Accurate urban road centerline extraction from VHR imagery via multiscale segmentation and tensor voting

Guangliang Cheng, Feiyun Zhu, Shiming Xiang*, Ying Wang, Chunhong Pan

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, No. 95, Zhongguancun East Road, Haidian District, Beijing, China

A R T I C L E   I N F O

Article history:
Received 5 November 2015
Received in revised form 29 February 2016
Accepted 15 April 2016
Communicated by Feiping Nie
Available online 11 May 2016

Keywords:
Road centerline extraction
Multiscale segmentation
Tensor voting
Non-maximum suppression
Fitting based centerline connection

A B S T R A C T

Accurate road centerline extraction from very-high-resolution (VHR) remote sensing imagery has various applications, such as road map generation and updating etc. There are three shortcomings of existing methods: (a) due to noise and occlusions, most road extraction methods bring in heterogeneous classification results; (b) morphological thinning is a fast and widely used algorithm to extract road centerline, while it produces small spurs; (c) many methods are ineffective to extract centerline around the road intersections. To address the above three issues, we propose a novel road centerline extraction method via three techniques: fused multiscale collaborative representation (FMCR) & graph cuts (GC), tensor voting (TV) & non-maximum suppression (NMS), and fitting based centerline connection. Specifically, FMCR-GC is developed to segment the road region from the image by incorporating multiple features and multiscale fusion. In this way, homogeneous road segmentation can be achieved. Then, TV-NMS is introduced to generate a road centerline network. It not only extracts smooth road centerline, but also connects the discontinuous ones together. Finally, a fitting based algorithm is proposed to overcome the ineffectiveness of existing methods in the road intersections. Extensive experiments on two datasets demonstrate that our method achieves higher quantitative results, as well as more satisfactory visual performances by comparing with state-of-the-art methods.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Accurate road extraction from remote sensing images is an essential preprocessing step for various applications, such as vehicle navigation [1], geographic information system (GIS) update [2] and intelligent transportation system [3] etc. However, it is time-consuming, costly and tedious to manually label the road area from the image. Thus, it is desired to find ways to automatically extract road areas from images. Although, recent researches on road extraction [4–7] have been proposed to address this challenging task, they are far from mature.

Due to recent advances in remote sensors, a large amount of high-resolution images become available, which exhibit more details about the earth surface. Various state-of-the-art approaches [8–12] have been proposed to tackle the object extraction task from the high-resolution remote sensing images. Yao et al. [8] proposed a coarse-to-fine model to detect the airport area from remote sensing images using target-oriented visual saliency and CRF. To extract man-made objects from high spatial resolution remote sensing images, Li et al. [9] introduced a fast level set evolution based approach. An SVM vector machine approach [10] was introduced to detect the cloud area in the remote sensing image. Yuan et al. [11] introduced an image segmentation algorithm for the remote sensing images by combining spectral and textual features. A new building extraction approach was introduced in [12], which fused the knowledge of shadow and urban area information.

Like various extraction tasks above, it is also an urgent task to extract road from high-resolution images for various applications. Although many existing methods (e.g. dynamic programming and snake based algorithm [13], template matching based algorithm [14] and hough transform based algorithm [15]) have achieved satisfactory performance for the low-resolution images. They are failed to extract road region from the high-resolution images. Compared with road extraction from low-resolution images, there are a number of difficulties to extract the road area from high-resolution remote sensing imagery. First, small objects can be observed and the images tend to be affected by noise. Thus the spectral signatures of road become more heterogeneous. Second, complex backgrounds and contextual objects, such as trees, buildings and vehicles on the roads, usually
appear in the high-resolution images. Finally, there are some road-like objects, such as buildings, rivers and parking lots, that may be misclassified as roads.

Most road area extraction methods [16,17] are developed to segment road region in pixel level. Due to noise and occlusions under vehicles and trees, these methods may bring in heterogeneous segmentation. For the task of road centerline extraction, many researchers [17–19] applied the morphological thinning algorithm. Although thinning-based approaches are fast and easy to perform, they may produce small spurs around the centerline. This will largely affect the construction of road network. Later, some regression based methods [20,21] are developed to alleviate this shortcoming. However, they are ineffective to extract the centerline around the road intersections.

To overcome the above shortcomings in the existing methods, we propose a novel three-stage based method to extract smooth and complete road centerline from very-high-resolution (VHR) remote sensing images: (1) Homogeneous road area segmentation; (2) Smooth and accurate road centerline extraction; 3) Centerline connection around the road intersections. The proposed method integrates three techniques, that is, fused multiscale collaborative representation (FMCR) & graph cuts (GC), tensor voting (TV) & non-maximum suppression (NMS), and fitting based centerline connection. Specifically, to obtain a homogeneous road area segmentation result, a FMCR-GC based road segmentation algorithm is proposed, which incorporates multiple features and multiscale fusion. Then, to gain smooth and accurate centerline network, a TV-NMS based centerline extraction algorithm is put forward. Finally, to well connect the centerline in the road intersections, a fitting based centerline connection algorithm is introduced.

The main contributions of the proposed approach are highlighted as follows:

- FMCR and GC are combined to obtain a homogenous road segmentation result. In FMCR, a novel road based feature is firstly proposed, which integrates spectral, structural and contextual road characteristics. This feature is in line with the human perception of road recognition.
- A new TV-NMS based centerline extraction method is introduced to extract the road network. It can not only extract smooth and accurate road centerline, but also connect the nearby discontinuous centerline due to unconnected regions in the segmentation result.
- To overcome the ineffectiveness of the existing centerline extraction methods in the intersection areas, a fitting based centerline connection algorithm is proposed to complete the unconnected centerline around the road intersections.
- A new and challenging road centerline extraction dataset is publicly available for further studies. It contains 30 VHR remote sensing images together with the corresponding centerline reference maps.

The remainder of this paper is arranged as follows. The related road extraction work is systematically reviewed in Section 2. In Section 3, the details of the proposed road area extraction and centerline extraction method are introduced. Experimental evaluations as well as detailed comparisons between our method and state-of-the-art methods are provided in Section 4. Finally, the conclusions will be outlined in Section 5.

2. Previous work

For VHR images, according to the extracted road results, the existing road extraction approaches can be classified into two classes: (1) road area extraction methods, (2) road centerline extraction methods. Road area extraction methods mainly depend on image classification and segmentation. Zhang et al. [16] proposed an integrated approach that combines k-means, fuzzy logical classifier and shape descriptors of angular texture signature. It can separate the roads from the parking lots that have been misclassified as roads. A new method for extracting roads based on advanced directional morphological operators was proposed in [22], in which Path Openings and Path Closings were introduced to extract structural pixel information. Yuan et al. [23] presented an automatic road extraction method for remote sensing images based on locally excitatory globally inhibitory oscillator networks. A multistage framework to extract roads from the high-resolution multispectral satellite image was introduced by Das et al. [6]. In this method, probabilistic support vector machines and salient features were used.

Recently, a convolutional neural network based algorithm was introduced to learn features from noisy labels in [24], in which the training labels were generated by applying an algorithm trained on a general image dataset. Mnih et al. [25] proposed a deep neural network method to extract urban road network from high-resolution images. In this method, unsupervised pretraining and supervised post-processing were introduced to improve the performance of the road detector substantially. A higher-order conditional random field (CRF) model was applied for road network extraction by Wegner et al. [26], in which the road prior was represented by higher-order cliques that connect sets of superpixels along straight line segments, and the road likelihood was amplified for thin chains of superpixels. Most popular and successful road centerline extraction methods consist of one or two processing steps: (1) classification and (2) centerline extraction. Zhu et al. [27] proposed a road centerline extraction approach, which was based on the binary-greyscale mathematical morphology and a line segment match algorithm. An accurate centerline detection and line width estimation method via radon transform was introduced in [28]. Gamba et al. [29] extracted the road centerline with the help of adaptive filtering and perceptual grouping. A novel road centerline extraction method was proposed in [17] by integrating multiscale spectral-structural features, support vector machines (SVMs) and morphological thinning algorithm. In recent years, Poullis and You [30] proposed a novel vision-based system for automatic road centerline extraction. This system integrated Gabor filtering, tensor voting and global optimization using graph-cuts into a unified framework. A novel system [31] was introduced to extract road centerline from high resolution images, in which probabilistic road center detection, road shape extraction, and graph-theory-based road network were utilized. Chaudhuri et al. [18] presented a semi-automatic road centerline extraction algorithm. To achieve this, directional morphological enhancement and directional segmentation were used to extract the road area, then thinning method was applied to extract the road network. An automatic road centerline extraction method was introduced by Miao et al. [21], in which potential road segments were obtained based on shape features and spectral features, followed by multivariate adaptive regression splines to extract road centerline.

Shi et al. [19] presented a two-step method for urban road extraction. First, spectral-spatial classification and shape features were employed to obtain road segmentation results. Then morphological thinning algorithm was used to extract centerline. An integrated urban main-road centerline detection method was introduced in [20]. Road extraction result was obtained by fusing the result of spectral-spatial classification and local Geary’s C method. Then, to extract smooth road centerline, local linear kernel smoothing regression algorithm was introduced. It relieves the shortcoming of thinning algorithm, while it can’t
work well in the intersection areas. Hu et al. [32] proposed a three-step road centerline extraction method. First, adaptive mean shift was used to detect road center points. Then, the salient linear features were enhanced via stick tensor voting. Finally, a weighted Hough transform was applied to extract the arc primitives of the road centerline. Sironi et al. [33] provided a new idea by defining the road centerline extraction task as a regression problem, which learns a scale space distance transform from the high-resolution image.

3. Road segmentation and centerline extraction

In this section, we propose a novel urban road centerline extraction method, which consists of three stages: road area segmentation, road centerline extraction, and road centerline connection. The flow chart of the proposed method is shown in Fig. 1. Specifically, in the 1st stage, FMCR and GC are combined to obtain an accurate road area segmentation result. Then, to eliminate those road-like segments from the segmentation result, a novel shape-based elimination algorithm is introduced. To obtain a smooth and accurate road centerline network, the 2nd stage introduces the tensor voting (TV) and non-maximum suppression (NMS) based road centerline extraction method. In the 3rd stage, to obtain a complete road centerline network, a fitting based centerline connection algorithm is utilized around the road intersections.

3.1. FMCR-GC based road area segmentation

In this subsection, we propose two techniques, i.e. fused multiscale collaborative representation and graph cuts (FMCR-GC), to classify the image into road class and non-road class. Specifically, there are three stages. First, we present a novel object-based multiple features, which integrate spectral, structural and contextual (SSC) features. Second, a FMCR-GC based framework is proposed to obtain an accurate road area segmentation result. Third, to remove the non-road objects (i.e. houses and parking lots) from the road class, a new shape-based elimination algorithm is introduced.

3.1.1. Object-based SSC feature extraction

In the VHR remote sensing image, the road segments tend to be elongated and under the occlusions of cars and trees etc. Based on this observation, we employ an object-oriented SSC feature to achieve the following two goals: (1) to reduce the side influence of occlusions and the spectral variations; (2) to extract the geometric characteristics of road segments. The basic assumption behind the object-based algorithm is that spatially adjacent pixels are prone to be grouped into spectrally homogeneous objects. Specifically, the simple linear iterative clustering (SLIC) [34] is employed to obtain the multiscale over-segmentation for VHR remote sensing images. SLIC adapts a local k-means clustering approach according to their color similarity and proximity in the image plane. There are three advantages to apply the SLIC for the over-segmentation task. SLIC not only performs more efficiently, but also achieves more satisfactory over-segmentation results compared with the state-of-the-art methods, for example graph-based algorithms [35,36] and gradient-ascent based algorithms [37,38] etc. Besides, there is only one parameter specifying the number of superpixels, which makes it extremely easy to use.

To fully exploit multiscale information from the image, we obtain multiscale segmentations with different superpixel numbers. Intuitively, superpixel number controls the segment size. A large superpixel number will averagely result in a small object size, and vice versa. As Fig. 1 shows, we get three different over-segmentation results for each image. For each over-segmentation result, we define the spectral attribute of an object as the average spectral value within this object. For the i-th object, the spectral attribute (SA) [17] is given by

\[
SA_i = \left\{ SA_i^r, SA_i^g, SA_i^b \right\}
\]

\[
SA_i^m = \frac{1}{N_i} \sum_{p \in \partial \theta_i} SP^m(p),
\]

where \(p\) is a pixel in the i-th object (1 \( \leq p \leq N_i\)) and \(SP^m(p)\) is the spectral value of pixel \(p\) in band \(m\); the superscripts \(r, g\) and \(b\) represent the RGB channels, respectively. To exploit the structural attributes for each object, we utilize the shape index (SI) [17,39]
and aspect ratio (AR) as follows:

\[
SL_i = \frac{\text{Per}_i}{4\sqrt{\text{Area}_i}},
\]

\[
AR_i = \frac{\text{Length}_i}{\text{Width}_i},
\]

where Per\(_i\), and Area\(_i\) denote the perimeter and area of the \(i\)-th object, respectively; Length\(_i\), and Width\(_i\), are the length and width of the minimum bounding box surrounding the \(i\)-th object.

The SI measures the smoothness of the object border and the AR describes the length-width ratio. Intuitively road regions tend to be elongated ribbons with large perimeters and small areas. Therefore, the SI and AR of road regions tend to be larger than the ones of non-road regions.

The aforementioned spectral and structural attributes for each object can be combined as a hybrid feature (HF), which is defined as

\[
HF^i_s = (SA^i_s, SL^i_s, AR^i_s),
\]

where \(HF^i_s\) is the hybrid feature for the \(i\)-th object at the \(s\)-th scale.

To enhance the discriminative power of each object, its spatially adjacent objects should be considered. In the proposed method, co-occurrence among objects is used as the high-level contextual feature. Intuitively, there are at least two road segments (i.e. left and right, or up and down) in the neighborhood of one road segment. This is the prior information for humans to recognize the road areas under the occlusions of trees and shadows. Our SSC feature employs this recognition scheme.

Fig. 2 illustrates how to extract the SSC feature for one center segment. There are three steps. First, we find out the neighboring segments (i.e. the blue border regions) for the center segment (i.e. the dark region). Second, we rank the neighboring segments according to their similarity values to the center segment, where we use the Euclidean distance as the similarity measure. Third, the HFs of both center segment and its top two neighboring segments are concatenated as the contextual feature for the center segment.

3.1.2. FMCR-GC based road area segmentation

In this subsection, FMCR and GC are combined to obtain an accurate road area segmentation result. Specifically, SSC feature is utilized by FMCR to gain the road probability of each segment. Then, to enhance the label consistency among the neighboring pixels, GC is introduced to incorporate the spatial information.

**FMCR based road likelihood:** Recently, sparse representation classification (SRC) [40] has been proposed for face recognition. SRC represents a testing sample by a sparse linear combination of training samples with \(\ell_1\)-norm constraint. In the remote sensing imagery, Chen et al. [41] applied a sparse framework for the hyperspectral image classification. A similar approach to SRC is the collaborative representation classification (CRC) [42–44]. CRC also represents a testing sample with the linear combination of training samples. However, contrary to the \(\ell_1\)-norm regularization in SRC, CRC employs an \(\ell_2\)-norm regularization. It provides competitive accuracy while with significantly lower computational complexity [44].

Consider a dataset with \(n\) training samples \(X = \{x_i\}_{i=1}^n\), where \(x_i \in \mathbb{R}^d\) and \(d\) is the feature dimensionality. Let \(y_i \in \{1, 2, \ldots, C\}\) be the class label, where \(C\) is the number of classes. For the road extraction task, we set \(C = 2\). Let \(n_1\) and \(n_2\) be the number of training samples for road class and non-road class, respectively.

For a testing sample \(x\), its corresponding representation coefficient \(\alpha\), based on all the training samples, could be obtained via

\[
\alpha^* = \arg\min_{\alpha} \|x - X\alpha\|_2^2 + \lambda \|I_{X_s} \alpha\|_2^2,
\]

where \(I_{X_s}\) is a biasing Tikhonov matrix between the test sample and all the training samples; \(\lambda\) is a global regularization parameter that balances the representation loss and the regularization term. We denote \(\alpha^*\) as the optimal representation vector of \(\alpha\) with \(n \times 1\) elements. Specifically, \(I_{X_s} \in \mathbb{R}^{n \times n}\) is designed in the following form:

\[
I_{X_s} = \begin{bmatrix}
\|x - x_1\|_2 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \|x - x_n\|_2
\end{bmatrix}.
\]

\(I_{X_s}\) is a diagonal matrix, whose diagonal value measures the discrepancy between a certain training sample and the testing sample. Intuitively, if the testing sample belongs to the road class, the discrepancies between the testing sample and those road-class training samples are small, while inversely the discrepancies are large between the testing sample and the non-road-class ones. Given a large regularization parameter \(\lambda\), to achieve a minimum objective in Eq. (5), the road-class testing sample should be more likely to be represented by road-class samples rather than non-road-class samples, thus \(\alpha^*\) tend to be sparse. The representation coefficient \(\alpha^*\) can be estimated in a closed-form solution as

\[
\alpha^* = \left(XX + \lambda I_{X_s} I_{X_s}ight)^{-1}X^T\hat{x}.
\]

After that, the training samples \(X\) are partitioned as road-class samples \(X_1\) and non-road-class samples \(X_2\), and the coefficient vector \(\alpha^*\) are partitioned as \(\alpha_1^*\) and \(\alpha_2^*\) accordingly. The residual between the approximation and the testing sample can be defined as

\[
R_i(\hat{x}) = \|X_1\alpha_1^* - \hat{x}\|_2^2.
\]

In this research, we get the class likelihood rather than the class label of each sample. Therefore, we define the road-class likelihood as

\[
p_r(i) = \frac{R_i(\hat{x})}{R_i(\hat{x}) + R_2(\hat{x})},
\]

and the non-road-class likelihood as \(p_{nr}(i) = 1 - p_r(i)\).

After obtaining the road likelihood for each object at a certain scale, all the pixels in the same object are given an identity likelihood value as the object. Then, we fuse all road likelihood maps from the three different scales into an integrated one. Finally, the road likelihood of each pixel \(x\) can be defined as

\[
p_r(x) = \max_{i \in \{1,2,3\}} p_r^i(x).
\]

We also tried other fusion strategies, such as mean, median and min. Their results are inferior to the max rule. Thus we use the max-fusion rule in our experiments.
GC based road area segmentation: In the remote sensing image, some road areas are under conditions of spectral variability as well as occlusions of trees and cars. To relieve the side effect of these conditions and get a coherence road extraction result, graph cuts (GC) algorithm is used for the road extraction task.

Given an image \( I \), the GC algorithm constructs an undirected graph \( G = (V, E) \), where \( V \) denotes the pixel set in the image and \( E \) is the set of undirected graph edges between neighboring pixels [45]. For the road extraction task, we define the label “1” for the road class and “0” for the non-road class. GC tries to minimize the following objective

\[
C(L) = C_r(L) + \alpha C_b(L),
\]

where \( L \) is a labeling set, \( C_r(L) \) and \( C_b(L) \) denote the regional term and boundary term, respectively; \( \alpha \) is a trade-off parameter balancing the two terms. In the road extraction problem, the regional term \( C_r(L) \) defines the individual penalty to classify each pixel into the road class. The boundary term \( C_b(L) \) describes the coherence between spatially neighboring pixels.

We have obtained the road likelihood map via aforementioned FMCR. The regional term is defined as

\[
C_r(L) = \sum_{i \in I} - \log (p_r(x_i)),
\]

where \( p_r(x_i) \) is the road likelihood probability of pixel \( x_i \). Intuitively, the spatially adjacent pixels tend to belong to the same class, thus the boundary term is defined to measure the label discontinuity among neighboring pixels, which is defined as

\[
C_b(L) = \sum_{i,j \in \mathcal{N}} m(l_{x_i}, l_{x_j}) \cdot \frac{1}{||x_i - x_j||^2 + \epsilon},
\]

where \( \mathcal{N} \) denotes a standard 8-neighborhood system, which contains all the unordered pairs of neighboring pixels. \( m(l_{x_i}, l_{x_j}) \) is the distance metric between the label \( l_{x_i} \) and \( l_{x_j} \), i.e. if \( l_{x_i} \) and \( l_{y_j} \) have different labels, we denote \( m(l_{x_i}, l_{y_j}) = 1 \), otherwise, we define it as \( 0 \). \( x_i \) and \( x_j \) are the RGB feature vectors of pixel \( x_i \) and \( x_j \); \( || \cdot ||_2 \) denotes the \( \ell_2 \) norm. To avoid a zero divisor, we add a small value \( \epsilon \) (typically \( \epsilon = 0.001 \)) to the denominator.

For binary labeling problem, the objective function in Eq. (11) can achieve the optimal solution via the mincut/maxflow algorithm in polynomial time [46]. As Fig. 1 shows, coherence road segmentation result can be obtained after the GC algorithm.

However, there are some road-like objects. Some strategies should be taken to remove those road-like objects.

### 3.1.3. Shape based elimination algorithm

After the process of the GC based road area segmentation, there still remains some road-like objects (see Fig. 3(b)). To address this issue, the road shape features should be employed to distinguish the potential road segments and road-like segments.

In general, roads have the following characteristics: (1) roads are connected and do not have small areas; In our experiments, we regard all the segments whose pixel number is less than a threshold \( K \) (i.e. 300) as non-road class. (2) Roads are elongated with long length and short width. Linear feature index (LFI) is defined to evaluate this characteristic, which can be denoted as

\[
LFI = \frac{L_{box}}{W_{box}},
\]

where \( L_{box} \) and \( W_{box} \) are the length and width of the bounding box (the red rectangle in Fig. 3(g)), respectively. Intuitively, road segments have large LFI values. However, the bounding box based algorithm may be failed as shown in Fig. 3(c) and (f), some road segments (the sketch map in Fig. 3(g)) may have small LFI values.

To overcome the shortcoming above, we propose a new ellipse-based road elimination algorithm (see the green ellipse in Fig. 3(g)). We use an ellipse to match each remaining segment after the area constraint. We define the new LFI as

\[
LFI_e = \frac{L_e}{W_e},
\]

where \( L_e \) and \( W_e \) are the major axis and minor axis of ellipse, respectively. In our experiments, we set the \( LFI_e \) as 3. Comparing Fig. 3(c)-(f), the real road segment is removed in Fig. 3(c) and (f) via bounding box constraint, while our ellipse-based elimination algorithm keeps it remained in Fig. 3(d) and (e). Thus, it demonstrates that the proposed algorithm is more effective than the bounding box based algorithm.

### 3.2. TV-NMS based road centerline extraction

After the above processes, there are still two issues with the gained road segmentation results. Due to the image noise and occlusions caused by trees and cars, the extracted road area always

---

**Fig. 3.** Comparison of elimination results between our ellipse-based algorithm and the bounding box based algorithm. (a) Original image. (b) Final segmentation result. (c) Elimination result via the bounding box based algorithm. (d) Elimination result by ellipse-based algorithm. (e) and (f) are the close ups of the rectangle region in (d) and (c) respectively. (g) The illustration of the ellipse-based algorithm and the bounding box based algorithm. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)
have an unsatisfactory property that is, discontinuities. Besides, for some real problems, we only consider whether there exists a road, and do not care about the width of the road. To relieve the side effect of discontinuity and to be more suitable for application, road centerline extraction has been an active research. The commonly used road centerline extraction algorithm is morphological thinning, which is fast and easy to implement. However, thinning based centerline extraction algorithm always produces many spurs and brings in many false positives, which reduce the smoothness and correctness of road network. To solve this problem, TV-NMS is introduced to extract road centerline, which can produce smooth centerline and complete the discontinuous road centerline.

Tensor voting (TV) was originally proposed by Guy and Medioni [47], and later presented a lot of improved versions [48,49]. It is a technique for perceptual grouping and robust extraction of lines and curves from images. TV aims at making the local feature measurements more robust by taking the neighboring measurements into consideration. In 2-D tensor voting, local image feature are encoded into a tensor and can be decomposed as a linear combination of stick tensor and plate tensor. Any 2-D tensor can be decomposed into the linear combination of stick tensor and plate tensor. For the scale factor can be set according to the road width, which will be discussed later.

To get the consistent road centerline, Non-Maximum Suppression (NMS) is applied on the resulting stick saliency map (see Fig. 5(b)). In the saliency map, the real road centerline positions tend to have local maximum values. NMS marks all the pixels whose intensity is not the maximal as zero within a certain local neighborhood. In this paper, the local neighborhood is a linear window [52] with certain length (i.e. 20 pixels), which is perpendicular to the local orientation. The NMS keeps only the locations that correspond to a local maximum in the line window, and set the locations that are not the local maximum as zero. In the experiments, we estimate the orientation using the Hessian matrix. As Fig. 5 shows, the TV-NMS method can extract smooth road centerline, while it is ineffective around the road intersection. To tackle this issue, some other measurements should be taken to complete the road centerline network.

3.3. Fitting based road centerline connection

TV-NMS based centerline extraction method can extract single-pixel-width road centerline well. However, it cannot extract the centerline around the road intersections properly (see Fig. 5(d)). To overcome this shortcoming, a new fitting based centerline connection method is proposed.

We get the junction map after the TV algorithm. As Fig. 6(a) shows, for each junction point (the red circle), we search its local area within a certain distance defined by the radius R. For example, there are three line segments in the radius areas in Fig. 6(a). For each segment in the radius, we calculate its middle location of all the centerline pixels, termed middle point. Then, we link the middle point and the corresponding center point of junction area (see the purple lines in Fig. 6(b)). In this subsection, we employ the linear fitting algorithm to connect the discontinuous road centerline around the intersection area. It is behind the assumption that road is straight at the intersection area, which is satisfied in most cases. As Fig. 5(e) shows our fitting based method can link the centerline well around the road intersection area.

4. Experiments and evaluation

In this paper, the definition of the “VHR” refers to the image with spatial resolution of 1–2 m per pixel. The road width in VHR...
image is 8–15 pixels. The corresponding road reference map is manually labeled.

To verify the effectiveness of the proposed method, extensive experiments, on the road centerline extraction from VHR remote sensing images, have been conducted on two datasets. The proposed method is also compared with other state-of-the-art methods.

4.1. Datasets description

This section introduces the information of two VHR image datasets for the road extraction experiments. It should be noted that few VHR urban road datasets are publicly available. Thus we collected 30 VHR images from Google Earth and labeled the road reference map and centerline reference map by a hand drawing method. This dataset will be publicly available for research.

Data #1: This dataset contains 30 VHR urban road images with a spatial resolution of 1.2 m per pixel. The road width is about 12–15 pixels. There are at least 600 × 600 pixels in each image. Most of the images are under the conditions of complex backgrounds and occlusions due to trees and cars. In addition, there are road-like objects (i.e. houses and parking lots) in the images. All these factors make it very challenging to obtain a satisfying road extraction results.

Data #2: This dataset is a publicly available road dataset provided by Turetken et al. [53]. It contains 14 images. The road width in this dataset is about 8–10 pixels. Some images are under the conditions of complex backgrounds and occlusions of trees. We manually labeled the centerline reference map of each image in this dataset.

4.2. Compared algorithms

To verify the performance, the proposed method is compared with four related methods. The main information of all these methods are summarized as follows:

1 Huang’s method (Huang): Huang and Zhang [17] introduce a novel road detection method based on multiscale structural features and support vector machine. Then, morphological thinning algorithm is used to extract the road centerline.

2 Miao’s method (Miao): Miao et al. [21] present a road extraction method based on spectral and shape features. Then, to overcome the shortcomings of morphological thinning algorithm, multivariate adaptive regression splines is introduced to get the smooth and accurate road centerline.

3 Shi’s method a (Shia): This method [19] integrates the spectral-spatial classification, shape features and local Geary’s C to extract road network. Then a morphological thinning algorithm is applied to extract centerline.

4 Shi’s method b (Shib): To get the road network, Shi et al. [20] fuse the spectral-spatial classification and homogeneous property based classification. After that, a local linear kernel smoothing regression algorithm is introduced to extract the centerline. It is a state-of-the-art method for road centerline extraction.

5 Proposed method with three scales (Proposed3): As Fig. 1 shows, the image is oversegmented with three different numbers of superpixels, such as 8000, 10,000 and 12,000. Then the following steps in Fig. 1 are performed to get the road centerline.

6 Proposed method with one scale (Proposed1): To investigate the effectiveness of multi-scale information fusion, the image is oversegmented with only one certain number of superpixels, for example 10,000. Other steps are performed as Proposed3 states.

It should be noted that the codes of the other four comparing methods are not readily available on the Internet. We implement
them according to the details described in their papers. Our implemented codes can achieve the same performance as the original paper presents. In the following experiments, we adjust the parameters to gain the best performance of the four methods for fair comparison. Among the four comparing methods, regression based centerline extraction algorithm is used in Miao and Shi\(^a\), thus they can obtain smooth centerline network. Meanwhile, Huang and Shi\(^b\) utilize the morphological thinning algorithm to extract road centerline, which will produce some small spurs.

### 4.3. Evaluation metrics

To comprehensively evaluate the proposed method, comparative experiments are conducted in two aspects: the comparison of road area extraction result and the comparison of the road centerline extraction result. Completeness (COM), Correctness (COR) and Quality (Q) [54] are employed as evaluation metrics to measure the both experiments as follows:

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

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

\[
Q = \frac{TP}{TP + FN + FP}
\]

where TP, FP and FN are the true positive, false positive and false negative, respectively.

For the evaluation of road area extraction result, we compare the extracted road map with the reference map in the corresponding locations. Due to the deviation between the manually labeled centerline and the real centerline, the comparison of centerline result is carried out by matching the extracted centerline map to the reference map using the so-called “buffer method”, in which every proportion of the network within a given buffer width \(\rho\) from the other is considered as matched [55]. That is, a predicted centerline point is considered to be a true positive if it is at most \(\rho\) distant from a reference centerline point.

### 4.4. Parameter setting

There are mainly two parameters in the proposed method, i.e. the balancing factor \(\alpha\) in GC and the scale factor \(\sigma\) in TV. In the experiments, we set other parameters as follows: The three different numbers of superpixels are set as 8000, 10,000 and 12,000 in Proposed\(^3\), while in Proposed\(^1\) we choose the number of superpixels as 10,000 in the case of no other special instructions. In FMCR, we randomly choose 50 positive samples and 50 negative samples from the ground truth. To obtain a sparse coefficients \(\alpha^s\), we set the weighting factor \(\lambda = 10\). We set the window size of neighborhood as 2\(r\) in NMS. In the fitting based centerline connection method, the width of searching window \(R\) is set to be \(\sigma\). For Data \#1, the road width is about 12–15 pixels, thus we choose 3 as the buffer width. While for Data \#2 we set the buffer width \(\rho\) as 2 pixels. Actually, in the experiments, we find that satisfactory results can be obtained for most images when setting \(\alpha = 0.8\) in GC. Thus we keep this parameter fixed in the following experiments. For the scale factor \(\sigma\) in TV, we get the optimal value for each image via cross validation, which will be discussed later.

### 4.5. Comparison of road area extraction results

In order to illustrate an intuitive comparison of different methods on road area extraction, we display the results by comparing the extracted map with the reference map in Fig. 7. Shi\(^b\) and Shi\(^a\) achieve better performance than other two comparing methods, while both of them are inefficient to extract the road under the occlusions of cars and trees. Huang’s method can extract the complete road network well. However, it brings in more false positives (the red areas in Fig. 7(b)). For Data \#1, the road width is about 12–15 pixels, thus we choose 3 as the buffer width. While for Data \#2 we set the buffer width \(\rho\) as 2 pixels. Actually, in the experiments, we find that satisfactory results can be obtained for most images when setting \(\alpha = 0.8\) in GC. Thus we keep this parameter fixed in the following experiments. For the scale factor \(\sigma\) in TV, we get the optimal value for each image via cross validation, which will be discussed later.

### 4.6. Comparison of road centerline result on Data \#1

We carried out experiments on all the 30 images in Data \#1. Due to space limit, we select 4 of them to illustrate the comparing

![Fig. 7. Visual comparisons of road area extraction results with four comparing methods. There are two rows and six columns of subfigures. Each row shows results on one dataset. From left to right: (a) original image, (b) result of Huang [17], (c) result of Miao [21], (d) result of Shi\(^a\) [19], (e) result of Shi\(^b\) [20], (f) result of Proposed\(^1\). From (b) to (f), true positive (TP) is marked in green, false positive (FP) in red, false negative (FN) in blue. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)](image)
results among the aforementioned methods in Fig. 8. By comparing the extracted road centerline map with its corresponding reference map, we can have following conclusions: (1) though morphological thinning based road centerline extraction methods (e.g. Huang’s method and Shi’s method) are fast and easy to perform, they produce small spurs, which reduce the smoothness and correctness of road network. (2) The regression based road centerline extraction methods (e.g. Miao’s method

Table 1
Quantitative comparisons with state-of-the-art methods in two datasets, where the red values marked in bold are the best and the bold blue values are the second best. (For interpretation of the references to color in this table caption, the reader is referred to the web version of this paper.)

<table>
<thead>
<tr>
<th>Methods</th>
<th>Image1</th>
<th></th>
<th>Image2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COM</td>
<td>COR</td>
<td>Q</td>
<td>COM</td>
</tr>
<tr>
<td>Huang</td>
<td>0.9479</td>
<td>0.7954</td>
<td>0.7621</td>
<td>0.9339</td>
</tr>
<tr>
<td>Miao</td>
<td>0.7690</td>
<td><strong>0.9875</strong></td>
<td>0.7203</td>
<td>0.2612</td>
</tr>
<tr>
<td>Shi\textsuperscript{a}</td>
<td>0.8722</td>
<td>0.9850</td>
<td><strong>0.8656</strong></td>
<td>0.8237</td>
</tr>
<tr>
<td>Shi\textsuperscript{b}</td>
<td>0.8760</td>
<td>0.9846</td>
<td>0.8642</td>
<td>0.8596</td>
</tr>
<tr>
<td>Proposed\textsuperscript{1}</td>
<td><strong>0.9264</strong></td>
<td><strong>0.9997</strong></td>
<td><strong>0.9261</strong></td>
<td><strong>0.9223</strong></td>
</tr>
</tbody>
</table>

Fig. 8. Visual comparisons of centerline extraction results on Data #1. There are four rows and eight columns of subfigures. Each row shows the results of one image. From left to right: (a) original image, (b) result of Huang [17], (c) result of Miao [21], (d) result of Shi\textsuperscript{a} [19], (e) result of Shi\textsuperscript{b} [20], (f) result of Proposed\textsuperscript{1}, (g) result of Proposed\textsuperscript{1}, (h) the reference map.
and Shi’s method) can extract relatively smooth centerline well. However, these methods can’t link the centerline well in the intersection areas (see subfigure (c) and (e) of Image5 in Fig. 8). (3) There are some false positives in Miao’s method, this is because some forest areas and bare soil are also tend to be homogenous regions in VHR remote sensing images, thus it is hard to choose an appropriate threshold to distinguish them with road areas. (4) Our proposed methods (Proposed1 and Proposed3) achieve more smooth and continuous road centerline than other comparing methods, and both of them work well in the junction areas. Besides, the centerline result of Proposed3 is more similar to the reference map than Proposed1, which demonstrates that multi-scale information fusion is suitable for the road centerline extraction task.

Quantitative comparisons of road centerline extraction result among different methods are illustrated in Table 2. As we shall see, in terms of COR, the proposed method (Proposed3) achieves the second best results after Shi’s method, while in terms of COM, the proposed method obtains bigger values than Shi’s method by a large margin. Thus our proposed method achieves relatively

### Table 2

Quantitative comparisons with state-of-the-art methods in Data #1, where the red values marked in bold are the best and the bold blue values are the second best.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Image3 COR</th>
<th>Image3 Q</th>
<th>Image4 COR</th>
<th>Image4 Q</th>
<th>Image5 COR</th>
<th>Image5 Q</th>
<th>Image6 COR</th>
<th>Image6 Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huang</td>
<td>0.970</td>
<td>0.627</td>
<td>0.963</td>
<td>0.802</td>
<td>0.765</td>
<td>0.906</td>
<td>0.783</td>
<td>0.724</td>
</tr>
<tr>
<td>Miao</td>
<td>0.966</td>
<td>0.722</td>
<td>0.985</td>
<td>0.820</td>
<td>0.810</td>
<td>0.983</td>
<td>0.791</td>
<td>0.780</td>
</tr>
<tr>
<td>Shi*</td>
<td>0.960</td>
<td>0.833</td>
<td>0.815</td>
<td>0.846</td>
<td>0.810</td>
<td>0.983</td>
<td>0.948</td>
<td>0.917</td>
</tr>
<tr>
<td>Shi*</td>
<td>0.920</td>
<td>0.903</td>
<td>0.971</td>
<td>0.948</td>
<td>0.900</td>
<td>0.983</td>
<td>0.948</td>
<td>0.934</td>
</tr>
<tr>
<td>Proposed1</td>
<td>0.969</td>
<td>0.911</td>
<td>0.979</td>
<td>0.933</td>
<td>0.895</td>
<td>0.922</td>
<td>0.950</td>
<td>0.897</td>
</tr>
<tr>
<td>Proposed3</td>
<td>0.965</td>
<td>0.943</td>
<td>0.974</td>
<td>0.917</td>
<td>0.876</td>
<td>0.957</td>
<td>0.950</td>
<td>0.910</td>
</tr>
</tbody>
</table>

Fig. 9. Visual comparisons of centerline extraction results on Data #2. There are eight rows and four columns of subfigures. Each row shows the results of one image. From left to right: (a) original image, (b) result of Huang [17], (c) result of Miao [21], (d) result of Shi* [19], (e) result of Shi* [20], (f) result of Proposed1, (g) result of Proposed3, (b) the reference map.
higher Q values, which is an overall evaluation index. The thinning based methods (i.e. Huang’s method and Shi’s method) and Miao’s method give relatively low values of COR and Q due to the appearance of false positives and small spurs. Proposed\(^1\) obtains the second best performance among all the methods, while it is inferior to the Proposed\(^3\) almost in each comparing items.

### 4.7. Comparison of road centerline result on Data #2

Visual comparisons of centerline extraction results on Data #2 among different methods are displayed in Fig. 9. We find that there are large areas of dense forests and grasslands in those images, which is hard to extract accurate road network from them for Miao’s method. Thus the centerline result of Miao’s method is incomplete (see Fig. 9(c)). Huang’s method and Shi’s method can extract relatively complete road network, while their methods produce small spurs due to the algorithm of centerline extraction. Shi’s method and Proposed\(^1\) achieve better performance than other methods (e.g. Huang’s method, Miao’s method and Shi’s method), while both of them are inferior to the Proposed\(^3\) in term of completeness and smoothness.

Table 3 presents the quantitative comparisons on Data #2 for different methods. As we shall see, Miao’s method achieves low values both in terms of COM and Q, because the road network is hard to extract under the conditions of large areas of forests and grasslands. Huang’s method and Shi’s method gain relatively large values in COM and small values in COR, which lead to obtain small Q values. Shi’s method and Proposed\(^1\) achieve similar performance in average among all the four images, while Proposed\(^2\) achieves higher values than both of them in terms of all the three metrics (e.g. COM, COR and Q). Specifically, the average Q value among all the images of the Proposed\(^3\) is 3% higher than the second best method.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Image7</th>
<th>Image8</th>
<th>Image9</th>
<th>Image10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COM</td>
<td>COR</td>
<td>Q</td>
<td>COM</td>
</tr>
<tr>
<td>Huang</td>
<td>0.804</td>
<td>0.773</td>
<td>0.853</td>
<td>0.872</td>
</tr>
<tr>
<td>Miao</td>
<td>0.641</td>
<td>0.557</td>
<td>0.612</td>
<td>0.641</td>
</tr>
<tr>
<td>Shi(^b)</td>
<td>0.889</td>
<td>0.839</td>
<td>0.752</td>
<td>0.921</td>
</tr>
<tr>
<td>Shi(^d)</td>
<td>0.923</td>
<td>0.977</td>
<td>0.902</td>
<td>0.857</td>
</tr>
<tr>
<td>Proposed(^1)</td>
<td>0.927</td>
<td>0.930</td>
<td>0.866</td>
<td>0.950</td>
</tr>
<tr>
<td>Proposed(^3)</td>
<td>0.964</td>
<td>0.955</td>
<td>0.922</td>
<td>0.940</td>
</tr>
</tbody>
</table>

#### Fig. 10. Quantitative comparisons of centerline extraction results with different scale parameters. COM, COR and Q are calculated when the buffer width \(\rho = 2\). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

continuous regions (see Fig. 11(c)–(e)). With the increase of scale parameter \(\sigma\) within a certain range, TV-NMS can extract centerline for the connected areas and link the centerline for the disconnected areas (see Fig. 11(f)–(j)). When \(\sigma\) is greater than a certain value, although TV-NMS can connect the discontinuous centerline well, it also brings in some false positives (see Fig. 11(k) and (l)). Those false positives lead to wrong navigation information in practical applications. Therefore, to get a smooth and complete road network, a suitable value for scale parameter should be carefully selected.

To give an intuitive comparison, the quantitative performances with different scale parameters are summarized in Fig. 10. In this figure, COM, COR and Q are calculated with the buffer width \(\rho = 2\). When \(\sigma\) is less than 12 pixels, COM, COR and Q are all increase along with the growth of \(\sigma\). Then with the increase of \(\sigma\), COR declines gradually, the overall metric (Q) reduces even though COM remains stable. We find that satisfactory results can be obtained with \(\sigma\) in a large range (e.g. 15–24 pixels). Specifically, the highest Q values are obtained when \(\sigma\) is set to be 15 or 18 pixels. It should be noted that the road width \(w\) of the test image is about 12 pixels, thus it is suitable to set the scale parameter as \(\sigma = 1.5\ w\).

### 4.9. Time comparison

The average running time among different methods in the classification stage and the centerline extraction stage are illustrated in Table 4. All the experiments are conducted on a
computer with Intel Core i5-3470 3.20 GHz CPU and 8GB RAM using Matlab 2013. As can be seen from the table, for an image of 1048x728, Huang’s method and the Proposed3 method take more time in the classification stage. For the centerline extraction time, morphological thinning methods (Huang’s method and Shi’s method) take least time, while they produce short spurs around the centerline, thus reducing the smoothness and correctness. The centerline extraction time of regression based methods (Miao’s method and Shi’s method) are longer than the proposed method. Though our multiscale version method (Proposed3) takes a bit longer overall time than other state-of-the-art methods (Miao’s method and Shi’s method), our one scale version (Proposed1) takes less time while gain better performance (see Tables 2 and 3) than them.

5. Conclusions

In this paper, we propose an accurate road centerline extraction method for VHR remote sensing images. The proposed method contains three stages: homogeneous road area segmentation, smooth and accurate road centerline extraction, and centerline connection around the road intersections. Specifically, in the 1st stage, to obtain an accurate road area extraction result, fused multiscale collaborative representation and graph cuts are introduced to incorporating the multiple features and spatial information. In the 2nd stage, to overcome the shortcomings of morphological thinning algorithm, tensor voting and non-maximum suppression algorithm are utilized to extract smooth and accurate road centerline. Finally, in the 3rd stage, to tackle the ineffectiveness of the existing methods around the road intersections, a fitting based centerline connection algorithm is proposed to complete the road network. In terms of both quantitative and visual performances, the proposed method achieves better results than all the other comparing methods. Moreover, the proposed method are not sensitive to the parameters, which can be tuned easily. As another contribution, a new and challenging road centerline extraction dataset for VHR remote sensing images will be publicly available for further studies.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants 91338202, 9138105, 61272049, 61305049 and 61370039.

References


Ying Wang received his B.S. degree from the Department of Information and Communications Technologies, Nanjing University of Information Science and Technology, China in 2005, and his M.S. degree from the Department of Automation Engineering, Nanjing University of Aeronautics and Astronautics, China, in 2008, and his Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, China, in 2012. He is currently an assistant professor at the National Laboratory of Pattern Recognition, Institute of Automation of Chinese Academy of Sciences. His research interests include computer vision, pattern recognition, and remote sensing.

Chunhong Pan received the B.S. degree in automatic control from Tsinghua University, Beijing, China, in 1987, the M.S. degree from the Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Beijing, in 1990, and the Ph.D. degree in pattern recognition and intelligent system from the Institute of Automation, Chinese Academy of Sciences, in 2000, where he is currently a Professor with the National Laboratory of Pattern Recognition. His research interests include computer vision, image processing, computer graphics, and remote sensing.