Multispectral Image Classification Using a New Bayesian Approach with Weighted Markov Random Fields

Shengxi Li, Ying Wang(✉), Jie Li, and Xinbo Gao

School of Electronic Engineering, Xidian University, Xi'an 710071, China
yingwang@xidian.edu.cn

Abstract. This paper presents a novel nonparametric supervised spectral-spatial classification method for multispectral image. In multispectral images, if an unknown pixel shows similar digital number (DN) vectors as pixels in the training class, it will obtain higher posterior probability when assuming DN vectors of different classes follow a type of statistical distribution. The proposed method assumes the DN vectors follow a Gaussian mixture distribution in each class. Particularly, we use Bayesian nonparametric method to adaptively estimate the parameters in Gaussian mixture model. Then, we construct an anisotropic multilevel logistic spatial prior to capture the spatial contextual information provided by multispectral image. Finally, simulated annealing optimization algorithm is used to accomplish the maximum a posteriori classification. The proposed approach is compared with recently advanced multispectral image classification methods. The comparison results of classification suggested that the proposed approach outperformed other classifiers in overall accuracy and kappa coefficient.

Keywords: Bayesian nonparametric model · Gaussian mixture model · Markov random field · Multispectral image classification

1 Introduction

Land cover, which could provide valuable information for understanding the nature of hydrological, geographical, agricultural, ecological, and socioeconomic systems, is an underlying variable. It could impact and connect many aspects of human life with physical environments [1]-[2]. As the rapid development of spatial, spectral and temporal resolutions of remote sensing image over the past years, multispectral image classification has become one of the most common approaches to extract land cover information in remote sensing.

Focusing on multispectral image classification, the strategies proposed in literature generally are categorized as unsupervised and supervised schemes [3]-[5]. Unsupervised methods investigate data statistics by subdividing the image into clusters of pixels with similar characteristics, e.g., iterative self-organizing data analysis (ISODATA) and K-means classification. They do not require labeled information provided by user, while the procedure may lose correlation between the clusters it found and classes user desired. For handling this problem, supervised techniques are

H. Zha et al. (Eds.): CCCV 2015, Part I, CCIS 546, pp. 168–178, 2015.
DOI: 10.1007/978-3-662-48558-3_17

characterized by finding explicit link between samples and classes. They have shown more promising accuracies in terms of image classification than unsupervised methods, e.g. the minimum distance classification (MinDC), the Mahalanobis distance classification (MDC), the maximum likelihood classification (MLC), the multinomial logistic regression (MLR) and support vector machine classification (SVM). With the development of statistical learning, the classification algorithms for multispectral image analysis in recently can also be divided into parametric and nonparametric schemes. The parametric method involves a fixed representation that does not grow structurally as more data are observed. Examples include MLC and ISODATA, which assume the digital number (DN) vectors of different classes follow the Gaussian distribution. In contrast, nonparametric method is based on representations that are allowed to grow structurally as more data are observed. In practice, the classification procedure often expects to introduce as few assumptions as possible, thus, the nonparametric approaches are often employed to "let the data speak" [6]. Recently, many spectral-spatial classification techniques have been proposed to impose the spatial information [7] for improving the result of spectral classification, such as mathematical morphological filters, composite kernels, graph kernels, partitional clustering and joint sparse representation. As a general statistical modeling method, Markov random fields (MRFs) have shown good performance in incorporating spatial information in remote sensing classifications [8]-[9].

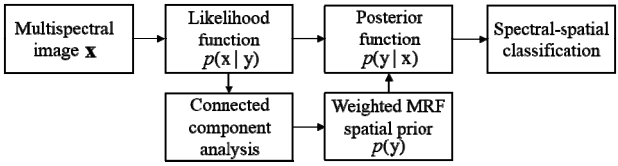


Fig. 1. Flowchart of proposed method.

In this paper, we proposed a novel nonparametric supervised spectral-spatial classification method. Firstly, considering the specificities and complexity of multispectral data and the fact that by using a sufficient number of Gaussians, almost all of the continuous density can be approximated to arbitrary accuracy [10], the Gaussian mixture distribution is constructed to describe the statistical properties of DN vectors in each class. Specially, the Bayesian nonparametric (BNP) method is used to adaptively estimate the parameters in Gaussian mixture model (GMM) and let the data determine the complexity of Gaussian mixture model by itself. A similar strategy has been designed in the field of brain MRI tissue classification [11]. Secondly and favorably, taking rich spatial information provided by multispectral image into consideration, an anisotropic multilevel logistic (MLL) spatial prior with region area information concerned is introduced to the maximum a posteriori (MAP) framework to obtain a more smooth and stable spectral-spatial classification result than what traditional isotropic multilevel logistic spatial prior gets. The flowchart of proposed method is presented in Fig. 1.

The remainder of this paper is organized as follows. Section 2 introduces the proposed nonparametric supervised spectral-spatial classification method in detail. Section 3 describes the data set we test on this paper and illustrates the performance of the proposed method in multispectral image classification. Conclusions are outlined in Section 4.

2 The Proposed Method

Similar to MLC method, we suggest that the DN vectors in multispectral image are regarded as random variable x, $\mathrm{x} \in R^d$ where d denotes the number of bands in multispectral image and y denotes the corresponding class label. For an unknown pixel x to be classified, according to the principle of maximum posterior probability, we employ Bayes' rules to arrange the unknown pixels into predefined classes.

$$p(\mathrm{y} \mid \mathrm{x}) = \frac{p(\mathrm{y})p(\mathrm{x} \mid \mathrm{y})}{p(\mathrm{x})} \propto p(\mathrm{y})p(\mathrm{x} \mid \mathrm{y}) \tag{1}$$

where $p(\mathrm{x} \mid \mathrm{y})$ is the likelihood function and $p(\mathrm{y})$ is the prior over the labels.

2.1 GMM Modeling for Likelihood Function

Given a remote sensing data set of a certain class $\{\mathrm{x}_i\}_{i=1}^N$, where N denotes the number of pixels in the training set. An observation x_i can be modeled as being generated from a Gaussian Mixture Model, which is described as:

$$\begin{aligned} c_i \mid \pi &\sim Multinomial(\cdot \mid \pi) \\ \mathrm{x}_i \mid c_i = k &\sim Gaussian(\cdot \mid \theta_k) \end{aligned} \tag{2}$$

where $\pi = (\pi_1, \pi_2, \ldots, \pi_K)$ denotes mixing coefficients, c_i denotes cluster label and θ_k stands for (μ_k, Σ_k) which are the mean vector and covariance matrix of each Gaussian component, respectively. Inspired by nonparametric statistics method, we assume the number of components in GMM is infinite, that is $K \to \infty$. Then BNP method is used to adaptively estimate the parameters in Gaussian mixture model.

Based on the infinite GMM (IGMM) assumption above, the parameters π and θ_k are defined as following:

$$\begin{aligned} \pi \mid \alpha &\sim Stick(\alpha) \\ \theta_k &\sim \mathrm{H} \end{aligned} \tag{3}$$

where α denotes concentration parameter in the Dirichlet distribution, $\pi \mid \alpha \sim Stick(\alpha)$ stands for:

$$\begin{aligned} \beta_k &\sim Beta(1,\ \alpha) \\ \pi_k &= \beta_k \prod_{l=1}^{k-1}(1-\beta_l),\ k \to \infty \end{aligned} \tag{4}$$

and $\theta_k \sim H$ stands for:

$$\begin{aligned} \mu_k &\sim Gaussian(\mu_0, \Sigma_k / \kappa_0) \\ \Sigma_k &\sim InverseWishart_{v0}(\Lambda_0^{-1}) \end{aligned} \tag{5}$$

We choose Gaussian distribution and Inverse Wishart distribution to describe the mean vector and covariance matrix in GMM [12] because they are conjugate priors for the Gaussian distribution, thus we can get a closed solution form for the posterior distribution for $C = \{c_i\}_{i=1}^N$ and $\Theta = \{\theta_k\}_{k=1}^K$ when $K \to \infty$.

Given an observation set $\{\mathrm{x}_i\}_{i=1}^N$, we intend to infer the parameters $\{\pi, C, \Theta\}$ based on their posterior distribution. However, the posterior distribution cannot be computed analytically. The Gibbs sampler, a widely used Markov Chain Monte Carlo (MCMC) method, is imposed as an alternative way to sample their posterior probabilities and the obtained samples will approximate the posterior distribution precisely [13]. Under IGMM, the posterior distribution for the unknown parameters is defined as [14]:

$$\begin{aligned} &p(C, \Theta, \pi, \alpha \mid \mathrm{x}) \propto \\ &p(\mathrm{x} \mid C, \Theta) p(\Theta \mid H) \prod_{i=1}^{N} p(c_i \mid \pi) p(\pi \mid \alpha) p(\alpha) \end{aligned} \tag{6}$$

Based on Eq. (3)-(5), we are able to integrate out π and get the posterior distributions for Θ and C. The infinite Gaussian mixture model with Chinese restaurant process sampler we used is a Matlab implementation provided by Wood et al [13]. Then the likelihood function $p(\mathrm{x} \mid \mathrm{y})$ in Bayes' formula of Eq. (1) has the following form:

$$p(\mathrm{x} \mid \mathrm{y}) = \sum_{k=1}^{K} \pi_k N(\mathrm{x} \mid \theta_k) \tag{7}$$

2.2 Anisotropic MLL for Spatial Prior

For an unknown pixel x_i to be classified, according to the principle of maximum likelihood (ML), which could calculate through Eq. (7), we can obtain y_i', where denotes the corresponding spectral classification label. Given an input multispectral image, a spectral classification map will be generated, then we use connected component analysis (CCA) to obtain connected region area R_i of each pixel, as shown in Fig. 2.

(a) (b)

Fig. 2. (a) Spectral classification label of each pixel. (b) Corresponding region area information of each pixel.

In a multispectral image, two adjacent neighboring pixels have a high probability belong to the same class. MRFs exploit the continuity of neighboring labels to impose spatial contextual information in Bayesian schemes in terms of the maximum a posteriori rule. In this paper, we construct an anisotropic MLL spatial prior to capture the spatial contextual information to constraint the class labels y. This spatial prior is an extension of the traditional isotropic MLL prior and belongs to the family of MRF.

Through the Hammersly-Clifford theorem, the prior $p(y)$ over class label obeys the Gibbs distribution, which has following form:

$$p(\mathrm{y}) = \frac{1}{Z} exp(-U(\mathrm{y})) \tag{8}$$

where $Z = \sum_{\mathrm{y}} exp(-U(\mathrm{y}))$ is a normalizing constant, $U(\mathrm{y}) = \sum_{c \in C} V_c(\mathrm{y})$ is the energy function summing the clique potentials $V_c(\mathrm{y})$ over all possible cliques C. As we know, the commonly used isotropic MLL model defines $V_c(\mathrm{y})$ as

$$V(\mathrm{y}_i, \mathrm{y}_j) = \begin{cases} -\beta, & if\ \mathrm{y}_i = \mathrm{y}_j \\ \beta, & otherwise \end{cases} \tag{9}$$

In this paper, we take the region area information which captured by spectral classification into the isotropic MLL. Then a new clique potential $V^w(\mathrm{y}_i, \mathrm{y}_j)$ which reflect the interaction between regions is defined as:

$$V^w(\mathrm{y}_i, \mathrm{y}_j) = \begin{cases} -\beta \dfrac{|R_j|}{|R_i|}, & if\ \mathrm{y}_i = \mathrm{y}_j \\ \beta \dfrac{|R_j|}{|R_i|}, & otherwise \end{cases} \tag{10}$$

where $|R_i|$ denotes the area of a region R_i. The penalty term $|R_j| / |R_i|$ could reflect the relative size information between two adjacent regions. The function of new clique potential $V^w(\mathrm{y}_i, \mathrm{y}_j)$ can be described as a relative large region have more strong impact than those relative small region over the class label and one needs to

provide more evidences for changing the label of a relative large region as that region may be a meaningful homogeneous region. Hence, we suppose the proposed anisotropic MLL spatial prior which named weighted MRF (WMRF) would lead to a more smooth and stable result compared with traditional isotropic MLL.

2.3 Computing the MAP Estimate via Simulated Annealing

Since it is difficult to maximize the joint probability of an MRF. The simulated annealing (SA) algorithm is used to compute the MAP estimate of final spectral-spatial classification map based on the spectral classification map [15] The SA algorithm for optimizing the global energy in the multispectral image can be summarized as an iteration of minimization of local energy function associated with randomly chosen pixels. The local energy function of a given pixel x can described as:

$$\begin{aligned} U(\mathrm{x}) &\propto -\ln p(\mathrm{y} \mid \mathrm{x}) \propto -\ln p(\mathrm{x} \mid \mathrm{y}) - \ln p(\mathrm{y}) \\ &\propto -\ln p(\mathrm{x} \mid \mathrm{y}) + U(\mathrm{y}) \end{aligned} \tag{11}$$

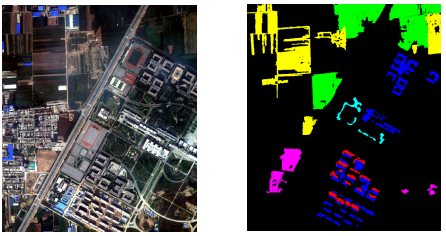
where $-\ln p(\mathrm{x} \mid \mathrm{y})$ is the spectral energy term obtained by Eq. (7), $U(\mathrm{y})$ is the spatial energy function computed over the local neighborhood via $V^{w}(\mathrm{y}_i, \mathrm{y}_j)$. In our work, an eight-neighborhood system is considered. The spectral spatial classification label is finally given by:

$$\hat{y} = \arg\min U(\mathrm{x}) \tag{12}$$

3 Experiments

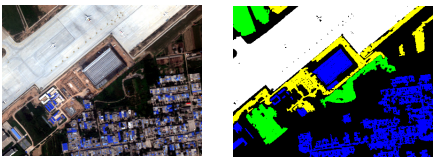
3.1 Data Sets

The multispectral image we used in this paper are acquired from SPOT6 satellite. It has four spectral bands that include blue band (0.455 to 0.525 μm), green band (0.530 to 0.590 μm), red band (0.625 to 0.695 μm) and near infrared band (0.760 to 0.890 μm). The spatial resolution is 2 m for each band. Two data sets are used to test the proposed classification algorithm and we also compare it with other advanced classification algorithms such as MinDC, MDC, MLC, MLR and SVM. These two data sets was acquired on September 30, 2012, covering Xidian University and Xianyang international airport in Xi'an, Shaanxi province. The image in University data set was 804×690 pixels and six classes of interests were considered: bare soil, building, meadow, water, shadow and gravel. The image in Airport data set was 324×426 pixels and four classes of interests were considered: bare soil, building, meadow and airport. Fig. 3 and Fig. 4 present false color images and reference data about two data sets. Training sets are randomly selected from reference data set and the size of training sets keep the ratio of 0.02 to test sets. Pixels from the training set are excluded from the test set in each case and vice versa.



(a) (b)

Fig. 3. Xidian University Area image. (a) Three band false color composite. (b) Reference data.

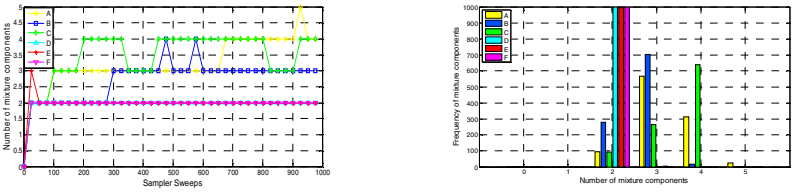


(a) (b)

Fig. 4. Xianyang Airport Area image. (a) Three band false color composite. (b) Reference data.

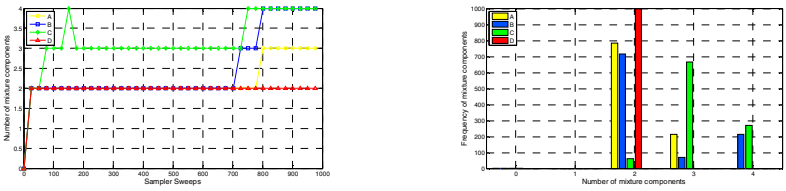
3.2 Experimental Results

The infinite Gaussian mixture model with Chinese restaurant process sampler experimental results of two data sets are presented in Fig. 5 and Fig. 6. Fig. 5(a) reflects the trend of latent clusters number in the process of BNP mixture sampler sweeps. Fig. 5(b) presents the statistic frequency histogram of each class, which describes the latent clusters' quantity of data. We use the maximum frequency of occurrence to denote the number of latent cluster. Namely, the number of latent cluster of bare soil, building, meadow, water, shadow and gravel are 3, 3, 4, 2, 2 and 2 respectively in the first data set. The number of latent cluster of bare soil, building, meadow and airport are 2, 2, 3 and 2 in the second data set, respectively.



(a) (b)

Fig. 5. BNP approach to sample the number of latent clusters. A. Bare soil; B. Building; C. Meadow; D. Water; E. Shadow; F. Gravel;



(a) (b)

Fig. 6. BNP approach to sample the number of latent clusters. A. Bare soil; B. Building; C. Meadow; D. Airport;

The data's frequency distributions and different Gaussian mixture fitting functions of DN values for building and meadow belonged to first data set are presented in Fig. 7. The data's frequency distributions and different Gaussian mixture fitting functions of DN values for bare soil and airport belonged to the second data set are present in Fig. 8. According to the fitting results we can know the number of latent clusters estimated by BNP mixtures is consistent with the data's real frequency distributions, and meanwhile, the Gaussian mixture distribution can describe the data set's statistical property approximately.

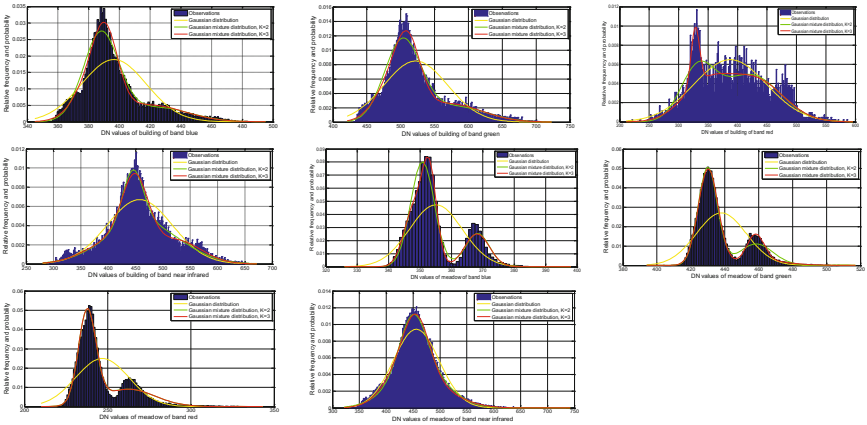


Fig. 7. Frequency distribution of DN values and its fitting likelihood functions by the Gaussian distribution, and Gaussian mixture distribution for bands blue, green, red and near infrared for building and meadow respectively.

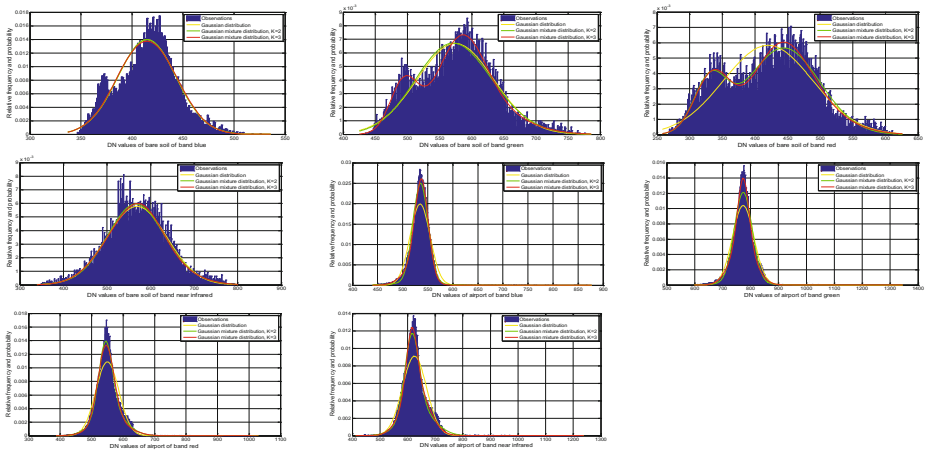


Fig. 8. Frequency distribution of DN values and its fitting likelihood functions by the Gaussian distribution, and Gaussian mixture distribution for bands blue, green, red and near infrared for bare soil and airport respectively.

The accuracy of proposed classification algorithm was assessed by the reference data using overall accuracies, and Kappa coefficient. The overall accuracies and the Kappa coefficients for MinDC, MDC, MLC, MLR, SVM and IGMM are presented in Table 1 and 2. Meanwhile, the overall accuracies for above spectral classification methods which imposing isotropic MLL and anisotropic MLL spatial prior named MLCMLL, MLRMLL, SVMMLL, IGMMMLL, MLCWMRF, MLRWMRF, SVMWMRF, IGMMWMRF are presented in Table 3 and 4. To obtain unbiased conclusions, the classification process was repeated 10 times with randomly selected different training and test sets and the average accuracies are given.

Table 1. Comparison of Spectral Classification Results in First Data set

	MinDC	MDC	MLC	MLR	SVM	IGMM
Overall Accuracy	68.31	77.61	80.94	71.64	84.67	**85.30**
Kappa coefficient	0.5744	0.7022	0.7421	0.6067	0.7861	**0.7996**

Table 2. Comparison of Spectral Classification Results in Second Data set

	MinDC	MDC	MLC	MLR	SVM	IGMM
Overall Accuracy	84.35	85.92	90.06	88.39	**93.80**	92.40
Kappa coefficient	0.7673	0.7952	0.8548	0.8267	**0.9080**	0.8877

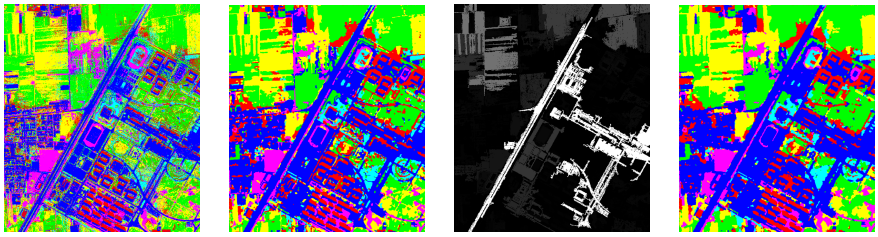
Table 3. Comparison of Spectral-Spatial Classification Results in First Data set

	$\beta = 1$	$\beta = 3$	$\beta = 5$	$\beta = 10$
MLCMLL	86.79	87.44	87.55	87.57
MLRMLL	73.88	74.11	74.16	74.18
SVMMLL	88.13	88.48	88.50	88.52
IGMMMLL	92.52	93.55	93.59	93.64
MLCWMRF	89.38	88.98	88.74	88.62
MLRWMRF	76.89	76.47	76.34	76.12
SVMWMRF	88.53	88.72	88.56	88.45
IGMMWMRF	**94.83**	**94.75**	**94.58**	**94.57**

Table 4. Comparison of Spectral-Spatial Classification Results in Second Data set

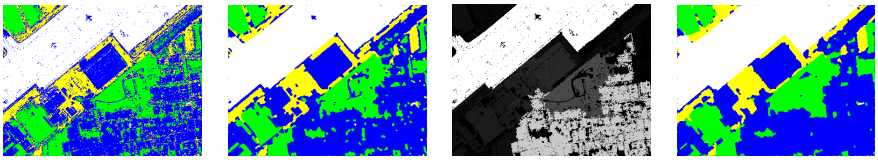
	$\beta = 1$	$\beta = 3$	$\beta = 5$	$\beta = 10$
MLCMLL	94.27	95.14	95.13	95.16
MLRMLL	91.69	92.15	92.18	92.17
SVMMLL	95.67	95.87	95.87	95.88
IGMMMLL	96.42	97.10	97.21	**97.37**
MLCWMRF	97.30	97.21	97.17	97.05
MLRWMRF	92.68	92.50	92.65	92.64
SVMWMRF	95.79	95.89	95.92	95.82
IGMMWMRF	**97.51**	**97.35**	**97.33**	**97.37**

In general, IGMM and SVM could produce relatively better overall accuracy and Kappa coefficient comparison with MinDC, MDC, MLC and MLR according to the spectral classification results. The proposed IGMM spectral classification method produce relatively better results can be explained as achieved a better statistics fitting precision at the cost of higher model complexity. The proposed anisotropic MLL prior achieve better performance than traditional isotropic MLL prior according to the spectral-spatial classification results. It can be explained as big size region has relatively higher impact to adjacent neighboring pixels than those small size region, thus the proposed anisotropic MLL prior could get a more smooth and stable results. Particularly, the proposed nonparametric supervised spectral-spatial classification algorithm outperformed than other recently advanced spectral-spatial classification method. The thematic maps obtained from the two data sets can be seen in Fig. 9 and Fig. 10. The gray scale in Fig. 9(c) and Fig. 10(c) denotes the region area information of corresponding pixel (light represents very big area, dark represents very small area), which captured by spectral classification.



(a) (b) (c) (d)

Fig. 9. Thematic maps obtained with the Xidian University data set: (a) IGMM method, (b) IGMMMLL method, (c) Area map, (d) IGMMWMRF method



(a) (b) (c) (d)

Fig. 10. Thematic maps obtained with the Xianyang International airport data set: (a) IGMM method, (b) IGMMMLL method, (c) Area map, (d) IGMMWMRF method.

4 Conclusion

This paper presents a new Bayesian approach with weighted Markov Random Fields for multispectral image classification. The proposed method shows competitive performance when compared with recent classification method, e.g. MinDC, MDC, MLC, MLR and SVM.

Acknowledgments. This research was supported partially by the National Natural Science Foundation of China (Grant Nos. 61125204, 61201294, 61201453 and 61432014), the Fundamental Research Funds for the Central Universities (Grant no. JB140225), and the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant nos. 20120203120009 and 20121401120015).

References

1. Foody, G.M.: Status of land cover classification accuracy assessment. Remote Sens. Environ. **80**(1), 185–201 (2002)
2. Townsend, P.A., Helmers, D.P., Kingdon, C.C., McNeil, B.E., de Beurs, K.M., Eshleman, K.N.: Changes in the extent of surface mining and reclamation in the Central Appalachians detected using a 1976–2006 Landsat time series. Remote Sens. Environ. **113**(1), 62–72 (2009)
3. Shackelford, A.K., Davis, C.H.: A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas. IEEE Trans. Geosci. Remote Sens. **41**(9), 1920–1932 (2003)
4. Insanic, E., Siqueira, P.R.: A maximum likelihood approach to estimation of vector velocity in Doppler radar networks. IEEE Trans. Geosci. Remote Sens. **50**(2), 553–567 (2012)
5. Ruiz, P., Mateos, J., Camps-Valls, G.: Bayesian active remote sensing image classification. IEEE Trans. Geosci. Remote Sens. **52**(4), 2186–2196 (2014)
6. Gershman, S.J., Blei, D.M.: A tutorial on Bayesian nonparametric models. Journal of Mathematical Psychology **56**(1), 1–12 (2012)
7. Fauvel, M., Tarabalka, Y., Benediktsson, J.A., Chanussot, J., Tilton, J.C.: Advances in spectral–spatial classification of hyperspectral images. Proc. IEEE **101**(3), 652–675 (2013)
8. Li, J., Bioucas-Dias, J.M., Plaza, A.: Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields. IEEE Trans. Geosci. Remote Sens. **50**(3), 809–823 (2012)
9. Moser, G., Serpico, S.B.: Combining support vector machines and Markov random fields in an integrated framework for contextual image classification. IEEE Trans. Geosci. Remote Sens. **51**(5), 2734–2752 (2013)
10. Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics). Springer, New York (2007)
11. da Silva, A.R.F.: A Dirichlet process mixture model for brain MRI tissue classification. Med. Image Anal. **2**, 169–182 (2007)
12. Nguyen, N.T., Zheng, R., Han, Z.: On identifying primary user emulation attacks in cognitive radio systems using nonparametric bayesian classification. IEEE Transactions on Signal Processing **60**(3), 1432–1445 (2012)
13. Neal, R.M.: Markov chain sampling methods for dirichlet process mixture models. Journal of Computational and Graphical Statistics **9**(2), 249–265 (2000)
14. Wood, F., Goldwater, S., Black, M.J.: A nonparametric bayesian approach to spike sorting. In: 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1165–1168 (2006)
15. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. **10**(6), 721–741 (1984)