A Multistage Method for Road Extraction from Optical Remotely Sensed Imagery

Rui $Xu^{1,2}$

¹ Key Lab of Spatial Data Mining & Information Sharing of Ministry of Education, Spatial Information Research Center of Fujian Province Fuzhou University No.523, Gongye Road, Fuzhou, 350002, P.R.China

² College of Information Science and Engineering Fujian University of Technology No.3, Xueyuan Road, University Town, Fuzhou, 350108, P.R.China dr.xurui@126.com

Received August, 2015; revised February, 2016

ABSTRACT. This paper presents a multistage framework for road extraction from remotely sensed imagery. The framework is a combination of multifeature-based mean shift, SVM classifier, and shape feature filter. Mean shift is first employed to cluster the remotely sensed images in the fusion of spectral-texture features and then a classifier of SVM is applied to classify the segmented images into two classes: the road class and the non-road class. Finally, the road class is refined by using road shape features. The experimental results show that this method is efficient in extracting road from remotely sensed imagery.

Keywords: Road extraction, Mean shift, Multi-dimension feature space, Classification, Shape feature.

1. Introduction. Roads are probably the most important information for Geographical Information System (GIS) applications (e.g. urban design, navigation, image registration, traffic management, emergency handling, etc). Road extraction from optical satellite images is an economical and efficient way to obtain and update a transportation database. Substantial work on road extraction has been accomplished since the 1970s in computer vision and digital photogrammetry, with pioneering works including Bajcsy and Tavakoli[1] and Quam[2]. In past decades, many algorithms have been developed to deal with road extraction from satellite images. The differences between these algorithms are closely related to the strategy adopted, e.g., image segmentation (often using a classifier), detection of a pair of parallel linear segments (using edge detector, Hough transform), snakes (active contours), morphological operations for cleaning and bridging discontinuities, matching of road templates, etc[3].

Many scholars have adopted region-based methods to extract road. The region-based methods first segment or classify the imagery into regions, followed by some rule to refine the extracted road networks. However, as there are misclassifications between roads and other spectrally similar objects, such as parking lots, buildings, crop fields, etc[3], the accuracy of classification or segmentation is far from satisfaction. Bogess[4] proposed an approach to identify roads in LANDSAT images using artificial neural networks (ANNs).

A feedforward neural network was trained with the spectral information. But the obtained results contain a lot of false alarms and disconnected road segments. Yuan and Cheriyadat [5] presented a simple and effective method which accurately segmented road regions using a weak supervision provided by road vector data and a factorization-based segmentation algorithm. This method largely depends on reliable vector data. In addition, the accurate registration of image data and vector data is also a challenging task. Amo *et al.*[6] proposed a combined semiautomatic approach consisting of region growing and region competition to extract road centerlines and sides. The initial seeds were given manually, and a region-growing-based model was applied to obtain a rough road approximation. This model is refined by the region competition algorithm. This approach performs well on extracting urban main-roads, but it is rather ineffective when it comes to larger curvature roads.

Recent road extraction research tends to use an overall workflow which incorporates a chain of several algorithms including image analysis, pattern recognition and high-level knowledge, and fruitful achievements have been made. We briefly highlight the representative work in the following. Song and Civco [7] designed a unique approach for road extraction utilizing pixel spectral information for classification and image segmentationderived object features was developed. First, SVM was used to classify the input image into road and non-road groups. Next, the road group image was segmented into geometrically homogeneous objects using a region growing technique. Finally, a thresholding process was performed to extract road features, which were further processed by thinning and vectorization to obtain road centerlines. Unsalan and Sirmacek[8] used probabilistic and graph theoretical to extract road networks from high-resolution remotely sensed images. Firstly, they extracted road primitives using canny operator. Secondly, they extracted the road shape by a binary balloon algorithm. Thirdly, they detected road centers by a probabilistic method. Shi et al. [9] combined spectral-spatial and shape features to urban main road. In the first step, they segmented the imagery into the road class and the non-road class using spectral-spatial classification. Then remotely sensed imagery was homogenized by using local Geary's C. As a final step, the road class was refined by using shape features. The experimental results indicated that the method they proposed could achieve a comparatively good performance in urban main road extraction.

Previous research indicates that no single-step approach can solve the road extraction problem completely. Most popular and successful methods rely on an overall workflow. According to this principle, we present a novel multistage methodology to detect the road network from optical remotely sensed images by integrating multifeature-based mean shift, SVM classification and shape feature filter. Firstly, we design an appropriate multi-dimension feature space to integrate texture features and spectral features which improves the robustness and accuracy of segmentation. Secondly, we design a set of suitably post processing stages to filter the undesired effects and thus improve the accuracy in road detection. Experiments show that the proposed method is effective and efficient in extracting road from remotely sensed imagery.

The remainder of this paper is organized as follows. The new approach is presented in Section 2. Experimental results are presented and discussed in Section 3. Conclusions are provided in Section 4.

2. Methodology. This study devises an overall workflow method to extract road from optical remotely sensed imagery. The proposed method consists of the following three steps.

(1) Mean shift clustering approach is used to segment the remotely sensed image.

R. Xu

Firstly, texture features and spectral features are extracted to form multi-dimension feature space. Secondly, mean-shift segmentation algorithm based on the multi-dimension feature space is used to segment the image into several parts.

(2) SVM classifier is used to classify the segmented imagery into two classes: the road class and the non-road class.

(3) False road segments are removed based on road shape features.

The details of each step are described in Section 2.1 to Section 2.3.

2.1. Mean shift segmentation. Mean shift is a simple iterative procedure that shifts each data point to the average of data points in its neighborhood. Twenty years after Fukunaga and Hostetlers pioneering work[10], Cheng [11] introduced the kernel function into the mean shift procedure, which made this method widely used in image segmentation.

2.1.1. Mean shift procedure. Assume that each data point x_i , i = 1, ..., n be a set of ddimensional points in the space \mathbb{R}^d , d is the dimension of the feature space, h is called the kernel bandwidth or window size, and determines the range of influence of the kernel located in x_i . The density at point x can be estimated by the Parzen window Kernel density estimator [12].

$$\hat{f}_{K}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^{d}} k \left(\left\| \frac{x - x_{i}}{h} \right\|^{2} \right)$$
(1)

Here, function k is called the profile of the spherically symmetric kernel K with bounded support, $c_{k,d}$ is the normalized constant which makes K(x) integrate to one, $||x|| \leq 1$, that satisfies.

$$K(x) = c_{k,d}k(||x||^2) > 0$$
(2)

The gradient of the density estimate are given by:

$$\hat{\nabla}f_K(x) \equiv \nabla \hat{f}_K(x) = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^n (x - x_i)k' \left(\left\| \frac{x - x_i}{h} \right\|^2 \right)$$
(3)

Let g(x) = -kI(x), then the gradient of the density estimator is written by

$$\nabla \hat{f}_{K} = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (x_{i} - x)g\left(\left\|\frac{x - x_{i}}{h}\right\|^{2}\right)$$

$$= \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} g\left(\left\|\frac{x - x_{i}}{h}\right\|^{2}\right) \left(\frac{\sum_{i=1}^{n} g\left(\left\|\frac{x - x_{i}}{h}\right\|^{2}\right)x_{i}}{\sum_{i=1}^{n} g\left(\left\|\frac{x - x_{i}}{h}\right\|^{2}\right)} - x\right)$$
(4)

The last part of (4) contains the adaptive mean shift vector:

$$M(x) = \left(\frac{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_{i}}{h}\right\|^{2}\right) x_{i}}{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_{i}}{h}\right\|^{2}\right)} - x\right)$$
(5)

Suppose $G(x) = c_{G,d}g(||x||^2)$, the vector is the difference between the weight mean using the kernel G(x) for weights and the center of the kernel. Then, we can rewrite expression

440

(4) as

$$\nabla \hat{f}_K = \frac{2c_{k,d}}{c_{G,d}h^2} \hat{f}_G(x) M(x) \tag{6}$$

which yields

$$M(x) = \frac{1}{2}h^2 c \frac{\nabla \hat{f}_K(x)}{\hat{f}_G(x)} \tag{7}$$

Formula (7) shows that the mean shift vector is proportional to the normalized gradient of the density estimate computed for kernel. Therefore, the mean shift vector points toward the direction of maximum density increase. This property is the basis of the mean shift clustering algorithm.

2.1.2. Construction of multi-dimension feature space. Mean shift is a clustering approach of feature space, the frequently used features in remotely sensed imagery including spectral, texture, shape, and edge. In this paper, we use spectral and texture features to form multi-dimension feature space.

As the same objects may have different spectrum, and different objects may have the same spectrum, the capability of segmentation is weakened if using spectral features only. Therefore, texture features are applied to improve the robustness of image segmentation. Grey Level Co-occurrence Matrices (GLCM) [13] is the most typical and representative analysis to characterize texture feature among statistical methods. In this paper, we select four GLCM statistics which include Contrast, Entropy, Angular second moment, and Correlation. Suppose the values (i, j) are pixel intensities, P(i, j) is the frequency in a reference matrix G that an element g that lies within a value range i has a neighbor within the value range j at an angle-distance metric δ , in a GLCM, the reference matrix G is the image. Thus, GLCM is a matrix of frequencies, where each element (i, j) is the sum of the number of times that pixel value i is some distance and angle δ from pixel intensity j. L is the number of the gray levels in the image under quantization. These statistics can be calculated as follows.

(1)Contrast (CON)

$$f_1 = CON = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i-j)^2 P(i,j)$$
(8)

Contrast (CON) is a measure of the intensity contrast between a pixel and its neighbor over the whole image which calculates the sum of diversity in the gray level image.

(2) Entropy (ENT)

$$f_2 = ENT = -\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i,j) \log_2 P(i,j)$$
(9)

Entropy (ENT) is a measure of the amount of information contained in the image. It reflects the degree of image texture complexity or disorder. The more complex texture, the greater entropy, and vice versa.

(3) Angular second moment (ASM)

$$f_3 = ASM = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P^2(i,j)$$
(10)

Angular second moment(ASM), which reflects image texture roughness, is the quadratic sum of each element in the co-occurrence. Rough texture energy means bigger energy,

441

R. Xu

while fine texture energy means smaller energy.

(4) Correlation (COR)

$$f_4 = COR = \frac{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - \mu_x)(j - \mu_y)P(i, j)}{\sigma_x \sigma_y}$$
(11)

where the mean (μ) and variance (σ) are given by $\mu_x = \sum_{i=0}^{L-1} i \sum_{j=0}^{L-1} P(i,j)$, $\mu_y = \sum_{j=0}^{L-1} j \sum_{i=0}^{L-1} P(i,j)$, $\sigma_x = \sum_{i=0}^{L-1} (i - \mu_x)^2 \sum_{j=0}^{L-1} P(i,j)$, $\sigma_y = \sum_{j=0}^{L-1} (j - \mu_y)^2 \sum_{i=0}^{L-1} P(i,j)$. Correla-

tion (COR) reflects the local texture correlation of an image. When the matrix elements are equal and uniform, the correlation value is big; conversely, when the matrix elements vary greatly, it is small.

Homogeneous property which reflects region consistency is closely related with the local information of remotely sensed imagery. It is an important feature of the road. In this paper, Moran's I [14] is used to measure image local homogeneous properties. Moran's I, which reflects the differing spatial structures of the smooth and rough surfaces, measures the spatial autocorrelation between a pixel and its neighboring pixels. Moran's I tends to measure image local homogeneous properties more precisely, so it is used in this paper.

Moran's I is calculated from the following formula.

$$I(d) = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left(z_i - \overline{z} \right) \left(z_j - \overline{z} \right)}{W \sum_{i=1}^{n} \left(z_i - \overline{z} \right)^2}$$
(12)

where, *n* is the number of spatial units indexed by *i* and j, w_{ij} is the weight at distance d, so that $w_{ij} = 1$ if point *j* is within distance *d* from point i, otherwise $w_{ij} = 0$. The z_i and z_j are pixel digital numbers, \overline{z} is the mean of the pixel digital numbers, $W = \sum_{i=1}^{n} w_{ij}$.

Obviously, the important parameter index distance d represents the context dependence of the calculation range. Fig. 1 shows the results of Moran's I on an image with different distance d.



FIGURE 1. Comparison with different d: (a) Original image, (b) d=3, (c) d=9, and (d) d=15.

In this paper, c_1 , c_2 and c_3 are the spectral features, f_1 , f_2 , f_3 and f_4 are GLCM statistics, I is Moran's I index. The normalized vector $(C_1, C_2, C_3, F_1, F_2, F_3, F_4, I)$ is used as feature vector.

442

2.2. **Detection of road regions.** The proposed mean shift method based on multidimension feature space can cluster the image into several regions effectively. Each region has uniform color which is calculated by the mean color of that region, Fig. 2(b) shows the clustering result by our mean shift method.

However, it is still difficult to extract road regions from the results of mean shift clustering based on traditional threshold method. Fig. 2(c) shows the result of traditional threshold method. To overcome the aforementioned shortcoming, this paper uses a SVM algorithm to detect road regions from segmented image.



FIGURE 2. (a) Original image. (b) The clustering result by our mean shift method. (c)Result of detecting road regions from Fig. 2(b) by traditional threshold method.

As SVM is a supervised classification method which is as good as or significantly better than other competing methods in most cases [15], we employ SVM to improve the robustness and the accuracy of extracting road regions. The SVM classification algorithm[16] is listed below.

Step1: Segment the remotely sensed image by our mean shift method.

Step2: For each pixel of the segmented image, spectral features are used as the input features.

Step3: Select 8% ground truth data of each class such as roads, parking lots, buildings, vegetation, bare land, etc. as the training samples from the segmented image.

Step4: Use Gaussian kernel, and select the parameters, namely, penalty parameter C and kernel parameter , to train the SVM classifier.

Step5: Classify the whole segmented image using the classifier trained in step 4.

Step6: Assign 1 to road class, and assign 0 to others.

2.3. Filtering by Shape Feature. Although the algorithm of classification is able to remove most false road regions from the segmented image, misclassified roads still exist. Further processing is needed to improve the reliability of road extraction. In general, roads have unique features: (1) Roads do not have small areas, so regions with small areas can be regarded as noisy and should be removed. (2) Roads are mostly elongated structures, with locally linear properties.(3) Road surface usually has homogeneous property which reflects region consistency, with occasional variations. Based on this homogeneous property, the inference is that roads should be located in homogeneous regions. Hence, road shape features can be used to filter false segments. These features can be measured by Area, Compactness, Slenderness and Length-width ratio, which are introduced as follows.

(1) Area refers to the number of pixels included in the region. Segments with small area values can be viewed as non-road class and be removed.

(2) Compactness is defined as $4 \cdot Pi \cdot A/P^2$, where P is perimeter of region and A is area of region. Compactness is in the range of (0, 1].

(3) Slenderness is defined as 2A/L, where L is the regional center line length.

(4) Length-width ratio is the aspect ratio of the minimum enclosing rectangle.

3. Experiments.

3.1. Experiment 1. The study area is a part of Calgary City image which was recorded by the Worldview-II optical sensor. The study area has a spatial dimension of $495^{*}430$ pixels. The spatial resolution is 0.5 m per pixel. Fig.3(a) shows the study area of experiment 1. The proposed mean shift algorithm using multi-dimension feature space is firstly applied to segment the study area, with the result shown in Fig.3(c). Fig.3(b) shows the result of the mean shift algorithm using spectral features only. Compared with the algorithm using spectral features, the proposed mean shift algorithm can produce more satisfactory segmentation. Secondly, aiming to obtain the road regions, we employ SVM classifier to classify the segmented imagery into two classes: the road class and the nonroad class. Fig.3(e) shows the result of the SVM method. The segmented image is also thresholded to the binary map. The obtained result contains many false road segments, which is shown in Fig.3(d). It is shown in the experiment that the presented SVM method is more effective than the threshold method in distinguishing road regions. Since road shape feature is unique, it is finally applied to remove false road segments, resulting in Fig.3(g). And Fig.3(g) is also the result by the method proposed in this study. It can be seen from Fig.3(h) that the roads extracted by the proposed method anastomose well with original roads. The result by the algorithm Song and Civco^[7] proposed is shown in Fig.3(f). Compared with Song and Civco 's[7], the proposed method could be more effective.

3.2. Experiment 2. In the second experiment, the experimental area is a part of Cosned'Allier City image downloaded from Google earth. This study area has a spatial dimension of 1024*768 pixels, it is shown in Fig.4(a). The steps of this experiment are the same as the previous one. Fig.4(b)-(c) show the segmentation results of the mean shift algorithm using spectral features and the proposed algorithm respectively. It is shown that the proposed mean shift algorithm can produce more satisfactory result. Compared with the result shown in Fig.4(d) and Fig.4(e), we can see clearly that the SVM method can effectively distinguish road groups while the threshold method is invalid. The filter using the shape features is also efficient in the experiment, which is shown in Fig.4(g). The superposition result of the original image and road centerlines is shown in Fig.4(f). As seen from Fig.4(f) and Fig.4(g), the proposed method has higher accuracy.

3.3. Analysis of the experiment. To quantify the performance of the proposed method, three accuracy measures [17] are used to evaluate it. These measures are: 1) completeness = TP/(TP+FN); 2) correctness = TP/(TP+FP); and 3)quality = TP/(TP+FP+FN). The variables TP, FN, and FP denote true positive, false negative, and false positive, respectively.

In Table 1, we make comparisons between our method and Song and Civco 's method in



FIGURE 3. (a) Original image. (b) Result of mean shift using spectral features. (c) Result of mean shift using multi-dimension feature space. (d) Result of distinguishing road regions from Fig.3(c) by method of threshold. (e) Result of distinguishing road regions from Fig.3(c) by method of SVM. (f) Result by the method of Song and Civco [7]. (g) Result of our proposed method. (h) The superposition result of the original image and the road extracted by the proposed method.

Completeness, Correctness and Quality. It indicates that the two methods achieve similar high completeness, but our method achieves higher accuracy in correctness and quality. In the first experiment, our method improves correctness and quality by 20.57% and 20.46%, respectively; and in the second one, our method also improves correctness and quality by 19.62% and 17.35%, respectively. The high accuracy of our method is partly due to the higher adaptability of the proposed mean shift clustering, and the subsequent SVM classifier and shape feature filter also improve the accuracy.

	Completeness (%)		Correctness (%)		Quality (%)	
	Song et al.	Proposed	Song et al.	Proposed	Song et al.	Proposed
Experiment 1	93.61	95.90	74.52	95.09	70.91	91.37
Experiment 2	94.94	93.94	74.22	93.84	71.14	88.49

TABLE 1. Comparison of the two different road extraction methods

4. **Conclusions.** In this paper, we have presented a multistage framework for accurate and reliable road extraction from remotely sensed imagery. Uniquely, this framework is a combination of mean-shift segmentation, construction of multi-dimension feature space, classification of SVM and shape feature filter. The proposed method is compared with the widely used method of Song and Civco[7]. The experimental results indicate that our method performs well with optical remotely sensed imagery and it can also be deployed to extract the road network in large scale imagery. There are two advantages about our method described as follows: Firstly, we improve the robustness and accuracy of segmentation by forming multi-dimension feature space. Mean shift which uses spectral values 446



FIGURE 4. (a) Original image. (b) Result of mean shift using spectral features. (c) Result of mean shift using multi-dimension feature space. (d) Result of distinguishing road regions from Fig.4(c) by method of threshold. (e) Result of distinguishing road regions from Fig.4(c) by method of SVM. (f) Result by the method of Song and Civco [7]. (g) Result of our proposed method. (h) The superposition result of the original image and the road extracted by the proposed method.

only cannot perform well because many land cover types have similar spectral reflectance. The experiments indicate our mean shift method based on multi-dimension feature space is accurate and effective. Secondly, traditional threshold method is invalid in distinguishing road regions, thus we use a SVM algorithm to detect road regions from segmented images which can produce satisfactory classification between road and non-road.

In the presented form, the proposed method is semi-automatic, which exists some limitations. First, many parameters, such as the kernel bandwidth, the distance of Moran's I and the threshold of shape features, need to be set by users. Second, SVM needs to be trained for each segmented image, which limits the applicability of the method in practice. Future work will therefore focus on an automatic threshold and parameter determination method.

Acknowledgment. This work is supported by the Education Department of Fujian Province Science and Technology Project (JA13220), the key project of science and technology of Fujian province (2014H6006).

REFERENCES

- R. Bajcsy and M. Tavakoli, Computer Recognition of Roads from Satellite Pictures, *IEEE Transac*tions on Systems, Man and Cybernetics, vol.6, no.9, pp. 623–637,1976.
- [2] L. H. Quam, Road Tracking and Anomaly Detection in Aerial Imagery, Image Understanding Workshop, London, pp.51–55, 1978.
- [3] S. Das, T. T. Mirnalinee, and K. Varghese, Use of Salient Features for the Design of a Multistage Framework to Extract Roads From High-Resolution Multispectral Satellite Images, *IEEE Transactions on Geoscience and Remote Sensing*, vol.49, no.10, pp.3906–3931, 2011.

- [4] J. Boggess, Identification of roads in satellite imagery using artificial neural networks, MS Thesis, Mississippi State University, Mississippi State, 1993.
- [5] J. Yuan and A. M. Cheriyadat, Road segmentation in aerial images by exploiting road vector data, Proc. of the 4th International Conference on Computing for Geospatial Research and Application, San Jose, USA, pp.16-23, 2013.
- [6] M. Amo, F. Martinez, and M. Torre, Road extraction from aerial images using a region competition algorithm, *IEEE Transactions on Image Processing*, vol. 15, no. 5, pp.1192–1201, 2006.
- [7] M. Song and D. Civco, Road Extraction Using SVM and Image Segmentation, *Engineering and Remote Sensing*, vol.70, no.12, pp.1365–1372, 2004.
- [8] C. Unsalan and B. Sirmacek, Road network detection using probabilistic and graph theoretical methods, *IEEE Transactions on Geoscience and Remote Sensing*, vol.50, no.11, pp.4441–4453, 2012.
- [9] W. Z. Shi, Z. L. Miao, Q. Wang, and H. Zhang, Spectral-spatial classification and shape features for urban road centerline extraction. *IEEE Geoscience and Remote Sensing Letters*, vol.11, no.4, pp.788–792, 2014.
- [10] K. Fukunaga and L. Hostetler, The estimation of the gradient of a density function with applications in pattern recognition, *IEEE Transactions on Information Theory*, vol.21, no.1, pp.32–40. 1975.
- [11] Y. Cheng, Mean shift, mode seeking, and clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.17, no.8, pp.790–799, 1995.
- [12] D. Comaniciu and P. Meer, Mean Shift: A robust approach toward feature space analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, no.5, pp.1–17, 2002.
- [13] R. M. Haralick, K. Shanmugam, and I. Dinstein, Texture features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, vol.3, no.6, pp.610–621, 1973.
- [14] P. A. P. Moran, Notes on continuous stochastic phenomena, *Biometrika*, 37, pp.17–23, 1950.
- [15] C. J. C. Burges, A tutorial on support vectormachines for pattern recognition, Data Mining Knowl. Discov., vol.2, no.2, pp.121–167, 1998.
- [16] W. Z. Shi, Z. L. Miao, and J. Debayle. An integrated method for urban main-road centerline extraction from optical remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing*, vol.52, no.6, pp.3359–3372, 2014.
- [17] C. Wiedemann, C. Heipke, H. Mayer and O. Jamet, Empirical evaluation of automatically extracted road axes, *Proc. of CVPR Workshop on Empirical Evaluation Methods in Computer Vision*, Los Alamitos, USA, pp.172-187, 1998.