Scene classification using combined spectral, textural and contextual information

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

[a]Department of Resource Management, National Defense Management College, NDU, 150, Ming-An Rd, Jon-Ho, Taipei 235, Taiwan
[b]Physics Department, Naval Postgraduate School, Monterey, CA, 93943, USA

ABSTRACT

A classification scheme incorporating spectral, textural, and contextual information is detailed in this paper. The gray level co-occurrence matrix (GLCM) is calculated to generate texture features. Those features are then subjected to a selection process for joining with spectral data in order to evaluate their discrimination capability in classification performance. The classification result is further enhanced with contexture in terms of a refined Markov random field (MRF) model. Multiscale edge features are derived to overcome the bias generally contributed by the presence of edge pixels during the MRF classification process. The smooth weighting parameter for the refined MRF model is chosen based on the probability histogram analysis of those edge pixels. The maximum a posterior margin (MPM) algorithm is used to search the solution. The joining of texture with spectral data produces a significant enhancement in classification accuracy. The refined MRF-model with a soft version line process, in comparison with the traditional MRF model, successfully restricted the commonly found over-smoothed result, and simultaneously improved the classification accuracy and visual interpretation.

Keywords: GLCM, MRF, MPM, texture, contexture, wavelet, multiscale edge.

1. INTRODUCTION

Image classification is a major goal for the field of remote sensing. 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 analysis of satellite-derived imagery. Advances in computer processing speed make techniques such as texture processing practical to implement in pursuit of improvement in classification accuracy. [1, 2] A trend for incorporating additional data into the classification pool is certainly triggered by the concept that the image properties adopted to identify land-use/land-cover types, in addition to tonal information, may also involve other attributes such as texture, context, pattern, size, shadow, and site. Among those features, textural and contextual information have drawn the widest interest for automatic remotely sensed imagery interpretation.[3]

Recent work has looked at texture in SAR imagery in conjunction with image spectra to conduct imagery classification[4, 5, 6.] More recently, high resolution remotely sensed imagery has been analyzed to derive texture. Analysis in conjunction with spectral information to conduct classification enhances the classification accuracy[7]. Texture can be described as the joint tonal variation within a prescribed area. Inside an image of interest, the texture will be related to the variations of pixel gray values within the image or predefined windows. Normally, one may use the term roughness or smoothness to define the characteristics of surface texture, and the resulting measurements are thus called texture features. A related concept is context, which 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 a boundary area. If such contextual information could be well modeled, the classification accuracy could be improved significantly[8].

* brandt@rs590.ndmc.edu.tw, olsen@nps.navy.mil

In order to extract texture features, there are numerous approaches available, such as those based on gray level co-occurrence matrix (GLCM)[9] , multiplicative autoregressive random fields[10], frequency domain filtering in terms of Fourier transform or cosine transform[11], fractal dimension[12], and wavelet transform [5, 13, 14] . These methods each provide different advantages and drawbacks in terms of computational burden and the capability of discriminating texture patterns. In previous work, a series of experiments were conducted to test the effectiveness of the texture extraction techniques noted above with the exception of the wavelet transform. The studies showed that GLCM is the best approach in terms of its higher texture discrimination accuracy [3, 8]. In a recent study by the authors, it was found that GLCM also gives better results than wavelet based approaches (submitted for publication).

In addition to texture, contextual information is also important. Incorporating contextual information 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 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 run. There is a more elegant way of modeling context and achieving higher accuracy. A class of contextual model that is of particular interest is based on Markovian Random Fields (MRF) [8, 15, 16] . The MRF is used to construct an a priori model in the Bayesian sense so as to accomplish the Maximum a Posteriori (MAP) estimate [15]. Such a MAP solution often provides more satisfactory results than a Maximum Likelihood (ML) classifier [3] .

Classification errors are common for boundary (edge) pixels. In the case of MRF, if we incorporate those edge pixels into a smoothing operation, classification errors are likely to result because the edge pixels are poorly defined in their identities. If smoothing is forced across those edge pixels, it will eventually bring about an over-smooth outcome. To avoid such an error, one needs to accurately identify those edge pixels and reduce their contribution during the MRF classification process. That is, one should systematically restrict the smoothing over the edge pixels. Determination of meaningful edges (i.e. real boundaries) is not trivial. Also, the edges derived from different resolutions (or so-called scales) may represent different levels of significance depending on the application and the scene properties. For edge detection, subtle variations may be detected at high resolution (i.e. small scale), while as resolution becomes coarser (i.e. in large scale), only sharp variations are found [17]. We thus need an objective and meaningful manner to gather edge information in multiscales and incorporate these into the classification process. There are many edge detection techniques available[17, 18, 19]. Of particular interest to us is the wavelet-based multiscale edge detection [3]. We will use this method to extract multiscale edges and to perform edge fusion to output higher confidence edge patterns for input into the MRF classification model.

A well-established model cannot execute effectively without the relevant parameters being specified. The determination of MRF model parameters thus is an important issue. The most widely known techniques for estimating the MRF smooth weighting parameters are known as the coding method[20] and least square fit method[21]. These approaches rely on the complete understanding of image neighborhood configurations. Unfortunately, for most supervised classification procedure conducted in remote sensing, the neighborhood configurations are not available. In most supervised classification cases, the parameter estimate in MRF model thus usually determined is trivial and is likely to considerably reduce the effectiveness of the model [8]. For instance, if the MRF model parameters chosen are too small, the smoothing effect will be weak and the results will be close to pixel-wise ML classification. However, if one adopts large MRF model parameters, the model will produce an over-smooth result, which in turn could produce unexpected classification patterns [22, 23]. Moreover as the model becomes more complicated (as with the involvement of edge information in this study), the parameter should be determined with more care. We thus developed a method in order to cope with the parameter estimation issue in MRF model.

Accordingly, the main objectives of this study mainly are twofold: namely, to evaluate the effect of joint spectral and spatial information contributing to the classification accuracy, and to develop a methodology to incorporate the edge information with objectively selected MRF model parameters into the classification process in order to achieve reasonable smoothing in MRF classification mechanisms. The paper is organized as follows, in Section II, the theoretical basis regarding wavelet transform, edge detection, GLCM texture, and MRF, respectively, are briefly described. Section III presents the experimental framework, the feature selection strategy, edge fusion, and the selection of MRF parameters. Experimental results obtained by the proposed methods are reported in Section IV, along with comparisons to those classification results generated by spectral, texture, and alternative MRF-MPM algorithms. Finally, concluding remarks are given.

2. THEORETICAL BACKGROUND

The classification process proceeded in this study involves the spectral, textural and edge information using a MRF model. In this section, the theoretical background for these processes is briefly described. A more in-depth presentation is presented in a paper submitted by the authors for publication elsewhere. We begin with a discussion of the wavelet transform, followed by wavelet-based texture and edge detection. The following sections addresses the basic concept of GLCM and the fundamentals of MRF model.

2.1 Wavelet Transform

The continuous wavelet transform decomposes the signal on a base of elementary functions called *wavelets*. Such a base is formed by translations and dilations of a kernel function ψ called the *mother wavelet*. Formally, we define $a, b \in \Re$, as the dilation step and translation step of translation and dilation processes, respectively. The base is then given by

$$\psi_{a,b} = D_a T_b \ \psi \qquad (1)$$

where $T_b\psi(x) = \psi(x-b)$ and $D_a\psi(x) = |a|^{-1/2}\psi(x/a)$ are the translation and dilation operators, respectively[24,25].

Once the wavelet has been defined, the continuous wavelet transform of a function $f(t)$ can be generated as

$$W_{\psi_{a,b}} f(t) = \langle f(t), T_b D_a \psi \rangle = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt \qquad (2)$$

where $\overline{\psi}$ denotes the complex conjugate of ψ [2]. The computation of the wavelet transform of $f(t)$ for various a and b provides a local representation of $f(t)$, and the information content is represented by the wavelet coefficient $W_{\psi_{a,b}} f(t)$. Our choice of wavelet ψ is a quadratic wavelet, which allows discrete computation using only a few filter coefficients, i.e., a relatively small support [17]

2.2 Multiscale Edge Detection with Wavelet

Mallat and Zhong show that, for edge detection, the edge occurs at the local maxima of the wavelet transform modulus $\left|W_{\psi_{a,b}} f(t)\right|$ for a signal at scale a.[17]. 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 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.

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

$$W^V_{\psi_{2^r}} f(x,y) = \frac{1}{2^{2r}} f\psi^V\left(\frac{x}{2^r}, \frac{y}{2^r}\right) . \qquad (3)$$

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

$$M_{2^r} f(x,y) = \sqrt{\left|W^H_{\psi_{2^r}} f(x,y)\right| + \left|W^V_{\psi_{2^r}} f(x,y)\right|} \qquad (4)$$

and the associated phase

$$A_{2^r} f(x,y) = \tan^{-1}\left(W^V_{\psi_{2^r}} f(x,y) \Big/ W^H_{\psi_{2^r}} f(x,y) \right). \qquad (5)$$

The edge points are then identified as the pixels with locally modulus maxima in one-dimension neighboring pixel along the direction $A_{2^r} f(x,y)$ [17]

2.3 Gray Level Co-occurrence Matrix (GLCM)

The Gray Level Co-occurrence Matrix is a standard technique for extracting texture characteristics [9]. The technique works by forming a window on the image, and then calculating the frequency of co-occurrence for the pixel values (DN). If the pixels' gray values range from m to n, then the matrix will be of dimension m by n, and the frequency of co-occurrence with pixel value i and j will be put into the entry (i, j). Note that the dimensionality of the matrix depends on the dynamic range of the data in the window, and the data are typically scaled to some fairly modest range of integers (e.g. 0-63). Once the counting for the co-occurrence frequencies of all the bi-pixels within the window have been completed, one can then design statistical measures to extract the characteristics of the matrix. The resulting measures reflect the gray value variation (i.e. texture) within the prescribed window. For example, if the area covered by the window is relatively smooth, then the resulting GLCM will hold peaks along the main diagonal. Likewise, if the pixels' values within the window are nearly random, then it will form a GLCM with similar frequency for all the entries. One may design some statistical measures to account for other bi-pixels' co-occurrence distributions. The GLCM has been shown to be very successful in capturing image textures [1, 9, 28, 29].

The window size, directions, and lag value selected for GLCM will vary according to problem at hand and the spatial scale for the features of interest. Normally the lag value of 1 is chosen (i.e. neighboring pixels) and four directions, namely horizontal, vertical, left diagonal, and right diagonal are used for forming GLCM.

In our study, after the GLCM is generated for each direction, eight statistical measures are used for texture extraction. The texture features are, mean, contrast, dissimilarity, homogeneity, angular second moment, entropy, variance, and correlation, respectively, following the definitions in Haralick et al.[9]. The four directions are then averaged to remove directional effects.

2. 5 Markov Random Field Models

Let $x = \{x_1, x_2, ..., x_n\}$ denotes an image with *n* pixels, and suppose there are *m* predetermined classes. Then $c = \{x_1 = c_i, x_2 = c_j, ..., x_n = c_k\}$ denotes the understanding of *x*, i.e., each pixel in *x* being assigned to one of *m* classes. *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

$$p(c|x)\,p(x) = p(x|c)\,p(c). \qquad (6)$$

Normally, prior probability $p(x)$ is assumed to be uniformly distributed, and one may adopt a *maximum a posteriori* (MAP) solution to perform estimate, then (6) can be solved to

$$\hat{c} = \arg\max_c \left[p(x|c)\,p(c) \right]. \qquad (7)$$

To resolve equation (7), one needs to compute both $p(x|c)$ and $p(c)$. 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). 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. In other words, a pixel, given its neighbors, is conditionally and statistically independent of the remainder of the image. Such an assumption provides a sufficiently flexibility while maintaining the mathematically tractable mechanism to model the prior probability $p(c)$. Accordingly, based on MRF, $p(c)$ is given by

$$p(c) = \prod_{i=1}^{n} p(c_i \left| c_{N_i} \right.) \tag{8}$$

where c_{N_i} denotes the neighboring pixels around the pixel *i*. The neighborhood configuration may contain second or higher orders of neighboring pixels 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 can be modeled by

$$p(c_i \left| c_{N_i} \right.) = \frac{\exp\left(\sum_{\{i,j\} \in N} \beta I(c_i = c_j) \right)}{Z} \tag{9}$$

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 classes 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.[22, 30]

The assumption for conditional independence between pixels is also known[30], then equation (7) yields:

$$\hat{c} = \arg\max_{c} \left[\prod_{i=1}^{n} p(x_i \left| c_i \right.) p(c_i \left| c_{N_i} \right.) \right] \tag{10}$$

Equivalently, by taking the log and matching use of the Gaussian assumption, it turns to

$$\hat{c} = \arg\max_{c} \left\{ \sum_{i=1}^{n} \left[-\ln \left| \mathrm{K}_c \right| - \left(x_i - m_c \right)^T \mathrm{K}_c^{-1} (x_i - m_c) \right] + \log p(c_i \left| c_{N_i} \right.) \right\} \tag{11}$$

.

where K_c is class-specific covariance matrix. The equation as shown in (11) is mainly to achieve *smoothness everywhere*. For real data, the scene is more likely to be analyzable in the sense of *piecewise smoothness*. In other words, there are always boundaries, and the boundary pixels are likely to be misclassified due to their uncertain nature. If equation (9) includes the boundary pixels during its smoothing operation, the results will tend to be misleading. To overcome such a misuse of boundary pixels, a *line process* can be introduced[22]. Equation (9) can be modified to include a line process as

$$\log p(c_i \left| c_{N_i} \right.) = \sum_{\{i,j\} \in N} \beta I(c_i = c_j)(1 - l_i) \tag{12}$$

where l_i = 1, if pixel *i* is recognized as an edge, and 0 otherwise. This result indicates that the smoothing will be *switched off* at the presence of an edge. In other words, the smoothing is not allowed to cross the boundaries. However, to make (12) useful, one has to determine the *edges* with higher confidence. Our method will be described below.

To solve equation (11), numerous algorithms can be adopted, such as Iteration Condition Mode (ICM) [30], Simulated Annealing (SA) [15], and Maximum a Posterior Margin (MPM) [31]. The classification algorithm in this study is mainly based on MPM due to its tractable computational burden. The results are comparable to SA, and better than ICM in terms of the higher classification accuracy achieved [3, 8]. The details of the algorithm are explained below.

.

3. EXPERIMENTS

3.1 Test Data Set

The study area known as Elkhorn Slough is located in the central California coast about 160 km south of San Francisco, California[32]. 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 of monitoring the evolution of the area.

The IKONOS imagery is "level 1" or "standard" imagery, and have not been orthorectified. Due to the size of the primary data set (10,000 samples by 20,000 lines), a test area (within the rectangle block as shown in Fig. 1) was extracted from the image in order to facilitate the analysis of the classification methodology effectiveness. 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. The panchromatic image was resampled to 4-m resolution and co-registered to multispectral imagery for use in texture analysis.

The test area shows a variety of landscape types. The agricultural fields around the area are planted with strawberries, broccoli, lettuce, and other similar crops. Many of the fields are well covered with vegetation late in this harvest season, while others were recently harvested or plowed in preparation for planting. The planted fields have color spectra similar to that of the grassland and forest area. The plowed or harvested fields, in turn, are difficult to discriminate from the dried out grass areas and other regions of barren soil in terms of spectral information alone. Some of the fields are covered with plastic, as part of the field preparation process, or as part if the normal cultivation process for strawberries. These areas, which are bright white, are difficult to discriminate from the residential sites. These spectral similarities motivate a desire to find additional means of discrimination.

In our classification experiment, eight information classes were chosen as shown in Table I. 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). The map was the product of a compilation of information from aerial photos and Global Positioning System assisted field observations. The ESF used a 50% thumb rule to qualitatively assign the class contained to each region defined by vector geo-referenced polygons. There are a total 71804 ground truth pixels available. Of those ground truth pixels, 15451 pixels are assigned to the training set to train the ML classifier, and 56353 pixels are used for classification performance evaluation.

Figure 1. Elkhorn Slough

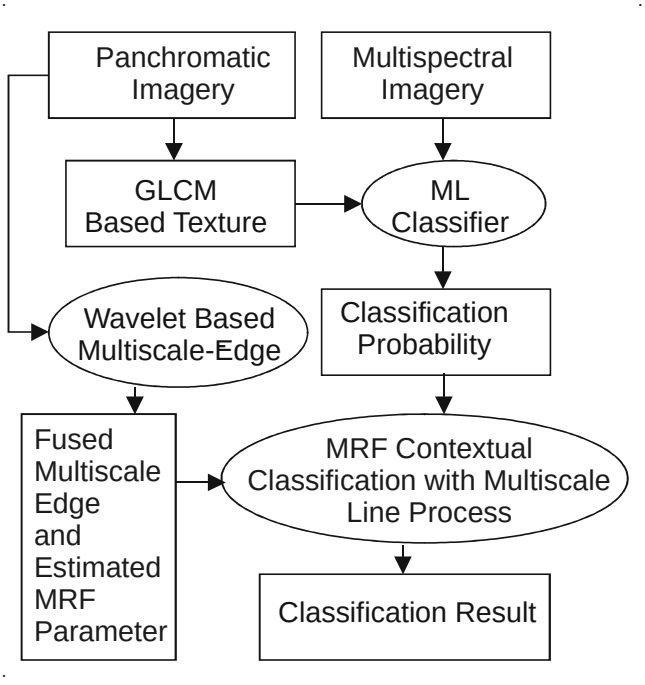


Figure 2 Processing diagram

3.2 Experimental Framework

The framework of the experiment design relating to the theoretical background described is now shown in Fig. 2. The panchromatic imagery is used to derive texture features in terms of GLCM. The texture features are combined with the multispectral imagery. The combined data set is then used as input to the maximum likelihood classifier. The resulting classification probabilities are then further refined with the fused multiscale edges (derived from panchromatic imagery) and the contextual classification using the MRF line process model. In order to achieve a less biased estimate of the MRF model smooth weighting parameter, we have used the classification probability histogram within the edge pixels to perform the parameter determination.

3.3 Feature Extraction and Selections

For texture extraction, it is important to determine the proper window size, since the texture may be scale dependent. Instead of using an ad hoc approach, the semi-varigram measure was adopted for window size determination. The semi-varigram is a useful statistical quantization, which can reflect the variation of data with different ranges (or lags). If we assume the landscape texture patterns are periodic, then the suitable window size for quantizing the texture pattern will occur around the lag where the semi-varigram achieves a maximum [3]. A 128 by 128 sub-image was subsetted from the panchromatic image for each class, using our knowledge of ground truth. The semi-varigram was then applied to these sub-images. It was found that the higher values of the semi-varigram measure were achieved for lags of 15 to 23. A window size of 17 was then chosen for calculating the GLCM.

The GLCMs were generated using the lag of one in each of the four canonical directions. Each GLCM calculation generates eight texture measures. The four directions (up/down, left/right) provide a set of thirty-two texture features or bands. The GLCM filters calculated in terms of different directions is sensitive to the varying orientation of the texture patterns, and we begin to have an excessive number of texture bands. The traditional approaches to reducing dimensionality were used including principle component analysis (PCA), maximum value preserved, and averaging. The results were subjected to visual evaluation. It was found that the simple average of four-orientation GLCM texture features produced the most satisfactory results among those methods. The direction-averaged features then served as texture candidates, which will be applied in the classification process through the procedure described in below.

The presence of a large number of derived bands could lead to poor classification results if care is not given to the contribution of these bands to the inter-class separability. Too many input bands can even decrease classification accuracy[33]. Also, too many input bands can cause a dramatic increase in computational burden. The feature selection process is therefore an important step. The methodology implemented in this study is based on the transformed divergence (TD) measure, notable for its good performance in feature selection[34, 35]. The TD value is related to the level of separability between classes. TD values vary from 0 to 2.0, corresponding to complete overlap and ideal separation between two classes, respectively. Rather then using the overall average of TD values among classes to conduct feature selection, our method is more focused on resolving the potential error contributed by low values of TD for class pairs.

Our feature selection strategy is as follows. First, the four multispectral bands are used to derive TD for each of the class pair in Table I. It was found that the separability between class pairs (1, 2), (1, 8), (4, 5), (4, 8), and (5, 6) are in poor separability with the values under 1.7, while for other class pairs all are above the value of 1.8. The selection of texture is then based on the improvement of average separability between these class pairs, and the four texture features are combined with the spectral bands to conduct classification. The texture features were selected by a strategy of comparing the TDs between class pairs with varying texture feature pairs. The feature pairs were sorted in descending order in terms of TD values. The winning feature pair with the highest average TD is then chosen as a texture feature to be used. The texture feature is combined with the 4 spectral bands, and the remaining texture features are again competed. The process is iterated with each winning texture feature added to the collection of final features. Using this process, the selected GLCM texture features are mean, homogeneity, contrast, and dissimilarity. GLCM textures are shown in Fig. 3.

Table 1

No.	Class Name
1	Cultivated Soil
2	Dry Grasses
3	Water
4	Wetland
5	Trees
6	Crop Vegetation
7	Bare Fields
8	Man-made Features (Building, Roads)

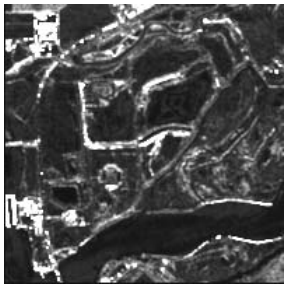
The edge detection method described above was applied, and the values for modulus (5), phase (6), and modulus maxima for each pixel in the wavelet transformed image were recorded. Three scale images from r=0 to 2 produced by wavelet transform are used to derive edges, and the edge image is generated by adopting the process similar to that used previously [18] (see Fig. 6). This was done on multiple scales (r = 0, 1, 2), and the results sifted for a final edge results. Details are given elsewhere.[38]

3.4 Parameter Determination

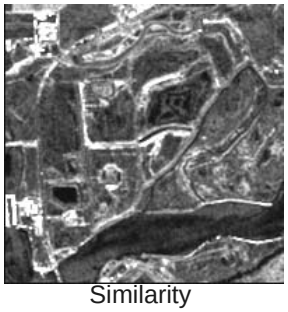
A model is not complete unless both the model formation and parameters are well defined. The MRF model described above provides a theoretically robust basis for modeling the context in spatial domain. Once the model is defined, the question then arises: how do we assign the suitable parameters into the model?

Two parameters need to be determined. The first, *li*, is the value for the occurrence of an edge on pixel *i*. The second is the smooth weighting parameter β. A rather complex process is adopted to match the spatial scales and the character of the boundaries found in this data set. Details are again provided elsewhere. [38]

Panchromatic Image

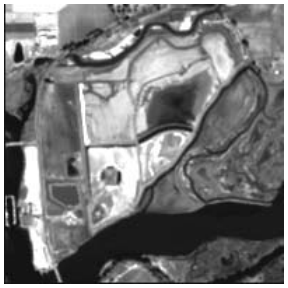


Contrast



Similarity

Homogeneity



Mean

Figure 3. GLCM results

Wavelet - r=1

Edges

Figure 4. Edge Detection

3.5. Classification Algorithm

After the model and parameters have been defined, the next step is to design the search approach to find the global or local minimum (or maximum) within the model. It was noted above that several classical algorithms[15, 30, 31] , are available for achieving such an aim. The Iterated Conditional Modes (ICM) described by Besag[30,] aims to find a local maximum, but the results are not quite comparable to the global methods[3]. Simulated Annealing (SA) aims to achieve a global maximum, but the computational burden is quite high[15]. An alternative optimization method called Maximer of the Posterior Marginals (MPM) is thus adopted in our study to find the maximum solution in equation (11) [31].

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 is present 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

$$p(c_r \mid x_i) = \frac{1}{\eta - \nu} \sum_{\nu+1}^{\eta} I(c_r) \qquad (13)$$

where $p(c_r \mid x_i)$ represents the probability of a class c_r given the observation x_i at pixel i, $1/(\eta\text{-}\nu)$ is the normalization term, while ν 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 ν to more than 20, respectively, stable results are likely to be reached.

Our experiments involve only the pair-wise neighborhood system in MRF. There is sufficient flexibility and classification power with such a neighborhood system[20]. It is noted that the edge pixels could contribute a bias within the neighborhood system. Accordingly, we make use of soft version of the product of the edge detection process. An iterative MPM process is then followed, beginning with the initial classification results from the ML classifier.

4. RESULTS AND DISCUSSIONS

4.1 Results of ML and the Joint Texture Analysis

Classification using IKONOS multispectral imagery with the ML classifier alone gives only modest success. The total classification accuracy of 65.9% with kappa statistic of 0.6 is achieved[37]. The corresponding classification confusion matrix (not given here, shows that there is a particular confusion involving class 8 (man-made features), and a modest confusion between classes 1 & 2 (cultivated soil and dry grass). The results of the ML classification thus leave considerable room for further improvement.

The inclusion of texture features contributes some improvement in classification accuracy. There is nearly 7% improvement from the use of GLCM texture features. (Overall Accuracy = 73.1%, Kappa Coefficient = 0.67) There was a particularly significant improvement in the assignment of the man-made features category. There was less improvement in the distinction between class 1 (cultivated soil) and class 2 (dry grass land). . This corresponds to our intuitive understanding of the two pairs of classes. More effort is needed to sort out some other potential texture measures in order to contribute more successful discrimination. Our texture selection strategy as shown in Section III-C is focused on resolving the serious classification confusion obtained from spectral data alone, and such an error-focusing method generates some improvements. Other feature selection strategies may contribute to different classification results.

Fig. 5(a) and (b) are parts of the classified images generated by multispectral imagery alone and with the addition of GLCM texture features, respectively. It can be seen that the ML classification using spectral data alone generates a busy looking result. After adding the texture features, the result, by visual inspection, is more correct. However, 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 mixed. Further refinement is necessary.

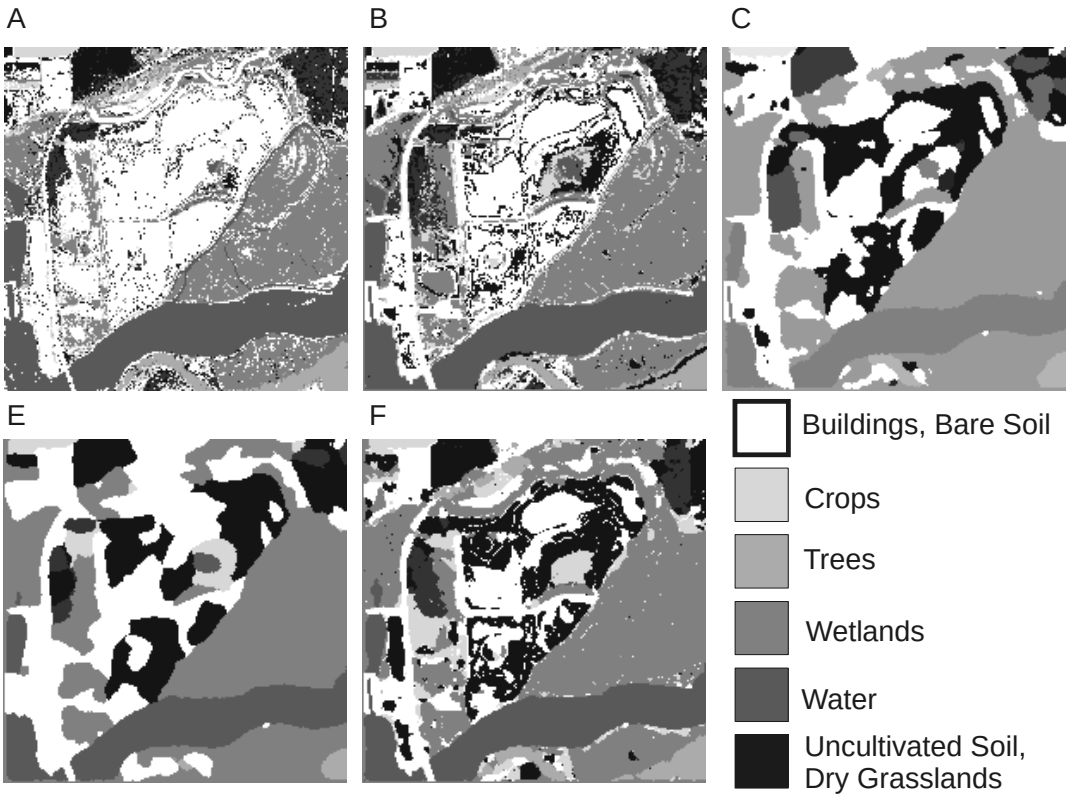


Figure 5. A) Classification using only MSI (4 color); B) MSI + textures; C) MSI, texture, with MPM algorithm, D) MPM w/o line process; E) MPM with Boolean line process

4.2 Results with Contextual Classification

The GLCM texture classification revealing the highest accuracy is used for further experiments in order to improve the visual interpretation and to evaluate the performance of joint contextual classification. The MPM algorithm with carefully set model parameters is used for finding the solution of maximum in MRF model. An accuracy of 78.56% (kappa 0.74) in overall classification is achieved. In comparison with spectral data alone, the increase in classification accuracy is around 13%, while there is nearly 5% enhancement comparing to classification using spectral and textural features.

Fig. 5(c) shows the same classification area as displayed in 5 (a) and (b). The visual quality clearly is an improvement over the classifications generated solely by spectral and texture features. For comparison purpose, the MRF model without the inclusion of the line process (i.e. equal smoothing for each pixel and its neighbors) and the MRF model with the Boolean line process (i.e. the smoothing is switched off in the presence of an edge), respectively, are both carried out in our experiments. The classification images are displayed in Fig. 5(d) and (e), respectively. In comparison with Fig. 5(c) and by analyzing the classification confusion table, Fig 5(d) reveals an over-smoothed result. Please note that on the lower-left part of Fig. 5(c), which is the pier area (a ++-like shape), our refined MPM algorithm retains the area boundaries very well and simultaneously contributes the smoothness to other areas where is no significant edge presence. In Fig. 5(d) the area boundaries have been smoothed out, which indicates that the smoothing is beyond control. Fig. 5(e) displays the image with smoothing switched off at the edge pixels. The result, by visual inspection, is clearly not satisfactory. The overall classification accuracy for the Boolean line process is 76.44% (kappa 0.70), while 75.11% (kappa 0.68) is achieved for the MRF model without the line process.

5. CONCLUDING REMARKS

In this study, we show the improvement in classification accuracy obtained from spectral imagery by adding texture and contextual information into the classification pool. The results reveal that, compared to the use of multispectral data alone, higher levels of accuracy can be obtained. The joining of texture features with spectral features, although bringing further improvement in classification accuracy, still shows problems with interpretation. However, with the involvement of contextual information, the enhancement in both the classification accuracy and visual interpretation, respectively, can be simultaneously achieved.

We have proposed a method to estimate the MRF smooth weighting parameter (from the stochastic perspective by analyzing the class probabilities assigned within edge pixels. Such a method is expected to dramatically outperform the traditional ad hoc approaches. The core of such a method is that to the extent edge pixels can be well managed, the classification results can achieve to a higher level of accuracy. We believe that only in the case that both model form and parameter are well designed that an effective result can be achieved. In this study, we present just for one MRF-based parameter estimate. The study might 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 for future study.

The MPM algorithm has been refined in this study with a soft version line-process to obtain more satisfactory classification results. The methodology presented here can systematically restrict the smoothing function and classification error due to the presence of an edge pixel. Our idea of bringing different smoothing functions with different significance for fused edges works quite well. The method effectively restricts the over smooth results commonly seen with the MRF-based model with higher values of neighborhood parameters. The methodology implemented here is combined with line features obtained in the off-line manner. It may be possible to develop a more efficient model to simultaneously achieve higher confidence of edge estimation and supervised classification.

REFERENCES

1. H. Anys, A. Bannari, D. C. He, and D. Morin, "Texture analysis for the mapping of urban areas using airborne MEIS-II images," Proceedings of the First International Airborne Remote Sensing Conference and Exhibition, ERIM, Strasbourg, France, 1994, pp. 231-245.
2. R. C. Olsen, J. Garner, E. V. Dyke, "Terrain classification in urban wetlands with high-spatial resolution multi-spectral imagery," Proceedings of SPIE, vol. 4881, pp. 686-691, 2002.
3. B. Tso and P. M. Mather, Classification Methods for Remotely Sensed Data. London: Taylor and Francis, 2001.
4. P. M. Mather, B. Tso, and M. Koch, "An evaluation of Landsat TM spectral data and SAR-derived textural information for lithological discrimination in the Red Sea Hills, Sudan," International Journal of Remote Sensing, vol. 19, pp. 587-604, 1998.
5. L. Bian, "Retrieving urban objects using a wavelet transform approach," Photogrammetric Engrineering and Remote Sensing, vol. 69, pp. 133-141, 2003.
6. M. Simard, S. S. Saatchi, and G. De Grandi, "The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest," IEEE Transactions on Geoscience and Remote Sensing, vol. 38, pp. 2310-2321, 2000.
7. M. Herold, X. H. Liu, and K. C. Clarke, "Spatial metrics and image texture for mapping urban land use," Photogrammetric Engineering and Remote Sensing, vol. 69, no. 9, pp. 991-1002, 2003.
8. B. Tso and P. M. Mather, "Classification of multisource remote sensing imagery using a Genetic Algorithm and Markov Random Fields," IEEE Transactions on Geoscience and Remote Sensing, vol. 37, pp. 1255-1260, 1999.
9. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Transactions on Systems, Man and Cybernetics, vol. 3, pp. 610-621, 1973.
10. A. H. Schistad, A K. Jain, and T. Taxt, "Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images," IEEE Transactions on Geoscience and Remote Sensing, vol. 32, pp. 768-778, 1994.
11. G. Sohn, and I. J. Dowman, "Extraction of buildings from high resolution satellite data," In Automated Extraction of Man-Made Objects from Aerial Space Images (III), E. Baltsavias, A. Grün, and L. Van Gool Editors, Lisse: A. A. Balkema, 2001, pp. 345-355.

12. H. L. Qiu, N. S. Lam, D. A. Quartrochi, and J. A. Gamon, "Fractal Characterization of Hyperspectral Imagery" Photogrammetric Engrineering and Remote Sensing, vol. 65, pp. 63-71, 1999.
13. C. Zhu and X. Yang, "Study of remote sensing image texture analysis and classification using wavelet," International Journal of Remote Sensing, vol. 19, pp. 3197-3203, 1998.
14. T. Ranchin, "Wavelet for modeling in remote sensing," Proceeding of SPIE, Wavelet Applications in Signal and Image Processing VII, Denver, Colorado, 19-21 July, 1999, pp. 208-218.
15. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and Bayesian restoration of images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 6, no. 6, 721-741.
16. G. Fan and X. G. Xia, "A joint multicontext and mutiscale approach to Baysian image segmentation. IEEE Transactions on Geoscience and Remote Sensing, vol. 39, pp. 2680-2687, 2001.
17. S. Mallat and S. Zhong, "Characterization of Signals from multiscale edges," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, pp. 710-732, 1992.
18. J. A. Canny, "A computational approach to edge detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679-698, 1986.
19. A. Rydberg and G. Borgefors, "Integrated method for boundary delineation of agricultural fields in multispectral satellite images," IEEE Transactions on Geoscience and Remote Sensing, vol. 39, no. 11, pp. 2514-2520, 2001.
20. H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 9, no. 1, pp. 39-55, 1987.
21. J. Besag, "Spatial interaction and the statistical analysis of lattice systems," Journal of Royal Statistics Society, vol. 36, pp. 192-236, 1974.
22. S. Z. Li, Markov Random Field Modeling in Computer Vision. New York: Springer-Verlag, 1995.
23. B. Tso, "An investigation of alternative strategies of combining spectral, textural, and contextual information for remotely sensed imagery classification," Ph. D. thesis, School of Geography, The University of Nottingham, 1997.
24. I. Daubechies, Ten Lectures on Wavelets. Philadelphia, PA: SIAM, 1992.
25. S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," IEEE Transactions on Pattern Analysis and Machine Learning, vol. 11, no. 7, pp. 674-693, 1989.
26. C. S. Lu, P. C. Chung, and C. F. Chen, "Unsupervised texture segmentation via wavelet transform," Pattern Recognition, vol. 30, pp. 729-742, 1997.
27. L. K. Soh and C. Tsatsoulis, "Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-occurrence Matrices," IEEE Transactions on Geoscience and Remote Sensing, vol. 37, pp. 780-795, 1999.
28. P. M. Treitz, P. J. Howarth, F. O. Rotunno, and E. D. Soulis, "Agricultural Crop Classification using SAR Tone and Texture Statistics," Canadian Journal of Remote Sensing, vol. 26, no. 1, pp. 18-29, 2000.
29. S. Arzandeh and J. Wang, "Texture evaluation of RADARSAT imagery for wetland mapping," Canadian Journal of Remote Sensing, vol.28, no. 5, pp. 653-666, 2002.
30. J. Besag, "On the statistical analysis of dirty pictures," Journal of the Royal Statistical Society, Series B, vol. 48, no. 3, pp. 259-302, 1986.
31. J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," Journal of the American Statistical Association , vol. 82, no. 397, pp. 76-89, 1987.
32. M. Silberstein and E. Campbell, Elkhorn Slough. CA: Monterey Bay Aquarium Foundation, 1989.
33. K. P. Price, X. Guo, and J. M. Stiles, "Optimal Landsat TM band combinations and vegetation indices for discrimination of six grassland types in eastern Knasas," International Journal of Remote Sensing, vol. 23, pp. 5031-5042, 2002.
34. P. H. Swain and S. M. Davis, Remote Sensing: The Quantitative Approach. New York: McGraw-Hill, 1978.
35. P. W. Mausel, W. J. Kramber, and J. K. Lee, "Optimum band selection for supervised classification of multispectral data," Photogrammetric Engrineering and Remote Sensing, vol. 56, pp. 55-60, 1990.
36. J. C. Bezdek, Pattern Recognition with Fuzzy Objective Algorithms. New York: Plenum Press, 1999.
37. R. Congalton, "A review of assessing the classifications of remotely sensed data," Remote Sensing of Environments, vol. 37, pp.35-46, 1991.
38. B. Tso and R. C. Olsen, “Remote Sensing Imagery Classification Using Combined Spectral, Textual, and Contextual Information, submitted for publication, IEEE TGRS, 2004.