A Semi-Supervised High-Level Feature Selection Framework for Road Centerline Extraction

Ruyi Liu, Qiguang Miao, Yi Zhang, Maoguo Gong, and Pengfei Xu

***Abstract*— Accurate road centerline extraction is very important for many vital applications. In the road extraction, the acquisition of labeled data is time-consuming; thus, there is only a small amount of labeled samples in reality. To solve the problem of limited labeled samples, a semi-supervised road centerline extraction is proposed, which incorporates high-level feature selection, Markov random field (MRF), and ridge transversal method. The proposed road extraction approach consists of three steps: multiple features extraction, semi-supervised road area extraction, and road centerlines extraction. To get more abstract and discriminative high-level features, we apply multiple-feature adaptive sparse representation in mid-level features in different views generated by different prototype sets. To obtain an accurate road area result, we combine the feature learning framework with MRF. Then, we integrate Gabor filters and nonmaxima suppression with the ridge transversal method to extract centerlines. It is verified the proposed method achieves comparable performance with the state-of-the-art methods in terms of visual and quantitative aspects.**

***Index Terms*— Ensemble projection (EP), feature learning, road centerline extraction, semi-supervised classification.**

I. Introduction

ROADS are of special importance among various man-made objects. Thus, it is desired and sometimes highly demanded to find ways to automatically extract a road network from remote sensing images. However, there are some difficulties in extracting a road network from remote sensing images [1]. Although various approaches have been proposed to address this challenging task, they are far from mature.

Manuscript received October 31, 2018; revised March 9, 2019 and June 18, 2019; accepted July 23, 2019. Date of publication August 30, 2019; date of current version April 22, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0807500, in part by the National Key Research and Development Program of China under Grant 238, in part by the National Natural Science Foundations of China under Grant 61772396, Grant 61472302, Grant 61772392, and Grant 61902296, in part by the Fundamental Research Funds for the Central Universities under Grant JBF180301 and Grant XJS190307, and in part by the Xi'an Key Laboratory of Big Data and Intelligent Vision under Grant 201805053ZD4CG37. *(Corresponding author: Qiguang Miao.)*

R. Liu and Q. Miao are with the School of Computer Science and Technology, Xidian University, Xi'an 710071, China, and also with the Xi'an Key Laboratory of Big Data and Intelligent Vision, Xi'an 710071, China (e-mail: ruyi198901210121@126.com; qgmiao@126.com).

Y. Zhang is with the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China.

M. Gong is with the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education, Xidian University, Xi'an 710071, China.

P. Xu is with the School of Information Technology, Northwest University, Xi'an 710069, China.

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2019.2931928

The road extraction task contains road area extraction and road centerline extraction. Various road area extraction techniques include knowledge-based methods, mathematical morphology, differential geometry, region competition, perceptual grouping, dynamic programming, and classification. Recently, classification has been widely applied to the road area extraction. Supervised classification relies on labeled samples for training the classifiers, where the classification accuracy heavily depends on the amount and quality of the training samples. Nevertheless, the annotations of such labeled samples are usually time-consuming and sometimes difficult to acquire in many real-world problems. In some cases where there is a small set of labeled samples and a large collection of unlabeled ones, some semi-supervised classification-based methods have been proposed, which jointly exploit labeled and unlabeled samples for training classifiers to improve classification accuracy. In the road centerline extraction problems, the morphological thinning algorithm [2], [3] and some regression-based centerline extraction algorithms [4], [5] cannot get satisfied performance, so there is a large research space in the road centerline extraction.

In this letter, a semi-supervised road centerline extraction method is proposed, which incorporates high-level feature selection, Markov random field (MRF), and ridge transversal method. We are inspired by the work of Dai and Van Gool [6], who introduced a feature learning approach based on ensemble projection (EP). An effective feature selection method aims to utilize a simple yet efficient criterion to determine the usefulness of features [7]–[9]. Road extraction is a typically two-class classification problem, i.e., road class and nonroad class. In fact, nonroad class contains many classes. It is not accurate to represent them as only one class. EP can exactly consider this diversity because it borrows ideas from ensemble learning to create an ensemble of diverse prototype sets. Features in different views can be obtained by EP. According to the level of fusion, multiview learning classification methods can be roughly categorized into two groups: 1) feature-level fusion and 2) classifier-level fusion [10]. To get more relevant and discriminative features, inspired by multiview learning, we propose a semi-supervised framework of high-level feature selection, in which the multiple-feature adaptive sparse representation (MFASR) [11] is introduced into EP instead of feature fusion based on the concatenation. After the classification, MRF is used to do postprocessing. It is a powerful tool that assumes that adjacent pixels are more likely to belong to the same class. To overcome the shortcomings of centerline extraction algorithms, Gabor filter, nonmaxima

1545-598X © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Xian Jiaotong University. Downloaded on April 24,2020 at 04:01:08 UTC from IEEE Xplore. Restrictions apply.

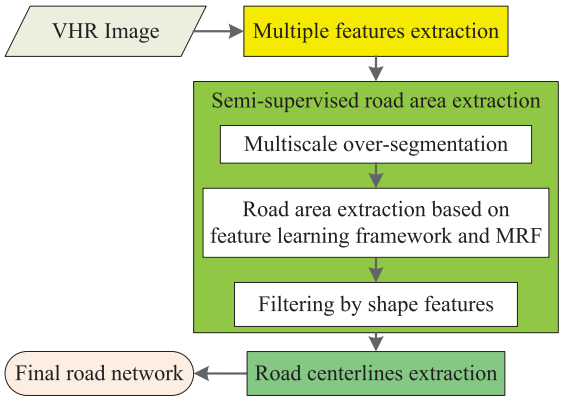

Fig. 1. Flowchart of the proposed method.

suppression, and ridge transversal method [12] are used to trace road centerlines.

In summary, the main contributions are listed as follows. First, we propose a semi-supervised framework of high-level feature selection for road extraction. Second, the perceptually homogeneous superpixel element is more appropriate for the human visual system, so we propose a superpixel-based feature learning framework. To the best of our knowledge, this is the first work that combines superpixel and EP, while previous works related to EP used square-image patches that cannot preserve edge information well. The remainder of this letter is arranged as follows. Section II presents the proposed methodology. Experimental evaluations and comparisons are reported in Section III. Conclusions are drawn in Section IV.

II. Methodology

The proposed road extraction approach consists of three steps: multiple features extraction, semi-supervised road area extraction, and road centerlines extraction. The flowchart of the proposed method is shown in Fig. 1.

A. Multiple Features Extraction

The main information of remote sensing image lies on the spectral dimension and spatial dimension. Therefore, three different kinds of features are employed as low-level features, including the spectral values, the extended multiattribute profile (EMAP), and 3-dimensional discrete-wavelet transform (3-D-DWT)-based feature. EMAP is widely used to capture the geometric and texture features. 3-D-DWT is adopted to generate spectral-spatial features, which have been validated to be more discriminative than the original spectral signature [13].

B. Semi-supervised Road Area Extraction

1) Multiscale Over-segmentation: Linear spectral clustering (LSC) is of linear computational complexity and high memory efficiency and is able to capture perceptually important global image properties [14]. Therefore, LSC is employed to obtain the multiscale oversegmentation for very high resolution (VHR) remote sensing images. We can get multiscale segmentation with different superpixel numbers. As shown in Fig. 2, we get three different oversegmentation results for each image. It can be seen that LSC superpixels have high boundary adherence. This is effective for suppressing the probability of misclassification of image pixels.

Fig. 2. Multiscale segmentation results with different superpixel numbers (the numbers increase from left to right).

2) Road Area Extraction Based on Feature Learning Framework and MRF: For each oversegmentation result, the spectral attribute of a superpixel is defined as the average spectral value within this superpixel. The EMAP feature of a superpixel is calculated as the mean value of the EMAP features in this superpixel. Similarly, 3-D-DWT-based feature of a superpixel is defined as the average 3-D-DWT values within this superpixel. Multiple features of superpixels are exploited by feature learning framework to obtain the high-level features and road probability of each segment. The feature learning framework can be shown in Fig. 3.

At a certain scale, we have obtained N superpixels after oversegmentation. For clarity, $X = \{x_i\}_{i=1}^{N}$ denotes the N superpixels, where x_i denotes the feature vector of superpixel i. In semi-supervised learning, only limited labeled samples are available. Suppose that labeled data are $D_l = \{(x_i, y_i)\}_i^l$ and unlabeled data $D_u = \{x_j\}_{j=l+1}^{l+u}$, where $y_i \in \{0, 1\}$ is its label.

Prototype sets originate from psychology and cognition fields. They are sampled automatically from all available data to represent a rich set of visual categories and attributes. We use max–min sampling and ensemble learning to get T prototype sets $P^{t,t\in\{1,\ldots,T\}} = \{(s_i^t, c_i^t)\}_{i=1}^{rn}$, where $s_i^t \in \{1, \ldots, l+u\}$ is the index of the i chosen superpixel and $c_i^t \in \{1, \ldots, r\}$ is the pseudolabel indicating to which prototype s_i^t belongs to. r is the number of classes in P^t and n is the number of superpixels sampled for each prototype (e.g., $r = 3$ and $n = 3$ in Fig. 3). For the skeleton, we randomly sampled m hypotheses, each hypothesis consists of r random-sampled superpixels and keeps the one having the largest mutual distance. Once the skeleton is created, the Min-step extends each seed superpixel to a superpixel prototype by introducing its n nearest neighbors (including itself). The pseudolabels are shared by all superpixels specifying the same prototype. Here, we used the EMAP feature of superpixels in max–min sampling, because it needs a low-dimensional feature to define neighborhoods. Clearly, information carried by a single prototype set P^t is quite limited, so we create T prototype sets. Ensemble learning benefits from the precision of its base learners and their diversity. For good precision, discriminative learning method is employed as the base learner. In order to learn precise projection function, we use the

Authorized licensed use limited to: Xian Jiaotong University. Downloaded on April 24,2020 at 04:01:08 UTC from IEEE Xplore. Restrictions apply.

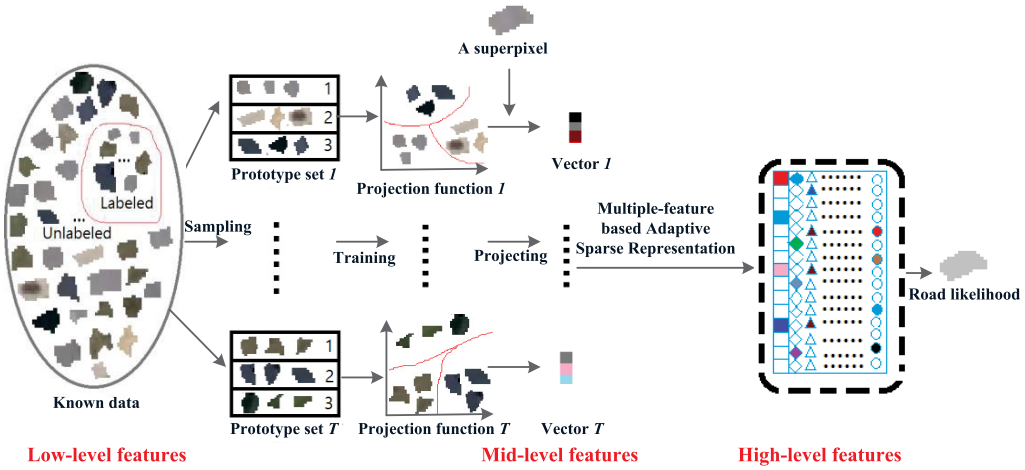

Fig. 3. Feature learning framework. The shapes with colors stand for the nonzero values, while the others represent the zero values.

concatenation of all the three features of superpixels. Logistic regression is the base learner to further ensure the accuracy.

After an ensemble of base learners is constructed, each base learner projects x to a similarity vector

$$S_j = (S_{j,1}, \ldots, S_{j,c}) \tag{1}$$

where $j \in \{1, \ldots, T\}$. $S_{j,c} (c \in \{1, \ldots, r\})$ measures the probability of x belonging to pseudo-label class c using base learner j. The diversity of the base learners is used to handle intraclass variance as different base learners capture different properties of the superpixel. A new superpixel representation $S^{\mathrm{MF}} = \{S_j\}_{j=1,2,\ldots,T}$ is obtained. Similarly, the corresponding T vector dictionary set $D^{\mathrm{MF}} = \{D_j\}_{j=1,2,\ldots,T}$ can be constructed by extracting superpixels from these vectors. Then, we use MFASR to utilize the similarities and diversities of multiple mid-level features (see Fig. 3). The multiple-feature-based adaptive sparse coefficients $\widehat{A} = [\widehat{\alpha}^1, \widehat{\alpha}^2, \ldots, \widehat{\alpha}^T]$ can be obtained by

$$\widehat{A} = \arg\min_{A} \sum_{k=1}^{T} \|S^k - D^k A^k\|_F, \quad \text{s.t.} \|A\|_{\text{adaptive},0} \le K_0 \tag{2}$$

where $\|A\|_{\text{adaptive},0}$ is the adaptive norm, which can select a number of the most representative adaptive sets from the multiple-feature sparse matrix A. Each adaptive set is denoted as the indexes of nonzero scalar coefficients, which belong to the same class in A. K_0 is the sparsity level. The optimization problem in (2) can be efficiently solved by an adaptive sparse algorithm that was introduced in [15]. After obtaining $\widehat{A}$, the joint residual errors can be calculated as

$$\underset{l \in \{0,1\}}{R_l} = \sum_{k=1}^{T} \left\| S^k - D_l^k \widehat{A}_l^k \right\|_F \tag{3}$$

where $\widehat{A}_l^k$ is the lth class subset of the kth feature's sparse vector, and $l = 0$ and $l = 1$ denote the nonroad class and the road class, respectively. Here, we get the road likelihood rather than the class label of each superpixel. Therefore, the road-class likelihood as

$$\text{prob\_road} = \frac{R_0}{R_0 + R_1}. \tag{4}$$

For a certain scale, all the pixels in the same superpixel are given an identity likelihood value as the superpixel. Similarly, we can get the probabilistic maps at different scales.

In general, adjacent pixels have the same label, so we enforce a spatially smooth label prior by treating the labels in y as an MRF. This prior encourages neighboring pixels to belong to the same class. Specifically, our segmentation model is given by

$$\widehat{y} = \arg\max_{y \in l} \left\{ \sum_{i=1}^{n} \log \text{prob\_road}(x_i) + \mu \sum_{(i,j) \in C} \delta(y_i - y_j) \right\} \tag{5}$$

where C is the set of cliques over the image and $\delta()$ is the unit impulse function. It should be noted that the pairwise interaction terms $\delta(y_i - y_j)$ obtain higher probability when neighboring labels are equal than when they are not equal. μ denotes the level of smoothness. In this way, the MRF prior encourages piecewise smooth segmentations. Therefore, three complementary road segmentation maps can be obtained. We fuse all road segmentation maps from the three different scales into an integrated one by a majority voting strategy.

3) Filtering by Shape Features: Although MRF has been used to postprocessing the classification maps, many road-like segments are also included in the final segmentation result. Therefore, we exploit the linear feature index (LFI) [1] to distinguish between potential road segments and road-like segments.

C. Road Centerlines Extraction

We extract road centerlines as follows: given the binary segmentation, we continuously filter the image with different scale Gabor filters. After continuous filtering, the real road centerline positions tend to have local maximum values. Then, nonmaxima suppression is applied to find a local maxima and a classic ridge transversal method [12] is applied to connect the local maxima. Finally, we remove small unwanted branches and use a linking procedure to connect some discontinuities.

Authorized licensed use limited to: Xian Jiaotong University. Downloaded on April 24,2020 at 04:01:08 UTC from IEEE Xplore. Restrictions apply.

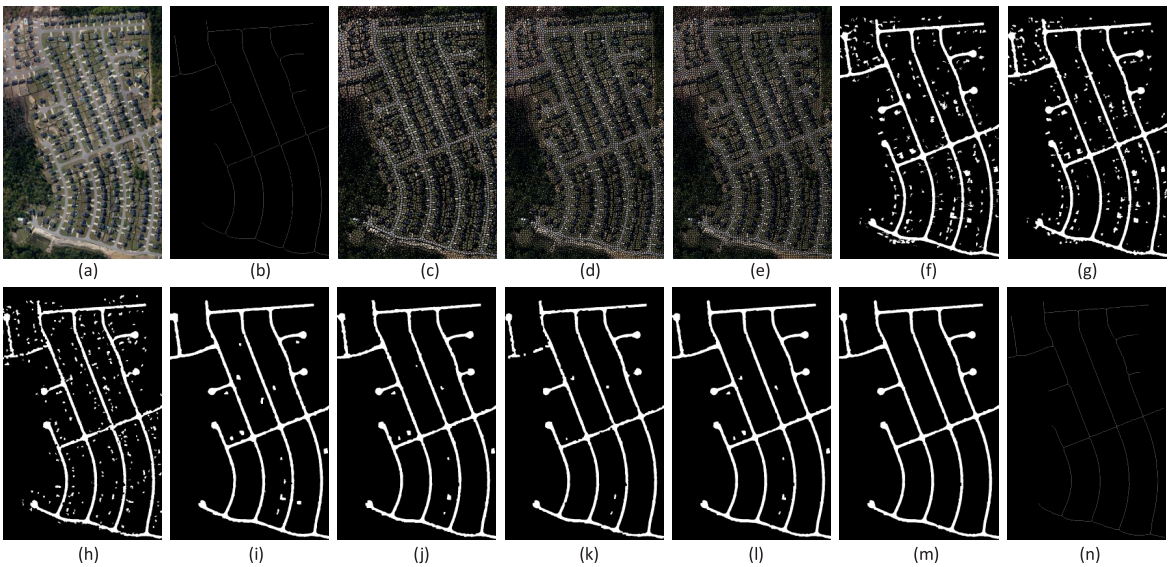

Fig. 4. Result of each step. (a) Aerial image. (b) Reference map. (c)–(e) Multiscale oversegmentation results with three different numbers of superpixels, such as 8000, 10 000, and 12 000. (f)–(h) Multiscale road area extraction results obtained by the feature learning framework. (i)–(k) Multiscale results refined by MRF. (l) Final result by majority voting. (m) Image after postprocessing based on shape features. (n) Road centerline results.

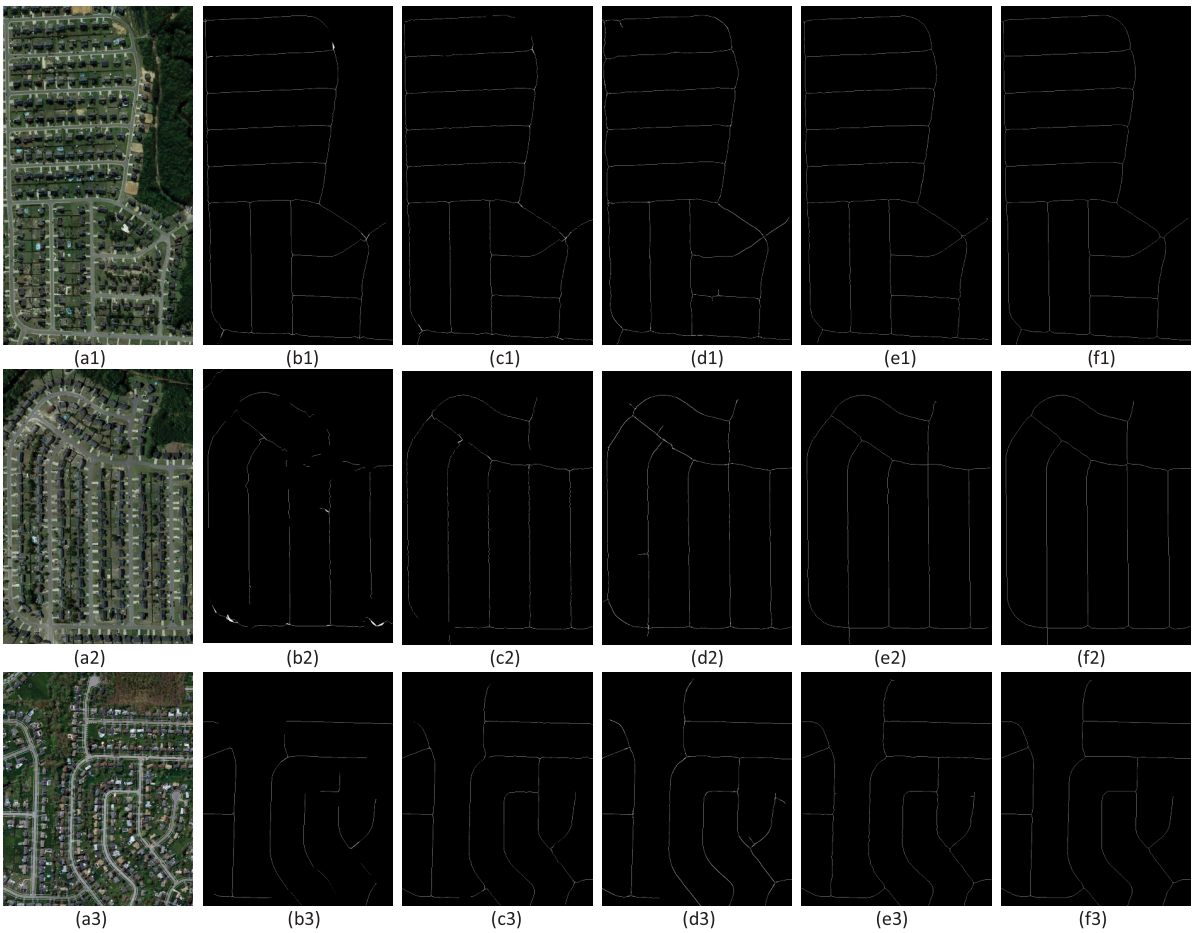

Fig. 5. Visual comparisons of road centerline extraction results. (a1)–(a3) Original image. (b1)–(b3) Results of Miao *et al.* [5]. (c1)–(c3) Results of Shi *et al.* [4]. (d1)–(d3) Results of Cheng *et al.* [1]. (e1)–(e3) Results of proposed method. (f1)–(f3) Reference maps. Due to the space limit, we only display three images.

III. Experimental Results

To demonstrate the performance, we do some experiments with aerial images and discuss our results. For evaluation, we use two public data sets in our experiments. One is the EPFL-data set, which is provided by Türetken *et al.* [16]. It contains 14 images. The road width in this data set

Authorized licensed use limited to: Xian Jiaotong University. Downloaded on April 24,2020 at 04:01:08 UTC from IEEE Xplore. Restrictions apply.

TABLE I
QUANTITATIVE EVALUATION RESULTS OF VARIOUS METHODS

Experiment	Methods	Com	Cor	Q
Fig. 5(a1)	Miao et al. [5]	87.16%	92.89%	81.71%
	Shi et al. [4]	93.71%	92.12%	86.76%
	Cheng et al. [1]	96.48%	94.92%	91.75%
	Proposed method	**96.66%**	**96.73%**	**93.61%**
Fig. 5(a2)	Miao et al. [5]	66.12%	72.01%	52.60%
	Shi et al. [4]	92.25%	94.07%	87.18%
	Cheng et al. [1]	**98.23%**	95.86%	94.24%
	Proposed method	97.40%	**99.29%**	**96.72%**
Fig. 5(a3)	Miao et al. [5]	78.60%	**98.31%**	77.55%
	Shi et al. [4]	94.22%	96.25%	90.89%
	Cheng et al. [1]	96.26%	96.59%	93.10%
	Proposed method	**98.99%**	98.23%	**97.26%**

is about 8–10 pixels. The other data set is developed by Cheng *et al.* [1], which contains 224 VHR urban road images with a spatial resolution of 1.2 m/pixel. The road width is about 12–15 pixels. Some images are under the conditions of complex backgrounds and occlusions of trees. The procedure of the proposed method is shown in Fig. 4. It can be seen that the proposed method can extract relatively smooth and continuous road centerline well.

For a fair comparison, the comparing methods are mainly unsupervised or semi-supervised. We have compared the proposed algorithm with three existing road extraction methods from the literature, i.e., Cheng *et al.* [1], Shi *et al.* [4], and Miao *et al.* [5]. Fig. 5 gives the comparison results of different road extraction methods. A vision comparison reveals that the quality of the proposed method is superior to those of the other three methods. To evaluate these methods quantitatively, the completeness (Com), correctness (Cor), and quality (Q) of each method is computed as follows:

$$\text{Com} = \text{TP}/(\text{TP} + \text{FN}) \tag{6}$$

$$\text{Cor} = \text{TP}/(\text{TP} + \text{FP}) \tag{7}$$

$$Q = \text{TP}/(\text{TP} + \text{FP} + \text{FN}) \tag{8}$$

where TP, FP, and FN are the true-positive, false-positive, and false-negative, respectively. The results are given in Table I. As we can see, the proposed method is comparable with or better than other methods, which gains the best performance among the comparing methods. It demonstrates that the high-level feature selection framework is suitable for the road extraction task. We also validate the efficiency of the semi-supervised high-level feature selection framework with CNN features. The CNN features perform significantly better than the rest. However, in order to get CNN features, we need a great amount of labeled data to train the neural network. If there is a small set of labeled samples, the proposed method is very appropriate.

IV. CONCLUSION

In this letter, a new method has been proposed to extract road centerline from high-resolution imagery accurately and smoothly. Specifically, in the first stage, spectral values, EMAP, and 3-D-DWT-based features are extracted. In the second stage, to get more abstract and discriminative high-level features, MFASR is applied to mid-level features generated by different prototype sets. After classification, MRF is used to do postprocessing to get accurate road area. Finally, we extract road centerlines by Gabor filter, nonmaxima suppression, and ridge transversal method. In terms of both quantitative and visual performances, the proposed method achieves better results than all the other comparing methods.

REFERENCES

[1] G. Cheng, F. Zhu, S. Xiang, Y. Wang, and C. Pan, "Accurate urban road centerline extraction from vhr imagery via multiscale segmentation and tensor voting," *Neurocomputing*, vol. 205, pp. 407–420, Sep. 2016.

[2] X. Huang and L. Zhang, "Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines," *Int. J. Remote Sens.*, vol. 30, no. 8, pp. 1977–1987, 2009.

[3] W. Shi, Z. Miao, Q. Wang, and H. Zhang, "Spectral–spatial classification and shape features for urban road centerline extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 4, pp. 788–792, Apr. 2014.

[4] W. Shi, Z. Miao, and J. Debayle, "An integrated method for urban main-road centerline extraction from optical remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3359–3372, Jun. 2014.

[5] Z. Miao, W. Shi, H. Zhang, and X. Wang, "Road centerline extraction from high-resolution imagery based on shape features and multivariate adaptive regression splines," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 13, pp. 583–587, May 2013.

[6] D. Dai and L. Van Gool, "Ensemble projection for semi-supervised image classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 2072–2079.

[7] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910–5922, Oct. 2018.

[8] Q. Wang, M. Chen, F. Nie, and X. Li, "Detecting coherent groups in crowd scenes by multiview clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published.

[9] Q. Wang, J. Wan, F. Nie, B. Liu, C. Yan, and X. Li, "Hierarchical feature selection for random projection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 5, pp. 1581–1586, May 2019.

[10] X. Xue, F. Nie, Z. Li, S. Wang, X. Li, and M. Yao, "A multiview learning framework with a linear computational cost," *IEEE Trans. Cybern.*, vol. 48, no. 8, pp. 2416–2425, Aug. 2018.

[11] L. Fang, C. Wang, S. Li, and J. A. Benediktsson, "Hyperspectral image classification via multiple-feature-based adaptive sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1646–1657, Jul. 2017.

[12] A. Hernandez-Vela *et al.*, "Accurate coronary centerline extraction, caliber estimation, and catheter detection in angiographies," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 6, pp. 1332–1340, Nov. 2012.

[13] X. Cao, L. Xu, D. Meng, Q. Zhao, and Z. Xu, "Integration of 3-dimensional discrete wavelet transform and Markov random field for hyperspectral image classification," *Neurocomputing*, vol. 226, pp. 90–100, Feb. 2017.

[14] Z. Li and J. Chen, "Superpixel segmentation using linear spectral clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1356–1363.

[15] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, Dec. 2014.

[16] E. Türetken, F. Benmansour, and P. Fua, "Automated reconstruction of tree structures using path classifiers and mixed integer programming," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 566–573.

Authorized licensed use limited to: Xian Jiaotong University. Downloaded on April 24,2020 at 04:01:08 UTC from IEEE Xplore. Restrictions apply.