Available online at www.sciencedirect.com

SCIENCE DIRECT

Remote Sensing of Environment 97 (2005) 127 – 136

Remote Sensing of Environment

www.elsevier.com/locate/rse

A contextual classification scheme based on MRF model with improved parameter estimation and multiscale fuzzy line process

Brandt Tso[a,*], Richard C. Olsen[b]

[a]National Defense Management College, Taiwan
[b]Naval Postgraduate School, USA

Received 21 January 2005; received in revised form 19 April 2005; accepted 24 April 2005

Abstract

A Markov random field (MRF) based method using both contextual information and multiscale fuzzy line process for classifying remotely sensed imagery is detailed in this paper. The study area known as Elkhorn Slough is an important natural reserve park located in the central California coast, USA. Satellite imagery such as IKONOS panchromatic and multispectral data provides a convenient way for supporting the monitoring process around this area. Within the proposed classification mechanism, the panchromatic image, benefited from its high resolution, mainly serves for extracting multiscale line features by means of wavelet transform techniques. The resulting multiscale line features are merged through a fuzzy fusion process and then incorporated into the MRF model accompanied with multispectral imagery to perform contextual classification so as to restrict the over-smooth classification patterns and reduce the bias commonly contributed by those boundary pixels. The MRF model parameter is estimated based on the probability histogram analysis to those boundary pixels, and the algorithm called maximum a posterior margin (MPM) is applied to search the solution. The results show that the proposed method, based on the MRF model with the multiscale fuzzy line process, successfully generates the patch-wise classification patterns, and simultaneously improved the accuracy and visual interpretation.

Keywords: MRF; Fuzzy; MPM; Line process; Contextual

1. Introduction

Image classification is an important part in many remote sensing applications. Beginning with the Earth Resources Satellite (LANDSAT-1), spectral imagery has been the primary tool for scene classification. With the advent of higher spatial resolution systems (IKONOS, Quickbird, SPOT-1), other techniques begin to offer promise for the analysis of satellite derived imagery. In recent years, the progress of computer capabilities makes spatial feature processing techniques practical to implement in pursuit of improvement in classification accuracy (Olsen et al., 2002). A trend for incorporating spatial data into the classification pool is certainly triggered by the concept that, for the same land-use/land-cover types, they not only reveal similarity in spectral reflectance, but should also contain certain relation in spatial domain. Contextual information is one kind of such spatial relationship and has drawn our particular interest for remotely sensed imagery interpretation shown in this study. Contextual information, or so-called context for simplicity, may be defined as how the probability of presence of one object (or objects) is affected by its (their) neighbors. Generally, in remote sensing land-use/land-cover classification, a pixel labeled as forest is likely to be surrounded by the same class of pixels unless that pixel is located in boundary area. If such contextual information can be well modeled, the classification accuracy may be improved significantly (Khedam & Belhadj-Aissa, 2003; Mather, 1999).

Incorporating contextual information into classification process can be done in different ways. One simple method of adopting context is to use majority voting within a prescribed window. In such a method, the central pixel will

* Corresponding author. Tel.: +886 2 22222137x8441; fax: +886 2 22250488.
E-mail address: brandttso@yahoo.com.tw (B. Tso).

0034-4257/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2005.04.021

be changed to the class that occurs most frequently in the window. This is a common post-processing technique after a pixel-based classifier has been implemented. There are more elegant ways of modeling such contextual behavior. A class of contextual model known as Markovian random fields (MRF) can be useful for modeling context in a more precise way (cf. Fan & Xia, 2001; Geman & Geman, 1984). The MRF is used to construct a priori probability in Bayesian sense so as to accomplish the Maximum a Posteriori (MAP) estimate during the classification process. Such a MAP solution often provides more satisfactory results than Maximum Likelihood (ML) classifier (Li, 1995).

Even though MRF is commonly robust in its classification performance, errors frequently occur upon boundary (edge) areas. If those boundary pixels are not well defined and controlled during the classification process, the resulting classified image will eventually reveal an over-smooth outcome (i.e. loss of significant details and generating too large patches, cf. Wang & Wang, 2004). To avoid such errors, one has to accurately identify those edge pixels and reduce their contribution to the classification pool. Once those edge pixels have been accurately identified, the success of pattern recognition can be significantly enhanced (Wei & Gertner, 2003).

To identify meaningful edges (i.e. real boundaries) should draw our further concern since edges hold resolution-dependant nature. Specifically, the edges derived from different resolutions (or so-called scales) may show different significant levels (i.e. the likelihood to be a real edge) depending on the application being undertaken and scene properties (Mather, 1999). If one could develop an objective way to gather those edge information from multiscales while simultaneously take those edges' various significances into account, the edges being incorporated into the classification process will be more accurate and naturally higher probability of success in classification can be achieved. There are many edge detection techniques available (cf. Bian, 2003; Canny, 1986; Mallat & Zhong, 1992; Rydberg & Borgefors, 2001). Of particular interest to us is the wavelet-based edge detection method developed by Mallat and Zhong (1992) due to its robustness in multiscale edges extraction. In the later experiment shown, once the edges have been quantified from different scales, edge fusion is then performed to generate combined multiscale fuzzy edge patterns for inputting into the MRF classification model.

A model is unable to show its full effectiveness if the relevant parameters are not accurately defined. In the case of MRF model, the most widely known technique for estimating the model parameters is the coding method (Derin & Elliott, 1987; Elliott et al., 1984) and least square fit method (Besag, 1974). The success of these approaches relies on the complete understanding of image neighborhood configurations (Ibáñez & Simó, 2003). Unfortunately, in practical sense, the neighborhood configurations are difficult to acquire. The parameter within MRF model is therefore somehow determined in trivial and thus considerably restricts the model capabilities (Tso & Mather, 1999, 2001). A more efficient method is clearly required to cope with such parameter estimation issue. We propose an approach based on the probability histogram analysis to the edge pixels to successfully perform parameter estimates which will be introduced below.

2. Theoretical background

2.1. Multiscale edge detection using wavelet

Mallat and Zhong (1992) show that, for edge detection, the edge occurs at the local maxima of the wavelet transform modulus $|W_{\psi_{a,b}}f(t)|$ for a signal, where $W_{\psi_{a,b}}f(t)$ denotes wavelet transform of a function $f(t)$, ψ called the *mother wavelet*, and a, b as the dilation step and translation step of translation and dilation processes, respectively. The detection of local modulus maxima is done via an adaptive sampling that finds the sharp variation points. By varying the parameter a, one can obtain the distribution of edges across multiple scales. For most purposes, as both the computational time and practical usage are concerned, the dyadic sequence (2^r) rather than continuous scale parameter is chosen. When implemented for an image at scale 2^r, the edge detection in two-dimensional case, two wavelets used for the horizontal (H) and vertical (V) direction transforms, respectively, are required, i.e.

$$W^H_{\psi_{2^r}}f(x,y) = \frac{1}{2^{2r}} f\psi^H\left(\frac{x}{2^r},\frac{y}{2^r}\right) \tag{1}$$

$$W^V_{\psi_{2^r}}f(x,y) = \frac{1}{2^{2r}} f\psi^V\left(\frac{x}{2^r},\frac{y}{2^r}\right) \tag{2}$$

At each scale, the modulus of the gradients derived by wavelet transform is given by

$$M_{2^r}f(x,y) = \sqrt{|W^H_{\psi_{2^r}}f(x,y)| + |W^V_{\psi_{2^r}}f(x,y)|}, \tag{3}$$

and the associated phase is shown as

$$A_{2^r}f(x,y) = \tan^{-1}\left(W^V_{\psi_{2^r}}f(x,y)/W^H_{\psi_{2^r}}f(x,y)\right) \tag{4}$$

The edge points are then identified as the pixels with locally modulus maxima in one-dimension neighboring pixel along the direction $A_{2^r}f(x,y)$ (Mallat & Zhong, 1992).

2.2. Markov random field models

Let $x=\{x_1,x_2,\ldots,x_n\}$ denote an image of n pixels, and suppose $c=\{c_1,c_2,\ldots,c_n\}$ denoting the understanding of x, i.e., c_i is the class to which pixel i belongs. Then x is called a random field. To construct the relationship between x and c, one may use the Bayesian paradigm, which holds the following conditional probability relation

$$p(c|x)p(x) = p(x|c)p(c). \tag{5}$$

Normally, to acquire c, one may adopt a maximum a posteriori (MAP) solution as

$$\hat{c} = \arg\max_{c} [p(x|c)p(c)]. \quad (6)$$

In general, $p(x|c)$ is modeled in terms of the Gaussian distribution, while $p(c)$, the prior probability about the understanding of image c, can be modeled based on the Markov Random Field (MRF) (cf. Besag, 1986; Li, 1995). The Markov assumption states that the conditional distribution of a pixel given all the other pixels in the image is identical to the conditional distribution of the pixel given the *neighboring* pixels. Accordingly, based on MRF, following equation sustains

$$p(c) = p\left(c_i|c_{S-\{i\}}\right) = p(c_i|c_{N_i}) \quad (7)$$

where c_{N_i} denotes the neighboring pixels around the pixel i, and S-$\{i\}$ is the set difference denoting all the pixels in the image except pixel i. The neighborhood configuration may contain second or higher orders of neighboring pixel arrangements. In the case of pair-wise second order neighborhood (i.e. the nearest horizontal, vertical, and triangular neighboring pixels) defined in MRF, the prior probability is modeled by

$$p(c_i|c_{N_i}) = \exp\left(\sum_{\{ij\in N\}} \beta \cdot I\left(c_i = c_j\right)\right) \Big/ Z \quad (8)$$

Here, β is the Gibbs distribution parameter expressing the strength of how an occurrence of class c_i for pixel i is affected by its neighborhood c_{N_i}. $I(A)$ is an indicator function for an event A to occur. Z is a normalization constant making $p(\cdot)$ a proper distribution (Besag, 1986). Note that Eq. (8) is mainly to achieve *smoothness everywhere* within an image. However, for real data, the scene is more likely to be analyzable in the sense of *piecewise continuous or smoothness*. In other words, there are always boundaries, and the boundary pixels are likely to be misclassified due to their uncertain nature. To cope with this issue, Eq. (8) is refined to include a so-called *line process* and in terms of taking logarithm domain as

$$\log p(c_i|c_{N_i}) = \sum_{\{i,j\}\in N} \beta I\left(c_i = c_j\right)(1 - l_i) \quad (9)$$

where l_i=1, if pixel i is recognized as an edge, and 0 otherwise. The above equation indicates that the contextual effect will turn off as an edge is encountered. In other words, the smoothing is not allowed to cross the boundaries. However, to make Eq. (9) useful, one has to determine the edges with higher confidence. This issue will be treated later.

To solve Eq. (6), numerous algorithms can be adopted, such as Iteration Condition Mode (ICM) (Besag, 1986), Simulated Annealing (SA) (Geman & Geman, 1984), and Maximum a Posterior Margin (MPM) (Marroquin et al., 1987). SA is a global optimization algorithm, which, however, suffers considerable computational burden. ICM can obtain the solution within very short term, but it normally only achieves to local minimum (or maximum). This study therefore adopts MPM as alternative for solving the Eq. (6) due to MPM's tractable computational burden and higher quality results. The practical application of the MPM algorithm relies on an important assumption that a Markov chain exists over m^n, where n denotes the number of pixels within an image, m is the total number of classes. Once the number of state transitions within a Markov chain has reached a steady state, the marginal posterior probability can be approximated by counting the number of times that each class occurs at each pixel in a series of configurations. The approximation made by MPM method adopts the essence of Markov chain Monte Carlo techniques, and is expressed by

$$p(c_r|x_i) = \frac{1}{\eta - v} \sum_{v+1}^{\eta} I(c_r) \quad (10)$$

where $p(c_r|x_i)$ represents the probability of a class c_r given the observation x_i at pixel i, $1/(\eta - v)$ is the normalization term, while v denotes the minimal iterations that an MRF required to reach a stable state and η denotes the maximal iteration number that we wish a MRF to perform transitions, and $I(c_r)$ is the Boolean indication function pointing c_r to occur. In practice, if one sets parameter η to more than 200 and v to more than 20, respectively, stable results are likely to be reached. (comparable to SA, cf. Tso & Mather, 2001).

3. Methodology

3.1. Test data and framework

The study area known as Elkhorn Slough is located in the central California coast about 160 km south of San Francisco, California, USA (Silberstein & Campbell, 1989). The study imagery for the scene was captured by the IKONOS satellite on October 23, 2002. The Elkhorn Slough is an important natural reserve in a largely agricultural/urban area. Satellite imagery can provide a convenient means for monitoring the evolution of the area. A test area (within the rectangle block as shown in Fig. 1) was extracted from the image to proceed the classification methodology analysis. For IKONOS multispectral imagery, the test area is 1024 by 1024 pixels, while the corresponding area in the 1-m resolution panchromatic image is 4096 by 4096 pixels in size. Eight information classes were chosen as shown in Table 1. The ground truth used to select the training set and later to evaluate the classification accuracy is based on the ground truth map provided by the Elkhorn Slough Foundation (ESF). There are totally 71,804 ground truth pixels available. Of those ground truth pixels, 15,451 pixels are assigned to the training set to train the ML classifier, and 56,353 pixels are used for accuracy evaluation.

The framework of the experiment is shown in Fig. 2. Wavelet transform technique (Mallat & Zhong, 1992) is applied to detect multiscale edges from panchromatic image.

Fig. 1. IKONOS panchromatic imagery of the study area around Elkhorn Slough C.A., USA.

Those multiscale edges are furthered through a fusion process to reveal their different significance levels. The multispectral imagery are input to the maximum likelihood classifier to obtain the first stage classification probabilities, and later refined by contextual information based on MRF fuzzy line process model with both well estimated MRF model parameter and fused line features.

3.2. Line feature extraction and multiscale edge fusion

According to the wavelet transform shown in Eqs. (1) (2) (3) and (4), three scale images with parameter r set as 0, 1, and 2, respectively, are used to derive edges. The resulting edge patterns are shown in Fig. 3(a) and (b). It should be recognized that, rather than arbitrarily determinate a pixel as

Table 1
Selected information classes for classification experiments

No.	Class name
1	Cultivated land
2	Dry grass land
3	Water
4	Wetland
5	Trees
6	Crop vegetation (strawberries, broccoli, and lettuce)
7	Naked range fields
8	Man-made features (buildings, roads)

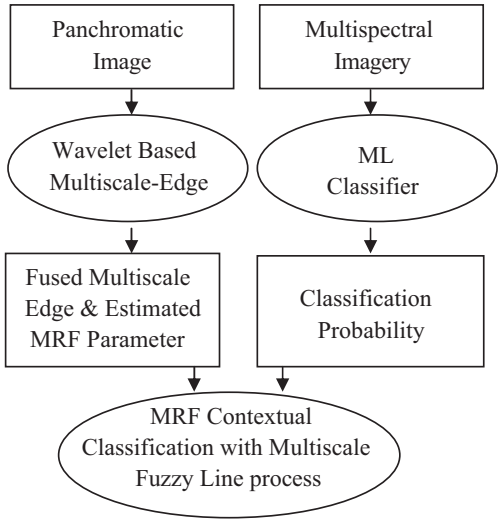

Fig. 2. Experimental framework.

an edge or not, edges do exhibit fuzziness in nature. Specifically, to an edge's occurrence, its significance (i.e. the likelihood to be a real edge) should be highly correlated to two factors, namely, how the variations between pixels are treated and image scales. In the lower scale image (such as scale $r=0$), even minor variations between pixels can be recorded. As the scale gets larger (e.g. $r=2$), only the sharper variations between pixels is detected. One may thus recognize that an edge simultaneously detected in both lower and higher scales should be more significant (i.e. more likely to be a real edge) than the edge only occurs in lower scale. Following such derivation, an edge fusion process is then adopted to address such concerns so as to reflect the edges' different significance levels. In such a way that the edges being incorporated into the classification process will be more accurate and naturally higher probability of success in classification can be pursued. The edge fusion process is based on the following proposed rules:

1. A pixel in the fused image is assigned with a value of '2', if the pixel is detected as an edge in the scale 2 and accompanied by at least scale '0' or '1';
2. A pixel in the fused image will be assigned with a value of '1', if the pixel is detected as an edge both in scale '0' and '1';
3. A pixel is assigned with a value of '0' if the pixel is detected as an edge only in scale '0'; and
4. A pixel is assigned to the value of '255' (in 8-bit image storing format) if the pixel is not an edge pixel in any scales.

The ways to the value assignment shown above is to provide as a basis to quantify the edge's significance. These values are then dedicated into contextual classification process through a fuzzy mechanism as will be described later.

The rules described in Rules 1 and 2 warrant further discussion. We consider the edges in the higher scales are

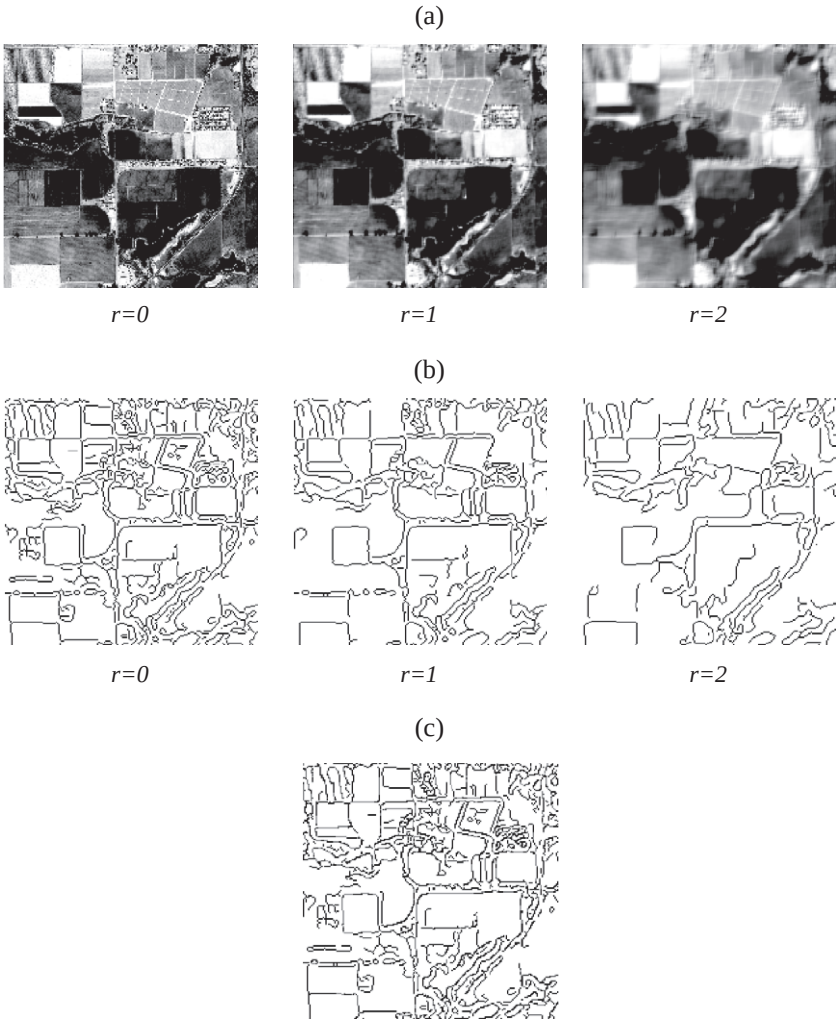

Fig. 3. (a) Parts of original images with scale $r=0$, 1, and 2, respectively, (b) edge detection results corresponding to each scale, and (c) fused edge image. Please note that different significance of the edges are represented in terms of different gray levels.

significant only if such edges (for example, scale 2) are reinforced from the lower scales (for example, scale 1 or 0). In this way, the misidentification of significant edges can be eliminated. The larger scale images could encounter potential distortion contributed by the wavelet transform process and can cause edges coexisting as neighbors within the fused image. An edge thinning process is thus applied to the fused edge image for correction purpose. Some other techniques (Steger, 1998) may also be used for improving such distortion due to scale. The choice of the edge location during the thinning process is according to where the lower value occurs within the pairing of edge pixels. It is recognized that, in the case of pair-wise edge coexistence, the edge detected by lower scale images should be more accurate in spatial location. Upon the thinning process, some isolated edge pixels are also eliminated. This, in turn, reduces the misidentification risk to an edge made by spurious noise. The fused edge image is finally shown in Fig. 3(c). Please note that the edges derived from three scales (i.e. r from 0 to 2) shown here is just for demonstration purposes; one may naturally extend the method to develop edges from more scales to fulfill particular needs.

3.3. Parameter estimation

A model is not complete if both the model form and the relating parameters are not well defined. The MRF model previously described provides a theoretically robust basis for modeling the context in spatial domain. However, how to assign the suitable parameters into the model remains as a serious problem. Recall that Eq. (9) showing two parameters, namely l_i and β, need to be determined. The l_i is normally Boolean which only concerns whether a pixel i is an edge ($l_i=1$) or not ($l_i=0$) and totally disregards the different significance levels relating to the edge. However, according to the nature of multiscale edges as described above, the parameter l_i adopted here is thus refined through a fuzzified process (Bezdek, 1999) to facilitate the quantization to the edges' significance in a

more objective way. The fuzzy membership function can quantify the level of significance about an object or event belonging to a fuzzy subset. In our case, in order to reflect different significance among multiscale edges, a sigmoid fuzzy membership function is chosen to this end and is expressed as

$$f(\alpha) = \frac{1}{1+\exp(-\alpha)} = l_i \tag{11}$$

where α denotes the pixel values shown in the fused edge image (according to the rules shown in previous section). In other words, when an edge pixel i with value of 0 occurs in the fused edge image, the pixel is assigned to the significance value of 0.5. The higher the edge values are, the higher the l_i value will be. This also indicates that the edge is in higher significance (i.e. more likely to be a real edge).

The determination of parameter β is more bothersome. Here, we assume the MRF is isotropic (i.e. orientation insensitive) and homogeneous (i.e. location insensitive). If underlying MRF is anisotropic or inhomogeneous, then more β values need to be estimated. The most popular techniques for β parameter estimation such as the ML estimate (Derin & Elliott, 1987; Elliott et al., 1984) or least square estimate (Besag, 1974), require a priori realization to the image. Such a realization denotes the complete understanding to all the possible neighborhood configurations within an image. In the case of supervised classification to real world scene, the training data is hard to well reflect such contextual information. This makes those estimation algorithms hard to implement and of less utility.

In order to overcome the difficulty in β determination, we develop an approach which adopts a more stochastic perspective rather than one which focuses on contextual neighborhood. The core of the method here is to determine a suitable value for β so as to adequately preserve the edge pixels to the certain extent during the contextual classification. For all the edge pixels, the corresponding logarithm probability differences between the *winning* class (rating 1st) and other rating classes (from 2nd to 8th) are calculated and the resulting histogram plots are shown in Fig. 4. The horizontal axis represents the probability difference values d (in dB unit) between the winning class and other rating classes, while the vertical axis indicates the counts for the appearances of each difference value. The logarithm differences d in Fig. 4 can be regarded as the *aid* (equivalent to $\sum\beta I(c_i=c_j)$ in Eq. (9)) required by those non-winning classes (rating from 2nd to 8th) to become the winning classes appearing upon those edge pixels. If such aid is suitably determined, the number for the edge pixels being smoothed can be well controlled according to our intent. In the following, Eq.

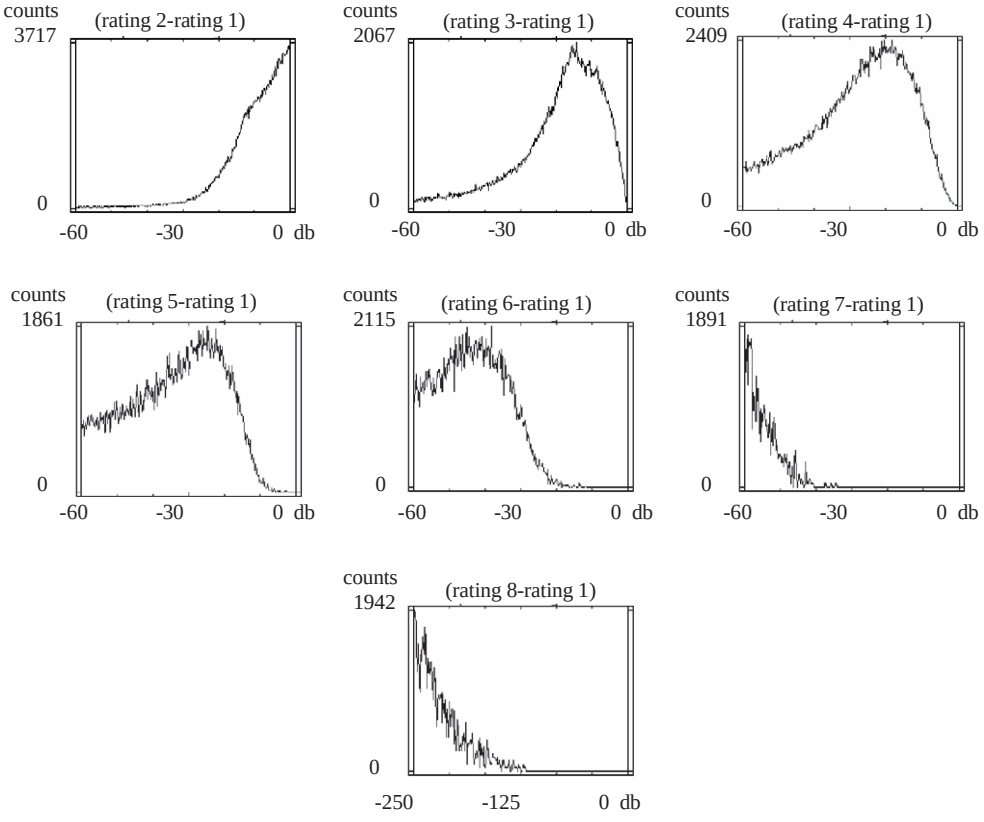

Fig. 4. The histogram of logarithm probability difference between the *winning* classes and different rating (from the 2nd) classes for the edge pixels.

(12) expresses the relationship between the aid value and the expected smoothed edge pixels.

$$\sum_{k=2}^{m}\frac{k-1}{m}\left(\Theta_k^d-\Theta_{k+1}^d\right)\cong\left[\text{Total edge pixels}-\sum_{\alpha=0}^{2}f(\alpha)\Omega_\alpha\right]. \quad (12)$$

where m is the total number of information classes, $f(\alpha)$ is defined in Eq. (11), Ω_α expresses the total number of edge pixels that hold value α, $\Sigma f(\alpha)\Omega_\alpha$ then denotes the number of the *valid* edge pixels, and Θ_k^d denotes the number of edge pixels whose logarithm probability differences between the winning class and the kth rating class (as shown in Fig. 4) fall under the *aid* value d, respectively. The right hand side on Eq. (12) then denotes that the tolerated number of the edge pixels can be smoothed. We also define $\Theta_{m+1}^d=0$. The factor $(\Theta_k^d-\Theta_{k+1}^d)$ specifies, under the *aid* value d, the number of pixels in which the winning class can be overturned by either the 2nd rating or other rating classes ($(k-1)$ classes in total). If one assumes that the probability of the neighborhood occurrence for all the classes is the same, the factor $(k-1)/m$ then denotes the probability that the winning class is overturned by one of the $(k-1)$ classes. For instance, in the case of $k=4$ and $m=8$, the factor $(\Theta_4^d-\Theta_5^d)$ then denotes under the *aid* value d the number of pixels in which the winning class (the 1st rating) can be overturned by either the 2nd, 3rd, or 4th rating class. Then 3/8 will be the probability for the winning class being overturned by either the 2nd, 3rd, or 4th rating class. In this study, the total number of edge pixels in the fused image with α equal to 0, 1, and 2 are 40,241, 26,550, and 17,450, respectively, and following the right side condition set in Eq. (12), the expected number of edge pixels being overturned are around 30,000 pixels. As shown in Fig. 5, it is found that with $d=-21$ the condition shown in Eq. (12) is fulfilled; this value is then selected. Once the *aid* value is determined, one can turn back to estimate the parameter β.

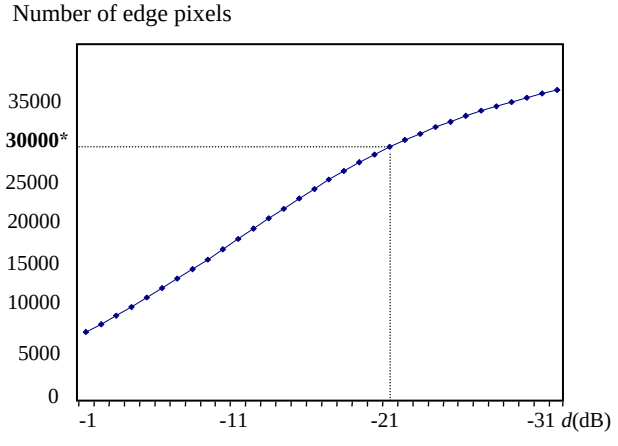

Fig. 5. The determination of d by Eq. (11).

Table 2
Classification confusion matrices for (a) ML classifier, (b) MRF model with Boolean line process, (c) MRF model with fuzzy line process, and (d) MRF model without line process

Class No.	1	2	3	4	5	6	7	8	Total
(a)									
1	9702	536	22	55	212	33	26	369	10,955
2	3489	3284	0	4	47	153	0	24	7001
3	0	0	4530	210	0	0	0	5	4745
4	2	0	435	7344	1044	228	0	89	9142
5	0	0	0	275	2156	4900	0	58	7389
6	121	128	14	147	321	4626	0	343	5700
7	1	0	0	1	0	0	1541	305	1848
8	2237	308	551	2053	363	54	54	3953	9573
Total	15,552	4256	5552	10,089	4143	9994	1621	5146	56,353
Overall accuracy=(37,136/56,353) 65.89%									
Kappa coefficient=0.60									
(b)									
1	10,011	1927	371	624	68	25	195	674	13,895
2	3740	1969	0	161	5	153	0	0	6028
3	0	0	4368	95	0	0	0	0	4463
4	153	1	477	7291	1358	306	0	131	9717
5	0	0	239	110	2370	10	0	10	2739
6	1298	352	26	1462	51	9439	0	7	12,635
7	25	0	0	0	0	0	1426	11	1462
8	325	7	71	346	291	61	0	4313	5414
Total	15,552	4256	5552	10,089	4143	9994	1621	5146	56,353
Overall accuracy=(41,187/56,353) 73.08%									
Kappa coefficient=0.67									
(c)									
1	11,785	1499	141	439	204	329	474	538	15,409
2	1488	2711	0	0	0	42	0	49	4290
3	0	0	4581	0	24	0	0	0	4605
4	434	22	419	9348	1315	92	0	69	11,711
5	0	0	0	1	1327	2	0	12	1330
6	1320	3	367	131	591	9012	0	102	11,526
7	5	0	0	0	0	0	1147	0	1147
8	525	21	44	170	682	517	0	4376	6335
Total	15,552	4256	5552	10,089	4143	9994	1621	5146	56,353
Overall accuracy=(43,077/56,353) 78.56%									
Kappa coefficient=0.74									
(d)									
1	9375	1570	1125	1298	298	17	167	482	14,323
2	4235	2281	0	72	16	188	0	0	6792
3	0	0	4047	4	1	0	0	0	4052
4	2	3	237	7466	1268	822	0	66	9864
5	0	0	17	13	2253	17	0	1	2301
6	1685	99	112	938	85	8794	0	122	11,535
7	45	0	0	0	0	0	1453	40	1538
8	210	303	14	298	231	156	1	4435	5647
Total	15,552	4256	5552	10,089	4143	9994	1621	5146	56,353
Overall accuracy=(40,104/56,353) 71.16%									
Kappa coefficient=0.65									

Normally, within a 3 by 3 window, there are four kinds of edges, namely, ‘line boundary’, ‘corner’, ‘triple junction’, and ‘quadruple junction’. By checking the fused edge image, it is found that ‘line boundary’ and ‘corner’ hold enormous majority. Under such circumstances, within a 3 by 3 window, the number of information class equivalent to the central pixels can range from 1 to 6 (note that the central

pixel and other two pixels forming 'line boundary' or 'corner' are excluded). Under the assumption that, within a window, the probabilities for each number (from 1 to 6) of classes' occurrences are the same, according to Eq. (9), the calculation for β can be obtained by

$$21 = [\beta(1+2+3+4+5+6)]/6 \tag{13}$$

Eventually, we sort out $\beta = 6$, which is thus used in the classification experiments. The classification method called Maximer of the Posterior Marginals (MPM) is adopted to find the solution (Marroquin et al., 1987). It is made clear that the proposed method, in terms of analyzing the probability histograms to each edge pixel, can well determine how much smoothing strength (i.e. in terms of deriving the value β) is required by the pixel so as to achieve the controllable contextual classification results under user's expectation. Therefore, the method is flexible and practical and should overcome the shortage of currently existing parameter estimation methods and overwhelm the commonly trivially selecting β value applied to MRF based contextual classification.

4. Results and discussions

Preliminary result of classification using IKONOS multispectral imagery alone with ML classifier is modest. The total classification accuracy of 65.89% with kappa statistic (Congalton, 1991) of 0.6 is achieved. The corresponding classification confusion matrix is shown in Table 2(a). It can be seen that confusion errors between class pairs are quite obvious. Particularly, it is found that classes 1, 4, 5, and 6 suffer serious commission error, while omission errors occur upon classes 2, 5 and 8, respectively. The result of ML classification thus poses a considerable room for further improvements. Such a classification performance yet preserves the potential difficulty in interpreting the classified image in a meaningful way because the different class pixels are still mixing and resulting in a noisy image view as shown in Fig. 6(a). Further refinement is clearly required.

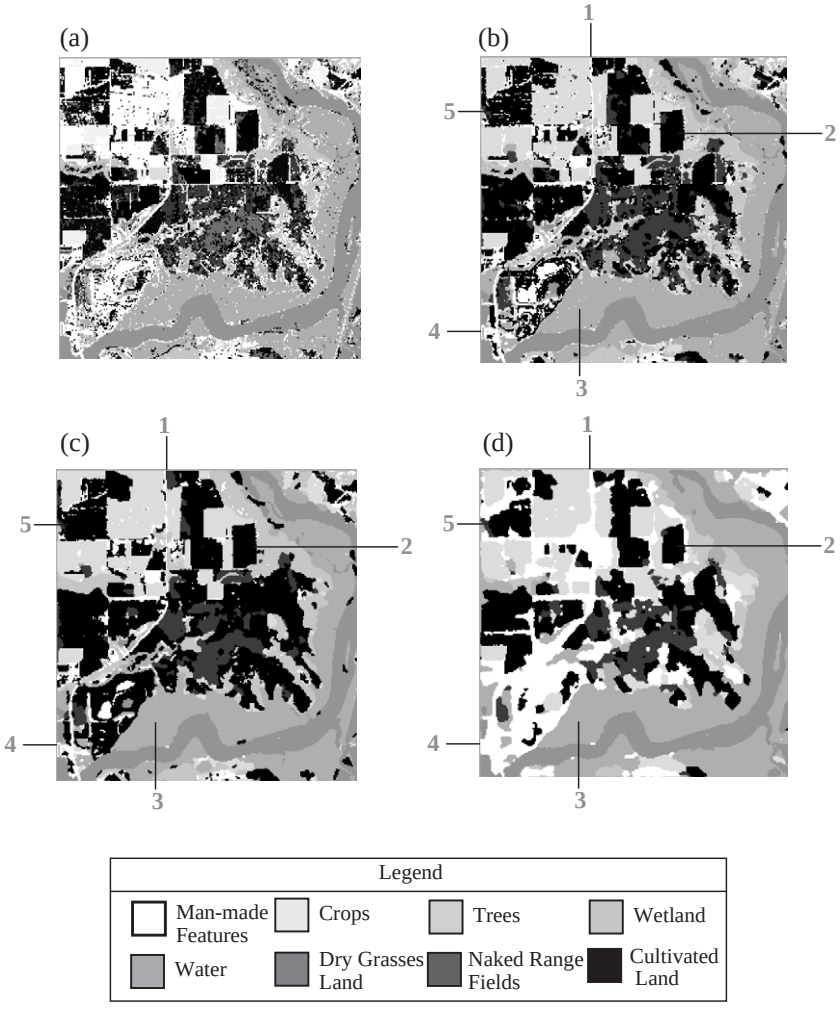

Fig. 6. Classification results achieved by (a) spectral data, (b) MPM with Boolean line process, (c) refined MPM algorithm, and (d) MPM without line process.

The MPM algorithm with carefully set model parameters as described previously is used for finding the solution in MRF contextual model. The result of classification confusion matrix generated by traditional Boolean line process is shown in Table 2(b). The overall accuracy of 73.08% (kappa 0.67) is achieved. The commission error for class 6 and omission error for class 5 and class 8 have been effectively reduced. Referring to the classified patterns shown in Fig. 6(b), it can be found that the linear features (marked as number 1, the roadway shown in white line across the image from top to bottom), boundaries (number 4, which is the boating area, ⊓-like shape) are well preserved, while for cultivated land (marked as numbers 2 and 5) and wetland area (number 3), the patterns are still disturbed by certain levels of noise.

When multiscale fuzzy line process is adopted, an accuracy of 78.56% (kappa 0.74) in overall classification is achieved. The commission error for classes 1 and 6 are successfully improved, while for both classes 5 and 8, the accuracies are also enhanced. Again, by referring to the classified pattern as shown in Fig. 6(c), the wetland (marked as number 3) and cultivated area (numbers 2 and 5) all are more patch-like and cleaner in comparison with Boolean line process, while the roadway (number 1) and boundary (number 4) are still well preserved. Such result demonstrates that the proposed fuzzy line process outperform the traditional Boolean line process mechanism. The efforts made for fuzzy edge fusion and parameter estimate are considered worthy.

The MRF model without inclusion of line process (i.e. equal smoothing to each pixel regardless of whether the pixels are edge or not) achieves an overall accuracy of 71.16% (kappa 0.65). The classification image is displayed in Fig. 6(d). In comparison with Fig. 6(b) and (c), Fig. 6(d) reveals an *over-smooth* result. It can be seen that the roadway (marked number 1) are no longer recognizable. The boundaries around the cultivated land (marked as number 2 and 5) and boating area (number 4) are also smeared. Such a result clearly indicates that the smoothing is beyond control. It, on the other hand, may demonstrate that the inclusion of line process is somehow necessary as the preservation for the detailed features and higher classification accuracy achievement are concerned.

Above all, it can be made clear that the proposed multiscale fuzzy line processing with carefully estimated MRF parameter outperform the MRF with Boolean line process (around 5% increase) and MRF without line process (around 7% enhancement). In comparison with traditional ML classification using spectral data alone, the demonstrated method achieves around 13% accuracy enhancement. The differences in accuracy between those classification results are tested by *Tau* (Ma & Redmond, 1995) to confirm the significance in classification accuracy improvements. We may conclude that the inclusion of contextual information can considerably improve the remotely sensed imagery classification performance and visual interpretation if the model is well defined and the relating parameter is carefully chosen.

5. Concluding remarks

In this study, we show the improvement in classification accuracy by adding contextual and edge information into the classification pool. The results reveal that, compared to the use of multispectral data alone, higher levels of accuracy can be obtained. Results reveal that, with the involvement of contextual information, the enhancements in both the classification accuracy and visual interpretation can be simultaneously achieved.

A method is proposed to estimate the MRF smooth weighting parameter β from the stochastic perspective which is based on the class probabilities corresponding to the edge pixels. The key point to the success of such a method is that, as far as the edge pixels are well managed, the contextual effect can then be controlled in a reasonable way, and naturally the higher accuracy can be expected. In this study, the approach for estimating one MRF-based parameter is presented. The study can be naturally extended to perform other more complicated MRF-based neighborhood parameter estimates (for example, in the anisotropic case or higher order neighborhood system), which may draw our attention in future study.

Acknowledgements

The authors are grateful to the reviewers' and editor's valuable comments which improved the quality of this paper. This study is supported by Remote Sensing Lab, Physics Department, Naval Postgraduate School, Monterey CA.

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society*, *36*, 192–236.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B*, *48*(3), 259–302.

Bezdek, J. C. (1999). *Pattern Recognition with Fuzzy Objective Algorithms*. New York: Plenum Press.

Bian, L. (2003). Retrieving urban objects using a wavelet transform approach. *Photogrammetric Engineering and Remote Sensing*, *69*, 133–141.

Canny, J. A. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *8*(6), 679–698.

Congalton, R. (1991). A review of assessing the classifications of remotely sensed data. *Remote Sensing of Environment*, *37*, 35–46.

Derin, H., & Elliott, H. (1987). Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *9*(1), 39–55.

Elliott, H., Derin, H., Cristi, R., & Geman, D. (1984). Application of the Gibbs distribution to image segmentation. *Proceedings of the International Conference on Acoustic, Speech and Signal Proceeding* (pp. 32.5.1–32.5.4). San Diego: IEEE.

Fan, G., & Xia, X. G. (2001). A joint multicontext and mutiscale approach to Bayesian image segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, *39*, 2680–2687.

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

Ibǎñez, M. V., & Simó, A. (2003). Parameter estimation in Markov random field image modeling with imperfect observations: A comparative study. *Pattern Recognition Letters*, *24*(14), 2377–2389.

Khedam, R., & Belhadj-Aissa, A. (2003). Contextual fusion by genetic approach applied to the classification of satellite images. In R. Goossens (Ed.), *Remote sensing in transition*. Rotterdam: Millpresse.

Li, S. Z. (1995). *Markov andom field modeling in computer vision*. New York: Springer-Verlag.

Ma, Z., & Redmond, R. L. (1995). Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing*, *61*, 435–439.

Mallat, S., & Zhong, S. (1992). Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*, 710–732.

Marroquin, J., Mitter, S., & Poggio, T. (1987). Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, *82*(397), 76–89.

Mather, P. M. (1999). *Computer processing of remotely-sensed images: An introduction* (2nd ed.). Chichester, UK: John Wiley and Sons.

Olsen, R. C., Garner, J., & Dyke, E. V. (2002). Terrain classification in urban wetlands with high-spatial resolution multi-spectral imagery. *Proceedings of SPIE*, *4881*, 686–691.

Rydberg, A., & Borgefors, G. (2001). Integrated method for boundary delineation of agricultural fields in multispectral satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(11), 2514–2520.

Silberstein, M., & Campbell, E. (1989). *Elkhorn slough*. CA, USA: Monterey Bay Aquarium Foundation.

Steger, C. (1998). An unbiased detector of curvilinear structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, 113–119.

Tso, B., & Mather, P. M. (1999). Classification of multisource remote sensing imagery using a genetic algorithm and Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, *37*, 1255–1260.

Tso, B., & Mather, P. M. (2001). *Classification methods for remotely sensed data*. London, UK: Taylor and Francis.

Wang, X., & Wang, H. (2004). Markov random field modeled range image segmentation. *Pattern Recognition Letters*, *25*(3), 367–375.

Wei, J., & Gertner, I. (2003). MRF-MAP-MFT visual object segmentation based on motion boundary field. *Pattern Recognition Letters*, *24*(16), 3125–3139.