Journal of
Applied Remote Sensing

Adaptive support vector machine and Markov random field model for classifying hyperspectral imagery

Shanshan Li
Bing Zhang
Dongmei Chen
Lianru Gao
Man Peng

Adaptive support vector machine and Markov random field model for classifying hyperspectral imagery

Shanshan Li,[a] Bing Zhang,[a] Dongmei Chen,[b] Lianru Gao,[a] and Man Peng[c]

[a] Chinese Academy of Sciences, Center for Earth Observation and Digital Earth, 100094 Beijing, China
ssli@ceode.ac.cn

[b] Queen's University, Department of Geography, K7L3N6 Kingston, Ontario, Canada

[c] Institute of Remote Sensing Application, Chinese Academy of Sciences, 100101 Beijing, China

Abstract. Markov random field (MRF) provides a useful model for integrating contextual information into remote sensing image classification. However, there are two limitations when using the conventional MRF model in hyperspectral image classification. First, the maximum likelihood classifier used in MRF to estimate the spectral-based probability needs accurate estimation of covariance matrix for each class, which is often hard to obtain with a small number of training samples for hyperspectral imagery. Second, a fixed spatial neighboring impact parameter for all pixels causes overcorrection of spatially high variation areas and makes class boundaries blurred. This paper presents an improved method for integrating a support vector machine (SVM) and Markov random field to classify the hyperspectral imagery. An adaptive spatial neighboring impact parameter is assigned to each pixel according to its spatial contextual correlation. Experimental results of a hyperspectral image show that the classification accuracy from the proposed method has been improved compared to those from the conventional MRF model and pixel-wise classifiers including the maximum likelihood classifier and SVM classifier. [DOI: 10.1117/1.3609847]

Keywords: hyperspectral; image classification; support vector machine; Markov random field.

Paper 10185R received Nov. 29, 2010; revised manuscript received May 30, 2011; accepted for publication Jun. 21, 2011; published online Jul. 15, 2011.

1 Introduction

Hyperspectral remote sensing imagery provides detailed spectral information by deploying hundreds of spectral bands for the same area. In theory, the high spectral resolution of hyperspectral imagery offers the potential for more accurate land-cover classification than that achieved by rough spectral resolution of traditional multi-spectral imagery. With the development of hyperspectral imaging sensors, the spatial resolution of hyperspectral imagery has also been significantly improved. Spatial resolution determines the level of spatial details on the Earth's surface that can be observed from remotely sensed data. Hyperspectral imagery with high spatial resolution offers the opportunity of exacting more detailed land use information than imagery with a medium or coarse spatial resolution. However, previous research has shown that improved spatial resolution does not always lead to better classification maps when employing conventional pixel-wise classification methods due to higher spectral intra-class variations (samples in the same class may have different spectral signatures) within land-cover units, which decrease their spectral separability and result in a "salt and pepper" appearance on the classification map.[1–3] The background noise and the mixed pixels blur edges between different classes.

1931-3195/2011/$25.00 © 2011 SPIE

Conventional spectral pixel-wise classification methods, such as minimum distance classification, maximum likelihood classification (MLC), and spectral angle mapping, conduct a classification by comparing the spectral similarity of each pixel with prior knowledge from training samples. These pixel-wise classification methods analyze data without incorporating spatial information, and remote sensing data are usually treated as unordered lists of spectral measurements without particular spatial arrangement.

Incorporating both spatial and spectral patterns has been identified as a useful approach to achieve better land cover classification for hyperspectral data. For example, contextual-based approaches have been developed to deal with intra-class spectral variations in order to improve pixel-based classification results.[4,5] In another application, context information was extracted as an additional feature in the image pre-processing and incorporated into image classification. Pixel relaxation labeling has also been used to modify initial labels based on the spatial relationship between the neighboring pixels.[6]

The Markov random field (MRF) contextual classifier is a commonly used contextual classification method that modifies the pixel-based discriminant function through an additional term that incorporates spatial information and recognizes the spatial corrections,[7] MRF-based models are effective in improving classification results.[8,9] In the widely used MRF classifier, the maximum likelihood decision rule is typically formulated as the minimization of a suitable energy function. However, there exist two limitations in conventional MRF for classifying hyperspectral images.

First, Gaussian probability density function is used to estimate the class probability of pixels on the scene in conventional MRF. The efficiency of maximum-likelihood classification depends on accurate estimation of the mean vector $\mathbf{m}$ and covariance matrix Σ for each spectral class, which needs a sufficient number of training samples of each class. Therefore, reliable estimation of statistical parameters of a high dimensional density function with a limited number of samples becomes a problem for hyperspectral imagery. Alternatively, a support vector machine (SVM) is attractive for its effectiveness to directly analyze hyperspectral data in the hyperdimensional feature space.[10] Some studies have been undertaken to integrate SVM with MRF for classification of remote sensing images. Liu incorporated the spatial-temporal information to monitor forest disease spread.[11] The algorithm first initialized the individual image by a spectral classification using SVM, and then a MRF model is used to model the spatial-temporal contextual prior probabilities. Bovolo[12] integrated the SVM with MRF to classify the synthetic aperture radar images, and SVM is used to provide the initial classification and prior probabilities for the proposed MRF-based method to segment the remote sensing images.[13] In this paper, we integrate SVM and MRF for hyerspectral image classification.

Second, conventional MRF treats neighboring impact the same for all pixels, which is not the case in real images. An adaptive weight parameter is proposed for each pixel according to its spatial contextual correlation.

The rest of the paper is organized as follows. Section 2 presents the background of a conventional MRF model. The proposed method is presented in details in the Sec. 3. An experiment using a real hyperspectral image is presented in Sec. 4. Finally, Sec. 5 concludes the paper.

2 Conventional Markov Random Field Model

Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m, \ldots, \mathbf{x}_N\}$ be a set of pixels in a hyperspectral image, while N is the total number of pixels in the image. Further, each $\mathbf{x}_m$ represents the spectral vector (with B bands) of a particular pixel m, and can be expressed as $\mathbf{x}_m = \{x_{mj}, j = 1, \ldots, B\}$, while x_{mj} represents the value (either reflectance or digital number) of pixel m for the band j. Supposing that there are total K classes, to decide into which class the pixel m should be classified to, the maximum likelihood classification is based on an estimated probability function $P(C_k|\mathbf{x}_m)$ for each class C_k ($k = 1, 2, \ldots, K$). By comparing all $P(C_k|\mathbf{x}_m)$ for different classes, pixel $\mathbf{x}_m$ is assigned to the class with the highest probability.

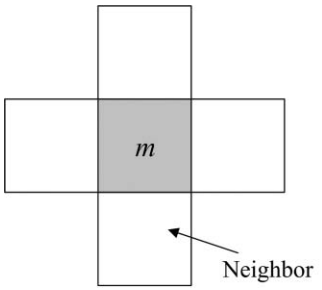

Fig. 1 First-order neighborhood of pixel *m*.

Since the pixel and neighbors are contextually dependent on the scene, adjacent pixels are likely to come from the same class. Therefore, when the spatial correlation of pixels is taken into consideration, the objective is to find the class C_k that maximizes $P(C_k|\mathbf{x}_m, C_g)$ for pixel m, where $C_g$represents the class to which pixel m's neighborhood pixels are assigned. The first-order neighborhood system is shown in Fig. 1.

According to the Bayes' theorem and total probability formula, $P(C_k|\mathbf{x}_m, C_g)$ can be expressed as

$$P(C_k|\mathbf{x}_m, C_g) = P(\mathbf{x}_m|C_k)P(C_k|C_g)\big/P(\mathbf{x}_m), \tag{1}$$

where $P(\mathbf{x}_m)$ is the probability of finding pixel $\mathbf{x}_m$ on the scene. Since it does not contribute to the decision concerning the correct label for pixel m, it can be removed from the expression. $P(\mathbf{x}_m|C_k)$ represents the probability of pixel m assigned to class C_k on the scene. $P(C_k|C_g)$ represents the conditional probability that pixel m is assigned to class C_k when its neighbors are assigned to class C_g. Suppose the class on pixel m assumes a multivariate Gaussian class conditional density function, then

$$P(\mathbf{x}_m|C_k) = \exp\left(-\frac{1}{2}\ln|\Sigma_k| - \frac{1}{2}(\mathbf{x}_m - \mathbf{m}_k)^T\Sigma_k^{-1}(\mathbf{x}_m - \mathbf{m}_k)\right), \tag{2}$$

where $\mathbf{m}_k$ and Σ_k is the mean vector and covariance of class C_k obtained from training data.

The probability $P(C_k|C_g)$ represents the conditional probability that pixel m is assigned to class $C_k$when its neighbors are assigned to class C_g, and it is called the spectral component of the MRF model. According to the property of MRFs, this conditional probability can be expressed in the form of a Gibbs distribution as

$$P(C_k|C_g) = \frac{1}{\mathbf{Z}}\exp\{-U(C_k)\}, \tag{3}$$

in which $U(\omega_{cm})$ is called an energy function and Z is the normalizing constant. Based on the Ising model,[14]

$$U(C_k) = \sum_g \beta[1 - \delta(C_k, C_g)], \tag{4}$$

where $\delta(C_k, C_g)$ is the Kroneker delta function and $\beta > 0$ is a parameter with a value fixed by the users when applying the MRF to control the influence of the neighbors. Drop item $1/\mathbf{Z}$, conventional MRF probability function can be expressed as

$$P_k(\mathbf{x}_m) = -\frac{1}{2}\ln|\Sigma_k| - \frac{1}{2}(\mathbf{x}_m - \mathbf{m}_k)^T\Sigma_k^{-1}(\mathbf{x}_m - \mathbf{m}_k) - \beta\sum_g[1 - \delta(C_k, C_g)]. \tag{5}$$

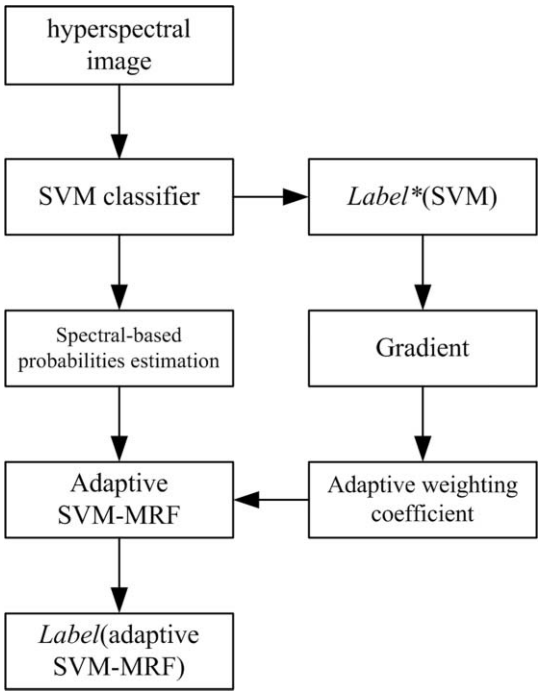

Fig. 2 Flowchart of the adaptive SVM-Markov random field classification approach scheme.

The final classification result of pixel m is assigned to the class that maximizes the discriminant function $P_k(\mathbf{x}_m)$.

3 Adaptive Support Vector Machine-Markov Random Field Model

From Eq. (5), MRF-based probability function can be expressed in a more general form as

$$P_k(\mathbf{x}_m) = a(C_k) + \beta b(C_k), \tag{6}$$

where $a(C_k) = -(1/2)\ln|\Sigma_k| - (1/2)(\mathbf{x}_m - \mathbf{m}_k)^T \Sigma_k^{-1}(\mathbf{x}_m - \mathbf{m}_k)$, $b(C_k) = \sum_g [1 - \delta(C_k, C_g)]$ in the conventional Markov random field. $a(C_k)$ represents the spectral measure probability of pixel m belonging to class C_k, $b(C_k)$ is the spatial measure probability of pixel m belonging to class C_k, and β is the weighting coefficient of spatial measure. Therefore, the MRF discriminant function consists of two main parts: the spectral and the spatial estimators. The MRF-based approach first applies a spectral-based pixel-wise classification to get the initial classification. Then spatial contextual information is used to refine the initial classification. These two processes are iteratively applied over the full scene until the class labeling is stabilized. The flowchart of the improved MRF classification approach scheme is shown in Fig. 2.

3.1 Spectral Estimator Based on Support Vector Machine

An SVM is especially advantageous for estimating statistical parameters in the presence of heterogeneous classes and small training sets. As shown in Fig. 3, it can directly and effectively analyze hyperspectral data in a hyper-dimensional feature space by the optimal separating hyperplane without the need of any feature-reduction procedure.[10] Therefore, an SVM is proposed for estimating the spectral-based probabilities in the MRF for hyperspectral image classification.

To derive an estimation of the probabilities $P(\mathbf{x}_m|C_k)$, the posterior probability from SVM's outputs was calculated according to an improved implementation[15] of Platt's

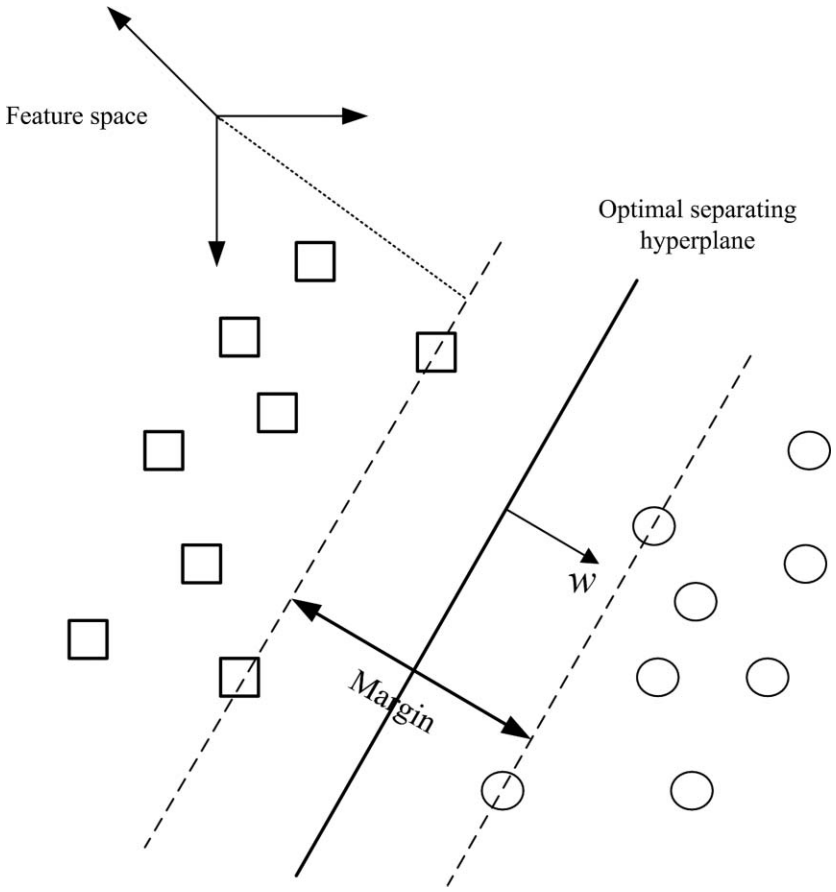

Fig. 3 Optimal separating hyperplane in SVMs of feature space.

posterior probabilities.[16] A sigmoid function can be fitted to the discriminant function

$$P(C_k|\mathbf{x}_m) = \frac{1}{1+\exp[S_k f(\mathbf{x}_m) + R_k]}, \tag{7}$$

where S_k and R_k are the sigmoid parameters computed on training samples set $(\mathbf{x}_i, C_k)$ by minimizing the cross-entropy error function.

$$\begin{cases} \text{minimize} : -\sum_k t_k \log(P(C_k|\mathbf{x}_i)) + (1-t_k)\log(1-P(C_k|\mathbf{x}_i)) \\ \text{subject to} : t_k = \dfrac{C_k+1}{2} \end{cases} \tag{8}$$

in which t_k represents target probabilities from the training set. The minimization can be performed using a model-trust minimization algorithm.[17]

According to Bayes' theorem,

$$p(\mathbf{x}_m|C_k) = \frac{p(\mathbf{x}_m)}{p(C_k)} p(C_k|\mathbf{x}_m) \tag{9}$$

in which $p(C_k)$ represents prior probability of the scene labeling and $P(\mathbf{x}_m)$ is the probability of pixel $\mathbf{x}_m$ on the scene, they do not affect the decision rule. Therefore, the spectral term of MRF discriminant function based on the probabilistic SVM pixel-wise classifier is

$$a(c) = -\ln(1+\exp[S_k f(\mathbf{x}_m) + R_k]). \tag{10}$$

3.2 Adaptive Weighting Coefficient of Spatial Estimator

The spatial term of the conventional MRF is a measure of spatial support for the pixel's likelihood of belonging to class C_k. In the conventional MRF approach, the same value of β represents the identical influence of neighbors for each pixel. Therefore, a larger value of β improves the classification accuracy more in homogeneous regions, but pixels at class boundaries are at the risk of overcorrection,[18] which blurs edges between different fields.

The weighting coefficient of the spatial estimator should be assigned to each pixel according to its spatial contextual correlation. When pixel m locates in a homogeneous region, then all of the neighbors are labeled the same as the pixel. The spatial correlation assumption is strong. A high value of weighting coefficient should be applied. On the contrary, pixel m lies in the class boundary, applying a low weighting coefficient, even zero, can avoid reduction of the likelihood that pixel m belongs to class C_k based on the spectral estimator. As a result, the appropriate weighting coefficient could be determined for each pixel according to the location of the pixel on the scene. The edge information is proposed to integrate into the spectral estimator of MRF.

As shown in Fig. 2, in order to differentiate weight coefficient of pixels in class boundary and homogeneous area, class label (*Label**) gradient of pixel is calculated using Laplacian of Gaussian (LoG) edge detection operator for its insensitivity to noise. It combines Gaussian smooth filter and Laplacial technique,[19] which can be represented as

$$\mathrm{LoG}(x, y) = -\frac{1}{\pi\sigma^4}\left[1 - \frac{x^2 + y^2}{2\sigma^2}\right] e^{-\frac{x^2+y^2}{2\sigma^2}}, \tag{11}$$

where σ is the Gaussian kernel of width.

The gradient result $\nabla(\mathbf{X}) = \{\rho_j \in R,\ j = 1, 2, \ldots N\}$, ρ_j represents the class label gradient of each pixel. A large gradient value usually means that the central pixel is located close to or in class boundaries, whereas a small gradient value represents that the central pixel is in a homogeneous area.

According to the relationship between the weighting coefficient and the gradient of the pixel, we define the following "edge function" thus

$$\varepsilon(\mathbf{x}_m) = \frac{\alpha}{\alpha + \rho_m}, \tag{12}$$

where α is a parameter controlling the range of weighting coefficient and $\varepsilon(\mathbf{x}_m)$ is normalized into (0, 1). The optimal value of the parameter is experimentally derived. Therefore, the adaptive weighting coefficient for pixel m is

$$\beta_m = \beta\varepsilon(\mathbf{x}_m). \tag{13}$$

The spatial term is proposed

$$b(c) = -\beta_m \sum_g [1 - \delta(C_k, C_g)]. \tag{14}$$

As a result, the adaptive SVM-Markov random field discriminant function is

$$P_k(\mathbf{x}_m) = -\ln(1 + \exp[S_k f(\mathbf{x}_m) + R_k]) - \varepsilon(\mathbf{x}_m)\beta \sum_g [1 - \delta(C_k, C_g)]. \tag{15}$$

The MRF classification problem is solved through minimizing function (15). In this research, the simulated annealing algorithm with logarithmic scheme proposed by Geman and Geman has been used to reach the global minimum of Eq. (18).[20]

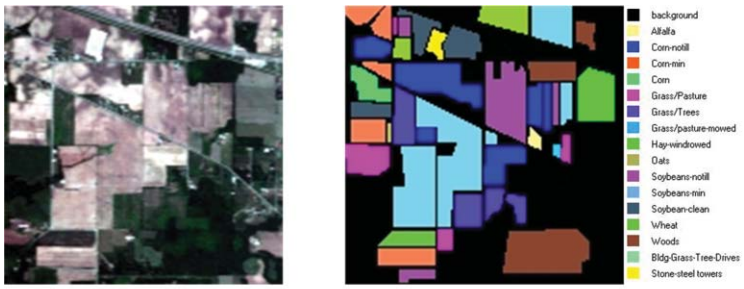

Fig. 4 (a) Three-band color composite. (b) Reference data.

4 Experiment

4.1 Hyperspectral Data

AVIRIS data covering an area of mixed agriculture and forestry landscape in the Indian Pine test site in North-western Indiana were used in the experiment. The test image has 145 by 145 pixels, with a spatial resolution of 20 m and 220 spectral channels. A total of 20 bands [104 to 108, 150 to 163, 220] are removed to exclude the water absorption region. The remaining 200 spectral channels were used for the experimental study.[21] A three-band false-color composite image and reference data are shown in Fig. 4. Sixteen classes of interest are considered in this experiment. Training samples have been randomly chosen from the reference map for each class. The details of classes and the number of training sizes for each class are listed in Table 1.

To test the effectiveness of the proposed approach, four classifiers were applied to classifying the hyperspectral image, including two pixel-based classifiers (maximum likelihood classifier and SVM) and two spectral-spatial classifiers (conventional MRF and adaptive SVM-Markov random field).

As shown in Table 1, compared with the number of spectral bands in the hyperspectral image, the number of reference samples is so small, especially for Alfalfa, Grass/Pasture-mowed, and Stone-steel towers, that classifiers based on covariance estimation cannot directly

Table 1 Overall and individual classification accuracies in percentage for hyperspectral image from different classification methods.

Class	Training samples	Test samples	MLC	SVM	Conventional MRF	Adaptive SVM-MRF
Alfalfa	10	44	77.78	50.00	94.44	**98.15**
Bldg-Grass-Tree-Drivers	100	280	65.53	53.16	66.05	85.00
Corn	100	134	58.55	66.24	91.88	**90.17**
Corn-min till	100	734	58.99	53.84	76.38	87.65
Corn-no till	100	1334	53.97	66.67	74.13	81.80
Grass/Pasture	100	397	86.32	89.34	92.15	**96.15**
Grass/Pasture-mowed	10	16	85.31	76.92	90.40	**96.15**
Grass/Tree	100	647	91.57	92.37	95.58	**99.20**
Hay-windrowed	100	389	96.93	98.57	97.55	**98.39**
Oats	10	10	85.00	100.00	92.40	**100.00**
Soybeans-clean till	100	514	78.99	52.77	88.93	89.09
Soybeans-min till	100	2368	54.67	84.68	63.21	**95.58**
Soybeans-no till	100	868	54.34	69.21	79.55	88.33
Stone-steel towers	50	45	94.74	90.53	96.84	**98.95**
Wheat	100	112	96.23	91.98	99.06	**99.53**
Woods	100	1194	84.93	97.14	96.29	**98.92**
Overall	–	–	64.68	77.82	80.42	**92.35**
Kappa	–	–	60.69	74.42	77.99	**91.27**

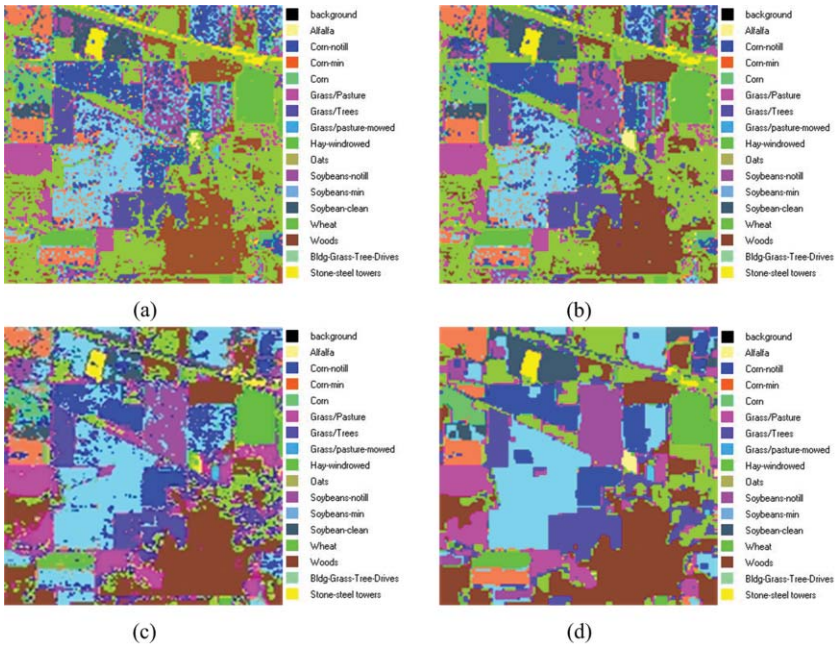

Fig. 5 (a) Maximum likelihood classification map. (b) Conventional MRF classification map. (c) SVM classification map. (d) Adaptive SVM-MRF classification map.

use original hyperspectral bands. Principle component transformation (PCT) was applied to reduce the dimension of spectral bands before classification. The first 10 components of PCT, which account for 99.75% of the original information in the original hyperspectral data, were then used in MLC and conventional MRF classifications.

In the experiment, the probabilistic one-versus-one SVM classification with the Gaussian radial basis function (RBF) kernel was applied. The optimal parameters C, which controls the amount of penalty during the SVM optimization, and γ, which controls spread of the RBF kernel, were chosen by five-fold cross validation in the probabilistic SVM.[22] Past experiments suggest that the MRF-based model is not sensitive to a particular setting of β.[23] As long as β is large enough, the accuracy of classification is quite stable. Our experiments found that the setting of $\beta \geq 5.0$ gives stable classification results. The adaptive optimal value of the parameter $\alpha = 10$ was experimentally determined.

4.2 Results

Figure 5 shows corresponding classification maps from four classifiers. It can be seen that the classification from spectral-spatial classifiers led to more homogeneous regions and less salt and pepper when compared to those from the pixel-based methods. Furthermore, a classification map from the adaptive SVM-Markov random field classifier can keep the spatial structure more complete and make edges sharper.

Table 1 summarizes the global and individual class-specific classification accuracies for the hyperspectral image, including overall average accuracies and kappa coefficient.[24] The overall accuracies from the maximum likelihood classifier and SVM were 60.69 and 74.42%, respectively, while the conventional MRF and adaptive MRF achieved overall accuracies of 77.99 and 91.27%. It is clear that classification accuracies are improved through the addition of contextual information. Moreover, adaptive SVM-MRF gives the best class-specific accuracies for most classes and its overall accuracy is 12% higher than conventional-MRF.

5 Discussion and Conclusions

The Markov random field process incorporates contextual information into the pixel-wise spectral classification discriminant function. Based on the analysis of the conventional MRF model, the approach of MRF-based is generalized as the polynomial combination of spectral and spatial terms. To solve the problems of conventional MRF for the classification of hyperspectral images, an adaptive SVM-Markov random field method has been proposed in this paper.

First, the conventional MRF uses the maximum likelihood estimation that cannot directly obtain reliable class statistic parameters with small training samples for hyperspectral image classification. A support vector machine is employed to provide posterior probabilities and improve the quality of initial spectral-based classification.

Second, the degree of the contextual relationship varied in different areas of image. In a homogeneous region, a pixel shows a strong spatial correlation with its neighboring pixels, so a large weighting coefficient can make the region become more homogeneous. On the other hand, the contextual relationship is weak in class boundaries and thus a small value is needed. In this paper, an edge detection operator is used to calculate the pixel label gradient value, which is applied to determine the appropriate weighting coefficient for each pixel. Therefore, the adaptive weight for the spatial term is beneficial in getting rid of noise in the homogeneous regions and preserving the class boundaries. The experimental results using the real hyperspectral image demonstrate the effectiveness of our proposed method.

Acknowledgments

The authors wish to thank Dr. D. Landgrebe of Purdue University, West Lafayette, Indiana, USA, for providing the AVIRIS data set and the documents. Funding of this research by National Basic Research Program of China (973 Program, 2009CB723902), National High Technology Research and Development Program 863(2008AA12Z113), and National Natural Science Foundation of China (40901225) is acknowledged.

References

1. S. Lennartz and R. Congalton, "Classifying and mapping forest cover types using Ikonos imagery in the northeastern United States," in *Proc. of the ASPRS 2004 Annual Conference*, pp. 23–28, CDROM, Denver (2004).
2. B. Zhang, X. Jia, Z. Chen, and Q. Tong, "A patch-based image classification by integrating hyperspectral data with GIS," *Int. J. Remote Sens.* **27**, 3337–3346 (2006).
3. Y. Zhao, L. Zhang, P. Li, and B. Huang, "Classification of high spatial resolution imagery using improved gaussian markov random-field-based texture features," *IEEE Trans. Geosci. Remote Sens.* **45**, 1458–1468 (2007).
4. D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *Int. J. Remote Sens.* **28**, 823–870 (2007).
5. J. Stuckens, P. R. Coppin, and M. E. Bauer, "Integrating contextual information with per-pixel classification for improved land cover classification," *Remote Sens. Environ.* **71**, 282–296 (2000).
6. J. Richards and X. Jia, *Remote sensing digital image analysis: an introduction*, Springer-Verlag, Berlin (2006).
7. S. Magnussen, P. Boudewyn, and M. Wulder, "Contextual classification of Landsat TM images to forest inventory cover types," *Int. J. Remote Sens.* **25**, 2421–2440 (2004).
8. R. Chellappa and S. Chatterjee, "Classification of textures using Gaussian Markov random fields," *IEEE Trans. Acoust., Speech, Signal Process.* **33**, 959–963 (1985).
9. T. Kasetkasem, M. K. Arora, and P. K. Varshney, "Super-resolution land cover mapping using a Markov random field based approach," *Remote Sens. Environ.* **96**, 302–314 (2005).

10. F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.* **42**, 1778–1790 (2004).
11. D. Liu, M. Kelly, and P. Gong, "A spatial-temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery," *Remote Sens. Environ.* **101**, 167–180 (2006).
12. F. Bovolo and L. Bruzzone, "A Context-sensitive Technique Based on Support Vector Machines for Image Classification," in *Proc. IEEE PReMI*, pp. 260–265, Springer, India (2005).
13. A. Farag, R. Mohamed, and A. El-Baz, "A unified framework for MAP estimation in remote sensing image segmentation," *IEEE Trans Geosci. Remote Sens.* **43**, 1617–1634 (2005).
14. J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis*, Springer, Berlin (2006).
15. H. Lin, C. Lin, and R. Weng, "A note on Platt's probabilistic outputs for support vector machines," *Mach. Learn.* **68**, 267–276 (2007).
16. J. Platt, "Probabilistic outputs for support vector machines," in *Advances in Large Margin Classifiers*, A Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, Eds. Cambridge, MA, MIT Press, pp. 61–74 (2000).
17. P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright, "Aspects of mathematical modelling related to optimization," *Appl. Math. Model.* **5**, 71–83 (1981).
18. X. Jia and J. Richards, "Managing the spectral-spatial mix in context classification using Markov random fields," *IEEE Geosci. Remote Sens.* **5**, 311–314 (2008).
19. G. Sotak, "The Laplacian-of-Gaussian kernel: A formal analysis and design procedure for fast, accurate convolution and full-frame output*," *Comput. Vis. Graph Image Process.* **48**, 147–189 (1989).
20. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *J. Appl. Stat.* **20**, 25–62 (1993).
21. S. Bahria, N. Essoussi, and M. Limam, "Hyperspectral data classification using geostatistics and support vector machines," *Remote Sens. Lett.* **2**, 99–106 (2011).
22. V. Vapnik and V. Vapnik, *Statistical Learning Theory*, Wiley, New York (1998).
23. H. Permuter, J. Francos, and I. Jermyn, "A study of Gaussian mixture models of color and texture features for image classification and segmentation," *Pattern Recogn.* **39**, 695–706 (2006).
24. G. Foody, "Status of land cover classification accuracy assessment," *Remote Sens. Environ.* **80**, 185–201 (2002).

Shanshan Li received his BS degree from Wuhan University, Wuhan, China, in 2004. He is currently working toward his PhD degree at Center for Earth Observation and Digital Earth (CEODE), Chinese Academy of Sciences. His current research interests include hyperspectral and high-resolution remote sensed image processing, spatial feather exaction, and remote sensing applications.

Bing Zhang is a professor and deputy director of Center for Earth Observation and Digital Earth (CEODE), Chinese Academy of Sciences. He received his MS degree from Peking University, Beijing, China, and received his MS and PhD degree from Institute of Remote Sensing Applications, Chinese Academy Sciences. He specializes in hyperspectral remote sensing and has more than 14 years of experience in this field. His research interests include development of physics-based models and image processing software for the use of hyperspectral remote sensing data in solving problems in geology, hydrology, ecology, and botany.

Dongmei Chen is an associate professor. Her research areas focus on the understanding and modeling of interactions between human activities and the physical environment by using GIS and remote sensing techniques and spatial modeling approaches from local to regional scales.

Lianru Gao is currently an associate professor of Center for Earth Observation and Digital Earth (CEODE), Chinese Academy of Sciences. He received his MS degree from Tsinghua

University, Beijing, China, and his MS and PhD degrees from Institute of Remote Sensing Applications, Chinese Academy Sciences. He specializes in hyperspectral feature extraction and target detection.

Man Peng received her MS and BS degrees from Wuhan University, Wuhan, China. She is currently working toward her PhD degree at Institute of Remote Sensing Applications, Chinese Academy Sciences. Her current research interests include digital close-range photogrammetric system image feature recognition, stereo image matching, and machine vision.