INTERNATIONAL JOURNAL OF REMOTE SENSING, 2018
VOL. 39, NO. 8, 2421–2440
https://doi.org/10.1080/01431161.2018.1425563

Unsupervised road extraction via a Gaussian mixture model with object-based features

Jiayuan Li, Qingwu Hu and Mingyao Ai

School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China

ABSTRACT

Automatic road extraction from remotely sensed images is an important and challenging task. This article proposes an unsupervised road detection method based on a Gaussian mixture model and object-based features. Our approach has five major stages, i.e. superpixel segmentation, feature description, homogeneous region merging, clustering via the Gaussian mixture model, and outlier filtering. In the third step, we present a graph-based region merging algorithm, in which the nodes of the graph are superpixels and edges are the similarities of intensity, colour, and texture. We also define two shape features, called deviation of parallelism (DoP) and narrow rate (NR), to automatically recognize road layer and filter outliers in the last step. We evaluated the proposed method on a variety of datasets, in which the Vaihingen dataset from the International Society for Photogrammetry and Remote Sensing Test Project is also included. Results demonstrate the power of our approach compared with some state-of-the-art methods.

ARTICLE HISTORY
Received 13 June 2017
Accepted 30 December 2017

1. Introduction

Road extraction is a key technique for many remote-sensing applications, such as geographic information system (GIS) construction, navigation, and emergency planning systems. It can significantly reduce time and labour compared with traditional manual methods. Over the past decades, various methods on road detection have been presented. Systematic review of the past works may be found in the literature (Quackenbush 2004; Mena 2003). In this article, we roughly categorize road extraction algorithms into semi-automatic and automatic methods, which is dependent on whether user interactions are required.

To sidestep the challenge of automatic road networks extraction, an effective way is to reduce the complexity of the problem with the aid of user-supplied information. A number of well-performing semi-automatic algorithms use seeds with directions to track the road. Gruen and Li (1997) proposed a three-dimensional road detection method, called Least-Squares B-spline Snakes (LSB-Snakes), by fusion of snakes model and least-squares framework with constraints of seeds. Hu, Zhang, and Tao (2004) semi-

CONTACT Qingwu Hu huqw@whu.edu.cn School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China

automatically detected road networks based on a piecewise parabola model and solved the parameters of the parabola by employing a least square template matching technique. In the works of Movaghati, Moghaddamjoo, and Tavakoli (2010), the geometric and radiometric properties of a road around the seed were modelled, and they tracked the road while updating the observation model via an extended Kalman filter (EKF) and a particle filter (PF). Unsalan and Sirmacek (2012) developed a road system which consists of three major stages: road centre extraction, road shape description, and road network formation. First, road primitives were detected with the assistance of users; then, road centres were extracted via kernel-based density estimation; and graph theory was used to represent the road shape for improving accuracy. Khesali et al. (2016) proposed a framework by fusion of radar satellite and high resolution optical images for semi-automatic road extraction. They developed two methods in their framework, one is based on neural network and another is knowledge based. Despite these methods being accurate and having the ability to can extract high quality road networks, their requirements of user input will dramatically reduce the efficiency and increase the labour cost. Furthermore, to incorporate these methods even with minimal user interactions into a fully automatic workflow is difficult.

A natural alternative to semi-automatic methods is automatic methods. The most straightforward automatic approaches are to extend semi-automatic ones with a road seed point detection stage. Barzohar and Cooper (1996) proposed geometric-probabilistic models for road detection. In their framework, roads were extracted by maximizing a posteriori probability of the Gibbs Distributions, in which starting points were selected based on intensity histograms. Hu et al. (2007) presented a road seeding algorithm based on rectangular approximations. They obtained the local homogeneous region of the seeds (called road footprint) with a spoke wheel operator and classified these footprints based on a Bayes decision model. However, the performance of these methods seriously relies on the seed generation stage. Classification-based methods, where remotely sensed image is segmented into road and non-road groups, have been proposed in (Mantero, Moser, and Serpico 2005; Song and Civco 2004). They usually share a common framework with stages of feature description and test set classification. In the first step, training and test set are represented by geometric or radiometric features, such as gradient, intensity, colour, length, shape and so on. Then, a classifier (e.g. Support Vector Machine (SVM), Bayes decision tree) is adopted to segment the image. Unfortunately, it is difficult to select suitable features, and samples of training sets are needed for supervised classification algorithms. In the work of Ghamisi and Benediktsson (2015), an effective feature selection strategy for road extraction that is based on particle swarm optimization and genetic algorithm is presented. Wang et al. (2015) introduced a deep convolutional neural network (DNN) for this task, which is a very famous deep learning technique in computer vision and machine learning. Mathematical morphology (MM), a widely used method, detects roads based on a shape structuring element. Guo, Weeks, and Klee (2007) exploited mean-shift clustering algorithm in Intensity–Hue–Saturation (HIS) colourspace for this task and used conditional mathematical morphology to improve the performance. Shi, Miao, and Debayle (2014) introduced a general adaptive neighbourhood mathematical morphology (GANMM) to perform spatial–spectral classification. However, MM-like methods are not sufficiently flexible for complicated road scenes, especially for images with multiple

types of structures. Knowledge-Based prior is also a powerful technique to produce road networks. For instance, Trinder and Wang (1998) proposed a knowledge-based model which includes the relationship between roads and the properties of roads. A knowledge prior about road junctions was introduced by Negri et al. (2006), and a Markov random field (MRF) was applied to describe the road networks. A higher-order probabilistic (abbreviated HOP) Conditional random field (CRF) is developed to model the observation prior in the literature (Wegner, Montoya-Zegarra, and Schindler 2013). Wegner developed (2015) a probabilistic representation of road networks. Traditional idea of minimum cost paths is combined with the CRF. However, this method generates nearly 15,000 superpixels in order to provide enough paths, which will significantly increase the processing time and physical memory.

In other studies, additional information (such as lidar data, multi-view images, and image- lidar fusion) is also explored for automatic road extraction. A multi-view image based method is developed by Hinz and Baumgartner (2003), in which redundancies were exploited to predict the occlusions and describe the roads. More recently, Hu et al. (2014) presented a lidar -based framework with impressive performance called MTH (mean shift, tensor voting, Hough transform), of which the key idea was to effectively extract geometric primitives of road candidates and to separate non-road regions from the roads. Ferraz, Mallet, and Chehata (2016) designed an approach for forested mountainous areas based on fine digital terrain models (DTMs). First, they adapted a supervised Random Forest classifier to extract potential road patches. Then, they built a graph to fill gaps created by occlusion. Finally, an object-based image analysis is applied.

In this article, we focus on the automatic road extraction task with only single image since lidar data are expensive. We address four research problems listed as follows: (1) how can we detect multiple object-based features to improve the robustness to noise, such as noisy pixels in road regions, grasslands and buildings with similar colour; (2) how can we represent the object-based features with a graph model for homogeneous region merging to make the geometric properties of roads more distinct; (3) how can we exploit a Gaussian mixture model (GMM) to cluster these object-based segments; (4) how can we describe the segments with two customized shape features for selecting the true road layer and filtering outliers?

2. The proposed road extraction approach

2.1. *Overview*

Figure 1 shows the schematic diagram of the proposed approach. The first stage, i.e. superpixel segmentation, is to segment the input image into small homogeneous regions by adopting an entropy rate superpixel (ERS) (Liu et al. 2011) algorithm. The pixels of a homogeneous region should have similar radiometric properties. Then, multiple features, including intensity, colour in YUV colourspace, texture by local binary patterns (LBP), are extracted to describe the superpixels. The subsequent steps are as follows: (1) homogeneous region merging based on graph model, in which the nodes of the graph are superpixels and edges are similarities of the multiple object-based features. The purpose of this stage is to eliminate the over-segmentation phenomenon and make the geometric properties of roads more distinct. (2) clustering via the GMM

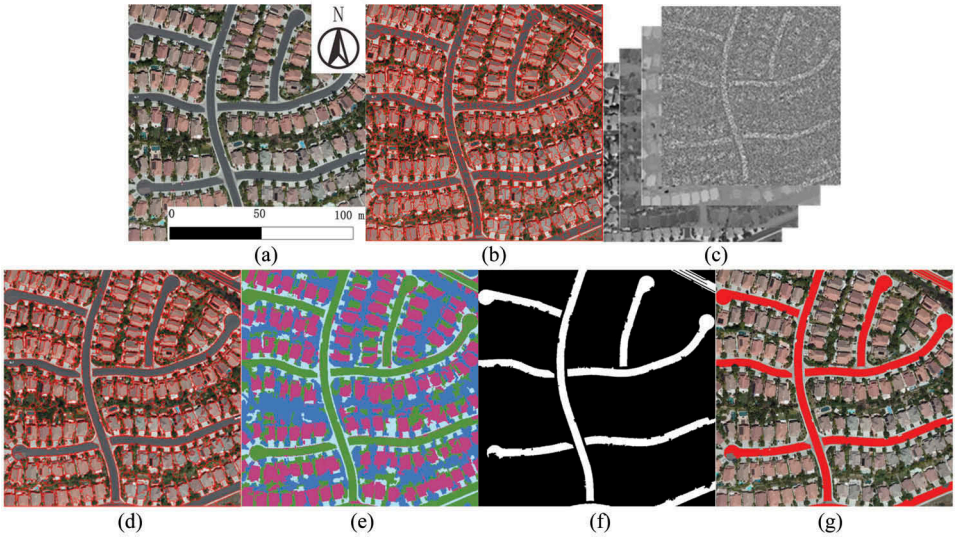

Figure 1. The schematic diagram of the proposed method. (a) input image; (b) superpixel segmentation via ERS; (c) detection of multiple features, including intensity, colour, and texture; (d) homogeneous region merging; (e) clustering via Gaussian mixture model (GMM); (f) outlier filtering; (g) output image with road layer.

(Zivkovic 2004) with the object-based feature vector. (3) shape feature construction, in which we design two geometric features called deviation of parallelism and narrow rate to automatically recognize road layer and filter outliers.

2.2. *Superpixel segmentation*

Superpixel segmentation is a popular pre-processing stage for many applications of computer vision and remote sensing such as image segmentation, image classification, object detection, and image interpretation. A superpixel is regarded as a perceptually uniform region, called homogeneous region in remote sensing. It is a group of spatially connected pixels with similar intensity, colour, and texture, and therefore these pixels are assumed to belong to the same object in the physical world.

The main advantage of using superpixel primitives instead of pixel primitives is computational efficiency since superpixel representation can largely reduce the number of primitives. For instance, in an N-classes unsupervised classification problem, the number of hypothesis for a superpixel representation is N^m, while the number of hypotheses for a pixel representation is N^n, where m and $n(m \ll n)$ are the number of superpixels and pixels, respectively. In addition, superpixels can be used for feature description in an object level which is more robust than pixels.

Many open-source superpixel segmentation algorithms have been proposed in recent years, such as simple linear iterative clustering (SLIC) (Achanta et al. 2012), ERS (Liu et al. 2011), etc. In this article, the ERS is adopted for this task. The ERS formulate the superpixel segmentation from a clustering perspective, for which a clustering objective function is developed:

$$\max_{A} H(A) + \lambda B(A) \text{ subject to } \mathrm{A} \subseteq \mathrm{E} \text{ and } N_{\mathrm{A}} \geq K, \tag{1}$$

where E and V denote the set of edges and the set of vertices of a graph $G = (V, E)$, respectively; K (1000 in our experiments) is the number of superpixels; A is a subset of edges. The goal of Equation (1) is to select a subset of edges $A \in E$ such that the graph $G = (V, A)$ consists of K connected subgraphs, which is a graph partition problem. N_A is the number of connected subgraphs. The first term $H(A)$ represents the random walk's entropy rate on graph $G = (V, A)$, which supports homogeneous regions – each segmented superpixel belongs to the same object in the physical world; $\lambda \geq 0$ is the weight of the balancing term $B(A)$; whereas the second term $B(A)$ represents the balancing term, which balances the sizes of homogeneous regions – each segmented superpixel contains a similar number of pixels. Figure 1(b) shows an example result of the ERS.

2.3. *Feature description*

To merge or to classify superpixels, similarity measured by feature vector should be defined. As a superpixel is a group of pixels with similar intensity, colour, and texture, these three types of low-level features are chosen to compute the similarity statistics.

We describe intensity and colour features in YUV colourspace since it minimizes the correlation between its three channels. In this model, Y stands for luminance channel (intensity); U and V stand for chromaticity channels (colour). Unlike traditional approaches, the median value is used instead of histogram technique for efficiency. For instance, the intensity and colour features of a superpixel are only a three-element vector for median value representation, while three 1D 256-bin histograms for histogram representation.

The texture feature is an important cue for coherent object-based region merging. The LBP (Ojala, Pietikäinen, and Harwood 1996) is a successfully applied operator for texture description. The central idea of LBP is to use the signs of differences with neighbouring pixels to describe image. It has been proved to be very efficient and invariant to monotonic illumination changes. Ojala, Pietikainen, and Maenpaa (2002) developed a variant of the LBP, called riu2-LBP, to extend the original LBP to be invariant to rotation and scale changes while not decreasing the computational efficiency of the original one. The riu2-LBP has been used for many computer vision applications, e.g. texture classification, video foreground/background segmentation, face recognition, and distinct region description. In our work, we introduce the riu2-LBP operator for texture description of superpixels, and the feature vector of each superpixel is a 10 bin histogram.

2.4. *Graph-based homogeneous region merging*

The ERS would segment an image into a fixed number of superpixel primitives. However, groups of nearby superpixel primitives are likely to belong to the same physical object due to the inherent property of over-segmentation. To eliminate this phenomenon, we develop a homogeneous region merging algorithm based on graph representation of these superpixel primitives.

Let $G = (V, E)$ be an undirected graph, where V represents the vertex set and E is a set of edges. In our algorithm, the vertices $v_i \in V$ are superpixel primitives, and the edges $(v_i, v_j) \in E$ are pairs of nearby superpixel primitives. A weight $w(v_i, v_j)$ is given to the edge $(v_i, v_j) \in E$ for measuring the similarity between superpixel v_i and superpixel v_j. As indicated earlier, we have obtained the feature vectors of superpixel primitives, thus, $w(v_i, v_j)$ is defined by

$$\begin{aligned} &w(v_i, v_j) = \alpha\left\|I_i - I_j\right\| + \beta\left\|C_i - C_j\right\| + (1 - \alpha - \beta)\left\|H_i - H_j\right\| \\ &\text{subject to } 0 < \alpha < 1,\ 0 < \beta < 1,\ 0 < (1 - \alpha - \beta) < 1, \end{aligned} \tag{2}$$

in which I, C, and H denote the intensity feature, colour feature, and texture histogram of the superpixel, respectively; α and β are coefficients (set to be 1/3 in our experiments). Note that all these three types of features are normalized to range of [0, 1].

The goal of this stage is to merge the similar superpixels. In other words, it can be regarded as a graph partition problem, in which superpixel set V is divided into disjoint subsets S such that each subset is a connected component s_i with similar superpixels. This means that the weights of edges inside a component s_i should be relatively low, while the weights of edges between two components s_i and s_j should be relatively large. Similar to the literature (Felzenszwalb and Huttenlocher 2004), internal difference and external difference are used in this algorithm. As defined by Felzenszwalb, the internal difference $Int(s_i)$ of component s_i is the maximum edge weight of s_i, and the external difference $Ext(s_i, s_j)$ between two components s_i and s_j is the minimum edge weight connecting s_i and s_j. Intuitively, if components s_i and s_j cannot be merged, the external difference $Ext(s_i, s_j)$ should be larger than at least one of the internal differences, $Int(s_i)$ and $Int(s_j)$. Thus, the steps of our homogeneous region merging algorithm can be summarized as follows, and a sample result is shown in Figure 1(d):

Step (1) Construct the graph $G = (V, E)$ with vertices representing the superpixel primitives and edge weights representing the similarities of nearby superpixels measured by intensity, colour, and texture features.

Step (2) Sort edge set E by non-decreasing order according to weights and set each superpixel to be an independent component s_i^0 of the initial subsets S^0. The internal difference η of each component is set to be 0.1 since the number of edges inside the component is 0.

Step (3) For the first edge (v_i, v_j), we assume that v_i belongs to component s_i^0 of $S^0$and v_j belongs to component s_j^0 of S^0. If $w(v_i, v_j) \leq \min(Int(s_i^0), Int(s_j^0))$ and $s_i^0 \neq s_j^0$ then components s_i^0 and s_j^0 can be merged to obtain subsets S^1. Otherwise $S^1 = S^0$. Do this step for the remainder edges and output the S^k as the result (k is the number of edges).

Step (4) Merge the superpixel primitives inside each component s_i^k to get a new superpixel and form feature vector of the new superpixel.

The main advantage of our homogeneous region merging algorithm is to make the geometric properties of roads more distinct, which is very important in the road layer recognition and outlier filter stage. The roads are long and parallel linear segments, however, superpixel segmentation would divide roads into small pieces, which will decrease the distinction between roads and other objects. Furthermore, homogeneous region merging algorithm will reduce the number of primitives.

2.5. Object-based clustering *via* Gaussian mixture model

After merging, neighbouring superpixels are likely to belong to different objects, while disjoint superpixels may be the same object. The goal of road extraction is to separate road areas from non-road areas. For this purpose, we first cluster these superpixel primitives into several classes based on the GMM such that roads are gathered into a category. The GMM, a parametric probability density function, is widely applied in computer vision and pattern recognition. It assumes that the distribution of the data can be modelled by a mixture of Gaussian distributions. Thus, the probability density function is a combination of Gaussian densities:

$$p(x|\Theta) = \sum_{t}^{i=1} \lambda_i g_i(x|u_i, \delta_i), \tag{3}$$

where function $g_i(\cdot)$ denotes Gaussian density; $\boldsymbol{x}$ is feature vector; $\boldsymbol{u}_i$ and $\boldsymbol{d}_i$ are the mean vector and the covariance matrix, respectively; t is the number of classes ($t = 4$ in our experiments). λ_i is the probability of a superpixel belonging to the i-th class and satisfies $(\sum_{i=1}^{t} \lambda_i) = 1$; $\mathbf{Q} = \{\lambda_i, \boldsymbol{u}_i, \mathbf{d}_i\}_{i=1}^{t}$.

We use the merged superpixels as clustering primitives and three types of features (intensity, colour, texture) to describe each superpixel. Thus, the dimension of feature vector x is 13, including 1 dimension of intensity, 2 of colour, and 10 of texture. This clustering algorithm will be very efficient since the total number of clustering primitives and feature dimension are very small (1000 and 13, respectively). See in Figure 1(e), the image is roughly classified into four classes, i.e. roads, buildings, grasslands, and others.

2.6. *Outlier filtering based on shape features*

Roads have two main geometric properties: (1) the edges of both sides of a road are almost parallel; (2) a road is an object which is long and narrow, like a linear feature. Based on these observations, we define two shape features, called deviation of parallelism (DoP) and narrow rate (NR), to automatically recognize the true road layer and remove outliers (e.g. some wrongly clustered buildings and grasslands in Figure 1(e)). The definition of DoP and NR are as follows:

The DoP is the deviation of the width of a merged superpixel. It reflects the parallelism of this superpixel. The smaller DoP represents that the edges of both sides of the superpixel are more parallel. The NR is the ratio of the length and the width of the superpixel. A superpixel with smaller DoP and larger NR is more likely to be a road segment.

In details, the contour edge of each superpixel primitive is firstly tracked. We sample n points $(p_0, \ldots, p_n)$ from the edge and calculate their corresponding normal direction $(d_0, \ldots, d_n)$ based on their neighbourhoods. Then, the points $(p_0, \ldots, p_n)$ are projected along their normal direction $(d_0, \ldots, d_n)$, intersecting at points $(q_0, \ldots, q_n)$. The length of a straight line $p_i q_i (i = 0, \ldots, n)$ is denoted by $l_{p_i q_i}$ (Figure 2). As can be seen, most sampled points are located at segments along the road direction, and their normal projection distances (the red-dotted line) are almost equal to the width of the road since the edges of both sides of a road are almost parallel. However, there are still some points on the

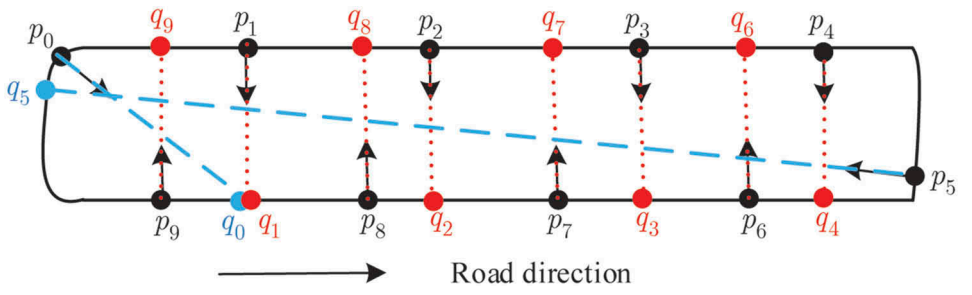

Figure 2. The illustration of constructing DoP and NR. First, the contour of a superpixel is tracked and some samples are generated (black dot); then, these samples are projected into the other side of the superpixel (red dot) and the normal distances $l_{p_iq_i}(i = 0, \ldots, n)$ can be computed. After removing outliers (blue-dotted line), the DoP and NR can be calculated by cleaned normal distances (red-dotted line). See text for details.

segments vertical to road direction, such as p_0 and p_5. Their normal projection distances $l_{p_0q_0}$ and $l_{p_5q_5}$ (blue-dotted line) are outliers for computing the DoP and NR. To filter outliers, the normal projection distances $l_{p_iq_i}(i = 0, \ldots, n)$ are sorted by non-decreasing order, then, 20% of maximum values and 20% of minimum values in $l_{p_iq_i}(i = 0, \ldots, n)$ are discarded, obtaining the cleaned normal projection distances $(l_0, \ldots, l_m)$. Finally, two shape features (DoP and NR) are defined. The DoP and NR reflect the parallelism and linearity of a road, respectively. The DoP is the maximum deviation between the cleaned normal projection distances $(l_0, \ldots, l_m)$ and the average distance $\bar{l}$ of $(l_0, \ldots, l_m)$. The NR is the ratio of the length and the width of a road. The width of a road is approximately equal to the average distance $\bar{l}$. Thus, the formulas of deviation of parallelism (DoP) and narrow rate (NR) are as follows:

$$\begin{cases} DoP = \frac{\max(\|l_i - \bar{l}\|)}{\bar{l}} (i = 0, \ldots, m) \\ NR = \frac{c/2 - \bar{l}}{\bar{l}} \end{cases}, \tag{4}$$

in which c is the perimeter of a road superpixel contour.

As mentioned earlier, we have classified the image into several categories. We calculate the average values of DoP and NR for each category layer and pick the layer with lowest DoP value and highest NR value as the road layer. Then, we remove the wrongly classified superpixel primitives in road layer which do not satisfy the following equation, and a sample cleaned road layer result is shown in Figure 1(f):

$$\begin{cases} DoP < \varepsilon \\ NR > \tau, \end{cases} \tag{5}$$

where ε and τ are threshold parameters.

To fill gaps created by shadows and occlusions in the road segments, we apply a simple but efficient gap-connecting strategy. The road centre lines are firstly extracted from the detected road layer. All the pixels of the centre lines are organized as a k-d tree structure and the end points of the centre lines are detected. Then, for each end point e_i, we search its neighbours N_{e_i}(the ones on the same centre line with e_i are discarded) in the k-d tree around a circular region with radius r ($r = 15\,m$ in our experiments if there

are no special instructions). If there is an end point in N_{e_i}, connecting them. Otherwise, end point e_i will be connected with the centre line with the most neighbours of e_i. Finally, the superpixels where the connected line segments located are added into the road layer. To obtain the cleanest road layer, shape features DoP and NR with better threshold are adapted again.

3. Experiments and evaluation

We evaluate the proposed method on three different remote-sensing datasets. The first one (Vaihingen) is provided by ISPRS Test Project for 3D Reconstruction and Urban Classification (Cramer 2010). This semantic labeling dataset consists of 33 tiles from digital aerial images acquired by an Intergraph/ZI DMC. These images are pan-sharpened colour infrared images with a ground resolution of 0.08m. The dataset is challenging since road detection task is seriously affected by shadows and occlusions. The second dataset (EPFL, École polytechnique fédérale de Lausanne (in French)) Turetken et al. 2013 consists of suburban ortho-images from Google Earth, captured by Turetken. Its ground resolution is about 1 m. We only use three images from the EPFL-dataset for comparison with two state-of-the-art methods, because these two methods only published their results of the selected three images. We also collect some images (Internet) with fewer shadows and occlusions from different scenes (e.g. flyovers, rural area, residential area, mountainous terrain, etc.).

3.1. *Evaluation measures and parameter study*

Three standard evaluation metrics widely adapted in the road detection task, i.e. completeness, correctness, and quality, are reported in this article. Their definitions are as follows:

$$\begin{cases} \text{Completeness} = \frac{(\text{TP})}{(\text{TP})+(\text{FN})} \times 100\% \\ \text{Correctness} = \frac{(\text{TP})}{(\text{TP})+(\text{FP})} \times 100\% \\ \text{Quality} = \frac{(\text{TP})}{(\text{TP})+(\text{FP})+(\text{FN})} \times 100\%, \end{cases} \tag{6}$$

where *TP*, *FN*, and *FP* are true positive, false negative, and false positive, respectively. True positive is the number of road pixels correctly identified; false negative is the number of road pixels wrongly identified; false positive is the number of non-road pixels identified as road pixels.

There are five main parameters in the proposed framework: K, η, t, ε, and τ. Parameter K is the number of superpixels. Large values of K will result in over-segmentation while too small values may produce under-segmented results. Parameter η is an internal difference threshold that decides if two superpixels can be merged. Parameter t is the number of clustering classes. Generally, the pixels of an image may be clustered into four classes, include roads, buildings, grasslands, and others. ε and τ are threshold parameters for shape features DoP and NR, respectively, which are used to filter outliers.

We study the parameters K, η, t, ε, and τ on the Internet dataset. We perform four independent experiments, where in each experiment, only one type of parameters is variable and the others are constant. The details can be found in Table 1. The results are reported in Tables 2–5.

Table 1. The details of parameter settings.

Experiment	Variable	Fixed parameters
Parameter K study	$K = [200,\ 500,\ 1000,\ 1500,\ 2000]$	$\eta = 0.1, t = 4, \begin{cases} \varepsilon = 0.3 \\ \tau = 1.5 \end{cases}$
Parameter η study	$\eta = [0.05,\ 0.1,\ 0.2,\ 0.3,\ 0.4]$	$K = 1000,\ t = 4, \begin{cases} \varepsilon = 0.3 \\ \tau = 1.5 \end{cases}$
Parameter t study	$t = [3,\ 4,\ 5,\ 6,\ 7,\ 8]$	$K = 1000,\ \eta = 0.1, \begin{cases} \varepsilon = 0.3 \\ \tau = 1.5 \end{cases}$
Parameter $\begin{cases} \varepsilon \\ \tau \end{cases}$ study	$\left[\begin{cases} \infty \\ 0 \end{cases}, \begin{cases} 0.3 \\ 1.5 \end{cases}, \begin{cases} 0.25 \\ 2 \end{cases}, \begin{cases} 0.2 \\ 2.5 \end{cases}, \begin{cases} 0.15 \\ 3 \end{cases}, \begin{cases} 0.1 \\ 3.5 \end{cases} \right]$	$K = 1000,\ \eta = 0.1, t = 4$

Table 2. The results of parameter K.

	K, $\eta = 0.1$, $t = 4$, $\varepsilon = 0.3$, $\tau = 1.5$				
Metric	200	500	1000	1500	2000
Completeness (%)	81.0	85.4	89.4	90.8	89.6
Correctness (%)	77.4	82.1	89.0	88.9	90.4
Quality (%)	65.5	72.0	80.5	81.6	81.8

Table 3. The results of parameter η.

	η, $K = 1000$, $t = 4$, $\varepsilon = 0.3$, $\tau = 1.5$				
Metric	0.05	0.1	0.2	0.3	0.4
Completeness (%)	82.2	89.4	92.6	76.6	79.4
Correctness (%)	93.8	89.0	82.3	76.8	67.8
Quality (%)	78.0	80.5	77.9	62.2	57.7

Table 4. The results of parameter t.

	t, $K = 1000$, $\eta = 0.1$, $\varepsilon = 0.3$, $\tau = 1.5$					
Metric	3	4	5	6	7	8
Completeness (%)	94.0	89.4	91.0	88.1	84.2	83.1
Correctness (%)	77.0	89.0	87.4	90.5	92.6	92.3
Quality (%)	73.4	80.5	80.4	80.6	78.9	77.7

Table 5. The results of parameter ε and τ.

	$\begin{cases} \varepsilon \\ \tau \end{cases}$, $K = 1000$, $\eta = 0.1$, $t = 4$, $\varepsilon = 0.3$, $\tau = 1.5$					
Metric	$\begin{cases} \infty \\ 0 \end{cases}$	$\begin{cases} 0.3 \\ 1.5 \end{cases}$	$\begin{cases} 0.25 \\ 2 \end{cases}$	$\begin{cases} 0.2 \\ 2.5 \end{cases}$	$\begin{cases} 0.15 \\ 3 \end{cases}$	$\begin{cases} 0.1 \\ 3.5 \end{cases}$
Completeness (%)	94.1	89.4	88.6	86.9	84.9	78.7
Correctness (%)	66.3	89.0	91.4	93.4	94.2	96.4
Quality (%)	63.7	80.5	81.7	81.9	80.7	76.5

According to the results, we can learn that: (1) small values of K perform much badly than large values due to the under-segmentation phenomenon. $K = 2000$ only achieves a slight improvement at the cost of much more running time compared with $K = 1000$. Because the sizes of the images in the three datasets are about 1000 × 1000 pixels, $K = 1000$ is enough to avoid over-segmentation

phenomenon. Thus, parameter K is relative to the image size, namely, large values of K should be selected if the image size is large. (2) the correctness is inversely proportional to η. The completeness curve rises first and then falls. Some road superpixels are not combined if η is small and some other objects are merged with roads if η is large. Both two cases may result in poor completeness performance due to the outlier filtering. In our experiment, $\eta = 0.1$ gets the best quality accuracy. (3) Parameter t has small influence to the results compared with other parameters. In our dataset, the pixels of an image can usually be clustered into four classes, thus, the quality accuracy decreases if t decreases from 4. The correctness increases while the completeness decreases when t increases from 4. (4) the shape feature constraints can largely improve the quality performance. The correctness is proportional to the feature constraints, while the completeness is inversely proportional. The best performance is achieved at $\{\varepsilon = 0.2\,,\ \tau = 2.5\}$; however, we choose $\{\varepsilon = 0.3\,,\ \tau = 1.5\}$ in our experiments. We can get better performance on the Vaihingen dataset when the shape feature constraints are relaxed, since this dataset suffers from shadows and occlusions which make the road structure irregular and complex. We fix $K = 1000$, $\eta = 0.1$,$t = 4$, and $\{\varepsilon = 0.3\,,\ \tau = 1.5\}$ for the following experiments.

3.2. *Vaihingen*

The road structure in Vaihingen dataset is irregular and complex. There are many main roads in this data, with shadows and occlusions. Shape features of roads are less distinct and some buildings are similar to roads in colour and texture, making road detection problem more challenging. Some visual results are depicted in Figure 3. As can been seen, main roads are detected well by our method. However, due to the effect of shadows and occlusions, some road pixels are missed (red labelled pixels in the figure). This is the main factor which affects the accuracy of the proposed method. Fortunately, shadows could be detected and removed in a preprocessing stage for practical applications. To make a fair comparison, there is no image preprocessing stage in all of our experiments.

We compare the proposed method with some state-of-the-art methods, including template matching (TM) (Hu and Tao 2005), HOP (Wegner, Montoya-Zegarra, and Schindler 2013), and DNN (Wang et al. 2015). There is an important reason for choosing these algorithms for comparison, i.e. these methods used the Vaihingen as the test dataset and reported their corresponding quantitative evaluation results in their works. So, we can easily compare the proposed method with their reported results. Quantitative results are reported in Table 6.

The proposed method without connecting ranks 3, 1, and 2 in terms of completeness, correctness, and quality compared with other methods, respectively. With connecting, our method achieves the best in all these three metrics. In our methods, we apply a simple connecting strategy. This strategy achieves 10.0%, 4.2%, and 10.4% growth rates in terms of completeness, correctness, and quality, respectively. The completeness of DNN is higher than our method without connecting and the correctness is almost the same. Compared with HOP, its completeness is slight better than the proposed method without connecting while its correctness and quality are

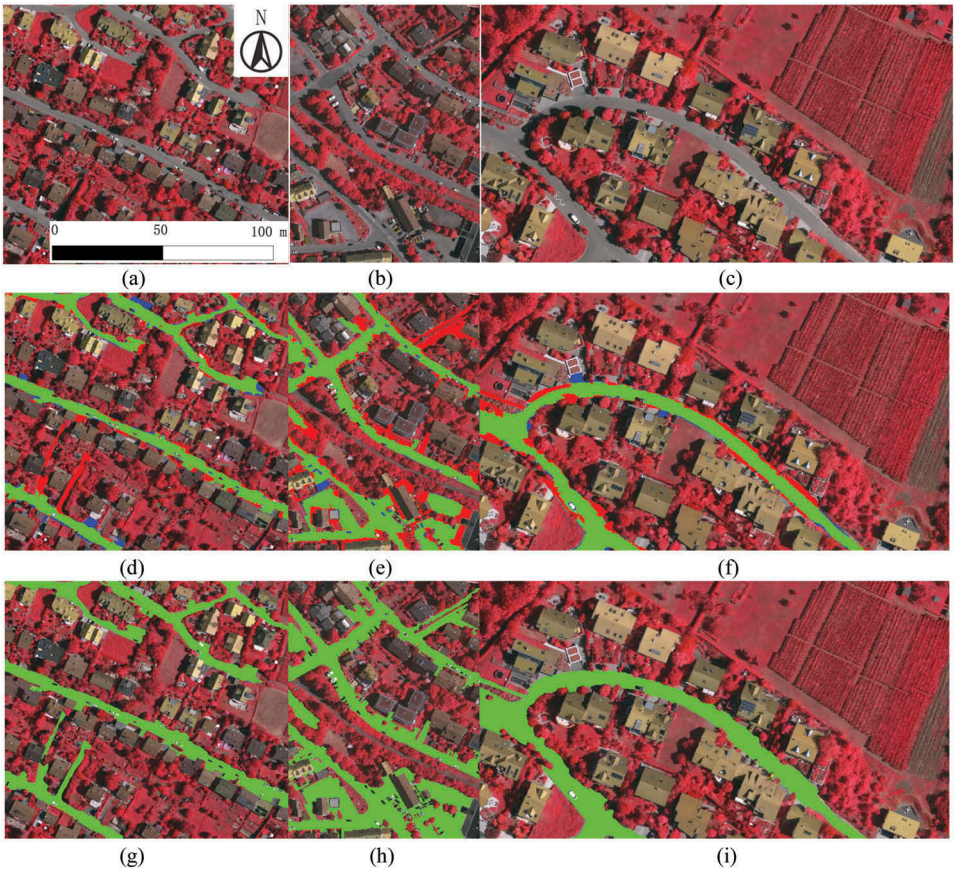

Figure 3. Road networks extracted in three patches of the Vaihingen. First row: input images; second row: results; third row: ground truth. Green true positives, blue false positives, red false negatives.

Table 6. Quantitative evaluation on the Vaihingen dataset.

Method	Completeness (%)	Correctness (%)	Quality (%)
TM	62.1	49.5	37.9
HOP	69.4	75	55.6
DNN	74.4	83.2	64.7
Proposed	66.2	83.3	58.3
Proposed + connecting	**76.2**	**87.5**	**68.7**

worse than ours. The reason is that our method adopts shape features to filter outliers for guarantee of correctness. The shadows and occlusions will divide the roads into irregular pieces, which leads to false negatives and decreases the completeness. Fortunately, a simple connecting strategy will make our method definitely better than HOP. DNN also gets very impressed correctness and quality accuracy. If a post-processing stage is performed on the DNN results, such as connecting or grouping in MTH, DNN may be even better than the proposed method with connecting. However, the major limitations of DNN are that it needs a large number of

labeled training sets and relies on high performance graphic processing unit (GPU) devices, which largely increase the artificial and hardware costs.

3.3. *EPFL*

The images of EPFL dataset are with lower spatial resolution and in greyscale with no colour information. Fortunately, there are only overhanging trees on the roads. Shadows are fewer than in the Vaihingen test data. In addition, the road structure is more regular and simple. To compare with two state-of-the-art methods, i.e. Turetken's method (Turetken et al. 2013) and Wegner's method (Wegner, Montoya-Zegarra, and Schindler 2015), three example images with results and ground truth provided by the authors are chosen as the test data. In the feature description stage, only intensity and texture are used to describe superpixels since there is no colour information. Qualitative and quantitative results are shown in Figure 4 and Table 7, respectively.

From Figure 4, we can see that the proposed method and Wegner's method can effectively avoid false positives while Turetken's method produces much more gross errors (blue-labeled pixels in the figure). It is interesting to note that the Turetken's method provides the smoothest road boundary results. This could be expected, as pointed out in the literature (Wegner, Montoya-Zegarra, and Schindler 2015), Turetken's method estimates the width of roads to as a constraint, which works well for the datasets with unoccluded roads with nearly constant width, just like the EPFL dataset. Our method gives fewer undetected roads pixels (known as false negative) than Wegner's method (red labeled pixels in the figure). There are many unconnected small pieces in Wegner's road results. This may be caused by the large number of superpixels needed by the method. It generates nearly 15,000 superpixels in order to provide enough paths. However, it also divided the images into too small pieces in the meantime, which makes the extraction task more challenging. In addition, the large number of superpixels will significantly increase the processing time and physical memory.

From the analysis of Table 7, we can draw similar conclusions as from the visual results. In terms of correctness, our method and Wegner's method achieve 8.2% and 14.8% improvements compared with Turetken's method, since there are many false positives produced by Turetken's method. For this data, our method does not use colour information. Thus, some road pixels have the same intensity and texture with their surrounding lands, which leads to wrongly merged superpixels. This may be the reason why our method could not perform as well as Wegner's method. Turetken's method provides the smoothest road boundaries. So, its completeness is the highest which is slightly higher than the proposed method. Both Turetken's method and ours are much better than Wegner's method, achieving 8.7% and 8.3% gains, respectively. Quality combines both completeness and correctness metrics into a single one which reflects the overall performance. As can be seen, our method gets the best performance, achieving 6.5% and 1.5% gains, respectively. In addition, the standard deviation of our method is smaller than others, which means that our method is more robust and stable.

Figure 4. Road networks extracted in three patches of the EPFL. First row: original images; second row: Turetken's results; third row: Wegner's results; fourth row: our results; fifth row: ground truth. Green true positives, blue false positives, red false negatives.

3.4. *Internet*

The dataset collected from internet have fewer shadows and occlusions. Figures 5 (a–e) are provided by (Zhang and Yuan 2011), in which (*a*), (*b*), (*e*) from the 'Residential' category and (*c*), (*d*) from the 'Viaduct' category. These images are

Table 7. Quantitative evaluation on the EPFH dataset.

Method	Image	Completeness (%)	Correctness (%)	Quality (%)
Turetken	1	88.0	79.8	71.9
	2	93.3	69.9	66.6
	3	82.1	75.5	64.8
	Mean	87.8	75.1	67.8
	SD	5.6	5.0	3.7
Wegner	1	79.4	88.5	71.9
	2	75.0	90.1	69.8
	3	82.9	91.2	76.8
	Mean	79.1	89.9	72.8
	SD	4.0	1.4	3.6
Proposed	1	86.9	81.5	72.5
	2	85.8	82.1	72.2
	3	89.5	86.2	78.3
	Mean	87.4	83.3	74.3
	SD	1.9	2.6	3.4

collected from Google Earth, with ground resolution of 0.5 m. Figure 5(f) is a UAV image with ground resolution of 0.2 m. Figure 5(g) is provided by DigitalGlobe (DigitalGlobe, 2015). This image is acquired by WorldView-3 with 0.3 m ground resolution. Figure 5(h) is an aerial image with 0.3 m space resolution. Figure 5(i) is an IKONOS image with 2 m space resolution provided by Mayer (Mayer et al. 2006). The mountainous terrain region covered by this image is 3.2 km×3.2 km and the length of the roads in the image is over 10 km. All images except Figure 5(i) are pan-sharpened RGB images. Figure 5(i) is a grey image. Figure 5 and Table 8 give the qualitative and quantitative results, respectively.

Main roads are well extracted and the boundaries are also quite smooth. However, there are still a small number of road primitives undetected. Since on the one hand overhanging trees will affect the extraction algorithm and on the other hand some unmerged road superpixels with small DoP or NR will be removed as outliers, as in the case of Figures 5(b, h, and i). In addition, some roads are not found in our results. Fortunately, these road segments could be connected by performing a more effective post-processing stage such as the grouping in MTH. Table 8 reports that our method achieves almost 90% accuracy in completeness and correctness, and 80% accuracy in quality. This is a very attractive performance for practical applications with shadow removal preprocessing stage to reduce time and labour cost.

4. Discussion and conclusion

4.1. *Limitation*

As the experiments reflected, there are still some problems in our algorithm that need to be resolved. First, shadows and occlusions dramatically affect the detection accuracy of our method. For the Vaihingen dataset, only 76%, 87%, and 69% accuracy in completeness, correctness, and quality are achieved. In contrast, our method achieves almost 90% accuracy in completeness and correctness, and 80% accuracy in quality for Internet dataset. The major distinction between the two

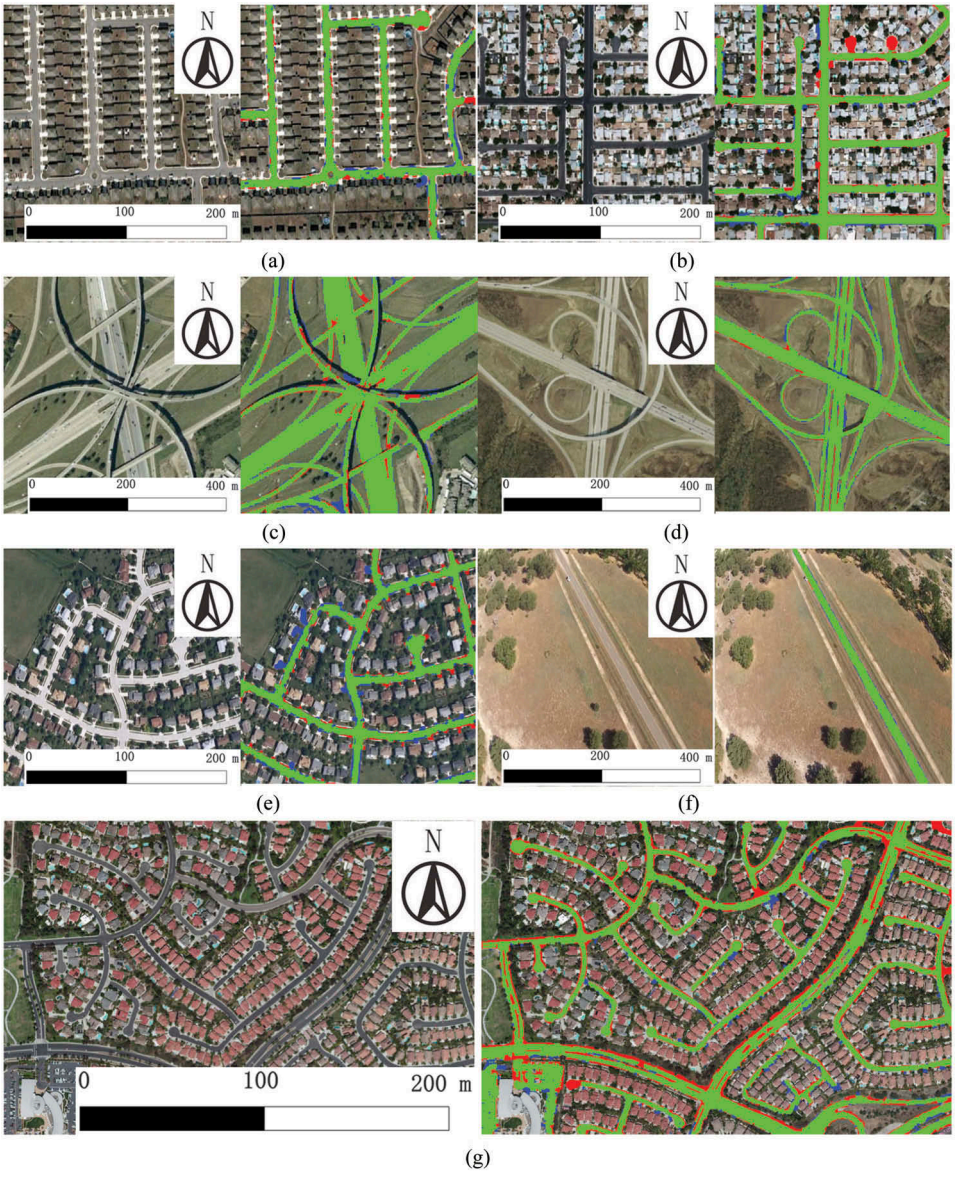

Figure 5. Road networks extracted in eight images of the Internet. Left images of (a)–(i): input images; right images of (a)–(i): our results. Green true positives, blue false positives, red false negatives.

datasets is the shadows and occlusions. So, in practical applications, shadow removal is necessary. Second, our method is not suitable for small-roads. The definition of 'small-road' relies on its width. The small-roads will be wrongly segmented in the superpixels generation stage. This is an inherent drawback of our method. Thus, our method is designed for high-resolution remote-sensing images road extraction task.

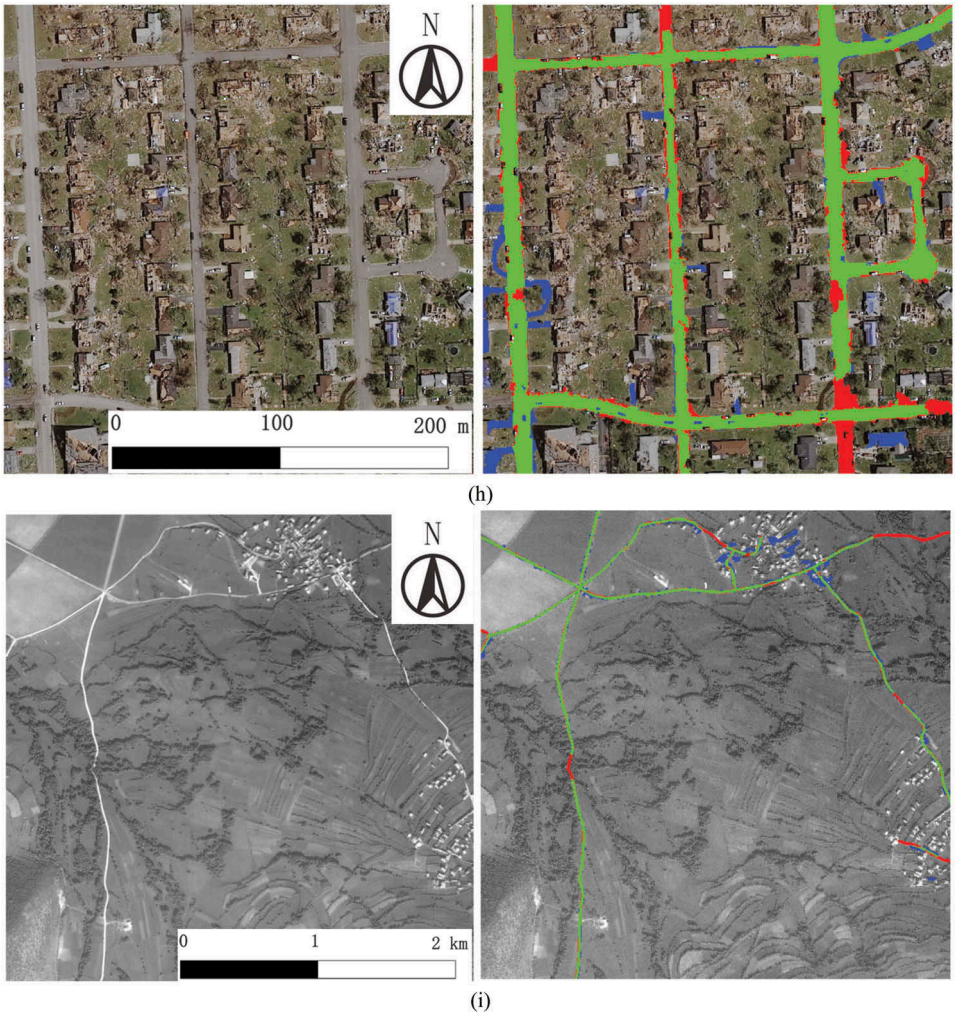

Figure 5. (continued).

Table 8. Quantitative evaluation on the Internet dataset.

Image	Completeness (%)	Correctness (%)	Quality (%)
1	90.2	91.1	82.9
2	87.3	93.4	82.2
3	93.4	92.0	86.4
4	93.8	94.3	88.8
5	88.7	84.0	75.9
6	99.1	92.1	91.4
7	76.1	93.5	72.3
8	80.6	86.5	71.6
9	91.8	77.9	72.8
Mean	89.0	89.4	80.5
SD	7.0	5.0	7.6

4.2. *Conclusion*

This article has proposed an unsupervised road detection method based on a single remotely sensed image, which achieves very impressing extraction performance. This method is fully automatic and no human interactions even for preparation of training sets are needed. It consists of five major stages, i.e. superpixel segmentation, feature description, homogeneous region merging, clustering via the GMM, and outlier filtering. First, images are segmented into object level. Second, over-segmented superpixels are merged based on three low-level object-based features. Then, these merged superpixels are classified into different categories via Gaussian mixture model. Finally, outliers are removed for improving accuracy. These steps are compact and each step is the basis of the next steps. The graph-based region merging algorithm could especially eliminate the over-segmentation phenomenon and make the geometric properties of roads more distinct. In addition, we make full use of shape properties of the roads and define two shape features, called deviation of parallelism and narrow rate, to recognize the roads. The experimental results on the ISPRS Vaihingen dataset and EPFL dataset demonstrate the power of our method, which show that the proposed method could achieve even better performance than some recent state-of-the-arts. We also show how good results could be obtained on datasets with fewer shadows, which reflects the potential of our method integrated with shadow detection algorithm for practical applications. As discussed earlier, some problems still need to be resolved, which will be the future work of us.

Acknowledgements

The authors would like to express their gratitude to the editors and the reviewers for their constructive and helpful comments for substantial improvement of this article.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National High-tech R&D Program of China (863 Program) [number 2013AA102401], the National Natural Science Foundation of China [number 41701528], the Fundamental Research Funds for the Central Universities [number 2042017KF0235], and the Key Technologies R&D Program of China [number 2015BAK03B04].

ORCID

Jiayuan Li http://orcid.org/0000-0002-9850-1668

References

Achanta, R., A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. 2012. "SLIC Superpixels Compared to State-Of-The-Art Superpixel Methods." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34 (11): 2274–2282. doi:10.1109/TPAMI.2012.120.

Barzohar, M., and D. B. Coope. 1996. "Automatic Finding of Main Roads in Aerial Images by Using Geometric-Stochastic Models and Estimation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18 (7): 707–721. doi:10.1109/34.506793.

Cramer, M. 2010. "The DGPF-Test on Digital Airborne Camera Evaluation–Overview and Test Design." *Photogrammetrie-Fernerkundung-Geoinformation* 2010 (2): 73–82. doi:10.1127/1432-8364/2010/0041.

DigitalGlobe. 2015. https://www.digitalglobe.com/.

Felzenszwalb, P. F., and D. P. Huttenlocher. 2004. "Efficient Graph-Based Image Segmentation." *International Journal of Computer Vision* 59 (2): 167–181. doi:10.1023/B:VISI.0000022288.19776.77.

Ferraz, A., C. Mallet, and N. Chehata. 2016. "Large-Scale Road Detection in Forested Mountainous Areas Using Airborne Topographic Lidar Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 112: 23–36. doi:10.1016/j.isprsjprs.2015.12.002.

Ghamisi, P., and J. A. Benediktsson. 2015. "Feature Selection Based on Hybridization of Genetic Algorithm and Particle Swarm Optimization." *IEEE Geoscience and Remote Sensing Letters* 12 (2): 309–313. doi:10.1109/LGRS.2014.2337320.

Gruen, A., and H. Li. 1997. "Semi-Automatic Linear Feature Extraction by Dynamic Programming and LSB-snakes." *Photogrammetric Engineering & Remote Sensing* 63 (8): 985–994.

Guo, D., A. Weeks, and H. Klee. 2007. "Robust Approach for Suburban Road Segmentation in High-Resolution Aerial Images." *International Journal of Remote Sensing* 28 (2): 307–318. doi:10.1080/01431160600721822.

Hinz, S., and A. Baumgartner. 2003. "Automatic Extraction of Urban Road Networks from Multi-View Aerial Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (1–2): 83–98. doi:10.1016/S0924-2716(03)00019-4.

Hu, J., A. Razdan, J. C. Femiani, M. Cui, and P. Wonka. 2007. "Road Network Extraction and Intersection Detection from Aerial Images by Tracking Road Footprints." *IEEE Transactions on Geoscience and Remote Sensing* 45 (12): 4144–4157. doi:10.1109/TGRS.2007.906107.

Hu, X., and C. V. Tao. 2005. "A Reliable and Fast Ribbon Road Detector Using Profile Analysis and Model-Based Verification." *International Journal of Remote Sensing* 26 (5): 887–902. doi:10.1080/01431160420002982 43.

Hu, X., Y. Li, J. Shan, J. Zhang, and Y. Zhang. 2014. "Road Centerline Extraction in Complex Urban Scenes from LiDAR Data Based on Multiple Features." *IEEE Transactions on Geoscience and Remote Sensing* 52 (11): 7448–7456. doi:10.1109/TGRS.2014.2312793.

Hu, X., Z. Zhang, and C. V. Tao. 2004. "A Robust Method for Semi-Automatic Extraction of Road Centerlines Using A Piecewise Parabolic Model and Least Square Template Matching." *Photogrammetric Engineering & Remote Sensing* 70 (12): 1393–1398. doi:10.14358/PERS.70.12.1393.

Khesali, E., M. J. V. Zoej, M. Mokhtarzade, and M. Dehghani. 2016. "Semi Automatic Road Extraction by Fusion of High Resolution Optical and Radar Images." *Journal of the Indian Society of Remote Sensing* 44 (1): 21-29. doi:10.1007/s12524-015-0480-2.

Liu, M. Y., O. Tuzel, S. Ramalingam, and R. Chellappa. 2011. "Entropy Rate Superpixel Segmentation." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2011, Colorado Springs, USA. 2097–2104. doi:10.1109/CVPR.2011.5995323.

Mantero, P., G. Moser, and S. B. Serpico. 2005. "Partially Supervised Classification of Remote Sensing Images through SVM-based Probability Density Estimation." *IEEE Transactions on Geoscience and Remote Sensing* 43 (3): 559–570. doi:10.1109/TGRS.2004.842022.

Mayer, H., S. Hinz, U. Bacher, and E. Baltsavias. 2006. "A Test of Automatic Road Extraction Approaches." *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences* 36 (3): 209–214.

Mena, J. B. 2003. "State of the Art on Automatic Road Extraction for GIS Update: A Novel Classification." *Pattern Recognition Letters* 24 (16): 3037–3058. doi:10.1016/S0167-8655(03)00164-8.

Movaghati, S., A. Moghaddamjoo, and A. Tavakoli. 2010. "Road Extraction from Satellite Images Using Particle Filtering and Extended Kalman Filtering." *IEEE Transactions on Geoscience and Remote Sensing* 48 (7): 2807–2817. doi:10.1109/TGRS.2010.2041783.

Negri, M., P. Gamba, G. Lisini, and F. Tupin. 2006. "Junction-Aware Extraction and Regularization of Urban Road Networks in High-Resolution SAR Images." *IEEE Transactions on Geoscience and Remote Sensing* 44 (10): 2962–2971. doi:10.1109/TGRS.2006.877289.

Ojala, T., M. Pietikäinen, and D. Harwood. 1996. "A Comparative Study of Texture Measures with Classification Based on Featured Distributions." *Pattern Recognition* 29 (1): 51–59. doi:10.1016/0031-3203(95)00067-4.

Ojala, T., M. Pietikainen, and T. Maenpaa. 2002. "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (7): 971–987. doi:10.1109/TPAMI.2002.1017623.

Quackenbush, L. J. 2004. "A Review of Techniques for Extracting Linear Features from Imagery." *Photogrammetric Engineering & Remote Sensing* 70 (12): 1383–1392. doi:10.14358/PERS.70.12.1383.

Shi, W., Z. Miao, and J. Debayle. 2014. "An Integrated Method for Urban Main-Road Centerline Extraction from Optical Remotely Sensed Imagery." *IEEE Transactions on Geoscience and Remote Sensing* 52 (6): 3359–3372. doi:10.1109/TGRS.2013.2272593.

Song, M., and D. Civco. 2004. "Road Extraction Using SVM and Image Segmentation." *Photogrammetric Engineering & Remote Sensing* 70 (12): 1365–1371. doi:10.14358/PERS.70.12.1365.

Trinder, J. C., and Y. Wang. 1998. "Knowledge-Based Road Interpretation in Aerial Images." *International Archives of Photogrammetry and Remote Sensing* 32: 635–640.

Turetken, E., F. Benmansour, B. Andres, H. Pfister, and P. Fua. 2013. "Reconstructing Loopy Curvilinear Structures Using Integer Programming." IEEE Conference on Computer Vision and Pattern Recognition, June 2013, Oregon, USA. 1822–1829. doi:10.1109/CVPR.2013.238.

Unsalan, C., and B. Sirmacek. 2012. "Road Network Detection Using Probabilistic and Graph Theoretical Methods." *IEEE Transactions on Geoscience and Remote Sensing* 50 (11): 4441–4453. doi:10.1109/TGRS.2012.2190078.

Wang, J., J. Song, M. Chen, and Z. Yang. 2015. "Road Network Extraction: A Neural-Dynamic Framework Based on Deep Learning and a Finite State Machine." *International Journal of Remote Sensing* 36 (12): 3144–3169. doi:10.1080/01431161.2015.1054049.

Wegner, J. D., J. Montoya-Zegarra, and K. Schindler. 2013. "A Higher-Order CRF Model for Road Network Extraction." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013, Oregon, USA. 1698–1705. doi:10.1109/CVPR.2013.222

Wegner, J. D., J. Montoya-Zegarra, and K. Schindler. 2015. "Road Networks as Collections of Minimum Cost Paths." *ISPRS Journal of Photogrammetry and Remote Sensing* 108: 128–137. doi:10.1016/j.isprsjprs.2015.07.002.

Zhang, X., and C. Yuan. 2011. "High-Resolution Satellite Scene Dataset" http://dsp.whu.edu.cn/cn/staff/yw/HRSscene.html.

Zivkovic, Z. 2004. "Improved Adaptive Gaussian Mixture Model for Background Subtraction." *Proceedings of the 17th International Conference on Pattern Recognition*, August 2004, Cambridge, UK. 2: 28–31. doi:10.1109/ICPR.2004.1333992