

# Multispectral image classification based on improved weighted MRF Bayesian



Zhaobin Cui, Ying Wang\*, Xinbo Gao, Jie Li, Yu Zheng

*Lab of Video and Image Processing Systems, School of Electronic Engineering, Xidian University, Xi'an, China*

## ARTICLE INFO

### Article history:

Received 19 November 2015

Received in revised form

17 March 2016

Accepted 23 March 2016

Available online 5 July 2016

### Keywords:

Bayesian nonparametric model

Gaussian mixture model

Markov random field

Multispectral image classification

## 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 certain type of statistical distribution. According to statistical characteristics about DN vectors, the proposed method assumes the vectors follow a Gaussian mixture distribution in each class. Particularly, adaptively Bayesian nonparametric method is developed to estimate the optimal settings in distribution model appropriately. Then, we construct an anisotropic hierarchical logistic spatial prior to capture the spatial contextual information provided by multispectral image. Finally, optimized simulated annealing algorithm is conducted to estimate the maximum a posteriori. The proposed approach is compared with state-of-the-arts classification methods of multispectral images. The comparison results suggested that the proposed approach outperformed in overall accuracy and kappa coefficient.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

As well known, land cover can provide abundant information for understanding the nature of hydrological, geographical, agricultural, ecological and socioeconomic systems. It could impact and connect many aspects of human life with physical environments [1,2]. Moreover, analyzing of land cover can characterize hydrologic response to land cover change, estimate ecosystem status and health, comprehend spatial patterns of biodiversity, and also develop management policies of land [2–4]. 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) [6] and K-means classification [7]. Wang et al. proposed a novel method named an improved ISODATA algorithm for hyperspectral image classification, which took into account the maximum and minimum spectrum of the image and determined

the initial cluster center by the stepped construction of spectrum accurately [6]. Wu et al. proposed a new remote sensing image classification algorithm based on K-means using HSV color feature, which is implemented by extracting three color features (hue, saturation, value) for K-means clustering [7]. Unsupervised methods 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 characterized by finding explicit relationship 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) [8], the Mahalanobis distance classification (MDC) [9], the maximum likelihood classification (MLC) [10], the multinomial logistic regression (MLR) [11] and the support vector machine classification (SVM) [12–14]. Duda et al. detailedly described the algorithm of minimum distance classification in Pattern Classification [8]. Richards et al. pointed out that the Mahalanobis distance classifier is relatively speaking one kind of fast classifier [9]. Sisodia et al. used supervised maximum likelihood classification for analyzing of remotely sensed image [10]. Li et al. proposed a new spectral-spatial hyperspectral image segmentation algorithm by using subspace multinomial logistic regression and markov random fields. The method employed MLR to learn the posterior probability distributions from spectral information and used markov random field to obtain the contextual information [11]. Melgani et al. introduced SVM to address the problem of

\* Corresponding author.

E-mail address: [yingwang@xidian.edu.cn](mailto:yingwang@xidian.edu.cn) (Y. Wang).

hyperspectral remote sensing images classification [13]. With the development of statistical learning, the classification algorithms for multispectral image analysis 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 [8] and ISODATA [6], 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. Modern classifiers such as decision trees (DT), boosting, nearest neighbor (NN) and support vector machine (SVM) methods are all nonparametric, which report significant performance. 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” [15]. These methods always focus on spectral information, which cover little spatial information. While as for remote sensing images, it exists “different body with same spectrum” or “same body with different spectrum” phenomenon, and that will restrain further improvement of classification accuracy. Recently, many spectral-spatial classification techniques have been proposed to impose the spatial information [16] for further improving the performance of spectral classification, such as mathematical morphological filters, composite kernels, graph kernels, partition clustering and joint sparse representation. As a general statistical modeling method, although markov random fields (MRFs) has shown good performance in incorporating spatial information in remote sensing classifications [11,17], these methods always tend to treat the neighbor pixels around core pixel equally, while ignoring the different contribution of these neighbor pixels in classification procedure. Furthermore, there are usually exist many small noisy area points which are so similar with DN vectors that it can lead to misclassification easily.

For handling the above problems, we proposed a novel nonparametric supervised spectral-spatial classification method in this paper. 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 [18], the Gaussian mixture distribution is constructed to describe the statistical properties of DN vectors in each class. It can fit the statistical properties of DN vectors better as getting benefit from the appropriate distribution assumption. Specially, following the similar problems in the field of brain MRI tissue classification [19], improved Bayesian nonparametric (BNP) method is developed to adaptively estimate the optimal parameters in Gaussian mixture model (GMM) and let the data determine the complexity of Gaussian mixture model by itself. While the traditional Gaussian model can not describe the statistical properties of data well. 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 designed to accurate the maximum a posteriori (MAP) framework. With the spatial area information, the proposed scheme can effectively eliminate the influence of noisy points and make the spectral-spatial classification results more smooth and robust. Finally, the optimal simulated annealing is implemented to compute the MAP estimate for obtaining final integral spectral-spatial classification map.

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

The traditional MLC method seems to be the most powerful decision rule available in the statistical learning field. Similar to MLC method, we suggest that the DN vectors in multispectral image are regarded as random variable  $x$ ,  $x \in R^d$  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.

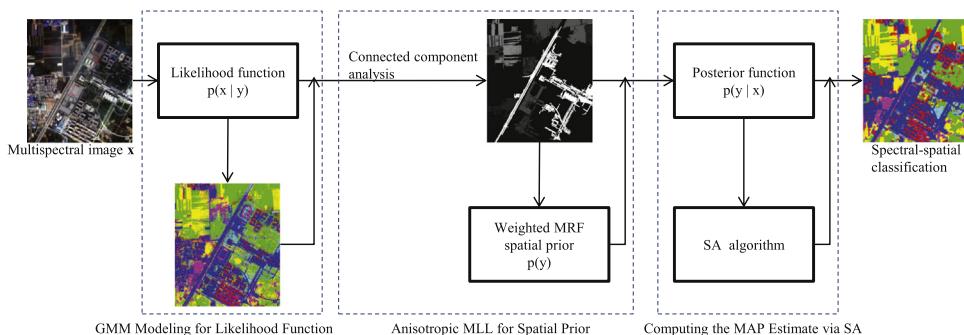
$$p(y|x) = \frac{p(y)p(x|y)}{p(x)} \propto p(y)p(x|y) \quad (1)$$

where  $p(x|y)$  is the likelihood function and  $p(y)$  is the prior over the labels. The flowchart of proposed method is presented in Fig. 1.

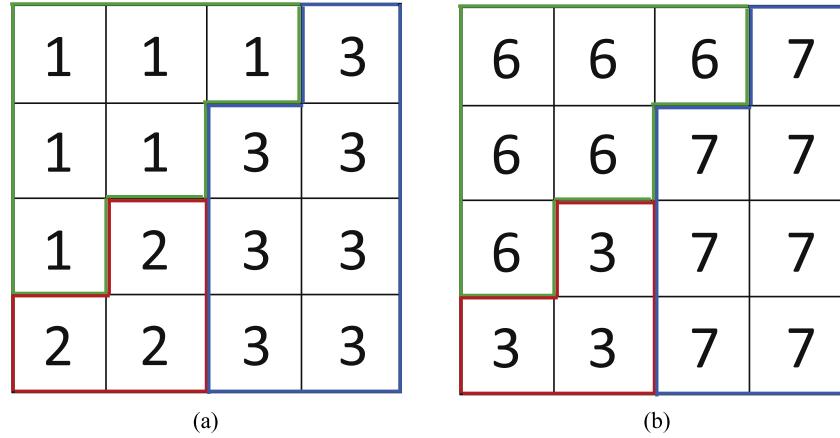
### 2.1. GMM modeling for likelihood function

The likelihood function which describes the spectral statistical properties of land objects plays an important role to statistical learning classification method. A well-performance likelihood function could get a better fitting precision and classification performance. Since the excellent fitting ability of Gaussian mixture distribution, we use it to describe the statistical properties of DN vectors in each class, as is shown in the first dotted box of Fig. 1.

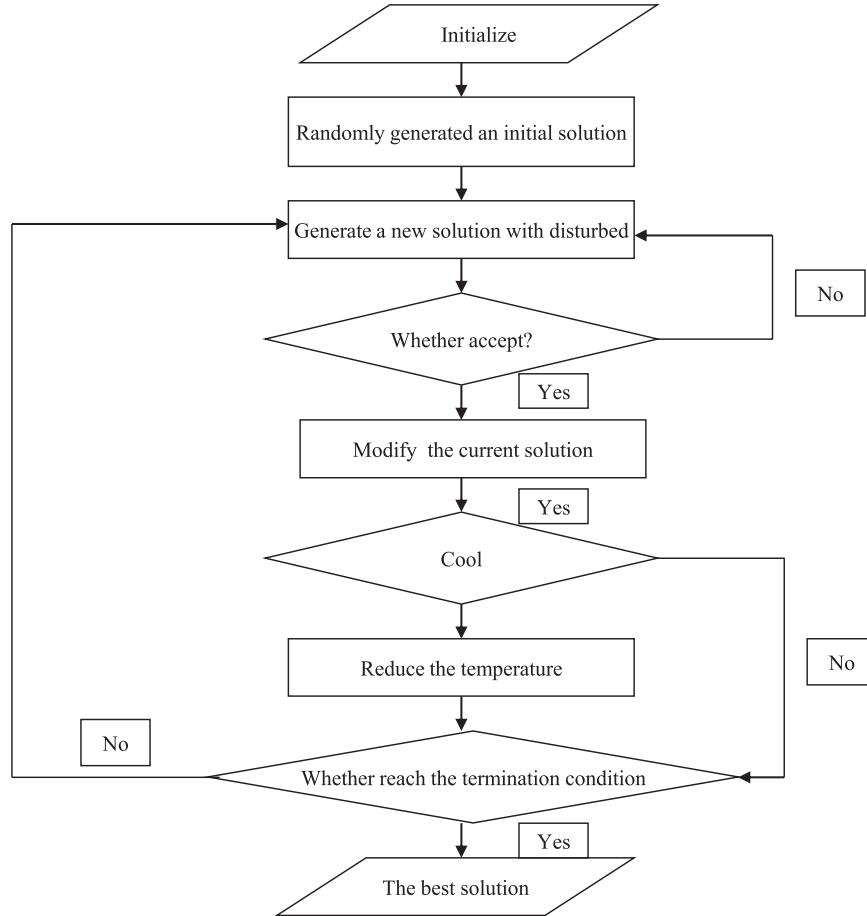
Given a remote sensing data set of a certain class  $\{x_i\}_{i=1}^N$ , where  $N$  denotes the number of pixels in the training set. An observation  $x_i$  can be modeled as being generated from a Gaussian Mixture Model, which is described as:



**Fig. 1.** Flowchart of proposed method. The first part constructs the GMM model for likelihood function which could well classify the spectral features with high accuracy; the second part designs the anisotropic MLL spatial prior for further improving the classification performance through combining the spectral and spatial information. And the third part optimizes the MAP estimate according to Simulated Annealing algorithm for obtaining spectral-spatial classification map. Then the final classification result can be obtained.



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



**Fig. 3.** Algorithm process of simulated annealing.

$$\begin{aligned} c_i | \pi &\sim \text{Multinomial}(\cdot | \pi) \\ x_i | c_i = k &\sim \text{Gaussian}(\cdot | \theta_k) \end{aligned} \quad (2)$$

where  $\pi = (\pi_1, \pi_2, \dots, \pi_K)$  denotes mixing coefficients,  $c_i$  denotes cluster label and  $\theta_k$  stands for  $(\mu_k, \Sigma_k)$  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  $K \rightarrow \infty$ . Then BNP method is used to adaptively estimate the parameters in Gaussian mixture model.

Based on the infinite GMM (IGMM) assumption above, the parameters  $\pi$  and  $\theta_k$  are defined as following:

$$\begin{aligned} \pi | \alpha &\sim \text{Stick}(\alpha) \\ \theta_k &\sim H \end{aligned} \quad (3)$$

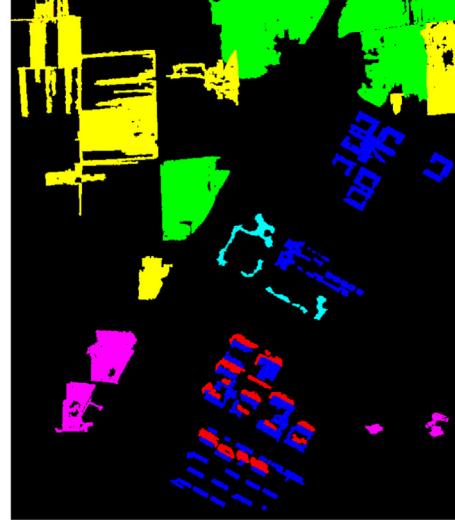
where  $\alpha$  denotes concentration parameter in the Dirichlet distribution,  $\pi | \alpha \sim \text{Stick}(\alpha)$  stands for:

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

and  $\theta_k \sim H$  stands for:



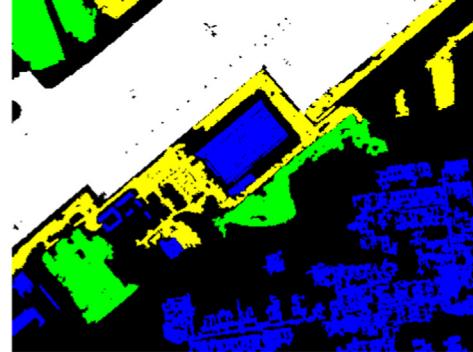
(a) Three band false color composite.



(b) Reference data.

**Fig. 4.** Xidian University area image.

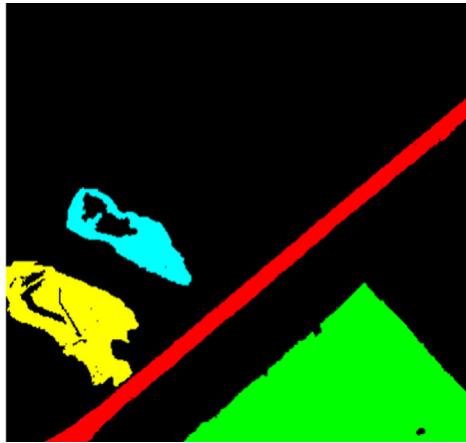
(a) Three band false color composite.



(b) Reference data.

**Fig. 5.** Xianyang airport area image.

(a) Three band false color composite.



(b) Reference data.

**Fig. 6.** Qujiang dam area image.

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

We choose Gaussian distribution and Inverse Wishart distribution to describe the mean vector and covariance matrix in

GMM [20] because they are conjugate priors for the Gaussian distribution. Thus we can get a closed solution form for the posterior distribution of  $C = \{c_i\}_{i=1}^N$  and  $\theta = \{\theta_k\}_{k=1}^K$  when  $K \rightarrow \infty$ .

Given an observation set  $\{x_i\}_{i=1}^N$ , we intend to infer the parameters  $\{\pi, C, \theta\}$  based on their posterior distribution. However, the posterior distribution cannot be computed analytically. The

**Table 1**

Information classes and training-test samples for the Xidian University data set.

Class	No	Samples	
		Train	Test
1	Bare soil	800	35,958
2	Building	500	17,921
3	Meadow	800	40,384
4	Water	200	2341
5	Shadow	200	4600
6	Gravel	300	8122
Total		2800	109,326

**Table 2**

Information classes and training-test samples for the Xianyang international airport data set.

Class	No	Samples	
		Train	Test
1	Bare Soil	300	10,870
2	Building	500	20,252
3	Meadow	300	10,445
4	Airport	800	36,761
Total		1900	78,328

**Table 3**

Information classes and training-test samples for the Qujiang dam data set.

Class	No	Samples	
		Train	Test
1	Road	200	5469
2	Water	500	15,782
3	Bare Soil	200	5067
4	Meadow	200	2676
Total		1100	28,994

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 [21]. Under IGMM, the posterior

distribution for the unknown parameters is defined as [22]:

$$p(C, \Theta, \pi, \alpha|x) \propto p(x|C, \Theta)p(\Theta|H) \prod_{i=1}^N p(c_i|\pi)p(\pi|\alpha)p(\alpha) \quad (6)$$

Based on Eqs. (3–5), we are able to integrate out  $\pi$  and get the posterior distributions for  $\Theta$  and  $C$ . The infinite Gaussian mixture model with Chinese restaurant process sampler we used is a Matlab implementation provided by Wood et al. [21]. Then the likelihood function  $p(x|y)$  in Bayes' formula of Eq. (1) has the following form:

$$p(x|y) = \sum_{k=1}^K \pi_k N(x|\theta_k) \quad (7)$$

The parameter estimation procedure of GMM modeling for likelihood function can be described in Algorithm 1.

#### Algorithm 1 Parameter Estimation

Input: Training data set  $\mathbf{x}$  and corresponding label set  $\mathbf{y}$

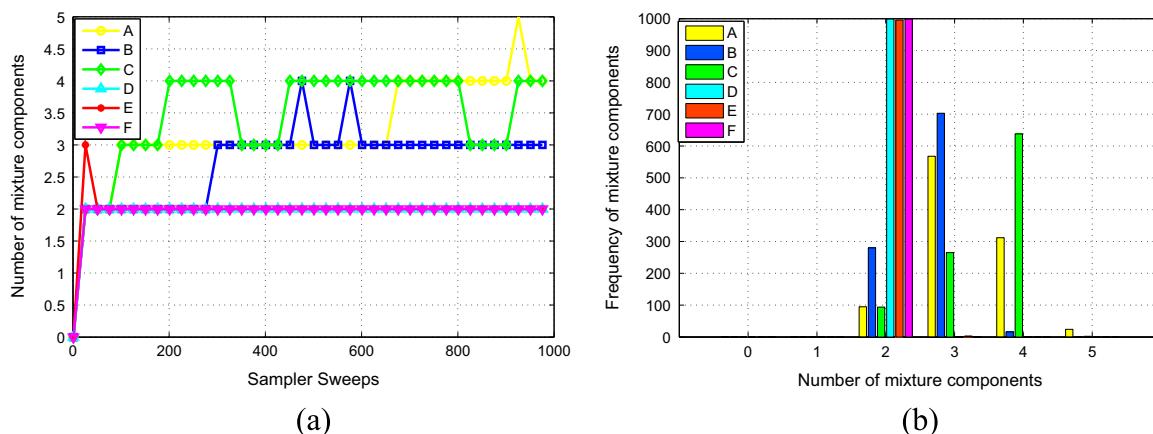
Output:  $\pi_k$  and  $\theta_k$  for each class

1. Initialize concentration parameter  $\alpha$  and base distribution  $H$
2. For a certain class data set  $\{\mathbf{x}_i\}_{i=1}^N$
3. Infer the parameters  $\{\pi, C, \Theta\}$  using Eq. (3)–(5)
4. Until all class data set have been estimated

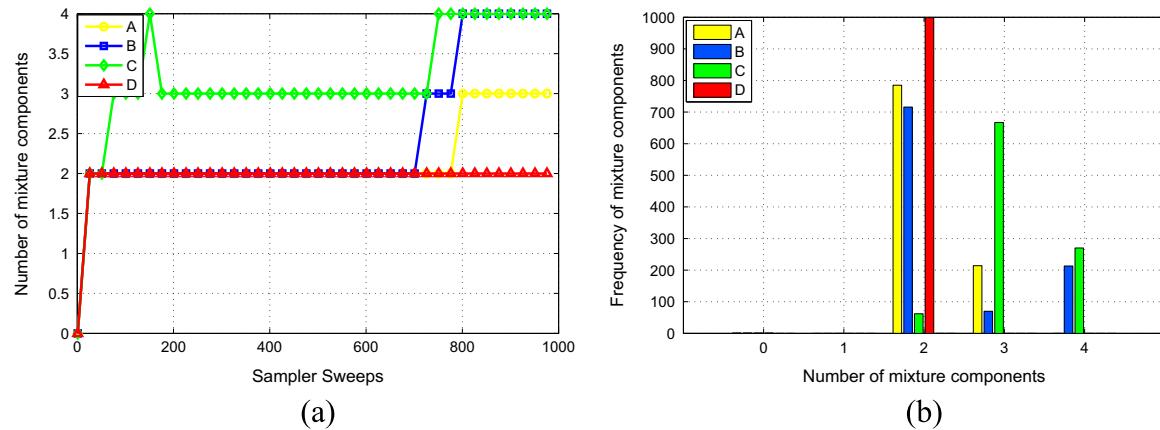
## 2.2. Anisotropic MLL for spatial prior

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  $y_i^*$ , which 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, we use the number of pixels in the connected component to denote the area information, which is shown in Fig. 2.

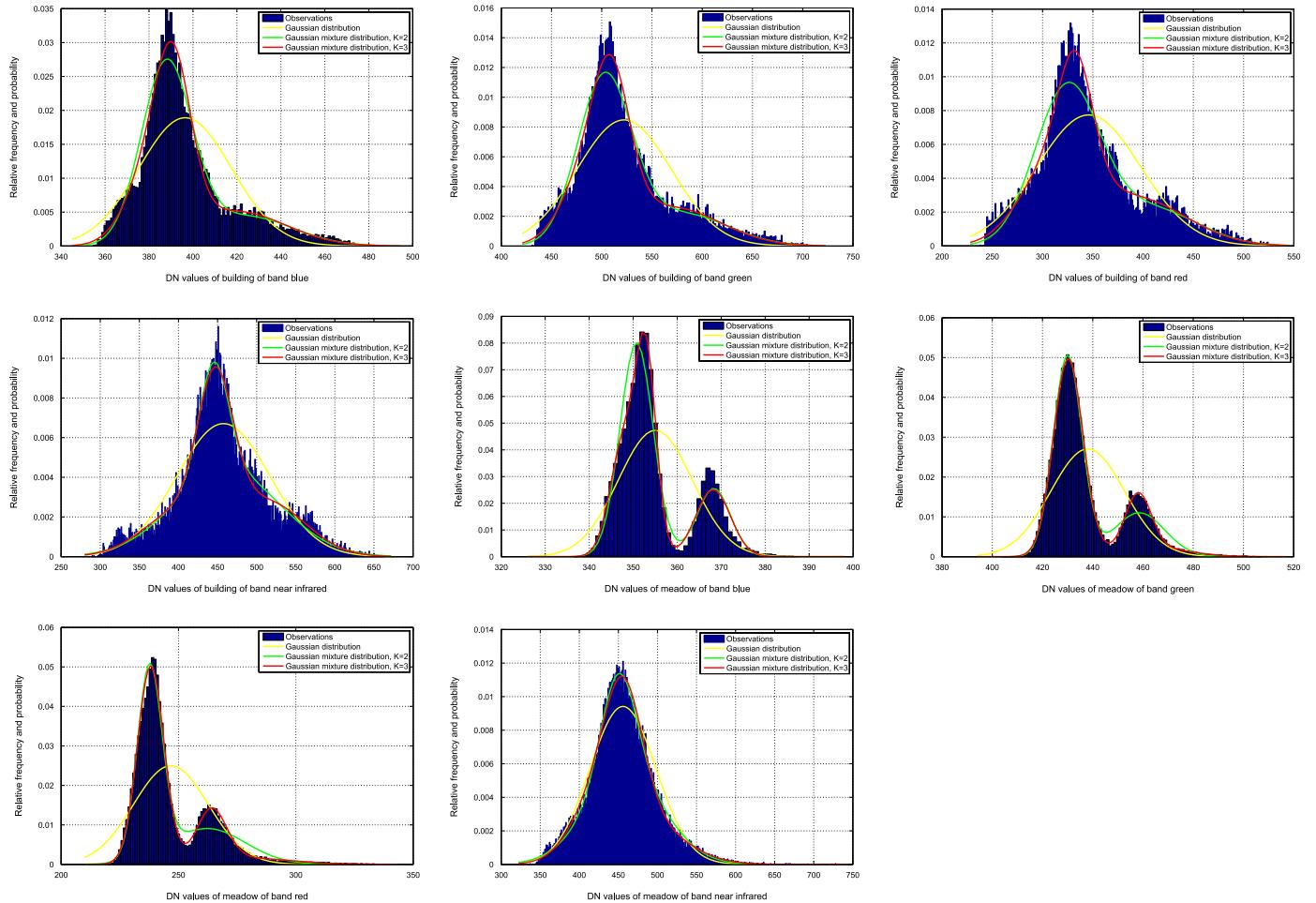
In a multispectral image, two adjacent pixels always possess a high probability that they belong to the same class. MRFs exploit the continuity of neighboring labels to impose spatial contextual information in Bayesian schemes in terms of the maximum a posteriori rule. As we all know, how to define a potential function is the most important part in MRFs model, different potential



**Fig. 7.** BNP approach to sample the number of latent clusters in the first data set. A. Bare soil; B. Building; C. Meadow; D. Water; E. Shadow; F. Gravel;



**Fig. 8.** BNP approach to sample the number of latent clusters in the second data set. A. Bare soil; B. Building; C. Meadow; D. Airport;



**Fig. 9.** 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

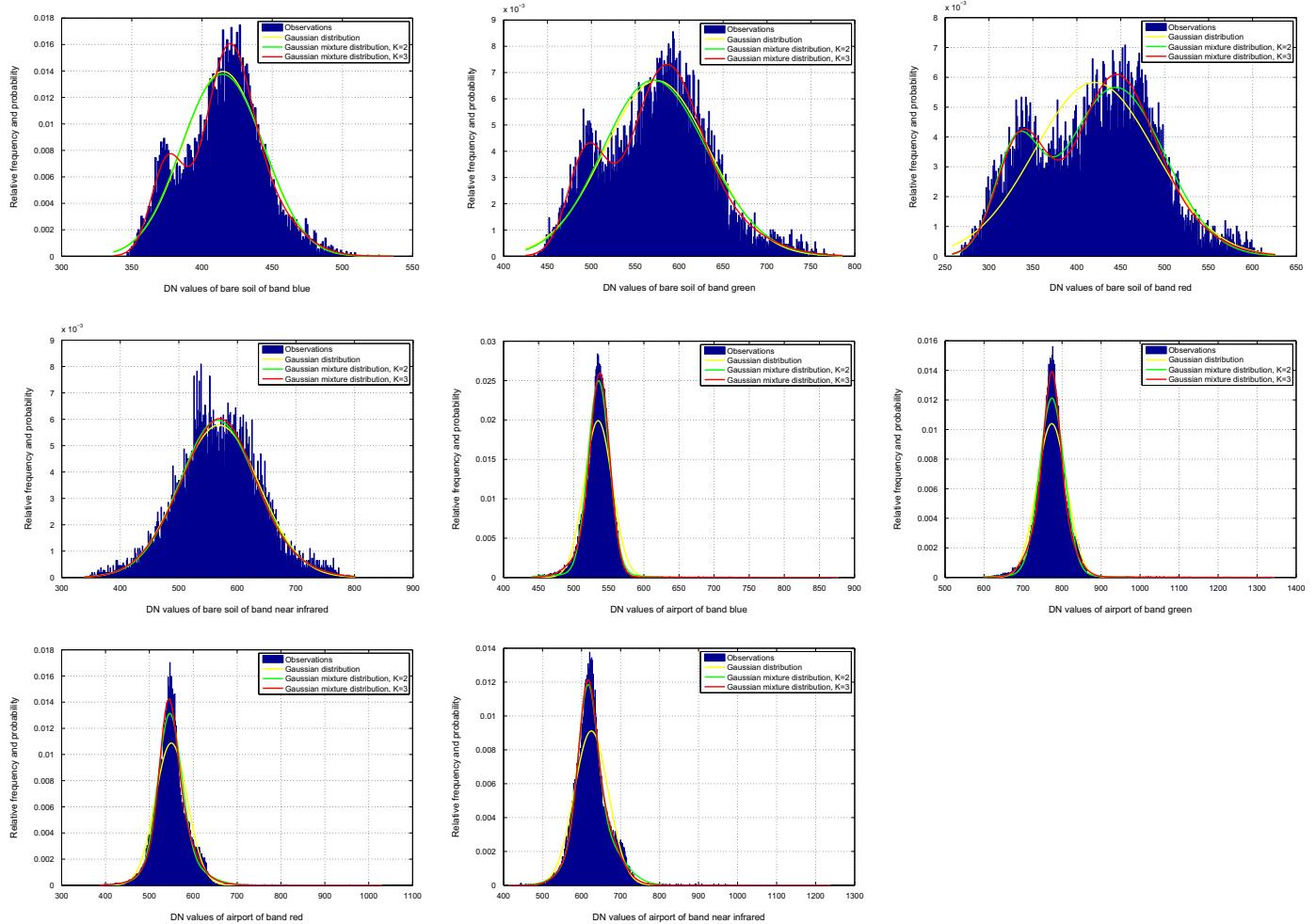
functions will produce different results. In this paper, we construct an anisotropic MLL spatial prior to capture the spatial contextual information to constraint the class labels  $y$ . This spatial prior is an extension of the traditional isotropic MLL prior and belongs to the family of MRF, which is shown in the second dotted box of Fig. 1.

Through the Hammersley-Clifford theorem, the prior  $p(y)$  over class label obeys the Gibbs distribution, thus the probability distribution of MRF can be solved by Gibbs distribution. The structural information of MRF model also can be described by Gibbs

distribution, which has following form:

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

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



**Fig. 10.** 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 4**  
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	<b>85.30</b>
Kappa coefficient	0.5744	0.7022	0.7421	0.6067	0.7861	<b>0.7996</b>

**Table 5**  
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	<b>93.80</b>	92.40
Kappa coefficient	0.7673	0.7952	0.8548	0.8267	<b>0.9080</b>	0.8877

**Table 6**  
Comparison of spectral classification results in third data set.

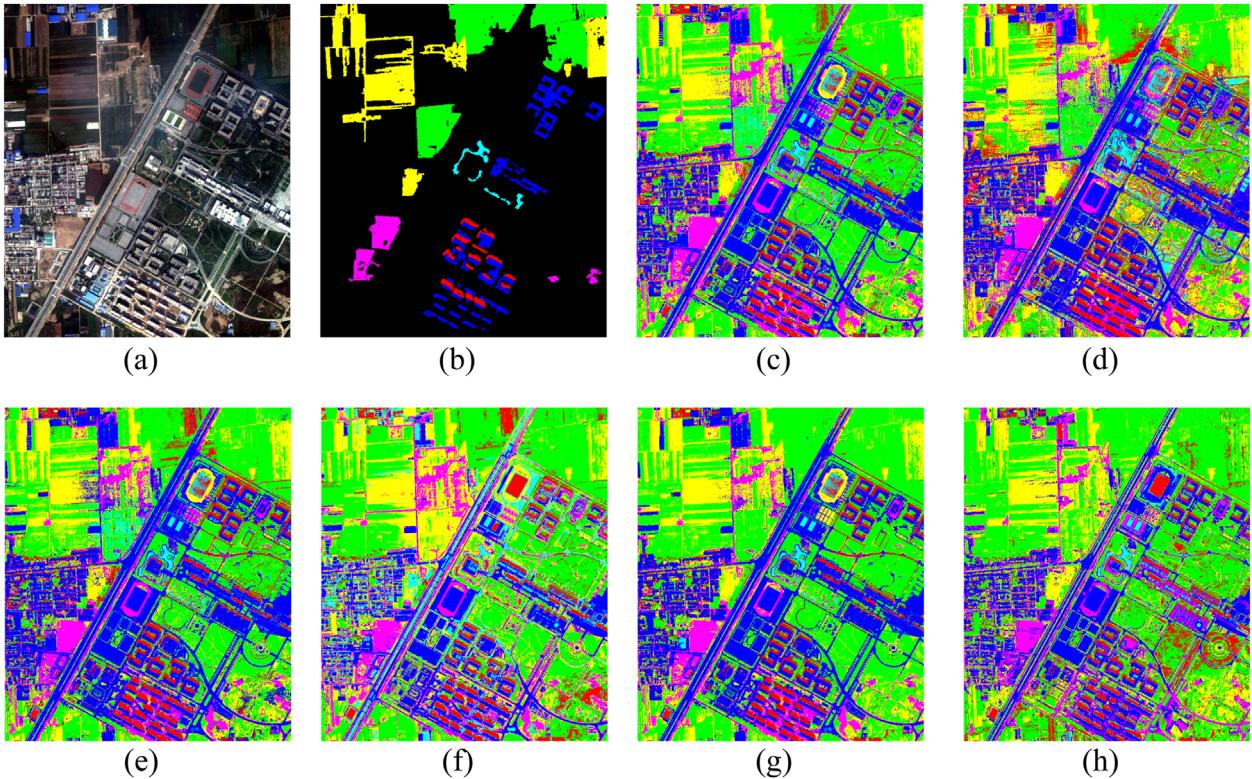
	MinDC	MDC	MLC	MLR	SVM	IGMM
Overall accuracy	92.32	95.41	95.75	93.55	97.29	<b>97.36</b>
Kappa coefficient	0.8786	0.9285	0.9334	0.8992	0.9570	<b>0.9583</b>

$$V(y_i, y_j) = \begin{cases} -\gamma, & \text{if } y_i = y_j \\ \gamma, & \text{otherwise} \end{cases} \quad (9)$$

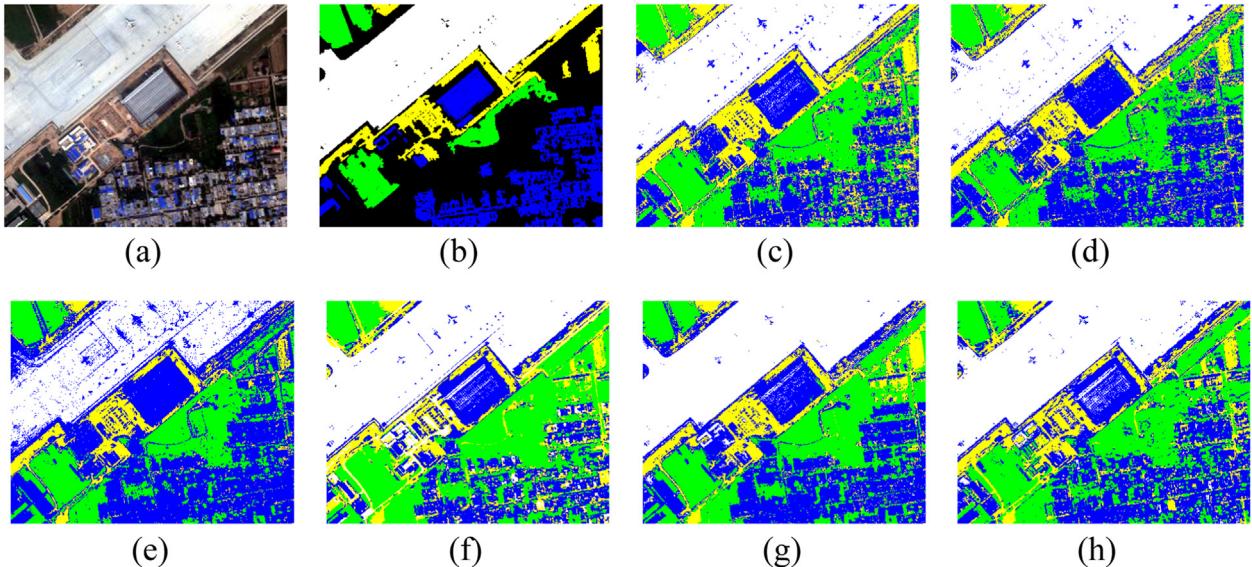
In the proposed method, we take the region area information which captured by spectral classification into the isotropic MLL. Then a new clique potential  $V^w(y_i, y_j)$  which reflect the interaction between regions is defined as:

$$V^w(y_i, y_j) = \begin{cases} -\gamma \frac{|R_j|}{|R_i|}, & \text{if } y_i = y_j \\ \gamma \frac{|R_j|}{|R_i|}, & \text{otherwise} \end{cases} \quad (10)$$

where  $|R_i|$  denotes the area of a region  $R_i$ . The proposed clique potential designs new weighted penalty term to introduce area information. The weight  $|R_j|/|R_i|$  could reflect the relative size information between two adjacent regions. The function of new clique potential  $V^w(y_i, y_j)$  can be described as a relative large region which has stronger impact than those relative small regions as for 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 smoother and more stable result compared with traditional isotropic MLL.



**Fig. 11.** Thematic maps obtained with the Xidian University data set: (a) Three band false color composite, (b) Reference data, (c) MLC method, (d) IGMM method, (e) MDC method, (f) MinDC method, (g) SVM method, (h) MLR method.

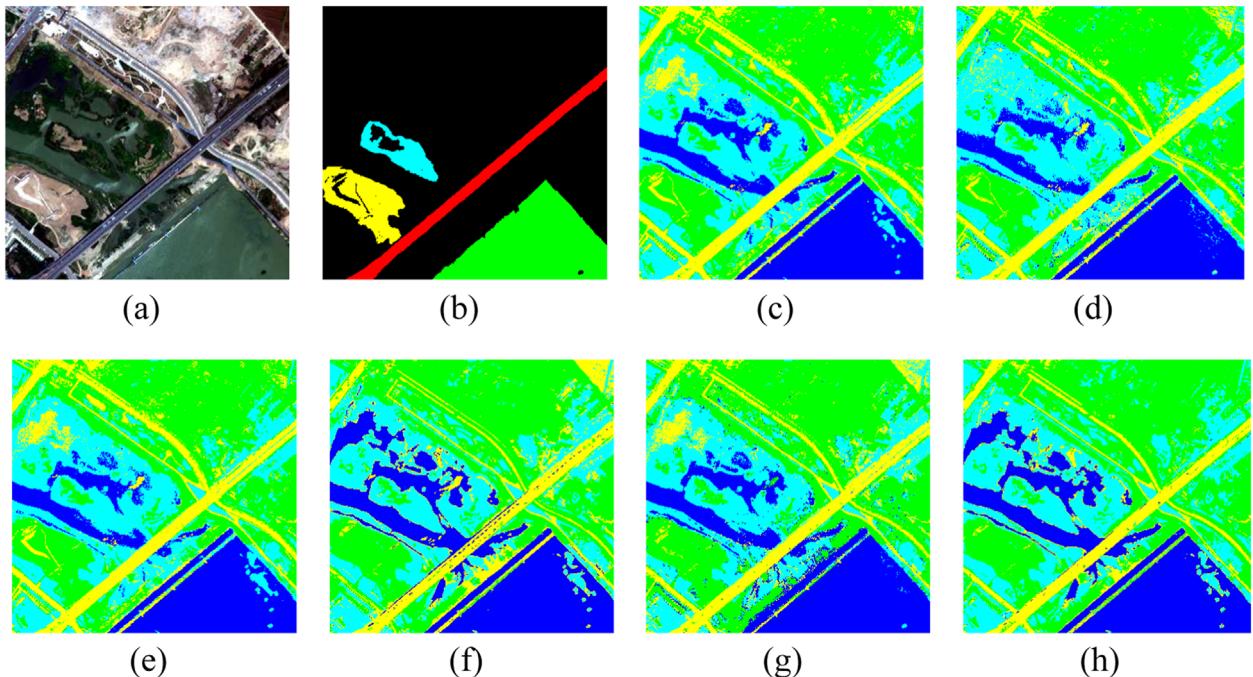


**Fig. 12.** Thematic maps obtained with the Xianyang International airport data set: (a) Three band false color composite, (b) Reference data, (c) MLC method, (d) IGMM method, (e) MDC method, (f) MinDC method, (g) SVM method, (h) MLR method.

### 2.3. Computing the MAP estimate via simulated annealing

Although it is difficult to maximize the joint probability of an MRF, the Simulated annealing algorithm (SA), whose search process is a kind of time-varying and the sudden jump probability finally tends to zero, can effectively avoid falling into local minimum and finally tends to the global optimal. In addition, it is a general optimization algorithm and has been widely used in

engineering. So it is used to compute the MAP estimate of final spectral-spatial classification map based on the spectral classification map [23]. 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 be described as:



**Fig. 13.** Thematic maps obtained with the Qujiang dam data set: (a) Three band false color composite, (b) Reference data, (c) MLC method, (d) IGMM method, (e) MDC method, (f) MinDC method, (g) SVM method, (h) MLR method.

**Table 7**  
Comparison of spectral-spatial classification results in first data set.

	MLCMLL	MLCEMLL	MLCWMRF	SVMMILL	SVMEMLL	SVMWMRF	IGMMMLL	IGMMEMLL	IGMMWMRF
Overall accuracy	87.38	85.69	89.10	88.50	87.59	88.56	93.29	91.34	<b>94.78</b>
Kappa coefficient	0.8267	0.8036	0.8497	0.8391	0.8263	0.8394	0.9037	0.8804	<b>0.9277</b>

**Table 8**  
Comparison of spectral-spatial classification results in second data set.

	MLCMLL	MLCEMLL	MLCWMRF	SVMMILL	SVMEMLL	SVMWMRF	IGMMMLL	IGMMEMLL	IGMMWMRF
Overall accuracy	95.14	94.42	97.11	95.87	95.61	95.98	96.39	95.94	<b>97.17</b>
Kappa coefficient	0.9285	0.9180	0.9571	0.9385	0.9346	0.9402	0.9466	0.9400	<b>0.9583</b>

**Table 9**  
Comparison of spectral-spatial classification results in third data set.

	MLCMLL	MLCEMLL	MLCWMRF	SVMMILL	SVMEMLL	SVMWMRF	IGMMMLL	IGMMEMLL	IGMMWMRF
Overall accuracy	96.08	95.55	98.47	99.08	98.81	<b>99.88</b>	98.56	98.16	<b>99.81</b>
Kappa coefficient	0.9387	0.9305	0.9758	0.9853	0.9811	<b>0.9982</b>	0.9771	0.9708	<b>0.9970</b>

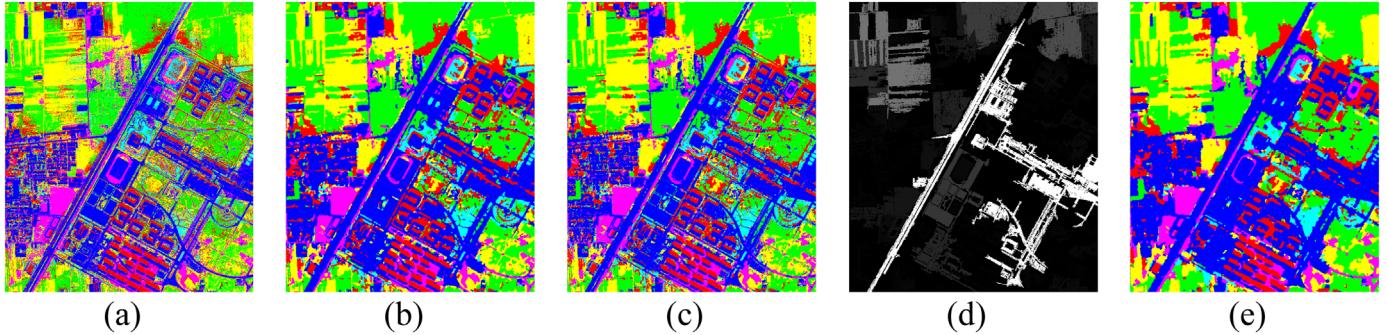
$$\begin{aligned} U(y_i|x) &= U(x|y_i) + U(y_i) \propto -[\ln p(x|y_i) + \ln p(y_i)] \\ &= -\ln \sum_{k=1}^K \pi_{ik} N(x|\theta_{ik}) + \sum_{j \in N_i} V^w(y_i, y_j) \end{aligned} \quad (11)$$

where  $U(x|y_i)$  is the spectral energy term obtained by Eq. (7),  $U(y)$  is the spatial energy function computed over the local neighborhood via  $V^w(y_i, y_j)$ . In our work, an eight-neighborhood system is considered. The spectral spatial classification label is finally given by:

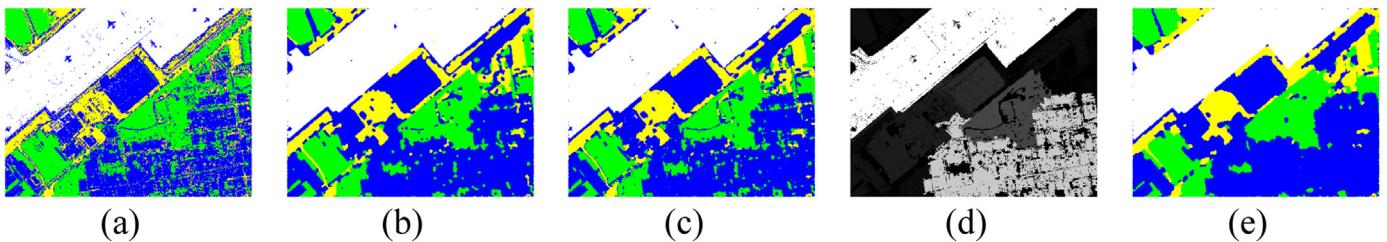
$$\hat{y} = \arg \min U(y_i|x) \quad (12)$$

According to Eq. (12), the final classification label can be achieved, the process of Simulated Annealing can be seen in Fig. 3. With introduction of spatial information, the classification procedure could obtain more accurate results. The spatial information actually can make the region more accordant with its original characteristics, which means properties within one region should be consistent. Therefore, the proposed method, which combines

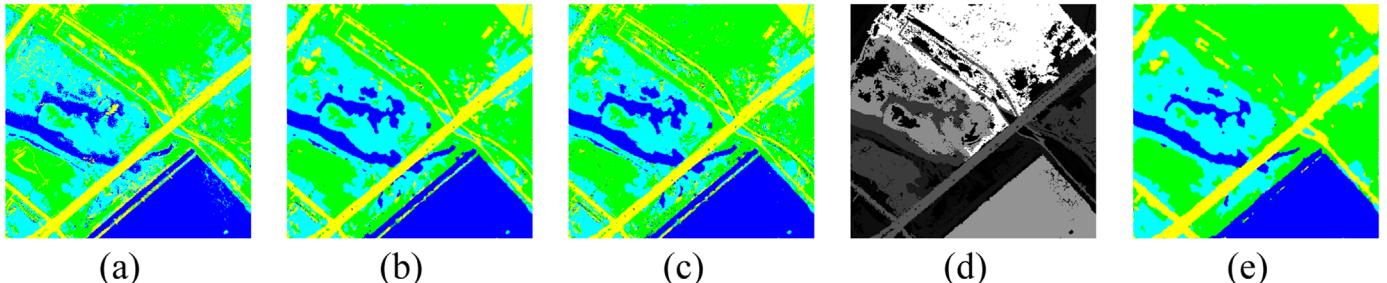




**Fig. 14.** Thematic maps obtained with the Xidian University data set: (a) IGMM method, (b) IGMMMLL method, (c) IGMMEMLL method, (d) Area map, and (e) IGMMWMRF method.



**Fig. 15.** Thematic maps obtained with the Xianyang International airport data set: (a) IGMM method, (b) IGMMMLL method, (c) IGMMEMLL method, (d) Area map, and (e) IGMMWMRF method.



**Fig. 16.** Thematic maps obtained with the Qujiang dam data set: (a) IGMM method, (b) IGMMMLL method, (c) IGMMEMLL method, (d) Area map, and (e) IGMMWMRF method.

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. 10. 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.

The accuracy of proposed classification method was evaluated by the reference data using overall accuracy, and Kappa coefficient. Overall accuracy is the ratio of the number of pixels which are classified correctly and the total number of pixels. And Kappa coefficient is an index of agreement between two images as a whole, the bigger the value of Kappa coefficient is, the better the classification results are. It is calculated by:

$$k = \frac{p_0 - p_c}{1 - p_c}$$

$$p_0 = \frac{N - FP - FN}{N}$$

$$p_c = \frac{TP + FP}{N} \times \frac{TP + FN}{N} + \frac{TN + FN}{N} \times \frac{TN + FP}{N}$$

here,  $N$  is the total number of pixels,  $FP$  is false positives,  $FN$  is false negatives,  $TP$  is true positives,  $TN$  is true negatives.

The overall accuracies and the Kappa coefficients for MinDC, MDC, MLC, MLR, SVM and IGMM are presented in Tables 4–6 to compare the spectral classification performance with different methods. The thematic maps of above methods obtained from the three data sets can be seen in Figs. 11–13. Meanwhile, the overall accuracies for above spectral classification methods which imposing isotropic MLL and anisotropic MLL spatial prior named MLCMLL, MLRMLL, SVMMMLL, IGMMMLL, MLCEMMLL, SVMEMMLL, IGMMEMLL, MLCWMRF, MLRWMRF, SVMWMRF, IGMMWMRF are presented in Tables 7–9 to compare the spectral-spatial classification performance with different methods. The EMLL is shorthand of multilevel logistic with gradient information which mentioned in [24]. To obtain unbiased conclusions, the classification process was repeated 10 times with randomly selected different training and test sets and the average accuracies are presented in our work.

From Tables 4–6 we can see that 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 produces relatively better results, which can be explained as achieving 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 in Tables 7–9. Tables 10 and 11 show the results of different values of  $\gamma$ ,  $\gamma$  is a experience value, which is generally between 1 and 10, different values have different results, we use four values as comparison. Though the value of  $\gamma$  is different, the proposed method is still outstanding. Since big size region has relatively higher impact to adjacent neighbor pixels than those small size region, the proposed anisotropic MLL prior could get a smoother and more stable result. Particularly, the proposed nonparametric supervised spectral-spatial classification algorithm outperformed other recently advanced spectral-spatial classification methods.

From Tables 12–14 we can see that the proposed method also achieves better classification in results in every class of three data sets. The thematic maps obtained from the two data sets can be seen in Figs. 14–16. The gray scale in Figs. 14(d), 15(d) and 16(d) denotes the region area information of corresponding pixels (light represents very big area, dark represents very small area), which captured by spectral classification. From Figs. 14–16 it can be seen that traditional MRF classification map contains a lot of noisy points, while the proposed method could obtain the classification map containing more homogeneous regions.

#### 4. Conclusions

In this paper, we present a new Bayesian approach with weighted Markov Random Fields for multispectral image classification. In the proposed method, we use Gaussian Mixture Model (GMM) take the place of traditional Single Gaussian Model (SGM) to fitting training data sets, also we construct an anisotropic MLL prior to capture the spatial contextual information to constraint the class labels. Experiments show that the GMM can fitting training data sets better than the SGM, and the anisotropic MLL actually lead to a smoother and more stable result compared with traditional isotropic MLL. We conduct a series of experiments on the new Bayesian approach, the proposed method shows competitive performance when comparison 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 no. 61571343), and the Fundamental Research Funds for the Central Universities (Grant no. JB140225).

#### References

- [1] G.M. Foody, Status of land cover classification accuracy assessment, *Remote Sens. Environ.* 80 (1) (2002) 185–201.
- [2] P.A. Townsend, D.P. Helmers, C.C. Kingdon, B.E. McNeil, K.M. de Beurs, K.N. Eshleman, 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) (2009) 62–72.
- [3] K. Shackelford, C.H. Davis, A hierarchical fuzzy classification approach for high resolution multispectral data over urban areas, *IEEE Trans. Geosci. Remote Sens.* 41 (9) (2003) 1920–1932.
- [4] E. Insanic, P.R. Siqueira, A maximum likelihood approach to estimation of vector velocity in Doppler radar networks, *IEEE Trans. Geosci. Remote Sens.* 50 (2) (2012) 553–567.
- [5] P. Ruiz, J. Mateos, G. Camps-Valls, Bayesian active remote sensing image classification, *IEEE Trans. Geosci. Remote Sens.* 52 (4) (2014) 2186–2196.
- [6] Qian Wang, Qingli Li, Hongying Liu, Yiting Wang and Jianzhong Zhu, An improved ISODATA algorithm for hyperspectral image classification, in: Image and Signal Processing, the 7th International Congress on IEEE, 2014, pp. 660–664.

- [7] Shulei Wu, Huandong Chen, Zhizhong Zhao, Haixia Long and Chunhui Song, An Improved Remote Sensing Image Classification Based on K-Means Using HSV Color Feature, in: Computational Intelligence and Security, the 10th International Conference on IEEE, 2014, pp. 201–204.
- [8] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, second ed., Wiley, New York, 2001.
- [9] J.A. Richards, X.P. Jia, *Remote Sensing Digital Image Analysis (An Introduction)*, Springer-Verlag, Berlin, 1999 (third revised and enlarged ed.).
- [10] P.S. Sisodia, V. Tiwari and A. Kumar, Analysis of Supervised Maximum Likelihood Classification for Remote Sensing Image, in: Recent Advances and Innovations in Engineering, International Conference on IEEE, 2014, pp. 1–4.
- [11] Jun Li, Jose M. Bioucas-Dias, A. Plaza, Spectral-Spatial Hyperspectral Image Segmentation Using Subspace Multinomial Logistic Regression and Markov Random Fields, *IEEE Trans. Geosci. Remote Sens.* 50 (3) (2012) 809–823.
- [12] J. Muñoz-Mari, F. Bovolo, L. Gómez-Chova, L. Bruzzone, G. Camps-Valls, Semisupervised one-class support vector machines for classification of remote sensing data, *IEEE Trans. Geosci. Remote Sens.* 48 (8) (2010) 3188–3197.
- [13] F. Melgani, L. Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, *IEEE Trans. Geosci. Remote Sens.* 42 (8) (2004) 1778–1790.
- [14] B. Waske, J.A. Benediktsson, Fusion of support vector machines for classification of multisensor data, *IEEE Trans. Geosci. Remote Sens.* 45 (12) (2007) 3858–3866.
- [15] S.J. Gershman, D.M. Blei, A tutorial on Bayesian nonparametric models, *J. Math. Psychol.* 56 (1) (2012) 1–12.
- [16] M. Fauvel, Y. Tarabalka, J.A. Benediktsson, J. Chanussot, J.C. Tilton, Advances in spectral-spatial classification of hyperspectral images, *Proc. IEEE* 101 (3) (2013) 652–675.
- [17] G. Moser, S.B. Serpico, Combining support vector machines and Markov random fields in an integrated framework for contextual image classification, *IEEE Trans. Geosci. Remote Sens.* 51 (5) (2013) 2734–2752.
- [18] C.M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, New York, 2007.
- [19] R.F. da Silva, A Dirichlet process mixture model for brain MRI tissue classification, *Med. Image Anal.* 2 (2007) 169–182.
- [20] N.T. Nguyen, R. Zheng, Z. Han, On identifying primary user emulation attacks in cognitive radio systems using nonparametric bayesian classification, *IEEE Trans. Signal Process.* 60 (3) (2012) 1432–1445.
- [21] R.M. Neal, Markov chain sampling methods for dirichlet process mixture models, *J. Comput. Graph. Stat.* 9 (2) (2000) 249–265.
- [22] F. Wood, S. Goldwater, and M.J. Black, A nonparametric bayesian approach to spike sorting, in: Engineering in Medicine and Biology Society, 28th Annual International Conference of the IEEE, 2006, pp. 1165–1168.
- [23] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell. PAMI* 6 (6) (1984) 721–741.
- [24] Y. Tarabalka, M. Fauvel, J. Chanussot, J.A. Benediktsson, SVM- and MRF-Based method for accurate classification of hyperspectral images, *IEEE Geosci. Remote Sens. Lett.* 7 (4) (2010) 736–740.



**Zhaobin Cui** received the B.Sc. degree in Information and Computing Science from Henan University of Technology in 2013. He is currently pursuing the Master's degree in signal and information processing with the VIPS Laboratory, School of Electronic Engineering, Xidian University.



**Ying Wang** received the B.Sc., M.Sc. and doctor degrees in signal and information processing from Xidian University, Xi'an, China, in 2003, 2006, and 2010 respectively. She is now an associate professor of Signal and Information Processing in Xidian University. Her research interests include medical image analysis, pattern recognition and computer-aided diagnosis.



**Xinbo Gao** (M'02-SM'07) received the B.Eng., M.Sc., and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively. From 1997 to 1998, he was a Research Fellow at the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Post-Doctoral Research Fellow at the Department of Information Engineering, the Chinese University of Hong Kong, Hong Kong. Since 2001, he has been at the School of Electronic Engineering, Xidian University. He is currently a Cheung Kong Professor of Ministry of Education, a Professor of Pattern Recognition and Intelligent System, and the Director of the

State Key Laboratory of Integrated Services Networks, Xi'an, China. His current research interests include multimedia analysis, computer vision, pattern recognition, machine learning, and wireless communications. He has published five books and around 200 technical articles in refereed journals and proceedings. Prof. Gao is on the Editorial Boards of several journals, including Signal Processing (Elsevier), and Neurocomputing (Elsevier). He served as the General Chair/ Co-Chair, Program Committee Chair/Co-Chair, or PC Member for around 30 major international conferences. He is currently a fellow of the Institution of Engineering and Technology.



**Yu Zheng** received the B.Eng. degree in electronic information engineering from Xidian University, Xi'an, China, in 2012. He is currently pursuing the Ph.D. degree in intelligent information processing with the VIPS Laboratory, School of Electronic Engineering, Xidian University. His current research interests include machine learning and computer vision.



**Jie Li** received the B.Sc., M.Sc. and Ph.D. degrees in Circuit and System from Xidian University, China, in 1995, 1998 and 2005 respectively. Since 1998, she joined the School of Electronic Engineering at Xidian University. Currently, she is an Associate Professor of Xidian University. Her research interests include computational intelligence, machine learning, and image processing. In these areas, she has published over 30 technical articles in refereed journals and proceedings including IEEE TCSVT, IJFS etc.