Object-based Conditional Random Fields for Road Extraction from Remote Sensing Image

Zhijian Huang[a, b], Fanjiang Xu[b], Lei Lu[c], Hongshan Nie[a]

[a] SPDF. School of Electronic Science and Engineering, National University of Defense Technology, Changsha, China (zhijian07@iscas.ac.cn)

[b]IIST. Key Lab. Institute of Software, Chinese Academy of Science, Beijing, China

[c]95980Troop, People's Liberation Army Air Force.

E-mail: zhijian07@iscas.ac.cn

ABSTRACT. To make full use of spatially contextual information and topological information in the procedure of Object-based Image Analysis (OBIA), an object-based conditional random field is proposed and used for road extraction. Objects are produced with an initial segmentation, then their neighbours are constructed. Each object is represented by three kinds of features, including the colour, the gradient of histogram and the texture. Formulating the road extraction as a binary classification problem, a Conditional Random Fields model learns and is used for inference. The experimental results demonstrate that the proposed method is effective.

1. Introduction

Road not only plays a central role in the vehicle navigation, but also is an important data layer in Geographical Information Systems(GIS). Automatic extraction of road can save time and labor to a great degree in updating road spatial database. The task of road extraction can be considered as a problem of binary classification. Most methods based either on individual pixel[1-3] or on constellations of homogeneous pixels, called object-based[4-6]. The latter has attracted more and more attention, since its good performance on anti-noise ability and making full use of spatial information. However, the conventional object-based approach is hard to make use of spatially contextual information and topological information, which are beneficial for accurate object extraction from remote sensing image.

Conditional Random Fields (CRF) [7, 8] offers a discriminative probabilistic graph model for fusing all kinds of contextual information, including the context of different scales, of the observation field and of the label field. Different from Markov Random Field, the CRF models the posterior probability directly, causing better predictive performance and faster inference speed[9]. The CRF also has the ability of fusing multi-feature, which is the characteristic of a reliable algorithm for object recognition and extraction. As a consequence, it has been used widely on natural scene understanding [10-12] and textual image analysis. Some researchers also try to use the CRF on remote sensing image analysis. However, all the works are based on pixels, which is easily disturbed by noise.

To take both advantages of the object-based method and the CRF, an object-based CRF for road extraction is proposed in this paper. Object is a set of connected pixels with certain homogeneous

characteristics. Objects are segmented with a Statistical Region Merging (SRM)[13] algorithm. The object neighborhood is different from the pixel neighborhood, and should be constructed before setting into the CRF. To capture the visual saliency of object in remote sensing image, not only the consistency of road region, but also the difference with its circumstance are considered with the CRF model. Three kinds of features, composing a 29 dimensions vector, are used for model training and inference. The results of experiment show that the proposed method is promising.

2. CRF Model for Road Extraction

2.1. CRF Model

The CRF is a discriminative model, proposed firstly by Professor Lafferty [7] at 2001. Different from Markov Random Field (MRF), CRF models directly the posterior probability given the entire observation.

An observation field is denoted by a set of random variables $\mathbf{x}=\{x_1, x_2, ..., x_n\}$, where x_i is a D dimensions feature vector of an object. A convention CRF is defined over $\mathbf{y}=\{y_1, y_2, ..., y_n\}$, where y_i takes a value from a label set $\mathbf{L}=\{l_1, l_2, ..., l_k\}$ corresponding to the set of object classes. With regards to road extraction, $k=2$, and means l_1 road and l_2 non-road respectively. According to [14], given data $\mathbf{y}$ the posterior probability of the CRF is a Gibbs distribution and can be written as:

$$P(\mathbf{x}\mid\mathbf{y},\boldsymbol{\theta})=\frac{1}{Z(\mathbf{y},\boldsymbol{\theta})}\exp\{-\textstyle\sum_{c\in C}\psi_c(\mathbf{x}_c,\mathbf{y},\boldsymbol{\theta})\} \tag{1}$$

where $Z(\mathbf{y},\boldsymbol{\theta})=\sum_{\mathbf{x}}\exp\{\sum_{c\in C}\psi_c(\mathbf{x}_c,\mathbf{y},\boldsymbol{\theta})\}$ is a normalizing constant called the partition function, $\boldsymbol{\theta}$ is the parameter vector to be learned by training and C is the set of all cliques. The term $\psi_c(\mathbf{x}_c,\mathbf{y},\boldsymbol{\theta})$ is known as the potential function of the clique. In this paper only the unary and pairwise potential are considered, so the equation (1) can be written as :

$$P(\mathbf{x}\mid\mathbf{y},\theta)=\frac{1}{Z(\mathbf{y},\theta)}\exp\{\overbrace{-\sum_{i\in s}\phi_i(x_i,\mathbf{y},\boldsymbol{\theta}_u)}^{unary}-\overbrace{\sum_{i\in s}\sum_{j\in\eta_i}\phi_{ij}(x_i,x_j,\mathbf{y},\boldsymbol{\theta}_p)}^{pairwise}\} \tag{2}$$

where $\boldsymbol{\theta}_u$ and $\boldsymbol{\theta}_p$ are the model parameters. The corresponding Gibbs energy is given by:

$$E(\mathbf{x})=-\log P(\mathbf{x}\mid\mathbf{y},\theta)+\log Z(\mathbf{y},\theta)=\textstyle\sum_{c\in C}\psi_c(\mathbf{x}_c,\mathbf{y},\theta)+\log Z(\mathbf{y},\theta) \tag{3}$$

So, the Maximum a Posteriori (MAP) is formulated as a problem of energy minimization:

$$\mathbf{x}^*=\arg\max_{\mathbf{x}\in\mathbf{L}}P(\mathbf{x}\mid\mathbf{y},\theta)=\arg\min_{\mathbf{x}\in\mathbf{L}}E(\mathbf{x}) \tag{4}$$

where $\mathbf{x}^*$ is the ground truth. As the partition function is constant and thus does not affect the solution of the optimization problem, it can be dropped for compactness.

2.2. Potential Functions

Theoretically, the Potential Function in CRF can be viewed as arbitrary local discriminative classifiers. This allows one to use domain-specific discriminative classifiers for structured data rather than restrict the potentials to a specific form. However, in order to be seamlessly integrated into the CRF models, the discriminative classifiers should have analytical formulation[15].

The unary potential is used to model and discriminate observations for single object. Contrary to conventional linear discriminant analysis, Logistic Regression (LR) mode requires fewer restrictive assumptions. In this case, the observations need not be normally distributed and linearly separable related to the class. Hence, the LR is usually utilized in the remote sensing image analysis. The LR model can be written as:

$$\phi_i(x_i,\mathbf{y},\boldsymbol{\theta}_u)=\log(\frac{1}{1+e^{-x_i\boldsymbol{\theta}_u{}^T y_i}})=\log(\sigma(x_i\boldsymbol{\theta}_u^T y_i)) \tag{5}$$

where $\sigma(x)=1/(1+e^{-x})$.

The pairwise potential emphasizes the spatial interaction, and the interaction strength depends on the observed objects. The ability of encoding contextual information is our focus to define a pairwise potential. In this paper, the Ising/Potts model is used to model the pairwise potential:

$$\phi_{ij}(x_i, x_j, \mathbf{y}, \boldsymbol{\theta}_p) = x_i x_j \boldsymbol{\theta}_p^T g_{ij}(y) \tag{6}$$

where $\boldsymbol{\theta}_p^T$ is the model parameter, and $g_{ij}(y)$ denotes the feature vector on adjacent objects i and j. The form in (6) acts as a data-dependent discontinuity adaptive model that will moderate smoothing when the data from the two objects are "different."

2.3. Training and inference

2.3.1. Training

For estimating the parameters $\boldsymbol{\theta} = \{\boldsymbol{\theta}_u, \boldsymbol{\theta}_p\}$, K independent identically distributed labelled training samples $\{x^k, y^k, k = 1, ..., K\}$ are available. The standard maximum-likelihood (ML) approach is utilized for training, i.e.

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \{\log(\prod_{k=1}^{K} p(x^k \mid \phi(y^k, \boldsymbol{\theta})))\} \tag{7}$$

The log likelihood of the CRF used in this paper is given by

$$L(\boldsymbol{\theta}) = \sum_{k=1}^{K} \left\{ -\sum_i \log(\sigma(x_i^k \boldsymbol{\theta}_u^T y_i^k)) - \sum_i \sum_j \log(x_i^k x_j^k \boldsymbol{\theta}_p^T g_{ij}(y^k)) + \log(z(y, \boldsymbol{\theta})) \right\} \tag{8}$$

Since there is not analytical solution of the (7), we used gradient descent method.

2.3.2. Inference

For binary classification, the MAP estimation can be computed exactly using the max-flow/min-cut type of algorithm if the probability distribution meets certain condition [16]. To release the restrictive condition, we use the loopy belief propagation (LBP), which is introduced in detail in[17].

3. Object and Neighbourhood Construction

An object is defined as a connected region with certain homogeneous characteristics. To get objects, it is necessary to utilize a segmentation algorithm, such as Normalized Cut Segmentation Algorithm[18] and Fractal Net Evolution Approach (FNEA)[19, 20]. The SRM is a state-of-the-art segmentation algorithm. It is able to capture the main structural components of image using a simple but effective statistical analysis, and it has the ability to cope with significant noise corruption, handle occlusions. Hence, the SRM algorithm is exploited[13] in this paper.

Different from pixel with simple 4-connective or 8-connective neighborhood, the neighborhood of object is more complex. A neighbor of object A is an object which has a common edge with object A. To construct neighborhood for each object, it is necessary to analyze the spatial topology relation. Tracing the contour of each object, its neighbors can be found out and reserved.

Based on the object neighborhoods, the pairwise potential of CRF is learned and used for inference. By this way, the contextual information carried by neighborhood objects can be made use of.

4. Features Extraction and Selection

To capture the characteristics of road objects, this paper selects three kinds of features, including the color, the gradient of histogram (GoH), and the Texton[21]. With these features, a 29-D feature vector for each object is constructed, shown in Table 1 in detail.

Though roads in the real world are constructed with various kinds of material, there are three typical road: the concrete, the asphaltum and the mud. Hence, the colors are important characteristics of road in remote sensing image. The color of each object consists of six elements,

including the red mean, the green mean, the blue mean, the hue mean, the hue standard deviation and the saturation mean.

Table 1. Features Used in The CRF

			Dimension
Color (6 dimension)	mean	Red channel	1
		Green channel	1
		Blue channel	1
		the saturation	1
		the hue	1
	standard deviation	the hue	1
GoH (6 dimension)	mean		1
	energy		1
	entropy		1
	the first three moment		3
Texton (17 dimension)	Gaussian Kernel	Red channel	3
		Green channel	3
		Blue channel	3
		X direction derivative	2
		y direction derivative	2
	Laplacians of Gaussian Kernel		4

The GoH can represent the smoothness of the object surface. Instead of incrementing the counts in the histogram, each count is weighted by the gradient magnitude at that pixel. If the object is smooth, the gradients will be very small and the mean magnitude of the histogram over all the bins will also be small. In contrast, if it is a textured region such as grassland, the histogram will have approximately uniformly distributed bin magnitudes. Finally, if super-pixel contains a few straight lines and/or edges embedded in smooth background, as is the case for the structured class, a few bins will have significant peaks in the histogram in comparison to the other bins[22]. The GoH of super-pixel is a six dimension vector including the mean of gradient magnitude, the energy, the entropy and the first three moment.

The textons mentioned in[21] is a kind of discriminative feature, which have been proven effective in categorizing. Using Mahalanobis distance, a dictionary of textons is learned by convolving a 17-dimensional filter bank with all the training images and running K-means clustering on the filter responses. For extracting the texton, each pixel in each image is assigned to the nearest cluster center. The slightly different here is that the texton is object level which is an average of its pixel level textons.

It is notable that all features mentioned above should be normalized for further calculation.

Figure 1. Examples of image database. The top row are the initial images, and the bottom row are the ground truths. The green regions are roads, and the green regions are non-roads.

5. Database Description

For training and inference, it is necessary to construct a database containing ground truth. We collect a group of images, and label the ground truth manually. The database contains 50 training and 100 test images in all, including QuickBird(download from: http://www.digitalglobe.com), IKONS(download from: http://www.geoeye.com/CorpSite/), SPOTS 5(download from: http://www.spotimage.com.cn/) .

The ground truths are manually described with the polygon drawing tool in ArcGis. They are saved as vector format. Road objects are labeled as 1 while non-road labeled as 0. Noted that we intentionally collect images with various backgrounds (e.g. urban, suburban and rural). The road in our image database also appear in various kinds of brightness, clarity, colors and shapes. In a word, we try our best to let the database be representative and be unbiased to any specified scene. Some examples of the database are show in the Figure 1.

6. Experiment and Discussion

In our experiments, the training image set includes 15 rural images, 15 suburban images and 20 urban images. The left 100 images are used for testing. During the training procedure, if above 50% area of an object is road in ground truth, the object is treated as road object. Some example of the result are showed in Figure 2, where the green regions are roads, and the black regions are non-roads.

Seen from the Figure 2, the scenes are complex. Roads appear in various kinds of shapes, such as interchange bridges and crossroads. The color and brightness of the roads are also not consistent. The roads are always occluded by trees, high buildings and big vehicles. The most serious is that there are many non-road regions which has the similar color, smoothness and texture with road region.

However, the results are exciting. Most of the road regions are extracted correctly. That should owe to the CRF model which considers the spatial contextual information and the topology information simultaneity.

There are also some false regions and missing regions. The false regions due to the situation that some non-road super-pixels are indeed similar with the road one, such as the bareland, the houses et al.. The extraction accuracy is also affected by the primary segmentation, i.e. the SRM. If the SRM merges a small road region with a non-road region (i.e. under-segment), the road region may miss in the last results.

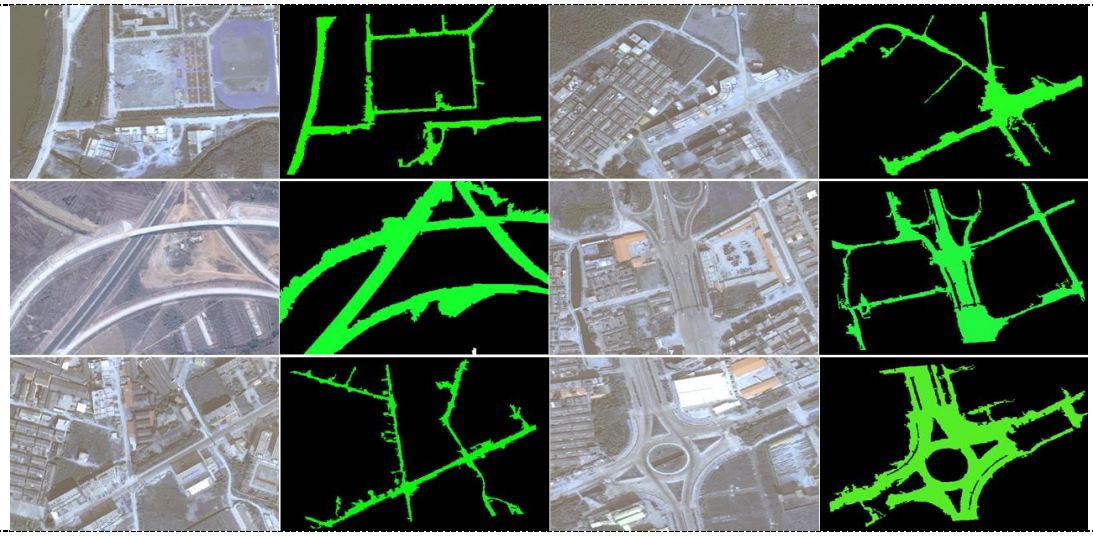

Figure 2. Results of experiment. The green regions are roads, and the black regions are non-roads.

7. Conclusion

To make full use of spatially contextual information and topological information in the object-based method, an object-based CRF model for road extraction is proposed. The SRM is used as an initial segmentation to produce objects. Three kinds of features (including the color and the GoH and the Texton) are exploited for model training and inference. The results of experiment show that the proposed method is effective.

Reference

[1] Amo M, Martinez F, and Torre M 2006 Road extraction from aerial images using a region competition algorithm *IEEE T. Image Process.* 15 1192-1201.
[2] Das S, Mirnalinee T T and Varghese K 2011 Use of Salient Features for the Design of a Multistage Framework to Extract Roads From High-Resolution Multispectral Satellite Images *IEEE T. Geosci. Remote*. 49 3906-3931.
[3] Mancini A, Frontoni E and Zingaretti P 2010 Road change detection from multi-spectral aerial data *ICPR 2010* 448-451.
[4] Zhang L, Zhang J, Zhang D, Hou X and Yang G 2010 Urban road extraction from high-resolution remote sensing images based on semantic model. 18th International Conference on Geoinformatics.
[5] M Song and D Civco 2004 Road extraction using SVM and image segmentation *Photogramm. Eng. Rem. S.* 70 1365-1371.
[6] H Xin and Z Liangpei 2009 Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines *Int. J. Remote Sens.* 30 1977-1987.
[7] J Lafferty, Andrew M and Fernando P 2001 Condition Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data *In International Conference on Machine Learning*.
[8] H M Wallach 2004 Conditional Random Fields: An Introduce *University of Pennsylvania CIS Technical Report.*
[9] Zhong P and Wang R 2006 Object Detection Based on Combination of Conditional Random Field and Markov Random Field *ICPR 2006* 160-163.
[10] Kohli P, Ladicky L and Torr Philip H S 2009 Robust higher order potentials for enforcing label consistency *Int. J. Comput. Vision* 82 302-324.
[11] Ladicky, *et al.*. Associative hierarchical CRFs for object class image segmentation. *IEEE ICCV 2009*, 739-746, 2009.
[12] Ladicky L, Russell C, Kohli P and Torr Philip H S 2010 Graph cut based inference with co-occurrence statistics *ECCV 2010* 239-253.

[13] R Nock and F Nielsen 2004 Statistical region merging. *IEEE T. Pattern. Anal.* 26 1452-1458.
[14] J M Hammersley and P Clifford 1971 Markov field on finite graph and lattices *unpublished*.
[15] Zhong P and Wang R 2007 A Multiple Conditional Random Fields Ensemble Model for Urban Area Detection in Remote Sensing Optical Images *IEEE Geosci. Remote*. 45 3978-3988.
[16] V Kolmogorov and R Zabin 2004 What energy functions can be minimized via graph cuts? *IEEE T. Pattern. Anal.* 26 147-159.
[17] S Jian, Nanning Z and Heung Y S 2003 Stereo matching using belief propagation. *IEEE T. Pattern. Anal.* 25 787-800.
[18] S Jianbo and J Malik 2000 Normalized cuts and image segmentation *IEEE T. Pattern. Anal.* 22 888-905.
[19] M. Baatz and A Schape 2000 Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation *Angewandte Geographische Information* 4 12-23.
[20] Benz U C, Peter H, Gregor W, Iris L and Heynen M 2004 Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information *ISPRS J. Photogramm.* 58 239-258.
[21] Shotton J, Winn J, Rother C and Antonio C 2006 TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation *ECCV 2006* 1-15.
[22] S Kumar and M Hebert 2006 Man-made structure detection in natural images using a causal multiscale random field IEEE CVPR 2003 119-126.