (Fig. 9b) and morphological closing is more effective in removing man-made features (Fig. 9c and d). The results of the post-processing stage are illustrated in Fig. 9h and i presenting an approximate complete road-network.

The proposed method is compared with [Sujatha and Selvathi \(2015\)](#) and [Maurya et al. \(2011\)](#) approaches, as shown in Figs. 10 and 11.

Sujatha and Selvathi approach ([Sujatha and Selvathi, 2015](#)) extracts incomplete road-network, as shown in Figs. 10b and 11b. It uses adaptive global threshold segmentation to construct connected components; however, intensity values of all roads are not consistent. The resultant image is further filtered using morphological opening/closing, but both of them depend on the size of structuring elements, which must be smaller than the main road, but slightly larger than unwanted path, which is a challenging problem. On the other hand, the occlusion of roads by tree or shadow explains disconnected components. The road is hidden under these areas. In order to cope with these situations, further processing is required to connect road-components.

The method of [Maurya et al. \(2011\)](#), which uses K-mean clustering as initial binary-map, can identify most of the main road, but it is unable to remove non-road noise in more complex urban scenes with tree, shadow, and other occlusions. Moreover, it selects most of parking lots and buildings as road-regions; although it eliminates all regions which have similar features to road features by



**Fig. 8.** 2nd example of the comparison in Massachusetts road data. (a) Test image with ground-truth, (b) The result obtained by Sujatha and Selvathi (2015) (Quality = 76.7%), (c) The result obtained by Maurya et al. (2011) (Quality = 63.7%), and (d) The result obtained by the proposed method (Quality = 91.0%). Red-lines highlight railways, which are misclassified by the compared methods. Green-circle presents one of the regions around intersections, which is misclassified by Sujatha and Selvathi and Maurya et al. approaches. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**

Comparison between the proposed methodology and other automatic methods in Massachusetts roads. It presents an average of applying the proposed method in 100 images.

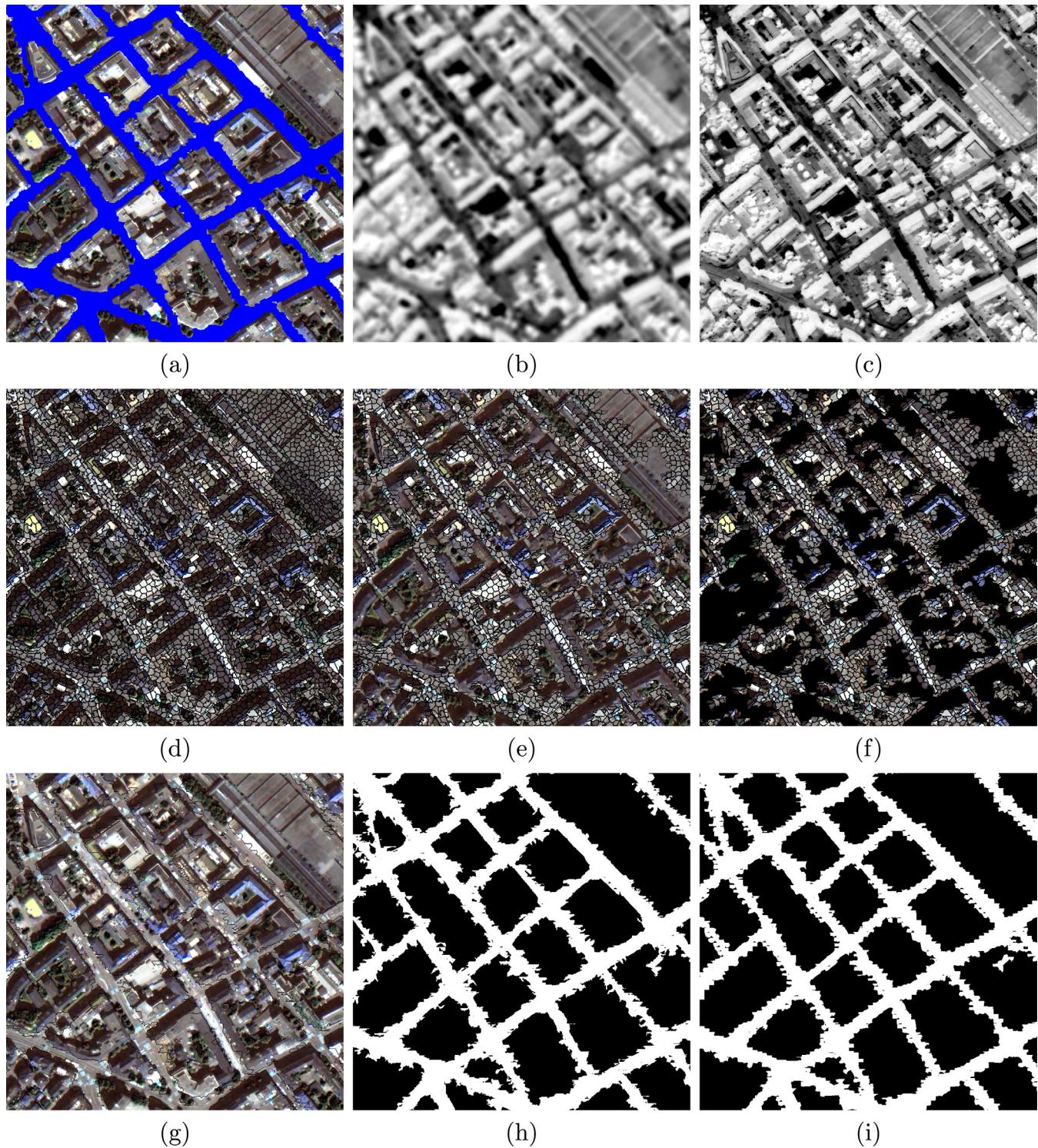
Methods	Completeness (%)	Correctness (%)	Quality (%)
Sujatha and Selvathi (2015)	$83.5 \pm 4.3$	$76.6 \pm 4.5$	$66.5 \pm 6.4$
Maurya et al. (2011)	$82.3 \pm 4.7$	$70.5 \pm 4.3$	$61.2 \pm 6.1$
Proposed method	$92.5 \pm 3.2$	$91.0 \pm 3.0$	$84.7 \pm 5.4$

computing areas of each connected component. However, in these images buildings/parking lots have large areas and it is difficult to deal with it, as shown in Figs. 10c and 11c.

A number of observations can be made when evaluating the results of the proposed method in Figs. 10 and 11. The road-regions extracted by our method are connected or less fragmented.

This is likely due to hierarchical merging functions that utilize various spectral and spatial properties of road-regions and it is, therefore, able to bridge the small gaps between the extracted road regions. Another observation to note is that our method yields cleaner results. This is the effect of the hierarchical splitting of non-roads (e.g., parking lots, building, etc.). Finally, our extracted road-regions are smoother than road-regions extracted from other two methods. It is a result of initial segment-size and compactness parameter which is used to balance between spectral and spatial features.

The extracted road-network is visually smooth, due to road-width. The road-width is larger/smaller than actual width because of two reasons: 1. Initial number of SLIC segments is not relatively large, although increasing the number of initial segments leads to inaccurate region-properties, which are used to merge road-regions. 2. Long-narrow regions may also be corrupted by many other sub-regions such as trees above the road surfaces and vehicles. One solution to solve the problem of vehicles is computing

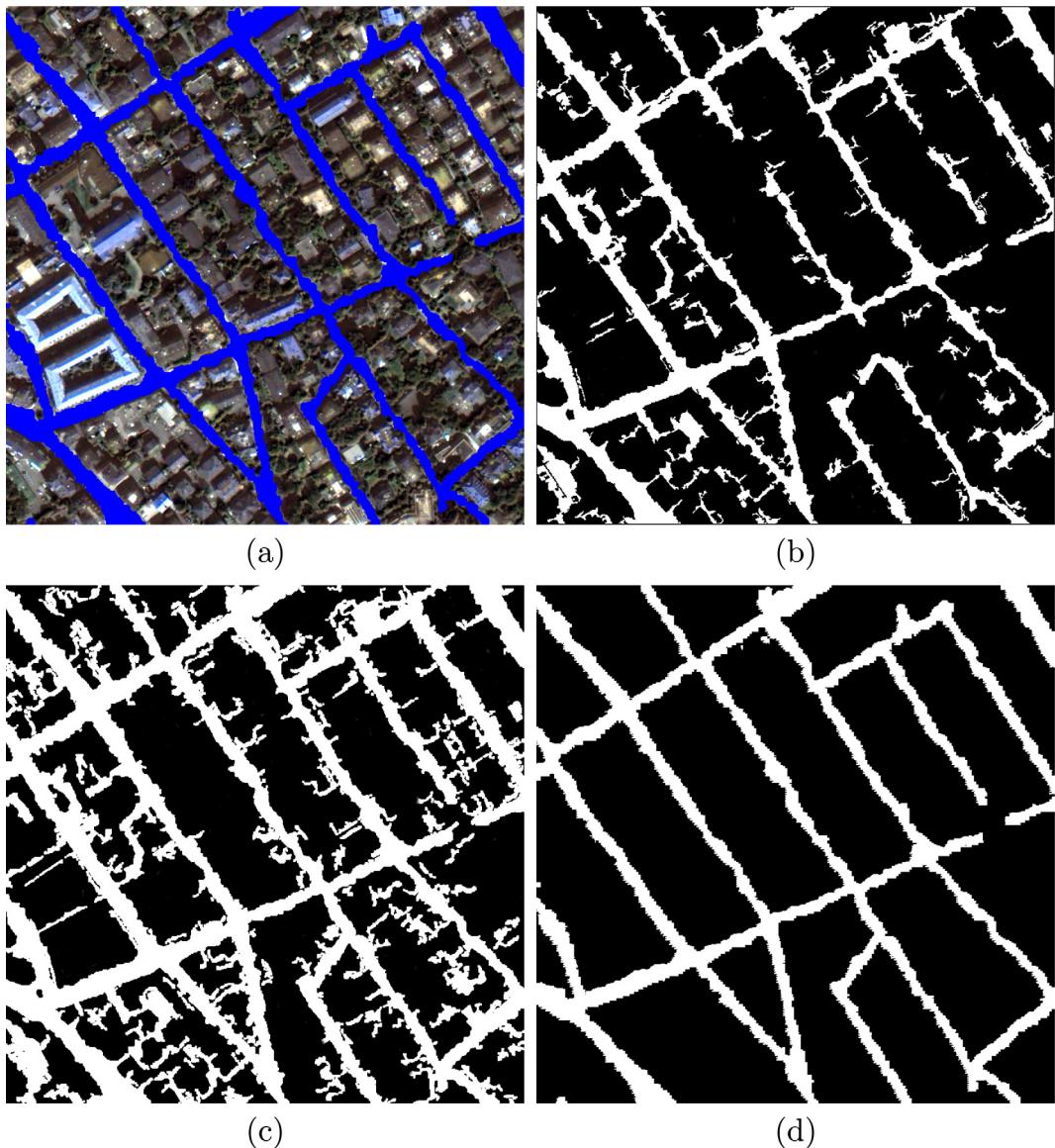


**Fig. 9.** An example of all stages of road extraction method from Zurich data. 1. Pre-processing (b and c): (b) PCA of Gabor-filtering, and (c) PCA of morphological-filtering, 2. Hierarchical graph-based segmentation (d, e, f and g): (d) RAG after SLIC segmentation, (e) RAG after Gabor and morphological filtering, (f) RAG after Gabor and morphology without background, (g) RAG after merging and splitting, 3. Post-processing (h and i): (h) After connecting, and (i) After regularizing.

the area of each sub-region in one superpixel and use it as splitting function in graph-based architecture.

Table 2 lists the quality metrics (completeness, correctness, and quality) for the proposed method, Sujatha and Selvathi (2015) and Maurya et al. (2011) approaches in Zurich data. It presents an average of completeness, correctness, and quality for the results of applying the proposed method in all images. The quality of the pro-

posed method exhibits noticeable results. The completeness, correction, and quality of the road network extraction are 93.4%, 90.9% and 85.4% respectively. Although the complexity of this data is high, completeness is higher than both compared methods and other six compared method (Maurya et al., 2011; Shi et al., 2014; Miao et al., 2014; Mena and Malpica, 2005; Jin and Davis, 2005; Hu et al., 2007; Huang et al., 2012), which are



**Fig. 10.** 1st example of the comparison in Zurich road data. (a) Test image with ground-truth, (b) The result obtained by [Sujatha and Selvathi \(2015\)](#) (Quality = 69.3%), (c) The result obtained by [Maurya et al. \(2011\)](#) (Quality = 81.6%), and (d) The result obtained by the proposed method (Quality = 89.9%).

compared with Sujatha and Selvathi approach ([Sujatha and Selvathi, 2015](#)). This demonstrates the necessity of the hierarchical grouping to combine road-primitives. This, in turn, implies that finding appropriate road primitives is an important prerequisite for a successful road network extractor.

### 3.2.3. Abu Dhabi data

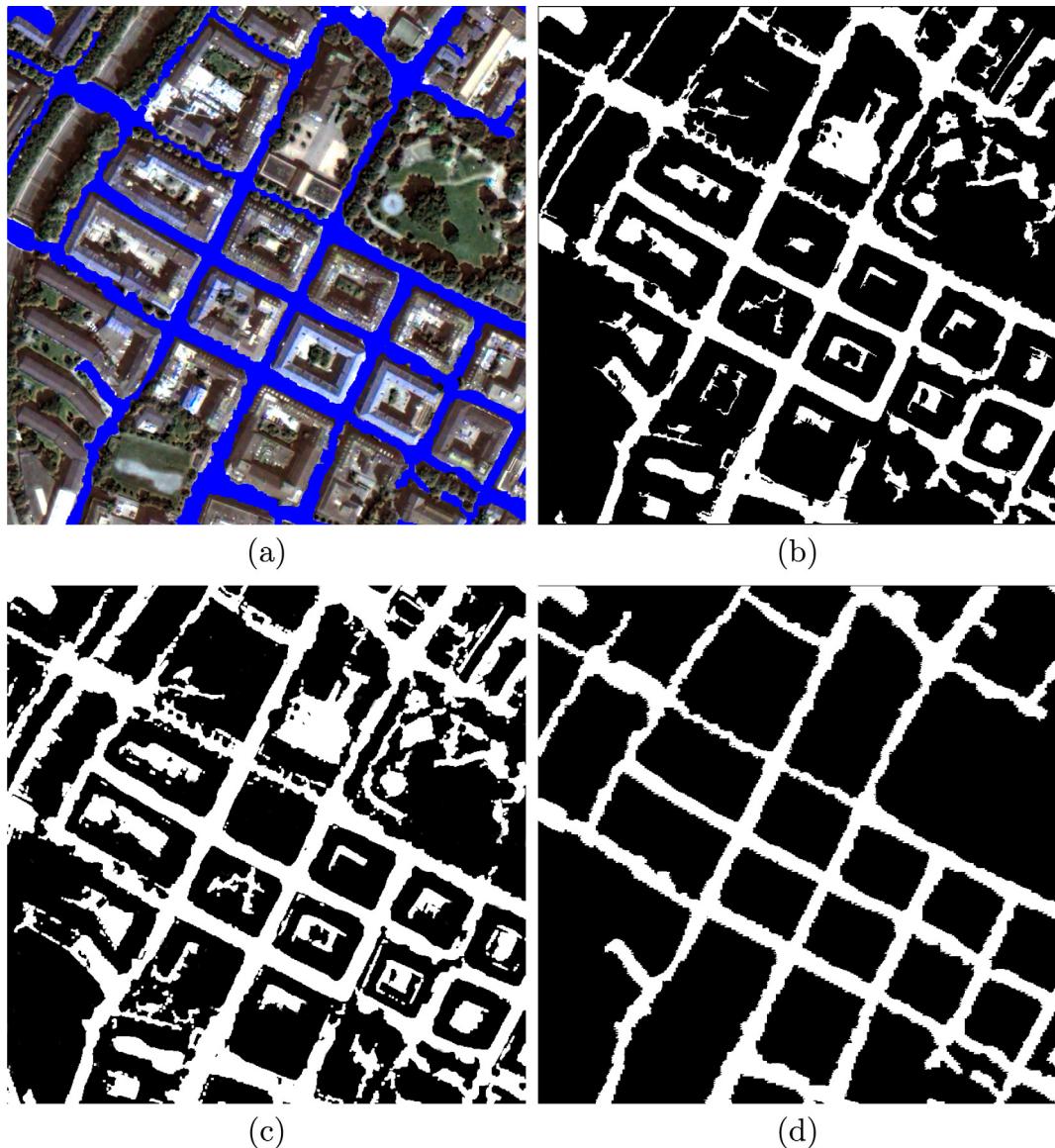
[Fig. 12](#) presents an example for extracting road-network from Abu Dhabi data. A test image with corresponding ground truth is given in [Fig. 12a](#) and results of the proposed method is presented in [Fig. 12b](#). The results demonstrate that the proposed method successfully extract most of road-regions, although of some false negative detection at adjacent road-regions.

[Table 3](#) lists the quality metrics (completeness, correctness and quality) for the proposed method, [Sujatha and Selvathi \(2015\)](#) and [Maurya et al. \(2011\)](#) approaches in Abu Dhabi data. The proposed method has the best performance compared to other methods.

## 4. Conclusion and future works

In this paper, a road extraction methodology has been presented for high-resolution imagery. The proposed method consists of the following steps: (1) Pre-processing to enhance the contrast between road and non-road regions by applying Gabor and morphological filtering, (2) Graph-based segmentation to merge road-regions and exclude all non-road regions, (3) Post-processing to connect all disjointed regions and regularize the shape of the final road-network.

The major contribution of this paper is that the proposed automatic method introduces the concept of hierarchical merging and then splitting in a multilevel graph-based framework. The proposed method was evaluated using different datasets. The results demonstrate that the proposed method is able to extract complete road-map and eliminate noise due to parking lots, buildings, tree, partial shadow over the road. The proposed approach is compared with two automatic road extraction methods. The experimental results have indicated that over 90% of road network has been



**Fig. 11.** 2nd example of the comparison in Zurich road data. (a) Test image with ground-truth, (b) The result obtained by Sujatha and Selvathi (2015) (Quality = 75.0%), (c) The result obtained by Maurya et al. (2011) (Quality = 82.9%), and (d) The result obtained by the proposed method (Quality = 89.6%).

**Table 2**

Comparison between the proposed methodology and other automatic methods in Zurich roads. It presents an average of applying the proposed method in all images.

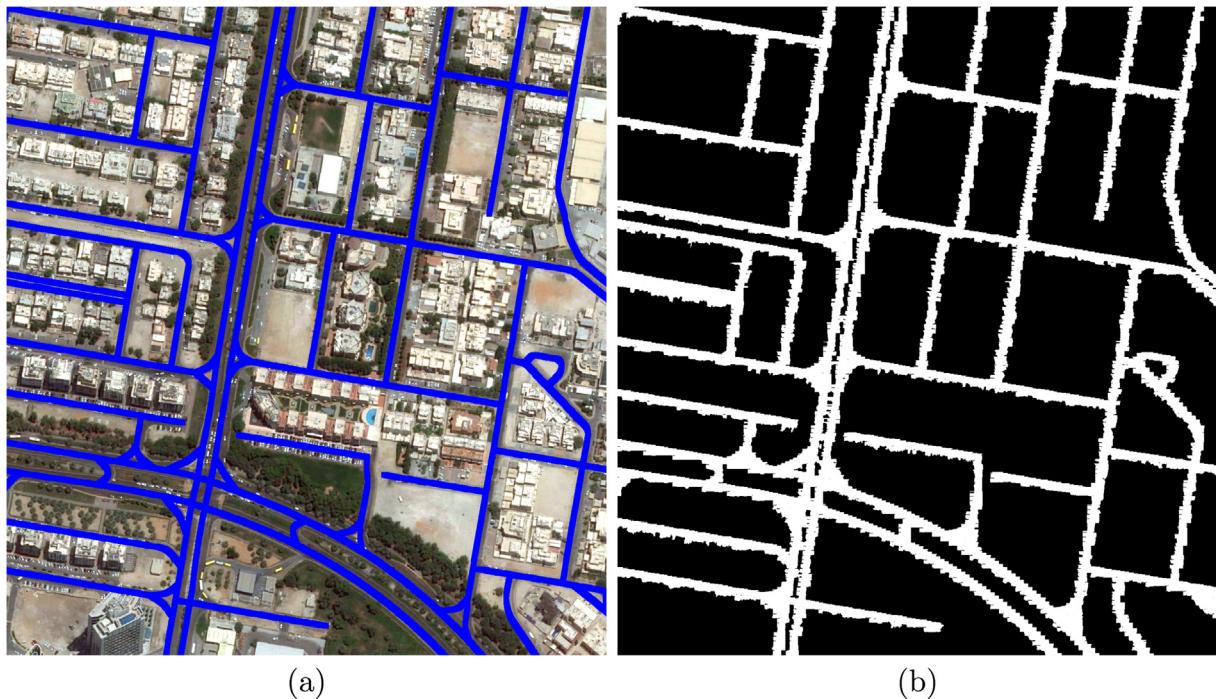
Methods	Completeness (%)	Correctness (%)	Quality (%)
Sujatha and Selvathi (2015)	$82.2 \pm 4.0$	$74.5 \pm 4.3$	$64.2 \pm 5.8$
Maurya et al. (2011)	$84.6 \pm 3.5$	$82.0 \pm 3.8$	$71.3 \pm 5.5$
Proposed method	$93.4 \pm 2.9$	$90.9 \pm 2.7$	$85.4 \pm 4.9$

correctly extracted. Moreover, the proposed method is superior to the compared methods and especially in complex urban areas.

Despite the efficiency of this method, the final road network is wider than what is annotated by ground-truth. The reason behind this is the initial segmentation. Although SLIC segmentation is one of the most efficient segmentation algorithms, which considers color similarity and location proximity, it is sensitive to the parameter setting such as number of segments and compactness. This segmentation method has been developed to overcome this prob-

lem by determining the most optimal compactness parameter based on maximization the homogeneity with a superpixel. In addition, the number of segments is determined by resolution of satellite imagery to restrict limitations of selecting the number of segments. In the future, we would like to overcome the problem of under-segmentation in the final roadmap. One another direction is to consider more information e.g., length of width to define additional splitting criteria.

Weighting and merging functions are used based on prior knowledge of road properties such as elongation and symmetry for extracting roads. These functions are selected based on empirical experiments. Other alternative functions could be used to have good segmentation results. For instance, Gaussian (Shi and Malik, 2000) or reciprocal, ratio of Gaussian to distance, or ratio of reciprocal to distance (Grady and Jolly, 2008) functions could be used as weighting functions with different dimensions based on the number of the used features. Also, using normalized minimum weight as a merging function based on the number of neighboring segments could be an alternative merging function.



**Fig. 12.** An example from Abu Dhabi data. (a) Test image with ground-truth, and (b) The result obtained by the proposed method (Quality = 90.5%).

**Table 3**

Comparison between the proposed methodology and other automatic methods in Abu Dhabi roads. It presents an average of applying the proposed method in 20 images.

Methods	Completeness (%)	Correctness (%)	Quality (%)
Sujatha and Selvathi (2015)	83.6 ± 3.7	76.1 ± 3.4	66.2 ± 5.0
Maurya et al. (2011)	84.1 ± 3.3	72.5 ± 3.5	63.8 ± 4.7
Proposed method	93.4 ± 2.8	91.4 ± 2.6	85.9 ± 4.8

The segmentation method is computationally expensive although the number of segments has been reduced from one level to another. This is because sequential merging and splitting methods and Dijkstra algorithm require additional computations depending on the size of the input images. For instance, computational time for extracting roads of Massachusetts image with dimension  $1500 \times 1500$  is 8 min and 10 s, Zurich image with dimension  $1000 \times 1150$  is 6 min and 14 s, Abu Dhabi image with dimension  $1500 \times 1500$  is 7 min and 30 s. Due to the time which was consumed for sequential processes, the proposed method was evaluated in only 100 and 20 images, which have the most discriminative features, from Massachusetts and Abu Dhabi datasets, respectively. This problem is a big motivation to design parallel algorithms instead of the slower sequential algorithms to evaluate the proposed method in all images.

Another future direction of this paper is adapting the proposed method to work with color infra-red (CIR) images, multispectral images with additional information from a larger number of bands, and with very-high resolution images. In addition, we will also explore using the results of the unsupervised method in semi-supervised framework for road-network extraction.

### Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

### Acknowledgments

The authors would like to thank V. Mnih for providing Massachusetts dataset and M. Volpi and V. Ferrari for providing Zurich data.

### References

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S., 2012. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Machine Intell.* 34 (11), 2274–2282.
- Akbari, D., Safari, A.R., 2013. Rule-based classification of a hyperspectral image using mssc hierarchical segmentation. *Int. Arch. Photogr. Remote Sensing Spatial Inform. Sci. (ISPRS) XL-1/W3*, 13–18.
- Auclair-Fortier, M.-F., Ziou, D., Armenakis, C., Wang, S., 2000. Automated correction and updating of roads from aerial ortho-images. In: Proceedings of the International Society Photogrammetry and Remote Sensing Conference, pp. 16–23.
- Bonnefon, R., Dhrt, P., Desachy, J., 2002. Geographic information system updating using remote sensing images. *Pattern Recogn. Lett.* 23 (9), 1073–1083.
- Cao, C., Sun, Y., 2014. Automatic road centerline extraction from imagery using road gps data. *Remote Sensing* 6 (9), 9014–9033.
- Cao, G., Wang, S., Liu, Y., 2015. An improved algorithm for automatic road detection in high-resolution remote sensing images by means of geometric features and path opening. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 1861–1864.
- Chaudhuri, D., Kushwaha, N.K., Samal, A., 2012. Semi-automated road detection from high resolution satellite images by directional morphological enhancement and segmentation techniques. *IEEE J. Select. Top. Appl. Earth Observ. Remote Sensing* 5 (5), 1538–1544.
- Cheng, G., Wang, Y., Gong, Y., Zhu, F., Pan, C., 2014. Urban road extraction via graph cuts based probability propagation. In: IEEE International Conference on Image Processing (ICIP), pp. 5072–5076.
- Cheng, G., Zhu, F., Xiang, S., Pan, C., 2015. Accurate urban road centerline extraction from VHR imagery via multiscale segmentation and tensor voting. Computing Research Repository (CRR) abs/1508.06163.
- Clausi, D.A., Jernigan, M.E., 2000. Designing gabor filters for optimal texture separability. *Pattern Recogn.* 33 (11), 1835–1849.
- Daugman, J., 1985. Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A* 2 (7), 1160–1169.
- Felzenszwalb, P.F., Huttenlocher, D.P., 2004. Efficient graph-based image segmentation. *Int. J. Comput. Vis.* 59 (2), 167–181.
- Grady, L., Jolly, M.-P., 2008. Weights and Topology: A Study of the Effects of Graph Construction on 3D Image Segmentation. Springer, Berlin, Heidelberg, pp. 153–161.

- Haykin, S., 2008. *Neural Networks and Learning Machines*. Prentice Hall.
- Huang, X., Zhang, L., 2009. Road centreline extraction from high resolution imagery based on multiscale structural features and support vector machines. *Int. J. Remote Sensing* 30, 1977–1987.
- Huang, Z., Zhang, J., Wang, L., Xu, F., 2012. A feature fusion method for road line extraction from remote sensing image. In: IEEE International Geoscience and Remote Sensing Symposium, pp. 52–55.
- Hu, J., Razdan, A., Femiani, J.C., Cui, M., Wonka, P., 2007. Road network extraction and intersection detection from aerial images by tracking road footprints. *IEEE Trans. Geosci. Remote Sensing* 45 (12), 4144–4157.
- Hu, X., Li, Y., Shan, J., Zhang, J., Zhang, Y., 2014. Road centerline extraction in complex urban scenes from lidar data based on multiple features. *IEEE Trans. Geosci. Remote Sensing* 52 (11), 7448–7456.
- Hunter, R.S., 1948. Accuracy, precision and stability of new photo-electric color-difference meter. *Proc. Thirty-Third Annu. Meet. Opt. Soc. Am.* 38 (12), 1092–1106.
- Idrißa, M., Achteroy, M., 2002. Texture classification using gabor filters. *Pattern Recognit. Lett.* 23 (9), 1095–1102.
- Jabari, S., Zhang, Y., 2013. Very high resolution satellite image classification using fuzzy rule-based systems. *Algorithms* 6 (4), 762–781.
- Jain, A.K., Farrokhnia, F., 1990. Unsupervised texture segmentation using gabor filters. In: IEEE International Conference on Systems, Man and Cybernetics, pp. 14–19.
- Jain, A.K., Farrokhnia, F., 1991. Unsupervised texture segmentation using gabor filters. *Pattern Recogn.* 24 (12), 1167–1186.
- Jin, X., Davis, C.H., 2005. An integrated system for automatic road mapping from high-resolution multi-spectral satellite imagery by information fusion. *Inform. Fusion* 6 (4), 257–273.
- Jin, H., Misra, M., Chung, E., Li, M., Feng, Y., 2012. Road feature extraction from high resolution aerial images upon rural regions based on multi-resolution image analysis and gabor filters. *Remote Sensing Adv. Tech. Platforms*, 388–414.
- Jirík, M., Ryba, T., Zelezny, M., 2011. Texture based segmentation using graph cut and gabor filters. *Pattern Recogn. Image Anal.* 21 (2), 258–261.
- Jolliffe, I., 2002. *Principal Component Analysis*. Springer Verlag.
- Li, Q., Chen, L., Li, M., Shaw, S.-L., Nüchter, A., 2014. A sensor-fusion drivable-region and lane-detection system for autonomous vehicle navigation in challenging road scenarios. *IEEE Trans. Vehicular Technol.* 63 (2), 540–555.
- Li, J., Jin, L., Fei, S., Ma, J., 2014. Robust urban road image segmentation. In: 11th World Congress on Intelligent Control and Automation (WCICA), pp. 2923–2928.
- Liu, B., Wu, H., Wang, Y., Liu, W., 2015. Main road extraction from zy-3 grayscale imagery based on directional mathematical morphology and vgi prior knowledge in urban areas. *PLoS ONE* 10 (9), 1–16.
- Maurya, R., Gupta, P.R., Shukla, A.S., 2011. Road extraction using k-means clustering and morphological operations. In: International Conference on Image Information Processing (ICIIP), pp. 1–6.
- Mena, J.B., 2003. State of the art on automatic road extraction for GIS update: a novel classification. *Pattern Recogn. Lett.* 24 (16), 3037–3058.
- Mena, J., Malpica, J., 2005. An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery. *Pattern Recogn. Lett.* 26 (9), 1201–1220.
- Miao, Z., Shi, W., Zhang, H., Wang, X., 2013. Road centerline extraction from high-resolution imagery based on shape features and multivariate adaptive regression splines. *IEEE Geosci. Remote Sensing Lett.* 10 (3), 583–587.
- Miao, Z., Wang, B., Shi, W., Wu, H., 2014. A method for accurate road centerline extraction from a classified image. *IEEE J. Select. Top. Appl. Earth Observ. Remote Sensing* 7 (12), 4762–4771.
- Mnih, V., 2013. Machine learning for aerial image labeling. Ph.D. thesis, University of Toronto.
- Mnih, V., Hinton, G.E., 2010. Learning to detect roads in high-resolution aerial images. In: Proceedings of the 11th European Conference on Computer Vision (ECCV): Part VI, pp. 210–223.
- Mokhtarzade, M., Zanj, M.V., 2007. Road detection from high-resolution satellite images using artificial neural networks. *Int. J. Appl. Earth Observ. Geoinform.* 9 (1), 32–40.
- Peng, B., Zhang, L., Zhang, D., 2011. Automatic image segmentation by dynamic region merging. *IEEE Trans. Image Processing* 20 (12), 3592–3605.
- Perciano, T., Tupin, F., Hirata, R., Cesar, R.M., 2011. A hierarchical markov random field for road network extraction and its application with optical and sar data. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 1159–1162.
- Rajeswari, M., Gurumurthy, K.S., Omkar, S.N., Senthilnath, J., Reddy, L.P., 2011. Automatic road extraction using high resolution satellite images based on level set and mean shift methods. In: 3rd International Conference on Electronics Computer Technology (ICECT), vol. 2, pp. 424–428.
- Russell, S.J., Norvig, P., 2003. *Artificial Intelligence: A Modern Approach*. Pearson Education.
- Saito, S., Aoki, Y., 2015. Building and road detection from large aerial imagery. In: Proceedings of Society of Photographic Instrumentation Engineers (SPIE) – The International Society of Optical Engineering, vol. 9405.
- Saito, S., Yamashita, T., Aoki, Y., 2016. Multiple object extraction from aerial imagery with convolutional neural networks. *J. Imaging Sci. Technol.* 60 (1), 010402-1–010402-9.
- Sheeren, D., Quirin, A., Puissant, A., Gancarski, P., Weber, C., 2006. Discovering rules with genetic algorithms to classify urban remotely sensed data. In: IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS), pp. 3919–3922.
- Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Machine Intell.* 22 (8), 888–905.
- Shi, W., Miao, Z., Wang, Q., Zhang, H., 2014. Spectral-spatial classification and shape features for urban road centerline extraction. *IEEE Geosci. Remote Sensing Lett.* 11 (4), 788–792.
- Shi, W., Miao, Z., Debayle, J., 2014. An integrated method for urban main-road centerline extraction from optical remotely sensed imagery. *IEEE Trans. Geosci. Remote Sensing* 52 (6), 3359–3372.
- Shu, Y., 2014. Deep convolutional neural networks for object extraction from high spatial resolution remotely sensed imagery. Ph.D. thesis, University of Waterloo.
- Sironi, A., Lepetit, V., Fua, P., 2014. Multiscale centerline detection by learning a scale-space distance transform. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2697–2704.
- Soille, P., 2003. *Morphological Image Analysis: Principles and Applications*. Springer-Verlag New York, Inc..
- Soille, P., 2006. Morphological image compositing. *IEEE Trans. Pattern Anal. Machine Intell.* 28 (5), 673–683.
- Sujatha, C., Selvathi, D., 2015. Connected component-based technique for automatic extraction of road centerline in high resolution satellite images. *EURASIP J. Image Video Processing* 2015 (08), 1–16.
- Tremeau, A., Colantoni, P., 2000. Regions adjacency graph applied to color image segmentation. *IEEE Trans. Image Processing* 9 (4), 735–744.
- Trimble-Definiens Joint Press Release, 2010. Trimble acquires definiens – earth sciences business to expand its geospatial portfolio.
- Unsalan, C., Sirmacek, B., 2012. Road network detection using probabilistic and graph theoretical methods. *IEEE Trans. Geosci. Remote Sensing* 50 (11), 4441–4453.
- Valero, S., Chanussot, J., Benediktsson, J.A., 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.
- Volpi, M., Ferrari, V., 2015. Semantic segmentation of urban scenes by learning local class interactions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 1–9.
- Wang, Y., Da, Y., Liu, X., Li, J., Huang, J., 2011. An object-oriented method for road damage detection from high resolution remote sensing images. In: 19th International Conference on Geoinformatics, pp. 1–5.
- Wegner, J.D., Montoya-Zegarra, J.A., Schindler, K., 2013. A higher-order crf model for road network extraction. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1698–1705.
- Yu, H., Yang, W., Xia, G.-S., Liu, G., 2016. A color-texture-structure descriptor for high-resolution satellite image classification. *Remote Sensing* 8 (3), 259.
- Zhang, J., Tan, T., 2002. Invariant texture segmentation via circular gabor filter. Proceedings of the 16th IAPR International Conference on Pattern Recognition (ICPR), vol. II. IEEE Computer Society, pp. 901–904.
- Zhou, S., Jiang, Y., Xi, J., Gong, J., Xiong, G., Chen, H., 2010. A novel lane detection based on geometrical model and gabor filter. In: IEEE Conference on Intelligent Vehicles Symposium (IV), pp. 59–64.