International Journal of Remote Sensing
Vol. 32, No. 24, 20 December 2011, 9843–9868

Taylor & Francis
Taylor & Francis Group

Technical Note

Classification of high-spatial resolution imagery based on distance-weighted Markov random field with an improved iterated conditional mode method

QIAN ZHANG, LIANGPEI ZHANG* and XIN HUANG
State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

(*Received 24 January 2010; in final form 2 November 2010*)

In this article, we propose a Distance-Weighted Markov Random Field (DwMRF) for classification of high–spatial resolution imagery. The proposed DwMRF integrates the spectral and spatial information of the image, and better coordinates the interaction between neighbours and the central pixels than the conventional Equal-weighted MRF (EwMRF). In addition, we propose a Serial Iterated Conditional Mode (SICM) method for the solution of the Markov Random Field (MRF) model. Experiments are conducted on three data sets: HYDICE data of the Mall in Washington, DC, HYMAP data of Purdue University and QuickBird data of Beijing. We compare the proposed DwMRF approach with other methods: the EwMRF, Maximum Likelihood Classification (MLC) and a multiresolution segmentation (Fractal Net Evolution Approach (FNEA)) method. Experiments show the DwMRF is robust and outperforms the other methods; furthermore, the proposed SICM method converges more rapidly than conventional Iterated Conditional Mode (ICM) and provides classification results comparable with the conventional ICM method.

1. Introduction

Recently, high–spatial resolution remote sensing images have become easy to acquire. They contain abundant detailed ground information such that complex characteristics of spectral and spatial information can be extracted for classification. However, in these images different objects may present spectral similarity or the same objects may have different spectral reflectance. Therefore, the traditional pixel-wise methods based on spectral information alone are not able to provide satisfactory results. In addition, the high–spatial resolution images contain abundant spatial information, which is very useful for identification of objects. Thus, many classifiers using both spectral and spatial information have been designed. In addition, it has been proved that the combination of spatial and spectral information in one classifier can improve the classification accuracy of high-resolution remote sensing images. Some spatial features are extracted and used for the classification of urban areas on high-resolution multispectral imagery (Huang *et al*. 2007). Huang *et al*. (2009) adopted the morphological texture features for mangrove forest mapping and species discrimination. Some methods utilize the spatial and spectral features to extract homogeneous objects, which

*Corresponding author. Email: zlp62@public.wh.hb.cn

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2011 Taylor & Francis
http://www.tandf.co.uk/journals
http://dx.doi.org/10.1080/01431161.2010.549854

can then be used to analyse the images, such as the Fractal Net Evolution Approach (FNEA) (Baatz and Schäpe 2000) and the adaptive mean-shift method (Huang and Zhang 2008).

The Markov Random Field (MRF) model plays an important role in the spectral–spatial classification for high-resolution images. The MRF usually combines with Bayesian theory used to establish *a priori* probability (Geman and Geman 1984, Li 1995, 2001). Jackson and Landgrebe (2002) constructed an adaptive Bayesian contextual classifier based on MRF; this classifier produces a high-quality classification map with significantly less pepper-salt noise than the pixel-wise method. In addition, the MRF model can easily integrate multi-source information (Gamba *et al*. 2007, Trianni and Gamba 2007). Tso and Olsen (2005) proposed a contextual classifier using both IKONOS panchromatic and multispectral data, which incorporates the multi-scale line features extracted from panchromatic data in the MRF model. This method improved the accuracy and visual results by avoiding over-smoothing because of the boundary information. Liu *et al*. (2005) used multi-temporal high–spatial resolution imagery for monitoring forest disease spread based on the MRF model, which models the spatial–temporal contextual prior probabilities of multi-temporal imagery. This method obtained better results than per-pixel, or single-data image classifications. Trianni and Gamba (2005) proposed a classifier based on the MRF model in urban areas, which integrates multi-source data including Landsat Thematic Mapper, ERS-1 and ASAR images.

Actually, the MRF is based on the principle that adjacent pixels are more likely to belong to the same kind of class. According to the first law of geography, close objects are more related than objects far apart. Therefore, in this study, we propose a Distance-weighted MRF (DwMRF), where the influences of neighbours for the central pixel are weighted according to distance.

On the other hand, there have been some methods available to solve the calculation of MRF, such as the Iterated Conditional Mode (ICM). However, the disadvantage of the ICM method is that it only reaches the steady state but cannot achieve convergence. Trianni and Gamba (2005) stopped the iterations when the percentage of changed pixels was smaller than a user specified value. Tso and Olsen (2005) assigned each pixel to the class with the highest frequency in the iterative process. However, in this case, when the iterations become steady, the labels of those pixels that do not achieve convergence will just swing between two classes; hence, it is difficult to determine the final results (Tso and Olsen 2005). In this context, we propose a Serial Iterated Conditional Mode (SICM), which allows rapid convergence, and is faster than the conventional ICM to reach the steady status.

This article is organized as follows. Section 2 introduces the basic theory of the MRF method, the proposed DwMRF and the SICM method. Experiments and the analysis of results are described in §3. Finally, conclusions are presented in §4.

2. Methodology

2.1 *Classification model based on Markov random field*

Here, the MRF classification model based on the Bayesian paradigm is illustrated simply. Let x_i denote the high-resolution image or feature vector with N pixels, and $\{i = 1, 2, \ldots, N\}$ denote the index of pixels or feature vectors. $p, \{p = 1, 2, \ldots, P\}$ is the number of dimensions of the image or feature vectors. $c = \{c_1, \ldots c_k, \ldots, c_K\}$ denotes the label level of the image, where $\{k = 1, 2, \ldots, K\}$, and K is the number of

classes. Here, the maximum *a posteriori* (MAP) is used to acquire the labels of image pixels; for any one pixel x, the formulation is represented as below:

$$c = \arg\max p(x|c)p(c). \tag{1}$$

$p(x|c)$ is the class-conditional distribution, in general modelled on the Gaussian distribution model; its energy function can be represented as equation (2) for pixel i. For the labels of any pixel i, the class-conditional distribution just depends on the spectral value of i; also $u_{\text{spectr}}(x_i, c_k)$ denotes the spectral cost term.

$$u_{\text{spectr}}(x_i, c_k) = \frac{1}{2}\ln|2\pi\,\Sigma_k| + \frac{1}{2}(x_i - \mu_k)^T\Sigma_k^{-1}(x_i - \mu_k), \tag{2}$$

where μ_k and Σ_k are the mean vector and covariance matrix of class k with dimension P, respectively; they can be obtained from training samples. In equation (1), $p(c)$ is *a priori* probability distribution of the label layer, and according to the Hammersley–Clifford theorem (Geman and Geman 1984) and the MRFs-Gibbs theorem (Li 2001), the global contextual information of i just depends on the local neighbourhood of i, denoted by N_i. Therefore, based on the MRF the *a priori* probability of the labelling $p(c)$ can be modelled as (Li 2001):

$$p(c) = \frac{1}{Z}\exp\{-u(c)\}, \tag{3}$$

$$u_{\text{sp}}(c(x_i)) = \sum_{j\in N_i}\beta I(c(x_i), c(x_j)), \tag{4}$$

$$I(c(x_i), c(x_j)) = \begin{cases} -1 \text{ if } c(x_i) = c(x_j) \\ 0 \text{ if } c(x_i) \neq c(x_j) \end{cases}, \tag{5}$$

where $c(x_i)$ is the label of pixel i, $u_{\text{sp}}(c(x_i))$ is called the spatial energy function (or spatial cost term) of pixel i in the label layer, Z is the normalizing constant, N_i is the spatial local neighbourhood of pixel i in the label layer and $i \notin N_i, j \in N_i$. Parameter β is controlling the strength between neighbourhood $j \in N_i$ and the pixel i in the label layer, which is decided by experience. Therefore, the function (1) can be solved by equation (6) or equation (7):

$$U(x_i, c(x_i)) = \arg\max(-u_{\text{spectr}}(x_i, c_k) - u_{\text{sp}}(c(x_i))), \tag{6}$$

$$U(x_i, c(x_i)) = \arg\min(u_{\text{spectr}}(x_i, c_k) + u_{\text{sp}}(c(x_i))). \tag{7}$$

In general, the ICM method is adopted to solve equation (7) to obtain the classification maps (Trianni and Gamba 2005, Tso and Olsen 2005). In this method, the initial label of the image is obtained by the pixel-wise Maximum Likelihood Classification (MLC); of course, the initial labels of the image can be acquired by other classifiers.

2.2 *Distance-weighted MRF*

In this section, the proposed DwMRF model will be defined in detail. Its basic principle is that the neighbours of pixels i in the neighbourhood N_i have different influences on the centre pixel i in the label layer. The proposed method uses the inverse of the Euclidian distance as the weight of interaction between neighbours and the centre pixel i.

$$I(c(x_i), \underset{j\in N_i}{c(x_j)}) = \begin{cases} -w_{ij} \text{ if } c(x_i) = c(x_j) \\ 0 \text{ if } c(x_i) \neq c(x_j) \end{cases}, \tag{8}$$

where w_{ij} is the weight of the interaction between i and j in the label layer. Figure 1 shows the square neighbourhood system of scale 3, i is the centre pixel. Figure 1(a) shows the neighbours of the pixel i, denotes $N_{i,j}\{j = 0, 1, \ldots, 7\}$. Figure 1($b$) shows the weight w_{ij} of $N_{i,j}$.

However, in order to keep the metric of the spectral cost term and the spatial cost term, this method redistributes the weight by the reciprocal of the Euclidean distance as follows:

$$W_{ij} = w_{ij} \cdot \frac{J}{w_{\text{all}}}, \tag{9}$$

$$w_{\text{all}} = \sum_{j\in N_i} w_{ij}, \tag{10}$$

where J is the number of neighbours, w_{all} is the summation of all the weights in the neighbourhood, J and w_{all} are constant values at a certain scale and W_{ij} is the regularized weight. Therefore, the neighbours near pixel i have stronger influence, and farther neighbours have a weaker influence. Then, function (8) becomes equation (11), and the spatial cost term is as follows (12):

$$I(c(x_i), \underset{j\in N_i}{c(x_j)}) = \begin{cases} -W_{ij} \text{ if } c(x_i) = c(x_j) \\ 0 \text{ if } c(x_i) \neq c(x_j) \end{cases}, \tag{11}$$

$$u_{\text{sp}}(c(x_i)) = \sum_{j\in N_i} I(c(x_i), c(x_j)). \tag{12}$$

$N_{i,0}$	$N_{i,1}$	$N_{i,2}$
$N_{i,3}$	i	$N_{i,4}$
$N_{i,5}$	$N_{i,6}$	$N_{i,7}$

(a)

$1/\sqrt{2}$	1	$1/\sqrt{2}$
1	i	1
$1/\sqrt{2}$	1	$1/\sqrt{2}$

(b)

Figure 1. The neighbours of pixel i at 3×3 window (a) and the distance-weighted weights of neighbours (b).

Equation (7) is substituted by the function (13), and the parameter β is substituted by parameter α. Parameter β in function (4) is just to control the interaction between pair-wise pixels in the neighbourhood of MRF, whereas the parameter α in function (13) can conveniently control the interrelations between spectral and spatial energy, where the sum of the weights of spatial and spectral energy is 1.

$$U(x_i, c(x_i)) = \arg\min_{c}((1-\alpha)\cdot u_{\text{spectr}}(x_i, c(x_i)) + \alpha \cdot u_{\text{sp}}(c(x_i))). \tag{13}$$

Experimental results show that the DwMRF method outperforms the conventional Equal-weighted MRF (EwMRF; as function (5)). In the EwMRF method, the interaction between pair-wise pixels is equal no matter what the distance between them is in spatial space. The value of parameter α can be obtained from experience; different scales have different values of α to obtain the optimal result. In general, as the scale increases, the optimal value of parameter α becomes smaller.

2.3 *Serial iterated conditional mode*

In general, the ICM method is usually employed to solve equation (13), which can quickly reach steady state. But the conventional ICM just reaches the steady state; it cannot reach convergence. In order to achieve convergence quickly, we propose an SICM method.

The process of the SICM is as follows. Suppose that c_l is the labels layer of the lth iteration result, obtained by the function (13). Figure 2 shows the neighbours of pixel i in a square neighbourhood system in the label layer in the iteration process, namely, $j, j = 0, 1, \ldots, 7$. Now, suppose the SICM is implementing the $l+1$ th iteration after the lth iteration and calculating the label of pixel i. Obviously, the pixels that are to the left and above pixel i, namely, $j, j = 0, 1, 2, 3$, denoted by $N_{i,\text{pre}}$, have labels in the $l+1$th iteration, denoted by $c_{i,j,l+1,\text{pre}}, j = 0, 1, 2, 3$. But the pixels that are at the right and below the pixel i, namely, $j, j = 4, 5, 6, 7$, denoted by $N_{i,\text{beh}}$, and pixel i do not have labels in the $l+1$th iteration, having just the labels in the lth iteration result, denoted by $c_{i,j,l,\text{beh}}, \{j = 4, 5, 6, 7\}$. And now when calculating the label for pixel i using the spatial energy term $u_{\text{sp}}(c)$, the neighbours' labels of i are $c_{i,j,l+1,\text{pre}}, j = 0, 1, 2, 3$ and $c_{i,j,l,\text{beh}}, \{j = 4, 5, 6, 7\}$.

It is obvious that in the proposed SICM method, the current pixel's label depends on the previous pixel's label obtained in the same iteration, so this method is called the SICM. The conventional ICM is called the Parallel ICM (PICM). The difference between the SICM and the conventional ICM is that the current pixel's label depends on its previous pixel's label obtained in the current iteration. The convergence of the

$c_{i,0,l+1}$	$c_{i,1,l+1}$	$c_{i,2,l+1}$
$c_{i,3,l+1}$	i	$c_{i,4,l}$
$c_{i,5,l}$	$c_{i,6,l}$	$c_{i,7,l}$

Figure 2. The neighbourhood of the pixel i at label layer.

central pixel is affected by its neighbours; therefore, the SICM can reduce the uncertainty of the PICM method and lead to a quick convergence. The initial labels can be obtained by any classifier.

Experiments show that the SICM method can reach convergence quickly, faster than the conventional ICM reaches steady state. And the difference between the classification accuracies is very small or even 0, and can be neglected.

3. Experiment and discussion

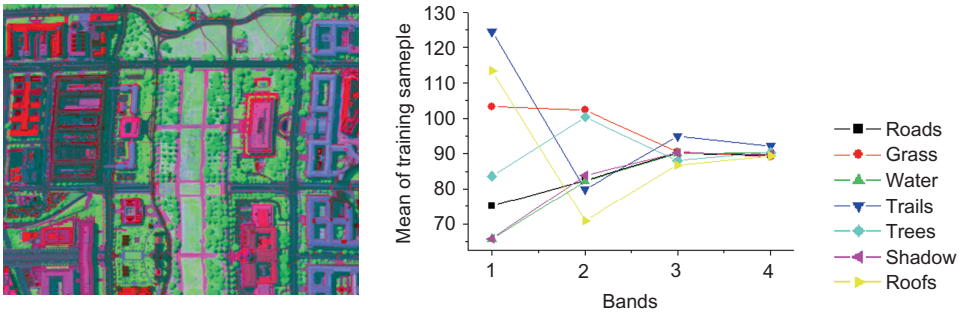
To verify the superiority of DwMRF, we compare it with the EwMRF, MLC and the widely used multiresolution segmentation method (Fractal Net Evolution Approach (FNEA)) (Baatz and Schäpe 2000). FNEA is the key technique of the well-known commercial software eCognition, which is an object-based analysis approach. For fair comparison, Support Vector Machine (SVM) is used to classify the segmented images.

Three high–spatial resolution data sets were employed to validate the proposed method: the HYDICE Washington, DC, image with 5 m spatial resolution, the HYMAP Purdue University image and the Beijing QuickBird image with 0.61 m spatial resolution. In experiments, the overall accuracy (OA) and Kappa coefficient (Ka) are used as the accuracy statistical parameters.

3.1 *The Washington, DC, data set*

The hyperspectral data set of the Mall, Washington, DC, was obtained by the HYDICE sensor, which has 210 bands in the 0.4–2.4 μm region of the visible and infrared spectrum. It retained 191 bands after the removal of the water absorption bands, and contains 1280 scan rows and 307 columns. This experiment adopts 4 bands extracted from the 191 bands by Principal Components Analysis (PCA). This data set has seven classes: roads, grass, water, trails, trees, shadow and roofs. Figure 3(*a*) shows one part of the first three principal components of the Washington, DC, data set. It has complex spectral information, as different classes have similar spectral information such as roofs, roads and trails, as shown in figure 3(*b*). Its training and test samples are listed in table 1.

3.1.1 Distance-weighted MRF. In general, as the scale increases, the spatial cost increases and the optimal parameter α becomes smaller, as table 2 shows. Experiments



(*a*) (*b*)

Figure 3. One part of Washington, DC, data set (*a*) and the histogram of training samples (*b*).

Table 1. Numbers of training and test samples.

Classes	Training	Test
Roads	110	6147
Grass	111	4622
Water	110	946
Trails	114	2401
Trees	110	5763
Shadow	111	2820
Roofs	111	5331

Table 2. Parameter α and scales of the highest statistical accuracy (OA and Ka) of DwMRF and EwMRF.

Scale	Method	α	OA (%)	Ka
3	EwMRF	0.70	95.01	0.9397
	DwMRF	0.65	94.95	0.9390
5	EwMRF	0.30	95.40	0.9445
	DwMRF	0.35	95.80	0.9492
7	EwMRF	0.15	95.17	0.9416
	DwMRF	0.15	95.29	0.9431
9	EwMRF	0.10	94.83	0.9374
	DwMRF	0.10	95.34	0.9436
11	EwMRF	0.10	92.15	0.9048
	DwMRF	0.10	93.90	0.9261

show that the appropriate scales are 3, 5, 7, 9 and 11, and the optimal parameter α of EwMRF and DwMRF are similar at the same scale (table 2). In addition, the McNemar's Test (McNemar 1947, Hollander and Wolfe 1999) was used and it was found that the DwMRF was significantly better than EwMRF when both got the best results. In table 2, the p-values are less than 0.00002, with the significance level being 0.05.

Both EwMRF and DwMRF obtained better results than the pixel-wise MLC and FNEA+SVM as shown in figure 4. The OA (Ka) of DwMRF and EwMRF are 95.51% (0.9457) and 94.12% (0.9288). The OA (Ka) of MLC is 81.57% (0.8255). Figure 4(*b*) shows the best results of FNEA with SVM: OA = 85.76% and Ka = 0.8282.

In MLC and FNEA+SVM (figure 4(*a*) and (*b*)), some shadows are misclassified as water, and roofs as trails. And there are many speckle errors in MLC. Although the grass and trails are well-extracted by the FNEA+SVM, some shadows and roads are wider than their actual values and many roofs are misclassified as roads. It is interesting to see that those errors were all corrected in both DwMRF and EwMRF (figure 4(*c*) and (*d*)), and the DwMRF are even better. For example, trails in the middle of the maps and some shadows were removed in EwMRF, but DwMRF identified them effectively.

Figure 5 shows that when α increases, the accuracies of DwMRF are always greater than those of EwMRF and the differences between DwMRF and EwMRF

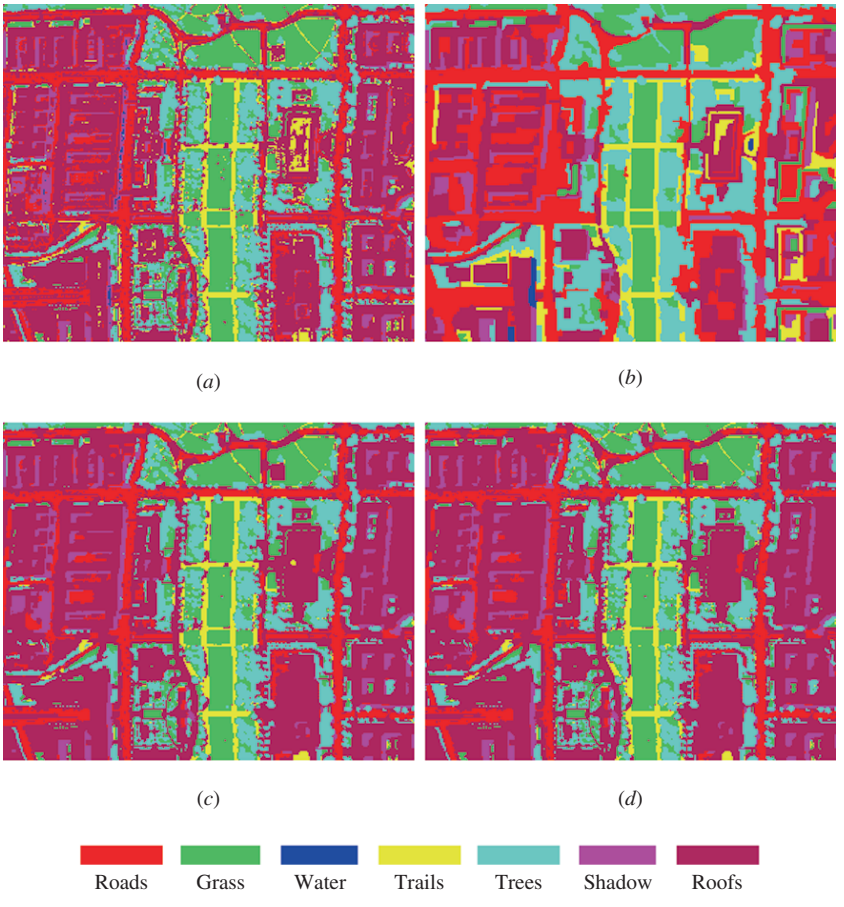


Figure 4. Classification maps of one part of Washington, DC, data set. (*a*) MLC, (*b*) FENA+SVM, S = 10, sw = 0.2, cw = 0.5, (*c*) DwMRF and (*d*) EwMRF with scale = 5, $\alpha = 0.4$.

are increasing. The classification maps in different scales presented in figure 6 further demonstrate the advantages of the DwMRF compared with EwMRF.

3.1.2 Multiscale distance-weighted Markov random field. To further prove the performance of DwMRF, we propose to use the multiscale distance-weighted (MDwMRF) classifier in this section. This method fuses two or three scales of DwMRF, as the spectral cost is the same in different scales of DwMRF for 1 pixel. Therefore, the multiscale information is fused in the decision level according to the spatial cost of each single scale DwMRF:

$$\overline{u_{s,\mathrm{sp}}(c)} = \frac{u_{s,\mathrm{sp}}(c)}{J(s)}, \tag{14}$$

$$c(x_i) = \underset{s,s\,=\,1,2,\ldots,S}{\arg\min}\,(\overline{u_{s,\mathrm{sp}}(c(x_i))}), \tag{15}$$

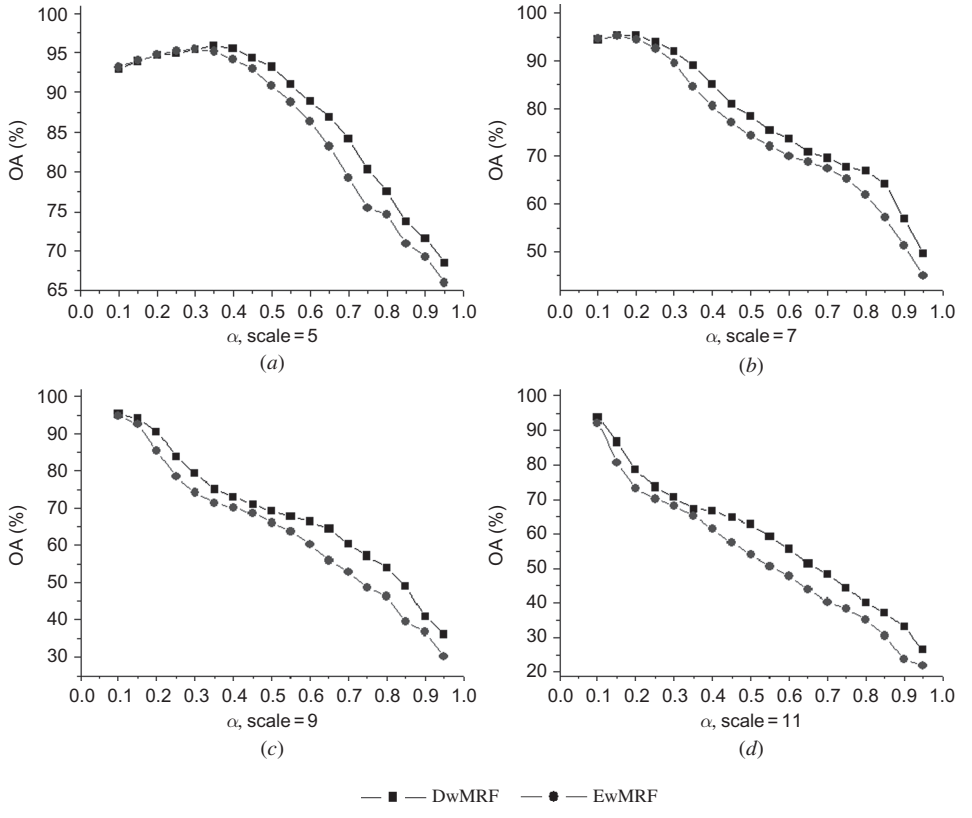


Figure 5. The OA of DwMRF and EwMRF at scales (*a*) 5, (*b*) 7, (*c*) 9 and (*d*) 11.

where scale $s, s = 1, 2, \ldots S$, is the number of the scales, $J(s)$ is the number of neighbourhoods in the scale s, $u_{s,\mathrm{sp}}(c(x_i))$ denotes the spatial cost in the scale s and $\overline{u_{s,\mathrm{sp}}(c(x_i))}$ is the normalized spatial cost for pixel i. The label of i can be decided according to equation (15), which comes from one of the single scale DwMRF classes where the normalized spatial cost is the minimum.

In figure 7(*a*), both MDwMRF and MEwMRF considered three scales (3, 5 and 7). In figure 7(*b*), scales 5 and 9 were fused. It is obvious that MDwMRF outperforms MEwMRF.

3.1.3 Serial iterated conditional mode. The experiments with the SICM and PICM methods are based on the DwMRF method. Table 3 shows the numbers of pixels that do not converge after the PICM reaches steady states, which are denoted as uncertain pixels. As the scale increases, the general trend in the numbers of uncertain pixels is decreasing with the same parameter α. Experiments proved that as the iteration continues after the PICM reaches the steady state, the labels of uncertain pixels will swing between two fixed classes.

Figure 8 shows the iterations of SICM and PICM of DwMRF at scales 3, 5, 7 and 9. The iterations of SICM are those which DwMRF needs to reach convergence, and the PICM iterations are the minimal iterations for the DwMRF to reach the steady

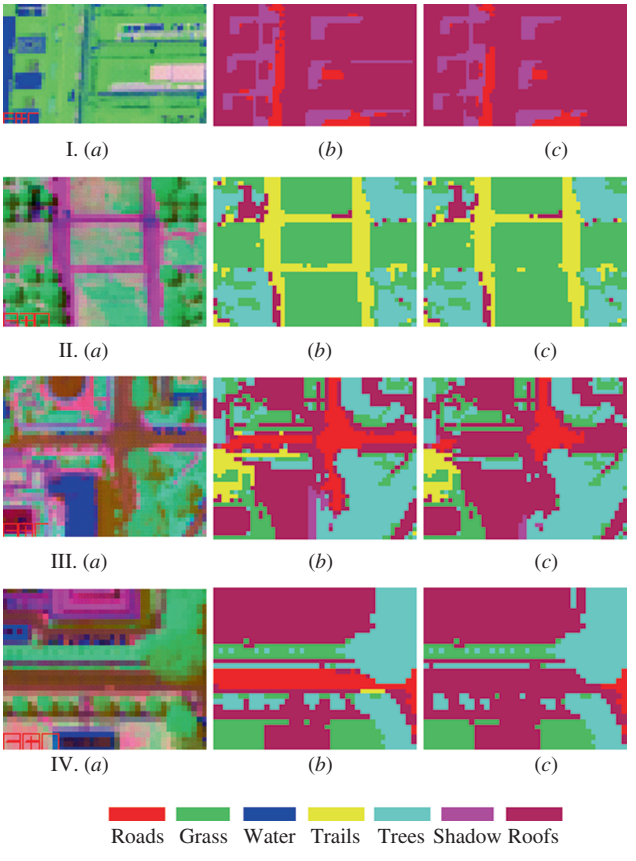


Figure 6. I–IV: (*a*) Four parts of Washington, DC, data set and the corresponding classification maps of (*b*) DwMRF and (*c*) EwMRF. I and II, scale = 5, $\alpha = 0.4$; III, scale = 7, $\alpha = 0.3$; and IV, scale = 9, $\alpha = 0.4$.

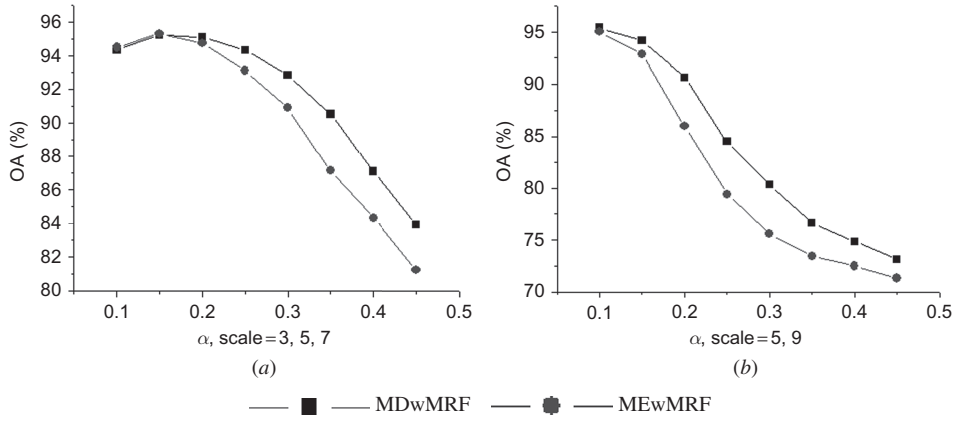


Figure 7. The OA of (*a*) MDwMRF scale = 3, 5 and 7 and (*b*) MEwMRF scale = 5 and 9.

Table 3. The numbers of uncertain pixels of PICM after they reached the steady state in EwMRF.

α	Scale = 3	Scale = 5	Scale = 7	Scale = 9	Scale = 11
0.1	29	44	34	28	16
0.2	114	83	79	38	21
0.3	250	160	80	48	29
0.4	362	150	92	62	44
0.5	592	195	93	70	33
0.6	880	194	96	61	38
0.7	1274	323	90	36	19
0.8	1797	343	119	34	19
0.9	2473	297	99	40	12

state. Besides, in the method of DwMRF, each sub-process of iteration of the PICM and SICM takes the same period of time, denoted as unit time, which is related to the size of the remote sensing data set and computer hardware. Figure 8 shows that the iterations of the SICM are almost always less than those of PICM at any scale. And the difference between them becomes larger with increasing parameter α and scales. When the parameter α is relatively large, the iterations of PICM are about twice that of SICM, or even more.

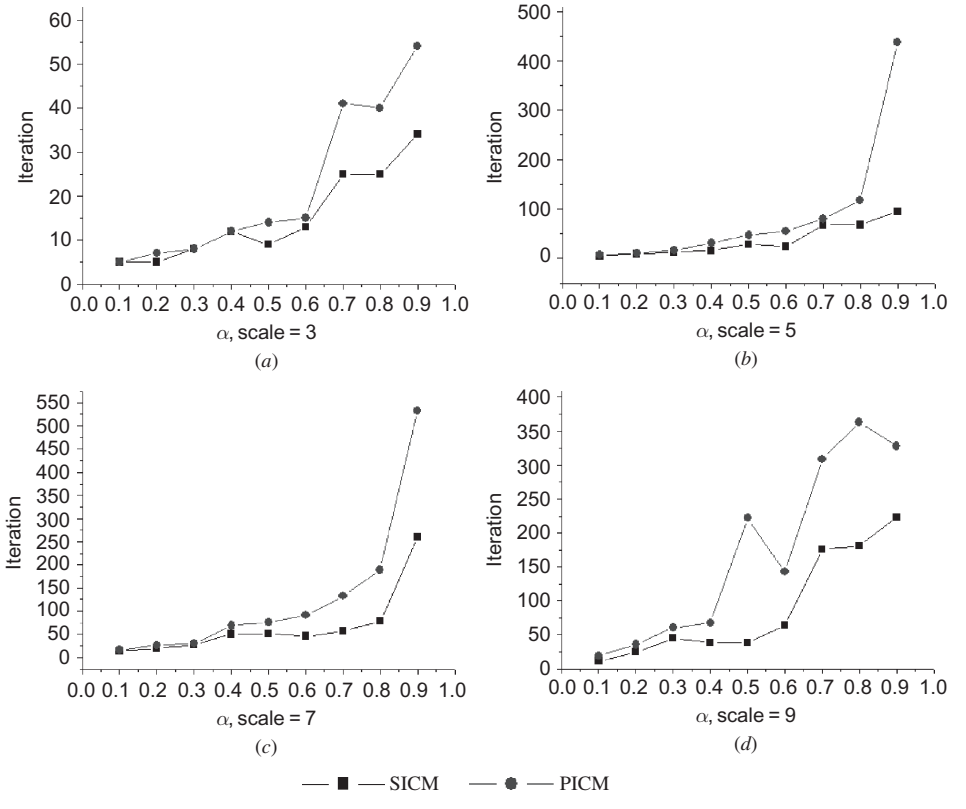


Figure 8. The iterations of SICM and PICM of DwMRF at scales (*a*) 3, (*b*) 5, (*c*) 7 and (*d*) 9.

SICM can produce as good results as PICM; table 4 proves that the accuracy statistics of SICM and PICM are similar or even the same, but with many fewer iteration steps.

3.2 *The Purdue University data set*

The hyperspectral data set of Purdue University, obtained by the HYMAP sensor, has 126 bands in the visible and infrared spectrum and is also of high-spatial resolution. This experiment adopts just four features extracted from the 126 bands by PCA. Figure 9(*a*) shows one part of the data set, which contains six classes: roads, grass, shadow, bare land, trees and roofs. It has complex spectral information in that different classes have similar spectral information such as roofs, roads and trails, and the same class has different spectral information such as roads and roofs, as shown in figure 9. This data set contains 377 scan rows and 512 columns; its training and test samples are listed in table 5.

3.2.1 Distance-weighted MRF. Table 6 shows that the optimal parameter α becomes smaller with increasing scale, and the optimal parameter α of EwMRF and DwMRF are similar at the same scale. Also, experiments proved that the appropriate scales are 3, 5, 7, 9 and 11. Once again, in order to verify whether DwMRF is significantly better than EwMRF when both obtain their best results (table 6), the McNemar's test was used. The *p*-values at different scales is lower than 0.0002, indicating that the DwMRF is significantly better than the EwMRF.

Figure 10 shows the best results of FNEA+SVM, DwMRF, EwMRF (table 6) and MLC. The OA (Ka) of MLC and FNEA+SVM are 86.58% (0.8255) and 80.28% (0.7421), respectively. In figure 10(*a*) and (*b*), grass sketched out by the blue oval boxes are misclassified as trees; roads sketched out by the black rectangle are misclassified as roofs; some bare land is polluted by pixels that are misclassified as roads. Besides, in FNEA+SVM (figure 10(*b*)), there are many grass areas that are misclassified as bare land, and areas of roofs and shadows are bigger than in reality. However, both DwMRF and EwMRF rectified them to a certain extent, and DwMRF was better than EwMRF. For example, roofs sketched out by the yellow rectangular box are lost in the EwMRF but well maintained in DwMRF.

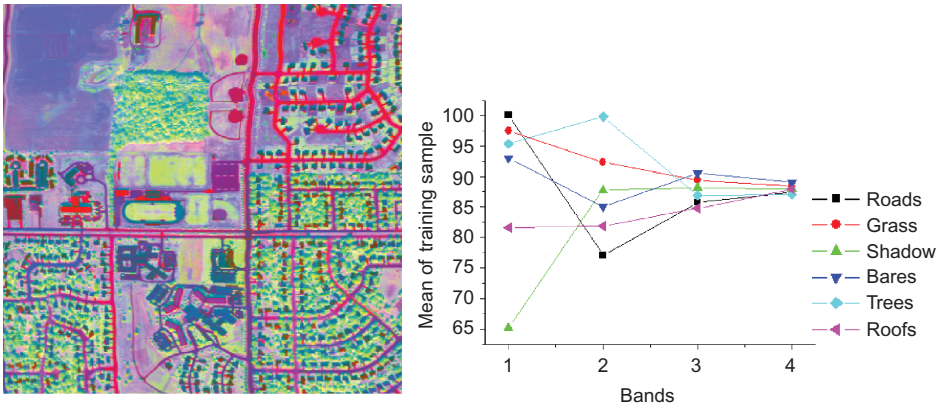
Figure 11(*a*) shows that the OA values of DwMRF are higher than those of EwMRF after they reach the best OA at scale 5. However, the OA value of DwMRF is always higher than EwMRF in the larger scales 7, 9 and 11. Figure 11 shows that the average difference of OA is increasing from scale 5 to scale 11.

The classification map also proves that DwMRF outperforms EwMRF, as shown in figure 12. Figure 12(*a*) presents the first three principal components of four principal component features extracted from original images by PCA. Four roofs are misclassified by EwMRF as roads in figure 12I(*c*); the grass in the middle of road and some roofs are also smoothed out by EwMRF in figure 12II(*c*). However, they are all preserved by the DwMRF to some extent (figure 12(*b*)).

3.2.2 Multiscale distance-weighed Markov random field. Figure 13 shows that the OA of MDwMRF is higher than MEwMRF, especially the combination of scales 5 and 9. This experiment proves that DwMRF outperforms the EwMRF method, not only with the single scales but also with multiscales.

Table 4. The OA (%)/ Ka of SICM and PICM in the DwMRF method.

	Scale = 3		Scale = 5		Scale = 7		Scale = 9		Scale = 11	
α	SICM	PICM	SICM	PICM	SICM	PICM	SICM	PICM	SICM	PICM
0.1	91.5/0.898	91.5/0.898	92.9/0.915	93.0/0.915	94.4/0.933	94.4/0.933	95.3/0.944	95.4/0.944	93.9/0.926	94.0/0.928
0.2	92.3/0.907	92.3/0.907	94.6/0.935	94.6/0.935	95.2/0.942	95.2/0.942	90.5/0.884	90.7/0.887	78.6/0.739	78.8/0.740
0.3	93.0/0.916	93.0/0.916	95.4/0.944	95.5/0.946	92.0/0.902	92.2/0.904	79.2/0.746	79.8/0.753	70.7/0.641	70.8/0.643
0.4	93.6/0.924	93.7/0.924	95.5/0.946	95.5/0.946	85.1/0.818	85.2/0.820	73.0/0.670	73.1/0.671	66.5/0.591	66.6/0.591
0.5	94.4/0.934	94.5/0.934	93.2/0.917	92.9/0.914	78.3/0.735	78.6/0.739	69.1/0.622	69.5/0.627	62.8/0.545	62.3/0.538
0.6	94.7/0.938	94.8/0.938	88.8/0.864	88.9/0.865	73.6/0.678	73.7/0.678	66.3/0.588	66.4/0.589	55.5/0.454	55.9/0.549
0.7	94.8/0.942	95.2/0.942	84.2/0.807	84.2/0.808	69.6/0.628	70.0/0.633	60.2/0.513	60.6/0.517	48.3/0.364	49.8/0.383
0.8	94.0/0.929	94.2/0.929	77.6/0.726	78.3/0.735	66.9/0.595	67.0/0.596	53.9/0.434	52.6/0.418	40.1/0.263	41.9/0.285
0.9	89.0/0.877	89.8/0.877	71.6/0.653	71.9/0.656	56.9/0.472	57.4/0.477	40.9/0.273	42.4/0.291	33.0/0.177	33.7/0.185



(*a*) (*b*)

Figure 9. One part of Purdue University data set (*a*) and the histogram of training samples (*b*).

Table 5. Numbers of training and test samples.

Classes	Training	Test
Roads	53	3812
Grass	53	4726
Shadow	53	900
Bares	37	344
Trees	49	2517
Roofs	53	4775

Table 6. Parameter α and scales of the highest statistical accuracy (OA and Ka) of DwMRF and EwMRF.

Scale	Method	α	OA (%)	Ka
3	EwMRF	0.85	93.04	0.9095
	DwMRF	0.85	92.87	0.9074
5	EwMRF	0.50	93.47	0.9152
	DwMRF	0.55	93.77	0.9175
7	EwMRF	0.35	92.88	0.9075
	DwMRF	0.30	93.45	0.9149
9	EwMRF	0.15	91.67	0.8918
	DwMRF	0.25	92.96	0.9085
11	EwMRF	0.10	90.34	0.8744
	DwMRF	0.10	92.07	0.8970

3.2.3 Serial iterated conditional model. Table 7 shows the number of uncertain pixels. Figure 14 shows the iterations of SICM and PICM. In general, as the scales and the parameter α increase, the difference in iterations is larger between SICM and PICM. Therefore, it is obvious that the method of SICM has the advantage of speed of calculation compared with the PICM method.

Table 8 shows that the OA and Ka of SICM and PICM are almost equal, except when the parameter α equals 0.8 and 0.9 at scales 7, 9 and 11. Experiments with the

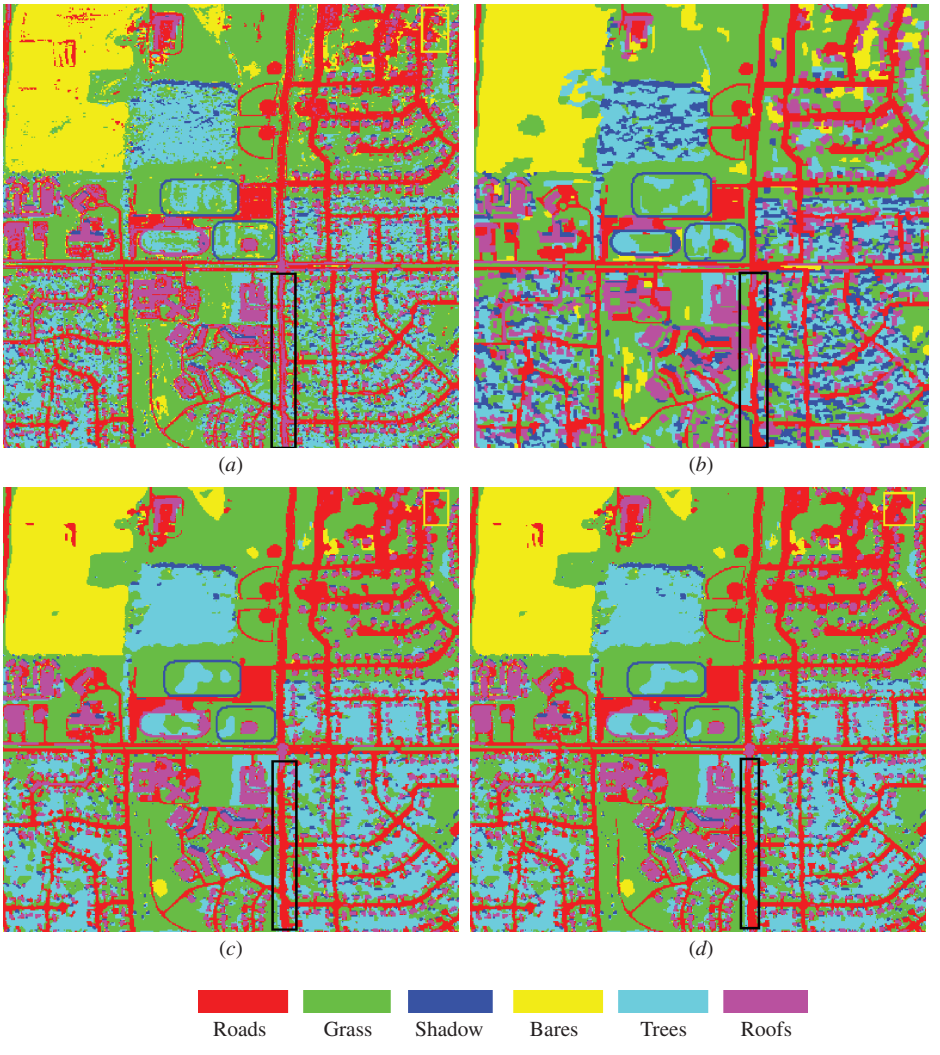


Figure 10. Classification maps of Purdue University data set. (*a*) MLC, (*b*) FNEA+SVM $S = 5$, sw = 0.2, cm = 0.5, (*c*) DwMRF, scale = 5, $\alpha = 0.55$ and (*d*) EwMRF, scale = 5, $\alpha = 0.50$.

Purdue University data set also prove that SICM outperforms PICM in the speed of computing and can also deliver results as good as PICM.

3.3 *The Beijing data set*

Figure 15(*a*) shows the high–spatial resolution data set of Beijing obtained by QuickBird. It consists of three multispectral bands (RGB) with 0.61 m resolution. Because roofs and roads used the same material, they have a similar spectrum and are difficult to distinguish, as shown in figure 15(*b*). The numbers of training and test samples are shown in table 9.

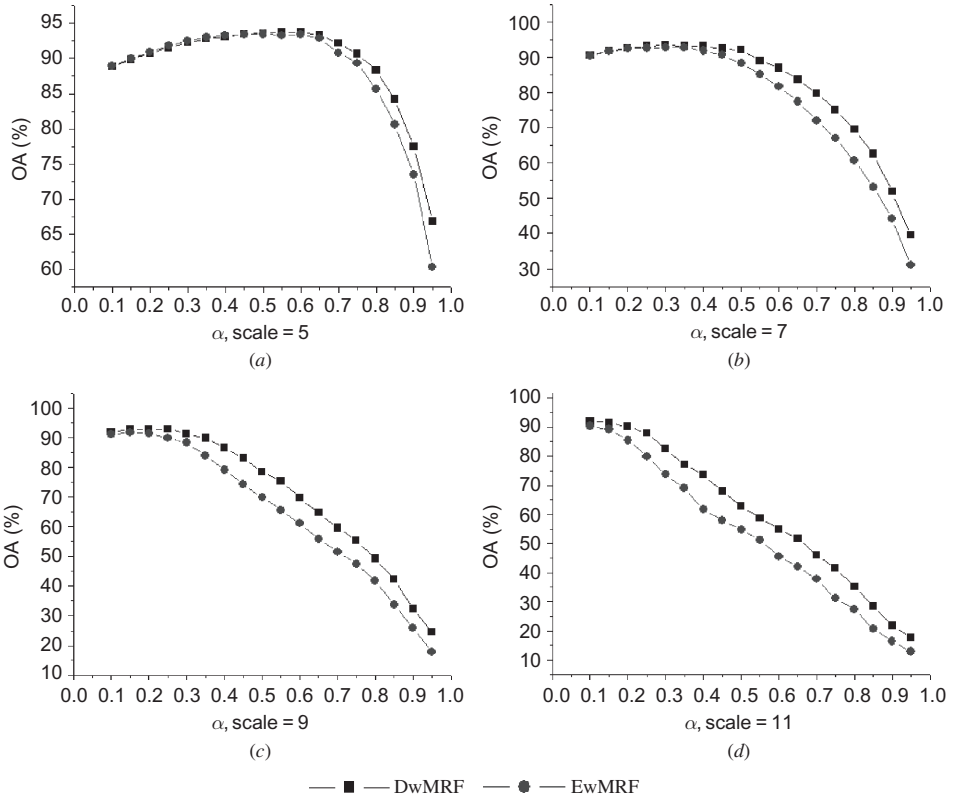


Figure 11. The OA of DwMRF and EwMRF at scales (*a*) 5, (*b*) 7, (*c*) 9 and (*d*) 11.

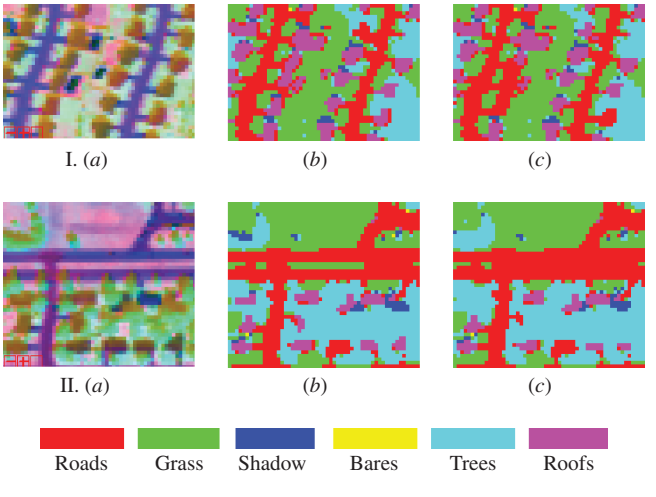


Figure 12. I–II: (*a*) Two parts of Purdue university data set and the corresponding classification maps of (*b*) DwMRF and (*c*) EwMRF with scale = 7, $\alpha = 0.5$.

3.3.1 Distance-weighted MRF (DwMRF). Table 10 shows the optimal parameter α of EwMRF and DwMRF is similar or even the same at the same scale, and the OA

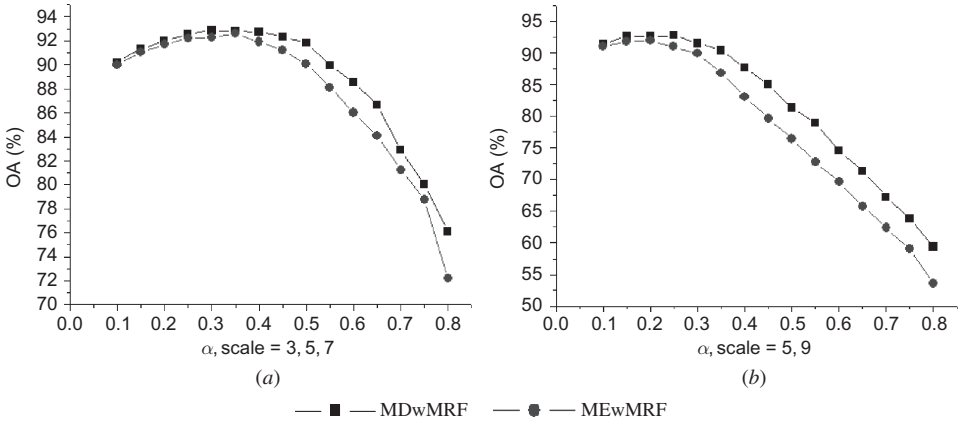


Figure 13. The OA of (*a*) MDwMRF, scale = 3, 5 and 7 and (*b*) MEwMRF, scale = 5 and 9.

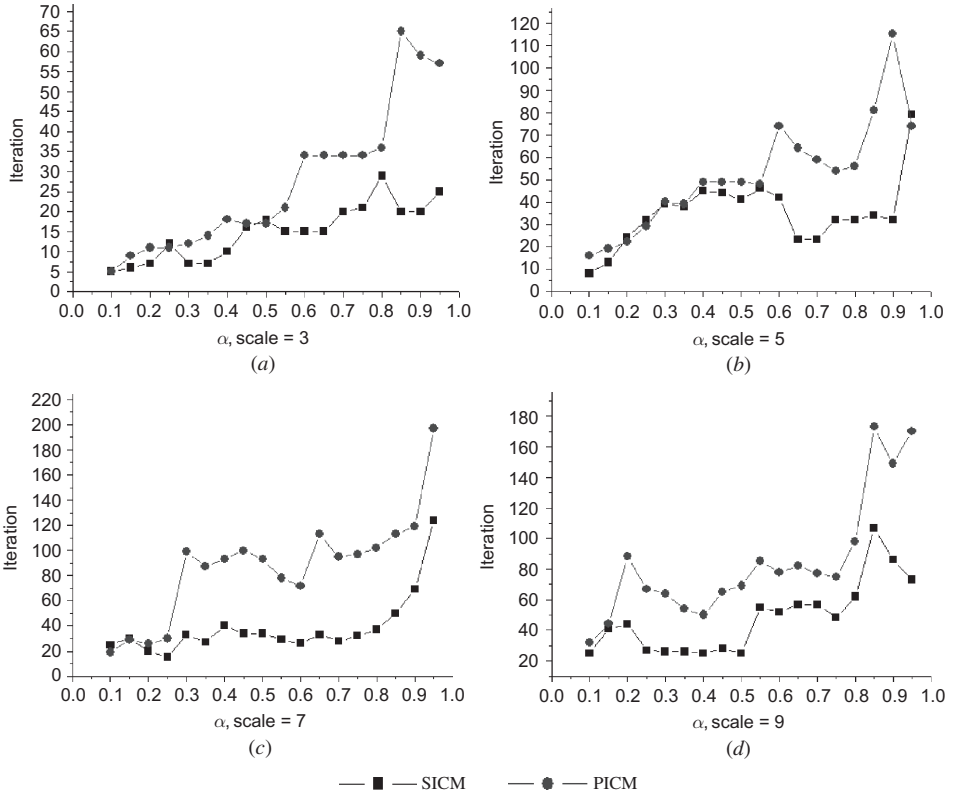


Figure 14. The iterations of SICM and PICM of DwMRF at scales (*a*) 3, (*b*) 5, (*c*) 7 and (*d*) 9.

and Ka of DwMRF are higher than EwMRF. The *p*-values of the McNemar's test is less than 0.00007, which proves that DwMRF is significantly better than EwMRF when both deliver their best results (table 10).

Table 7. The numbers of uncertain pixels of the PICM after they reached the steady state in the DwMRF.

α	Scale = 3	Scale = 5	Scale = 7	Scale = 9	Scale = 11
0.1	26	31	28	24	8
0.2	88	80	54	28	37
0.3	176	129	64	59	12
0.4	311	157	84	34	23
0.5	493	170	77	43	34
0.6	656	196	86	48	25
0.7	942	262	84	24	29
0.8	1422	265	95	40	15
0.9	2210	265	77	32	3

The best results of DwMRF and EwMRF at scale 5 (table 10) are much better than those of MLC and FNEA+SVM in figure 16. The OA (Ka) of the MLC and FNEA+ SVM are 65.30% (0.5700) and 74.10% (0.6708), and about 24.24% (0.2903) and 15.44% (0.1895) lower than DwMRF. DwMRF shows a slight improvement compared with EwMRF. Compared with DwMRF and EwMRF (figure 16(*c*) and (*d*)), there are too many roofs misclassified as roads in MLC (figure 16(*a*)). In FNEA+SVM (figure 16(*b*)), too many roofs are misclassified as bare land, and there are many misclassified points in the lower-right corner.

Figure 17(*a*) and (*b*) show the OA of DwMRF slightly lower than that of EwMRF before they achieve their best results. However, after they reached the optimal α, the OA of DwMRF is higher than that of EwMRF. At scales 9 and 11 (figure 17(*c*) and (*d*)), the OA of DwMRF is always higher than that of EwMRF. Figure 18 shows the classification map of DwMRF (OA = 83.46%, Ka = 0.7770) and EwMRF (OA = 76.66%, Ka = 0.6836) with the same scale 5 and parameter α of 0.6. It can be seen that some road and shadow were smoothed out in EwMRF. However, all were retained well in the DwMRF.

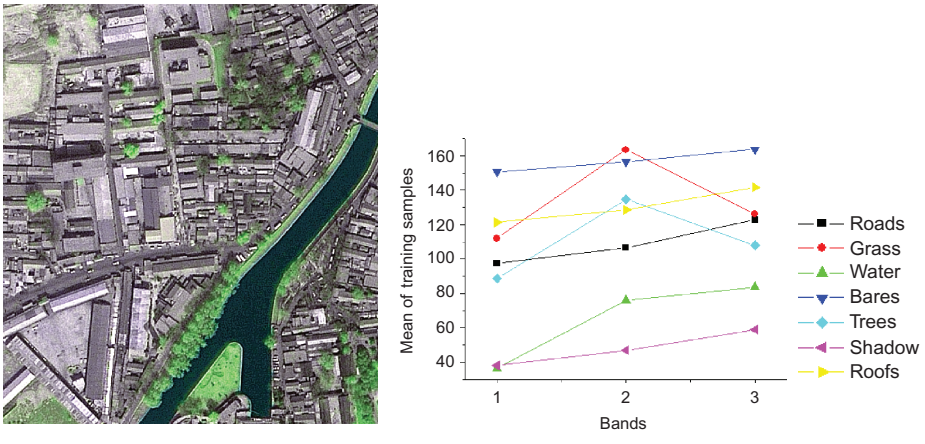
3.3.2 Multiscale distance-weighted Markov random filed. Compared with EwMRF, the advantages of DwMRF are still obvious in the multiscale MRF method. In the multiscale 3, 5 and 7 (figure 19(*a*)), the improvement of MDwMRF is obvious compared with MEwMRF. In multiscale 5 and 9, the MDwMRF is always higher than the MEwMRF (figure 19(*b*)).

3.3.3 Serial iterated conditional model (SICM). Table 11 shows that the general trend of the number of uncertain pixels is decreasing with the increase of scale. Figure 20 shows that the iterations of both PICM and SICM increase with the increase of scales and parameters α; and the iterations of PICM are about twice that of SICM, or even more.

The improved method SICM is not only much faster than the PICM in computing speed but can also reach convergence, and can obtain results as good as PICM. Table 12 shows that the OA and Ka both of SICM and PICM in the DwMRF method are similar and the differences between them are very small.

Table 8. The OA (%)/Ka of SICM and PICM in the DwMRF method.

	Scale = 3		Scale = 5		Scale = 7		Scale = 9		Scale = 11	
α	SICM	PICM	SICM	PICM	SICM	PICM	SICM	PICM	SICM	PICM
0.1	87.5/0.840	87.6/0.839	88.9/0.856	88.8/0.855	90.6/0.878	90.5/0.877	91.8/0.894	91.8/0.893	92.1/0.897	92.0/0.896
0.2	88.5/0.851	88.5/0.850	90.8/0.880	90.7/0.879	92.6/0.904	92.6/0.904	92.8/0.907	92.8/0.906	90.2/0.872	90.6/0.879
0.3	89.2/0.860	89.2/0.860	92.3/0.900	92.2/0.899	93.4/0.915	93.6/0.916	91.3/0.887	91.3/0.888	82.6/0.775	83.6/0.788
0.4	90.2/0.873	90.1/0.872	93.0/0.910	93.0/0.909	93.1/0.911	93.2/0.911	86.6/0.827	87.5/0.839	73.6/0.662	74.7/0.675
0.5	91.0/0.883	90.9/0.882	93.6/0.918	93.7/0.918	92.1/0.897	91.4/0.888	78.6/0.725	79.9/0.742	62.9/0.526	64.5/0.547
0.6	91.7/0.893	91.5/0.890	93.7/0.912	93.2/0.912	87.0/0.832	87.0/0.831	69.8/0.613	71.4/0.633	55.0/0.426	57.7/0.460
0.7	92.3/0.900	92.2/0.899	92.1/0.898	92.5/0.902	79.9/0.741	79.9/0.741	59.6/0.485	61.9/0.514	45.8/0.309	48.5/0.342
0.8	92.8/0.907	92.7/0.905	88.4/0.849	88.5/0.850	69.5/0.601	70.4/0.621	49.4/0.353	50.2/0.364	35.0/0.162	37.5/0.201
0.9	92.5/0.902	92.3/0.899	77.5/0.709	78.0/0.716	51.9/0.384	53.9/0.411	32.5/0.127	35.7/0.176	21.7/0.23	24.5/0.25



(*a*) (*b*)

Figure 15. One part of Beijing data set (*a*) and the histogram of training samples (*b*).

Table 9. Numbers of training and test samples.

Classes	Training	Test
Roads	44	2322
Grass	37	933
Water	37	547
Bares	43	1396
Trees	41	4644
Shadow	41	9677
Roofs	40	10 286

Table 10. Parameter α and the scale of the highest statistical accuracy (OA and Ka) of DwMRF and EwMRF.

Scale	Method	α	OA (%)	Ka
3	EwMRF	0.65	82.10	0.7690
	DwMRF	0.7	82.53	0.7745
5	EwMRF	0.3	88.70	0.8500
	DwMRF	0.4	89.54	0.8603
7	EwMRF	0.15	87.89	0.8378
	DwMRF	0.15	88.74	0.8502
9	EwMRF	0.1	83.88	0.7872
	DwMRF	0.1	87.61	0.8338
11	EwMRF	0.1	71.93	0.6154
	DwMRF	0.1	79.16	0.7170

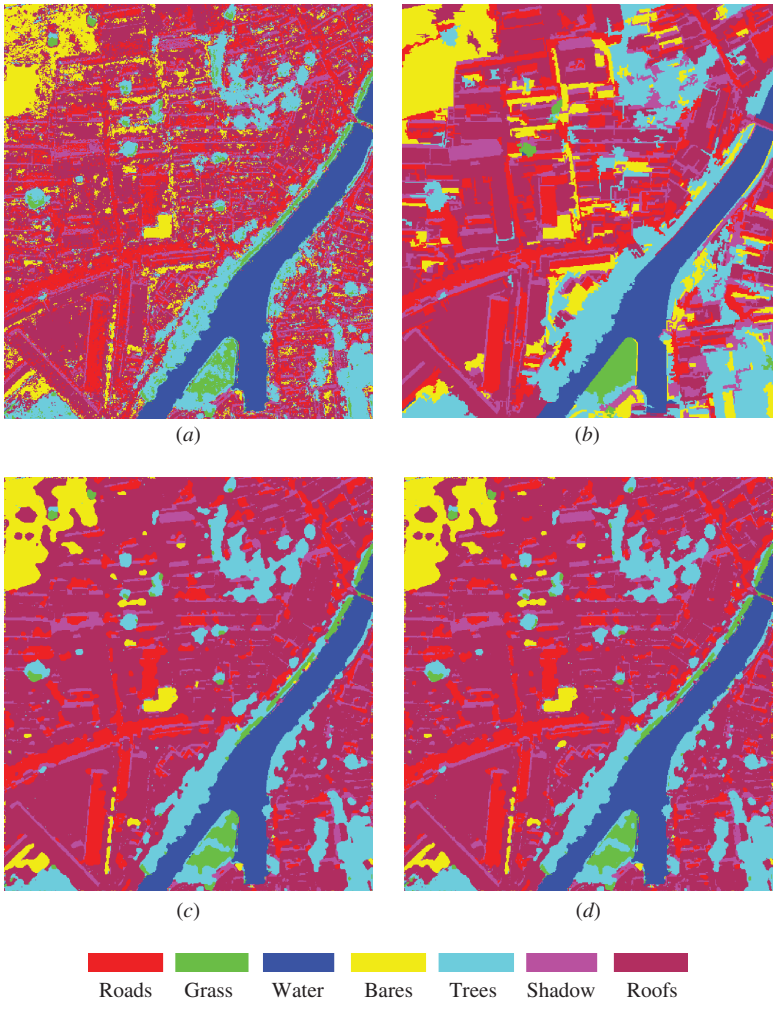


Figure 16. Classification maps of Beijing data set. (*a*) MLC, (*b*) FNEA+SVM, $S = 30$, $\mathrm{sw} = 0.3$, $\mathrm{cw} = 0.5$, (*c*) DwMRF, scale = 5, $\alpha = 0.4$ and (*d*) EwMRF, scale = 5, $\alpha = 0.3$.

4. Conclusions

Experiments show that the DwMRF obviously outperforms MLC and FNEA+SVM. Also proved is that the proposed DwMRF is more robust than the conventional EwMRF. Especially when the scale increases, the advantages of the proposed method become more prominent because DwMRF well coordinates the interaction between neighbours and the centre pixels. It can be seen that DwMRF could well maintain the integrity of objects, avoiding the displacement or smoothing out by EwMRF, especially with narrow objects. Experiments also proved that the advantages of DwMRF are more prominent when adopting multiscale.

In addition, experiments proved that the SICM can achieve convergence faster than the conventional ICM (PICM), and the results obtained from the SICM are

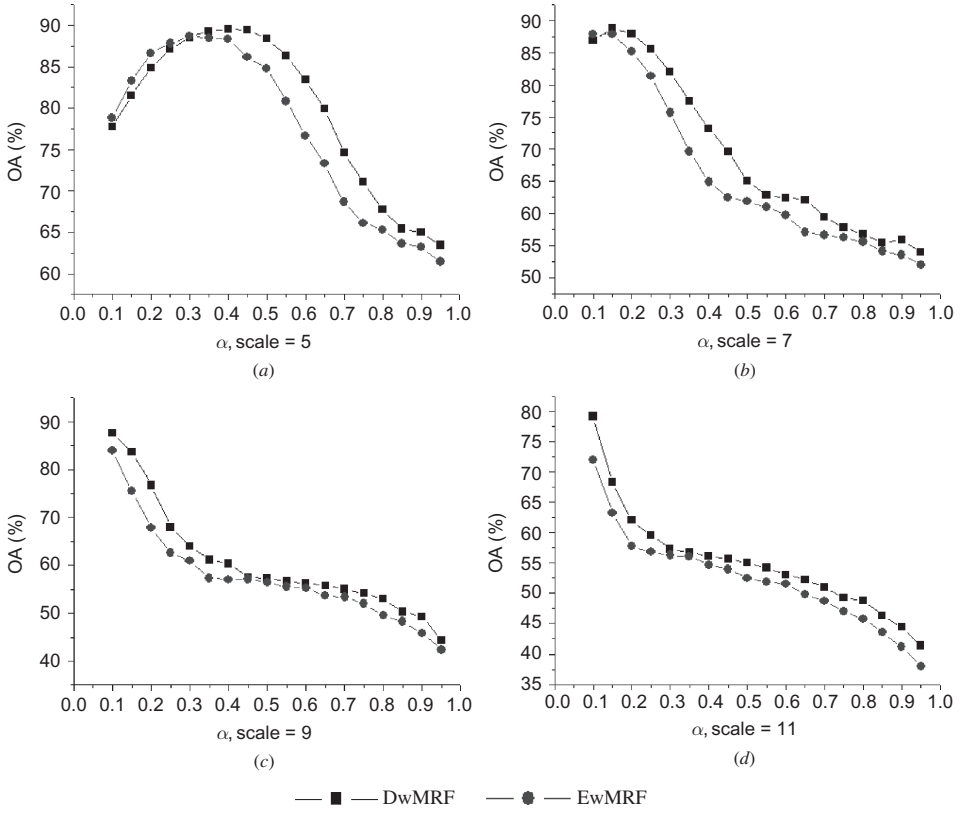


Figure 17. The OA of DwMRF and EwMRF of Beijing data set at scales (*a*) 5, (*b*) 7, (*c*) 9 and (*d*) 11.

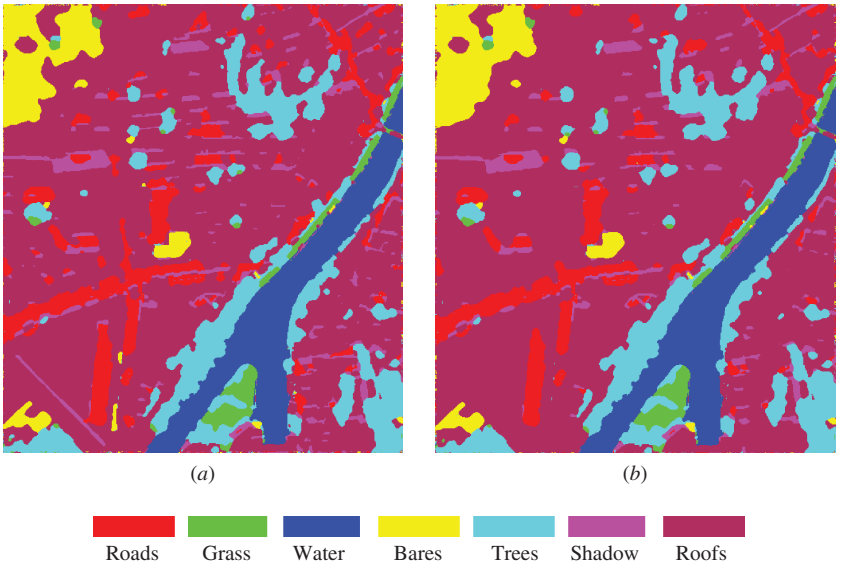


Figure 18. Classification maps of (*a*) DwMRF and (*b*) EwMRF at scale = 5 and $\alpha = 0.6$.

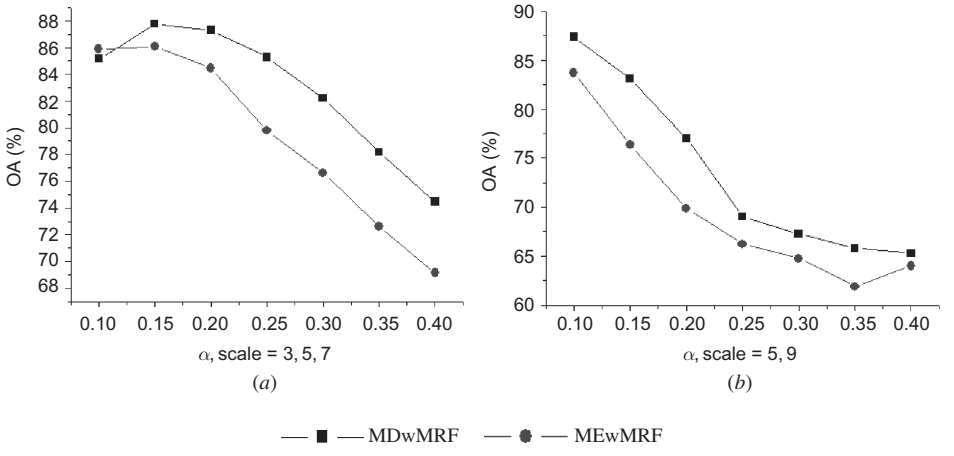


Figure 19. The OA of MDwMRF and MEwMRF at scales (*a*) 3, 5, 7 and (*b*) 5 and 9.

Table 11. The numbers of uncertain pixels of the PICM after they reached the steady state in the DwMRF.

α	Scale = 3	Scale = 5	Scale = 7	Scale = 9	Scale = 11
0.1	481	192	111	28	11
0.2	999	270	61	18	11
0.3	1604	245	58	10	9
0.4	2129	303	54	20	2
0.5	2673	249	39	18	10
0.6	3256	228	38	10	6
0.7	3806	198	0.29	8	4
0.8	4414	161	56	13	7
0.9	5005	192	74	17	8

comparable with the conventional ICM. It was also proved that it is meaningless to continue the iterations after PICM reaches its steady state, because the labels of uncertain pixels will then swing between two classes.

Experiments show that different scales have different optimal values for the parameter α, and in general, as scale increases, the optimal value of α is decreasing. However, the shortcoming is that as the parameter α increases after reaching the best result, the accuracies of DwMRF (also EwMRF) are low, especially when both scale and the parameter α are large, though DwMRF is much better than EwMRF. That is because when both scale and α are very large, spatial information plays a greater role than spectral information, leading to spatial information dominating the classifier, which easily leads to misclassification. Therefore, further work will focus on automatic selection of the scale and parameter α and will attempt to design a more stable classifier that is not sensitive to the parameters, especially the parameter of scale.

Table 12. The OA (%)/Ka of SICM and PICM in the DwMRF method.

	Scale = 3		Scale = 5		Scale = 7		Scale = 9		Scale = 11	
α	SICM	PICM	SICM	PICM	SICM	PICM	SICM	PICM	SICM	PICM
0.1	67.9/0.600	67.8/0.599	77.8/0.716	77.6/0.714	86.9/0.828	86.7/0.826	87.6/0.834	87.4/0.832	79.2/0.717	79.8/0.726
0.2	71.2/0.639	70.6/0.632	84.9/0.803	84.5/0.799	88.1/0.838	87.7/0.835	76.7/0.683	76.7/0.683	62.0/0.475	62.6/0.483
0.3	74.4/0.676	73.6/0.667	88.6/0.849	88.3/0.846	82.0/0.756	82.0/0.757	64.0/0.505	65.1/0.520	57.3/0.402	57.2/0.400
0.4	77.1/0.708	76.3/0.699	89.5/0.860	88.5/0.847	73.2/0.636	74.1/0.648	60.4/0.452	60.3/0.449	56.2/0.390	55.8/0.381
0.5	79.8/0.742	78.4/0.725	88.4/0.844	87.3/0.830	65.0/0.521	67.2/0.552	57.3/0.403	56.2/0.385	55.1/0.373	54.1/0.353
0.6	81.3/0.760	80.0/0.743	83.5/0.777	82.8/0.770	62.4/0.487	64.0/0.505	56.2/0.391	55.2/0.372	53.1/0.341	52.3/0.327
0.7	82.5/0.775	80.7/0.752	74.6/0.655	75.1/0.664	59.5/0.436	60.8/0.457	55.0/0.374	53.9/0.352	51.0/0.309	50.2/0.294
0.8	82.3/0.771	80.4/0.748	67.8/0.563	68.0/0.567	56.8/0.398	55.6/0.382	53.0/0.346	52.1/0.326	48.9/0.275	47.6/0.252
0.9	79.3/0.734	78.0/0.718	65.0/0.526	64.0/0.516	55.9/0.391	55.3/0.384	49.3/0.287	47.7/0.261	44.4/0.203	44.2/0.197

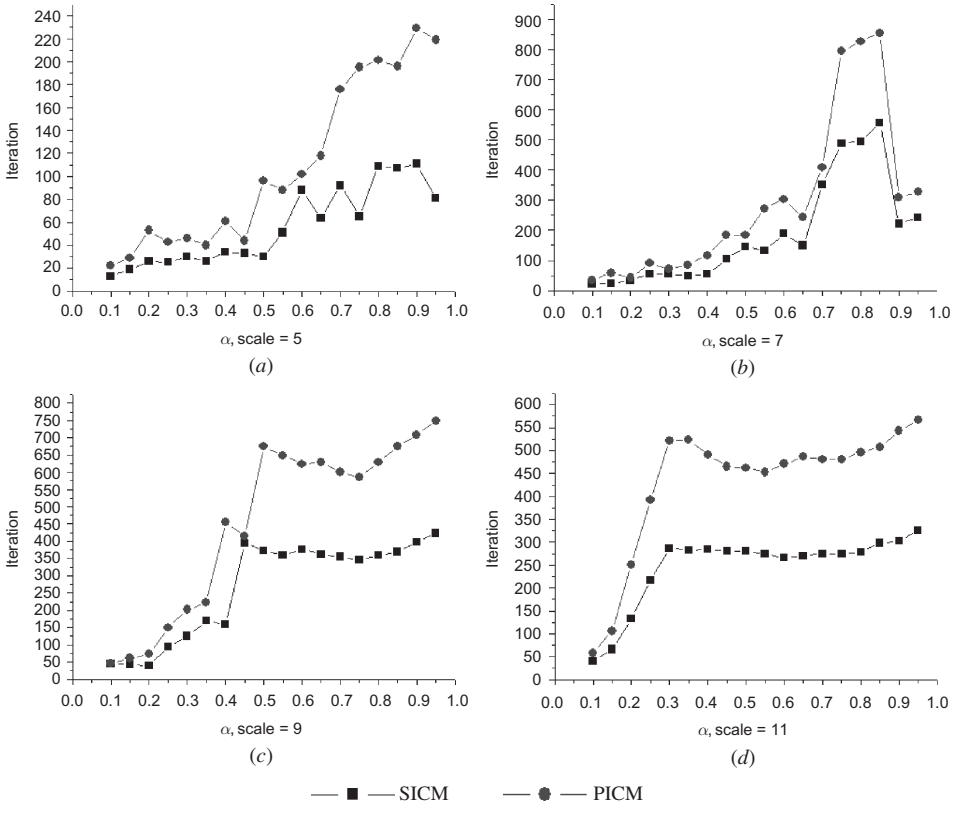


Figure 20. The iterations of SICM and PICM of DwMRF at scales (*a*) 5, (*b*) 7, (*c*) 9 and (*d*) 11.

Acknowledgements

This work was supported by the Major State Basic Research Development Program (973 Program) under Grant no. 2009CB723905, the National Science Foundation of China under Grant no. 40930532, the Fundamental Research Funds for the Central Universities under Grant no. 3101016 and the LIESMARS Special Research Funding.

References

Baatz, M. and Schäpe, A., 2000, Multiresolution segmentation an optimization approach for high quality multiscale image segmentation. In *Angewandte Geographische Informations-Verarbeitung XII*, J. Strobl, T. Blaschke and G. Griesebner (Eds.), pp. 12–23 (Karlsruhe, Germany: Weichmann Verlag).

Gamba, P., Dell'Acqua, F., Lisini, G. and Trianni, G., 2007, Improved VHR urban area mapping exploiting object boundaries. *IEEE Transactions on Geosciences and Remote Sensing*, **45**, pp. 2676–2682.

Geman, S. and Geman, D., 1984, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-6**, pp. 721–741.

Hollander, M. and Wolfe, D.A., 1999, *Nonparametric Statistical Methods*, 2nd ed., pp. 468–470 (New York: John Wiley & Sons).

HUANG, X. and ZHANG, L., 2008, An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery. *IEEE Transactions on Geosciences and Remote Sensing*, **46**, pp. 4173–4185.

HUANG, X., ZHANG, L. and LI, P., 2007, Classification and extraction of spatial features in urban areas using high resolution multispectral imagery. *IEEE Geosciences and Remote Sensing Letters*, **4**, pp. 260–264.

HUANG, X., ZHANG, L. and WANG, L., 2009, Evaluation of morphological texture features for mangrove forest mapping and species discrimination using multispectral IKONOS imagery. *IEEE Geosciences and Remote Sensing Letters*, **6**, pp. 393–397.

JACKSON, Q. and LANDGREBE, D., 2002, Adaptive Bayesian contextual classification based on Markov random fields. *IEEE Transactions on Geosciences and Remote Sensing*, **40**, pp. 2454–2463.

LI, S.Z., 1995, On discontinuity-adaptive smoothness priors in computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**, pp. 576–586.

LI, S.Z., 2001, *Markov Random Field Modeling in Image Analysis*, 2nd ed. (New York: Springer-Verlag).

LIU, D., KELLY, M. and PENG, G., 2005, A spatial–temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery. *Remote Sensing of Environment*, **101**, pp. 167–180.

MCNEMAR, Q., 1947, Note of the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, **12**, pp. 153–157.

TRIANNI, G. and GAMBA, P., 2005, A novel MRF model for multisource data fusion in urban areas. In *Proceedings of URSI General Assembly*, 23–29 October 2005, New Delhi, India (CD-ROM).

TRIANNI, G. and GAMBA, P., 2007, Boundary-adaptive MRF classification of optical very high resolution images. *IEEE International Geosciences and Remote Sensing Symposium*, pp. 1493–1496.

TSO, B. and OLSEN, R.C., 2005, A contextual classification scheme based on MRF model with improved parameter estimation and multiscale fuzzy line process. *Remote Sensing of Environment*, **97**, pp. 127–136.