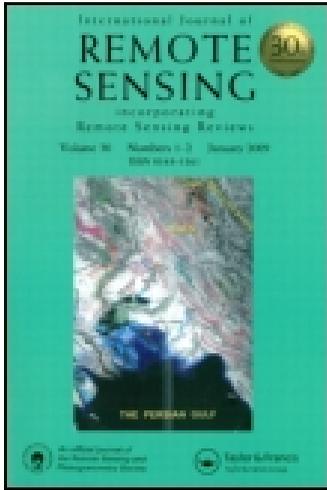


This article was downloaded by: [University of Virginia, Charlottesville]

On: 14 December 2014, At: 18:41

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



International Journal of Remote Sensing

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/tres20>

Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines

Xin Huang^a & Liangpei Zhang^a

^a The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, P. R. China

Published online: 30 Apr 2009.

To cite this article: Xin Huang & Liangpei Zhang (2009) Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines, International Journal of Remote Sensing, 30:8, 1977-1987, DOI: [10.1080/01431160802546837](https://doi.org/10.1080/01431160802546837)

To link to this article: <http://dx.doi.org/10.1080/01431160802546837>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &

Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Technical Note

Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines

XIN HUANG* and LIANGPEI ZHANG

The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, P. R. China

(Received 12 October 2007; in final form 8 May 2008)

This paper investigates road centreline extraction from high-resolution imagery. A novel road detection system is proposed based on multiscale structural features and support vector machines (SVMs). The salient aspects of the strategy are: (1) structural features are exploited because road objects are narrow and extensive, with large perimeters and small radii; (2) the object-based approach is used to extract multiscale information so as to reduce the local spectral variation caused by vehicles, shadows, road markings, etc.; (3) the hybrid spectral–structural features are analysed using the SVM classifier; and (4) multiple object levels are integrated because a multiscale approach can exploit the rich spatial information and detect multiscale road objects. Experiments were conducted on two IKONOS multispectral datasets and the results validated the proposed method.

1. Introduction

Road extraction using remote sensing data is important in applications such as Geographic Information System (GIS) updating, transportation analysis and urban planning. Many studies have reported the extraction of road networks using medium- or low-resolution satellite images (e.g. Landsat Thematic Mapper (TM) and SPOT). Roads on these images usually appear with widths of one or two pixels and some details of roads and trails cannot be observed in these low-resolution images. Therefore, traditional road extraction methods focus on line detection, such as the ‘snakes’ algorithm (Gruen and Li 1997), Hough transform (Dell’Acqua and Gamba 2001) and statistical methods (Barzohar and Cooper 1996, Tupin *et al.* 1998). In recent years, high-resolution satellite sensors with multispectral channels, such as Quickbird, IKONOS and SPOT-5, have provided richer spatial information; therefore, high-resolution imagery should potentially be useful for detailed road detection. However, in these images, small objects can be observed and hence the noise affecting road extraction increases (e.g. vehicles, shadows, markings and trees along the roads). Consequently, the spectral signatures of roads become more heterogeneous; moreover, some spectrally similar classes such as roofs and bare soils also lead to errors in road recognition. To address these issues, some efficient algorithms have been proposed recently to extract road networks from high-resolution imagery, such as adaptive directional filtering (Gamba *et al.* 2006), a contextual method based on the Gaussian pyramid (Binaghi *et al.* 2003), use of a

*Corresponding author. Email: huang_wlu@163.com

mean shift filter (Long and Zhao 2005), and morphological transform (Zhu *et al.* 2005).

This paper proposes a novel road extraction system based on multiscale object-based structural features and support vector machines (SVMs). As shown in figure 1, the system consists of three processing phases: (1) structural feature extraction; (2) multiscale fusion; and (3) post-processing. Each phase is described below.

- (1) The extraction of multiscale spectral–structural features is based on object-oriented segmentation. The object-based method can generalize the spectral information in a spatial neighbour and exaggerate the spectral distinction between spectrally similar classes (Wang *et al.* 2004). Furthermore, it can reduce the spectral variation and noise effects in road regions. In addition to spectral information, structural features are also extracted for each object so as to discriminate spectrally similar objects such as roads, roofs and bare soils. The combination of spectral–structural features is necessary because spectral information has proved inadequate for object recognition in high-resolution imagery (Huang *et al.* 2007).
- (2) The road objects for different scales are obtained by interpreting spectral and structural features using SVMs. A majority voting approach is then used to integrate the multiscale road information at the decision level.

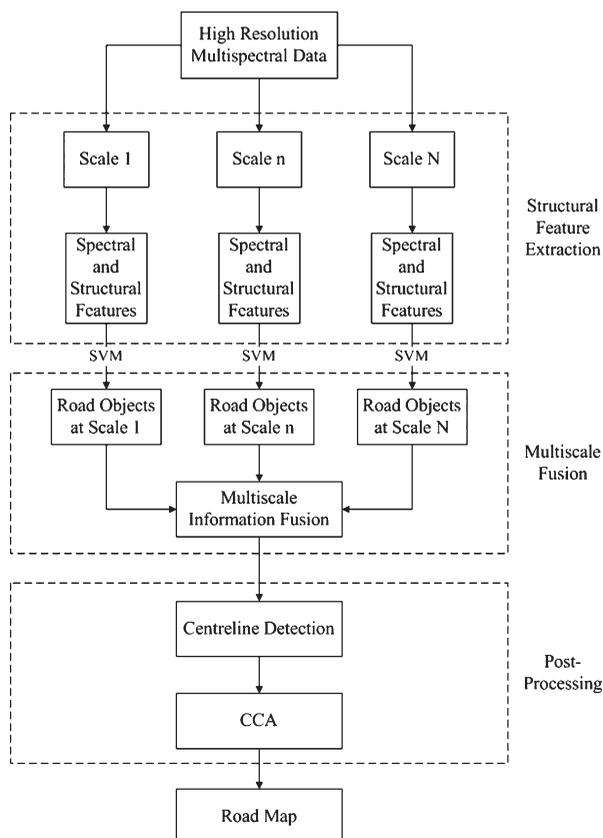


Figure 1. The flow chart of the proposed road extraction framework.

- (3) Post-processing includes centreline extraction and connected component analysis (CCA). The centreline is extracted by using a morphological thinning algorithm on the binary road objects and CCA is then used to remove small branches shorter than a user-defined threshold.

2. Methodology

2.1 Object-based spectral and structural feature extraction

In this study, the object-oriented algorithm is used to extract spectral and structural attributes of road regions. The basic idea is to group the spatially adjacent pixels into spectrally homogeneous objects first, and then conduct classification on objects as the minimum mapping units (Yu *et al.* 2006). The object-based approach is used for road extraction from high-resolution data because it can reduce the spectral variations during image segmentation, and at the same time apply geometrical features to road objects. In this research, the fractal net evolution approach (FNEA) (Hay *et al.* 2003) is adopted for multiscale segmentation. It uses fuzzy set theory to extract the objects of interest, at the scale of interest, segmenting images simultaneously at both fine and coarse scales (Hay *et al.* 2003). The FNEA is a bottom-up region-merging technique starting from a single pixel. In an iterative way, image objects are merged into larger ones at each subsequent step. The region-merging decision is made with local homogeneity criteria, which can be defined as

$$H = \sum_{b=1}^B W_b [N_{\text{merge}} \sigma_{\text{merge}} - (N_{\text{obj1}} \sigma_{\text{obj1}} + N_{\text{obj2}} \sigma_{\text{obj2}})] \quad (1)$$

where W_b controls the weight of band b ($1 \leq b \leq B$), N_{merge} , N_{obj1} and N_{obj2} represent the number of pixels within the merged object, object 1 and object 2, respectively. σ_{merge} , σ_{obj1} and σ_{obj2} are the corresponding standard deviations. When a possible merge of a pair of image objects is examined, the merge is performed when the heterogeneity H is below the scale parameter T (i.e. $H < T$). The segmentation process stops as soon as this condition cannot be met by any possible merge. T is a measure of the maximum change in heterogeneity, and hence it controls the segment size. A small-scale parameter will give rise to a small object size on average, while a large value will lead to a large object size on average.

The averaged spectral value within an object is defined as the spectral attribute of this object and, for object i , its spectral attribute (SA) is given by

$$\text{SA}(i) = \{\text{SA}^b(i)\}_{b=1}^B \quad \text{with} \quad \text{SA}^b(i) = \frac{1}{m} \sum_{p \in i} S^b(p), \quad (2)$$

where p represents a pixel within the object i ($1 \leq p \leq m$) and $S^b(p)$ is its spectral value. The structural attributes for each object are characterized by

$$\text{Shape Index (SI)} : \quad \text{SI}(i) = \frac{e(i)}{4 \sqrt{A(i)}} \quad (3)$$

$$\text{Compactness (COM)} : \quad \text{COM}(i) = \frac{L(i) \times W(i)}{A(i)} \quad (4)$$

$$\text{Density (DEN)} : \quad \text{DEN}(i) = \frac{\sqrt{A(i)}}{1 + \sqrt{V(X_i) + V(Y_i)}} \quad (5)$$

where $e(i)$, $A(i)$, $L(i)$ and $W(i)$ denote the perimeter, area, length and width for object i , respectively. $V(X_i)$ and $V(Y_i)$ are the variances of X and Y coordinates of all pixels forming the object i . It should be noted that in equation (4), for each object, the width and length are computed based on a bounding box approximation (eCognition User Guide 2002; www.definiens-imaging.com). In equation (5), $\sqrt{V(X_i) + V(Y_i)}$ is used to calculate the radius of the object.

The SI aims to describe the smoothness of the image object borders. Road objects have larger perimeters and small areas; hence they should have larger SI and COM values. DEN measures the compactness, and the more the form of an object is like a square, the higher its DEN value. Therefore, road features should have small DEN values because they are often modelled as continuous and elongated regions. Each object is characterized by combining the aforementioned spectral–structural attributes, and the multiscale vector can then be extracted by defining multiple scale parameters. The hybrid feature set is described as

$$F^T(i) = \{SA^T(i), SI^T(i), DEN^T(i), COM^T(i)\} \quad (6)$$

where $F^T(i)$ is the feature set of object i in scale T , and $SA^T(i)$, $SI^T(i)$, $DEN^T(i)$ and $COM^T(i)$ are the spectral–structural attributes.

2.2 The SVM for object-based road extraction

In this study, an SVM (Cortes and Vapnik 1995) was used to interpret the spectral–structural features and extract multiscale road objects. The SVM is intrinsically less sensitive to the distribution and dimensionality of the feature space, which makes it more suitable for complex input than some traditional classifiers. For instance, the maximum likelihood classifier (MLC) is not capable of achieving satisfactory results because the complex object-based features cannot be modelled as normally distributed. Moreover, because of the hybrid feature space and the spectrally similar objects in the images, the decision boundary should be nonlinear. The SVM has advantages in nonlinear recognition problems. It finds an optimal linear hyperplane in a higher dimensional feature space that is nonlinear in the original input space. The kernel trick avoids direct evaluation in the higher dimensional feature space by computing it through the kernel function with data vectors in the input space. To our knowledge, few studies have been reported on the use of the SVM and object-based structural features for road detection; therefore, it is interesting to evaluate its effectiveness.

The SVM was originally designed for binary classification; therefore, in this study, the binary SVM is implemented directly (roads and non-roads). The commonly used kernel functions are the radial basis function (RBF) and the polynomial function. In experiments, RBF kernels are used because they have been shown to be effective in many classification problems. An overview of the application of SVMs to remote sensing is given by Foody and Mathur (2006).

2.3 Multiscale information fusion

In high-resolution imagery, road features with different scales and sizes (e.g. highways, small roads and trails) can be observed in detail. As mentioned earlier, it

is important to choose an optimal scale parameter for object-based road recognition. However, a large-scale parameter is better at detecting large and wide road objects, while a small one is advantageous for detailed objects such as paths and trails. Therefore, in this study, multiscale object-based features were fused. The multiscale approach makes road detection more robust because a geostatistical analysis indicated that there was no single scale that would adequately characterize the range of textural conditions present in remote sensing images (Coburn and Roberts 2004).

As shown in figure 1, multiscale spectral–structural features were analysed using SVMs, and then objects for different scales were labelled as roads or non-roads. Let $P(T_n)$ be the attribute of a pixel p at scale n , and $O(T_n)$ the label of the corresponding object level [$p \in O(T_n)$], and therefore:

$$\begin{cases} P(T_n) = 1 & O(T_n) \text{ is a road - object} \\ P(T_n) = 0 & \text{else} \end{cases} \quad \text{with } 1 \leq n \leq N. \quad (7)$$

The multiscale information is then integrated by majority voting:

$$M_p = \sum_{n=1}^N P(T_n) \quad (8)$$

where M_p represents the number of times the pixel p has been detected as a road in multiscale levels. The fusion rule is defined as

$$P = \begin{cases} 1 & M_p \geq (N/2) \\ P(T_N) & \text{else} \end{cases} \quad (9)$$

where P is the final label for each pixel. Equation (9) shows that if the pixel has been detected as a road by at least $N/2$ scale levels, it is finally identified as a road. Otherwise, it is assigned to the attribute in the largest scale level [$P(T_N)$] because there is a lot of noise in small-scale levels and the road information is more reliable in the large-scale level.

2.4 Post-processing

Post-processing consists of two steps: pruning small regions and then extracting road centrelines. Connected component analysis (CCA) was used for small noise removal. CCA aims to group the pixels into connected components based on pixel connectivity (eight neighbours) and calculates some geometrical attributes for each component, such as area, perimeter, Euler number and orientation. In this research, the maximum length for each component was computed and the noise and small regions whose maximum length was under a threshold were deleted. In the experiments, the threshold was set to a small value to delete the noise and at the same time preserve the detailed roads. Afterwards, the morphological thinning algorithm was applied to locate road centrelines with a width of only one pixel.

3. Experiments and analysis

The proposed methodology was tested on two IKONOS multispectral datasets with three visible channels and 4-m spatial resolution. The SVM was implemented using OSU SVM toolbox 3.0. In equation (1), the weight for each band was set to the same value. The multiscale parameters T were chosen according to the characteristics of

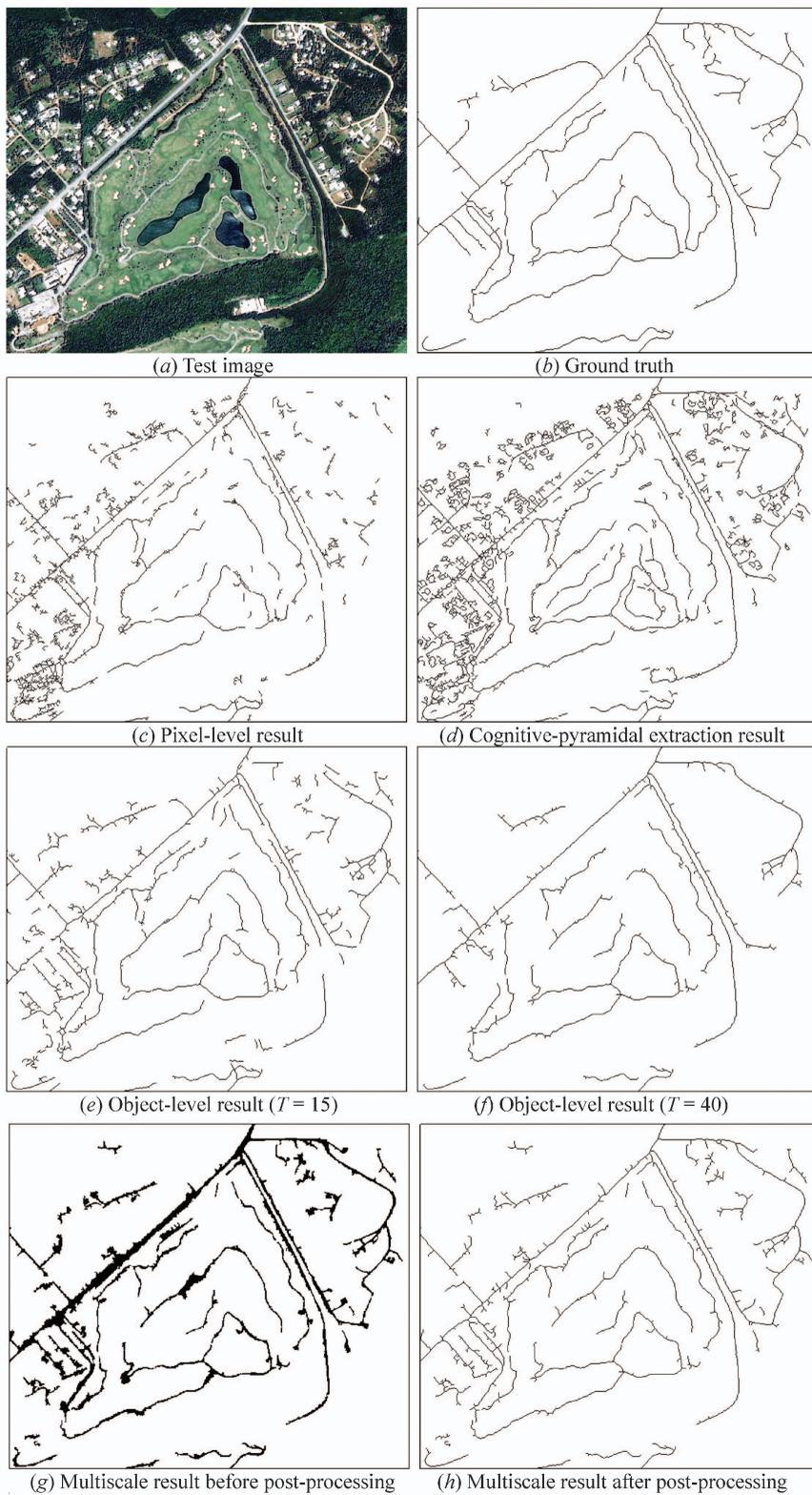


Table 1. The accuracy indexes of road detection for IKONOS experiment 1.

Feature levels	Pixel level	Cognitive-pyramidal approach	Object-based results				Multiscale fusion
			$T=15$	$T=25$	$T=30$	$T=40$	
Correctness (%)	42.6	62.6	73.6	85.1	87.5	71.9	93.1
CE (%)	71.7	78.2	42.6	27.7	20.9	12.9	18.7
OE (%)	57.4	37.4	26.4	14.9	12.5	28.1	6.9

the different sensors and the scenes. In this study, four scales were used for object-level feature extraction and multiscale fusion: $T=15$, 25, 30 and 40 (accordingly, $N=4$). The reason for not using a scale value over 40 is that large values cannot describe the actual shape of objects and can lead to the omission of some small roads. In the experiments, a multiscale cognitive pyramid algorithm (Binaghi *et al.* 2003) was implemented for comparison. The cognitive pyramidal approach is based on concentric windows and pyramidal resampling in an attempt to mimic human perception in identifying objects of different shapes and structures on different scales. In both datasets, four windows (3×3 , 7×7 , 13×13 and 25×25) were used for multiscale feature extraction and Gaussian pyramidal resampling. A multilayer perception (MLP) neural network with one hidden layer was then used to classify the multiwindow pyramids. The accuracy of road extraction is evaluated by both visual comparison and quantitative analysis. Three accuracy indices are used: correctness represents the fraction of extracted road length belonging to actual roads; omission error (OE) and commission error (CE) denote the fractions of pixels that are wrongly identified as background and that are wrongly classified as roads, respectively.

The road maps for the first experiment are shown in figure 2, and the accuracies are provided in table 1. Figure 2(a) is a test image of a rural region where the road objects show multiscale features with both wide highways and narrow trails. Moreover, some spectrally similar classes (e.g. small roofs and bare soil) may result in noise and misclassifications. Figure 2(b) is the manually extracted ground truth; 2(c) is the pixel-level result based on the SVM classification and post-processing; 2(d) is the result of the multiscale cognitive pyramids; 2(e) and 2(f) are object-level road maps of single scale ($T=15$ and 40, respectively); and 2(g) and 2(h) are the multiscale fusion results before and after the post-processing, respectively. It can be seen that the pixel-level results are sensitive to noise and many short spurious segments exist because the pixel-level approach cannot exploit the rich spatial information in high-resolution images. At the lower left quarter of figure 2(c), a roof region is wrongly identified as a road because they show similar spectral responses and cannot be discriminated using only pixel-level information. Errors can also be observed in figure 2(d). With respect to the object-based methods, roads appear more continuous and regular, and most of the noise segments are removed when spectral-structural attributes are exploited for each object. Comparing figures 2(e),

Figure 2. (a) A rural-region test image (535 by 461 pixels with 4-m spatial resolution). (b) The ground truth road network. (c) and (d) Road maps for pixel-level and multiscale cognitive pyramid approaches, respectively. (e) and (f) Object-level results for a single scale ($T=15$ and 40, respectively). (g) and (h) The multiscale fusion results before and after post-processing, respectively.



Figure 3. (a) A suburban test image (400 by 430 pixels with 4-m spatial resolution). (b) The ground truth road network. (c) and (d) Road maps for pixel-level and multiscale cognitive pyramid approaches, respectively. (e) and (f) Object-level results for a single scale ($T=15$ and 40, respectively). (g) and (h) The multiscale fusion results before and after post-processing,

2(f) and 2(h), it can be seen that small-scale segmentation is better at detecting detailed roads (e.g. zigzagging trails) while the large-scale approach is more effective at extracting the main roads (e.g. wide highways). As can be seen from figure 2(h), the multiscale fusion approach provides a more accurate road map than the single-scale algorithms. The quantitative statistics in table 1 confirm the visual analysis. The multiscale fusion algorithm gave the highest correctness and lowest omission errors. The smallest commission error was achieved by $T=40$ because the smaller the number of extracted roads, the lower will be the probability that they are wrongly identified.

Road extraction results for the second IKONOS test image are provided in figure 3 and table 2. The test area in figure 3(a) shows a suburban environment,

Table 2. The accuracy indexes of road detection for IKONOS experiment 2.

Feature levels	Pixel level	Cognitive-pyramidal approach	Object-based results				Multiscale fusion
			$T=15$	$T=25$	$T=30$	$T=40$	
Correctness (%)	33.1	51.9	60.1	77.2	83.0	70.0	90.0
CE (%)	71.7	68.1	46.6	34.6	25.1	33.5	22.3
OE (%)	66.9	48.1	39.9	22.8	17.0	30.0	9.9

where the road features include highway, trails and roads between buildings. Similar conclusions can be drawn from the visual analysis: detailed road regions were detected effectively using the small-scale approach while the large-scale approach gave better results for main roads and highways. In this experiment, the multiscale fusion gave the best results for all three indices, verifying that the information fusion algorithm is effective in exploiting the advantages in different scales.

Finally, a feature space analysis was conducted to demonstrate the efficiency of the structural attributes described in this study. Figure 4 shows the feature values and the separability for spectral and structural attributes. Figures 4(a) and 4(b) were obtained by averaging the feature values ($T=30$) within about 100 objects for each class. The statistics were normalized into $[0, 255]$ using a linear stretch. From these two figures, it can be seen that: (a) the roads, roofs and bare soil have similar spectral responses and hence the spectral approach often leads to errors and misclassifications; (b) the SI, COM and DEN attributes are very different between roofs and other objects; therefore, the road objects are well characterized by the structural features defined in equations (3), (4) and (5). Figures 4(c) and 4(d) show the separability analysis between roads and roofs using the Jeffries–Matusita (JM) distance (Richards and Jia 1999). The values of the JM distance indicate how well the selected class pairs are statistically separate. A high value indicates that the feature space can be well separated, while a low value indicates that the feature space is not well separated. The statistics in figures 4(c) and 4(d) confirm the conclusion that the structural information can enhance the separability of roads and other features and hence improve the extraction results.

4. Conclusion

In this study we proposed an effective framework for road centreline extraction from high-resolution imagery. The novelty of this paper consists in using (1) object-oriented spectral–structural information for road extraction based on SVMs, and (2) a multiscale information fusion approach. The road features are considered as objects instead of pixels; consequently, a more accurate road map can be delineated by combining both spectral and structural features. On the whole, the experiments on the two IKONOS datasets show that the multiscale fusion approach gives higher accuracies than all the single-scale levels and the multiwindow cognitive pyramid algorithm. The results verified that the proposed algorithm could efficiently exploit multilevel features and detect road objects of different scales and sizes.

Acknowledgements

We thank the editor and the two anonymous reviewers whose insightful suggestions have significantly improved this paper. This work is supported by the 863 High

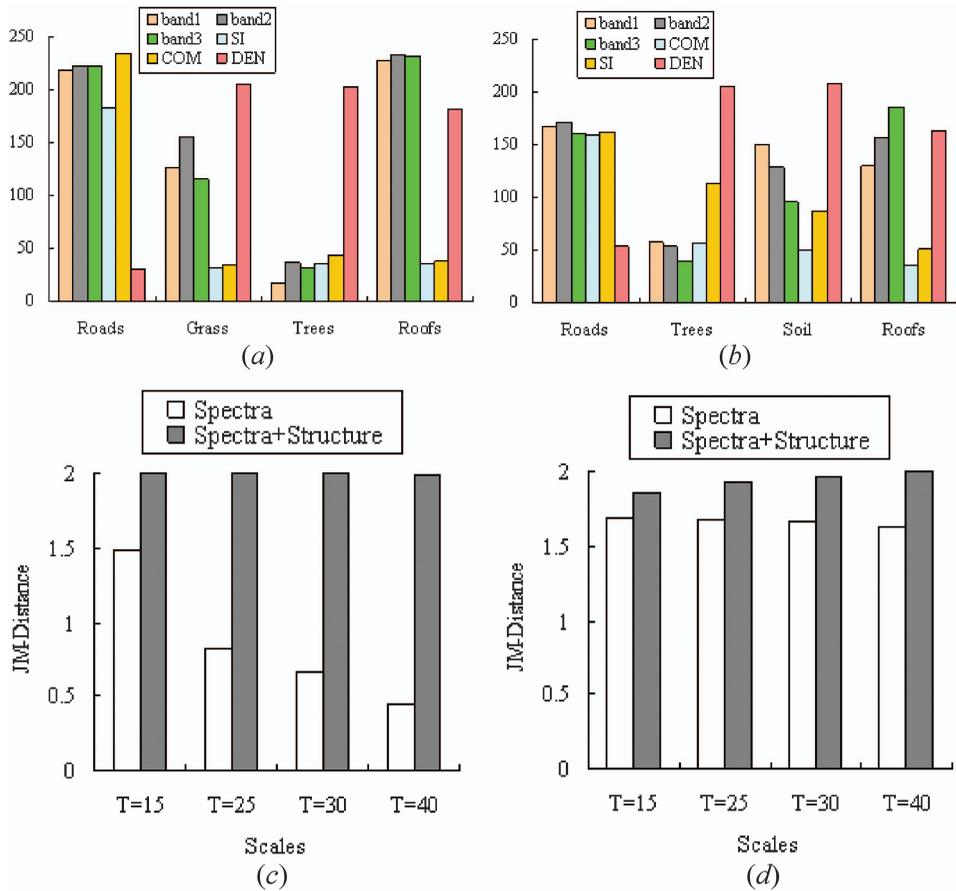


Figure 4. (a) and (b) The averaged feature values with $T=30$ for different information classes in tests 1 and 2, respectively. (c) and (d) The JM distance of spectral and spectral-structural feature space for datasets 1 and 2, respectively.

Technology Program of China under Grant 2007AA12Z148 and by the National Science Foundation of China under Grant 40771139.

References

- BARZOHR, M. and COOPER, D.B., 1996, Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**, pp. 707–721.
- BINAGHI, E., GALLO, I. and PEPE, M., 2003, A cognitive pyramid for contextual classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, pp. 2906–2922.
- COBURN, C.A. and ROBERTS, A.C.B., 2004, A multiscale texture analysis procedure for improved forest stand classification. *International Journal of Remote Sensing*, **25**, pp. 4287–4308.
- CORTES, C. and VAPNIK, V., 1995, Support vector networks. *Machine Learning*, **20**, pp. 273–297.
- DELL'ACQUA, F. and GAMBA, P., 2001, Detection of urban structures in SAR images by robust fuzzy clustering algorithms: the example of street tracking. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, pp. 2287–2297.

- FOODY, G.M. and MATHUR, A., 2006, The use of small training sets containing mixed pixels for accurate hard image classification: training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment*, **103**, pp. 179–189.
- GAMBA, P., DELL'ACQUA, F. and LISINI, G., 2006, Improving urban road extraction in high-resolution images exploiting directional filtering, perceptual grouping, and simple topological concepts. *IEEE Geoscience and Remote Sensing Letters*, **3**, pp. 387–391.
- GRUEN, A. and LI, H., 1997, Semi-automatic linear feature extraction by dynamic programming and LSB-snakes. *Photogrammetric Engineering and Remote Sensing*, **63**, pp. 985–995.
- HAY, G.J., BLASCHKE, T., MARCEAU, D.J. and BOUCHARD, A., 2003, A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS Journal of Photogrammetry and Remote Sensing*, **57**, pp. 327–345.
- HUANG, X., ZHANG, L. and LI, P., 2007, Classification and extraction of spatial features in urban areas using high-resolution multispectral imagery. *IEEE Geoscience and Remote Sensing Letters*, **4**, pp. 260–264.
- LONG, H. and ZHAO, Z., 2005, Urban road extraction from high-resolution optical satellite images. *International Journal of Remote Sensing*, **26**, pp. 4907–4921.
- RICHARDS, J.A. and JIA, X., 1999, *Remote Sensing Digital Image Analysis: An Introduction* (Berlin, Germany: Springer-Verlag).
- TUPIN, F., MAITRE, H., MANGIN, J.F., NICOLAS, J.M. and PECHERSKY, E., 1998, Detection of linear features in SAR images: application to road network extraction. *IEEE Transactions on Geoscience and Remote Sensing*, **36**, pp. 434–453.
- WANG, L., SOUSA, W.P. and GONG, P., 2004, Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery. *International Journal of Remote Sensing*, **25**, pp. 5655–5668.
- YU, Q., GONG, P., CLINTON, N., BIGING, G. and SCHIROKAUER, D., 2006, Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogrammetric Engineering and Remote Sensing*, **72**, pp. 799–811.
- ZHU, C., SHI, W., PESARESI, M., LIU, L., CHEN, X. and KING, B., 2005, The recognition of road network from high-resolution satellite remotely sensed data using image morphological characteristics. *International Journal of Remote Sensing*, **26**, pp. 5493–5508.