International Journal of Remote Sensing
Vol. 26, No. 10, 20 May 2005, 2113–2133

Taylor & Francis
Taylor & Francis Group

Combining spectral and spatial information into hidden Markov models for unsupervised image classification

B. TSO† and R. C. OLSEN‡

†Department of Resource Management, National Defense Management College, NDU, 150, Ming-An Rd, Jon-Ho, Taipei, 235, Taiwan; e-mail: brandt@rs590.ndmc.edu.tw

‡Physics Department, Naval Postgraduate School, Monterey, CA, 93943, USA; e-mail: olsen@monterey.nps.navy.mil

(*Received 12 January 2004; in final form 22 October 2004*)

Unsupervised classification methodology applied to remote sensing image processing can provide benefits in automatically converting the raw image data into useful information so long as higher classification accuracy is achieved. The traditional k-means clustering scheme using spectral data alone does not perform well in general as far as accuracy is concerned. This is partly due to the failure to take the spatial inter-pixels dependencies (i.e. the context) into account, resulting in a 'busy' visual appearance to the output imagery. To address this, the hidden Markov models (HMM) are introduced in this study as a fundamental framework to incorporate both the spectral and contextual information in analysis. This helps generate more patch-like output imagery and produces higher classification accuracy in an unsupervised scheme. The newly developed unsupervised classification approach is based on observation-sequence and observation-density adjustments, which have been proposed for incorporating 2D spatial information into the linear HMM. For the observation-sequence adjustment methods, there are a total of five neighbourhood systems being proposed. Two neighbourhood systems were incorporated into the observation-density methods for study. The classification accuracy is then evaluated by means of confusion matrices made by randomly chosen test samples. The classification obtained by k-means clustering and the HMM with commonly seen strip-like and Hilbert-Peano sequence fitting methods were also measured. Experimental results showed that the proposed approaches for combining both the spectral and spatial information into HMM unsupervised classification mechanism present improvements in both classification accuracy and visual qualities.

1. Introduction

Remotely sensed imagery interpretation, or more generally the term 'classification', is an important process, which translates the raw image data into more meaningful and understandable information. Normally, the classification process can be categorized into two broad categories known as supervised and unsupervised classification, respectively (Mather 1999). Supervised classification requires training data for each class selected in advance to train the classifier. The trained classifier is then used to identify the pixels in the imagery. Unsupervised classification mechanism automatically clusters the image data into several groups according to some predefined criterion or cost function (for example, clustering data based on minimum distance). Those groups are then mapped into classes. Unsupervised methods are more attractive in the sense that it does not require as much

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160512331337844

intervention. As the data volume collected by a variety of air-borne and space sensors increases dramatically, unsupervised classification can bring considerable benefit in enhancing imagery processing speed to the extent that a stable and high classification accuracy is achieved. Unfortunately, traditional unsupervised classification methodology, such as k-means clustering using imagery spectral data alone, does not generally produce high classification accuracy (Tso and Mather 2001). This is due to the fact that, in common and traditional clustering mechanisms, each pixel in an image is treated as spatially independent. This, in turn, makes the output image after clustering unlikely to form a patch-like and easily interpretable pattern. In order to achieve better classification outcomes, the validity of such an inter-pixel independency assumption should be concerned. More robust ways of modelling pixel interactions are needed. The focus of this study is thus to deal with the spatial modelling issue for unsupervised classification mechanisms.

In remotely sensed imagery, there are factors that cause neighbouring pixels to exhibit some level of mutual characteristics. Examples of such factors can be atmospheric interaction, the spatial and spectral resolution of a sensor, and the mechanism of the pixels being generated (e.g. SAR multi-look imagery). Also, when mapping the pixels to landscape patterns, if a pixel identified as 'forest', it will be most likely surrounded by the same class of pixels. If such a spatial interaction is well modelled, the classification accuracy can be improved (Tso and Mather 1999). We may, in short, use 'context' to represent such spatial relationships, which can be interpreted as how the probability of presence of one pixel (or pixels) is affected by its (their) neighbours.

Incorporating contextual information into the classification process can be achieved in different ways. One simple method of adopting context is to use so-called majority voting within a prescribed window. In such a method, the central pixel is forced to adopt the class that presents most frequently in the window. However, there is a more robust way of modelling context. A class of contextual model that is of particular interest is based on Markovian theory known as the hidden Markov model (HMM) (Baum and Petrie 1966, Baum and Egon 1967, Baum *et al.* 1970, Rabiner 1989). Since the development of the HMM, it has earned popularity in speech recognition (Huang *et al.* 1990, Cole *et al.* 1995). Applications of the HMM to image processing has been also growing (see, for instance, He and Kundu 1991, Viovy and Saint 1994, Li *et al.* 2000a, b, Gader *et al.* 2001, Runkle *et al.* 2001, Fjørtoft *et al.* 2003).

When an HMM is implemented for unsupervised image classification, the most straightforward method of incorporating pixels into HMM is by sweeping the image line by line or column by column, or some alternative scan sequence can be applied, such as the Hilbert-Peano scan (Abend *et al.* 1965, Skarbek 1992, Giordana and Pieczynski 1997, Fjørtoft *et al.* 2003). It is noted that the original theory of HMM assumes a one-dimensional linear dependent structure (Baum and Petrie 1966), while an image is normally 2-dimensional in a contextual sense. The spatial dependencies are thus not well modelled with the linear structure of HMMs, and may result in restrictions to the enhancement of classification accuracy. There are alternative strategies to build 2D HMMs, or so-called Markov meshes to pursue better fitness for such 2D spatial dependencies (Abend *et al.* 1965, Devijver 1985, Li *et al.* 2000a, b). However, using 2D HMMs requires developing more complicated algorithms for searching optimal model states. This may require making additional assumptions regarding models, and also may contribute heavier computational load

(Devijver 1985, Li *et al* 2000a). To overcome such drawbacks, rather than making changes to the HMM structure, we have developed methodology in terms of spatial and feature space manipulation within a linear HMM to pursue higher accuracy through unsupervised classification. In this way, the 2D information can be embedded into HMM, while the original one-dimensional linear structure of an HMM is still remained. The ways of converting 2D information of the remotely sensed imagery into one-dimension HMM are thus the main interest of this study.

This paper is organized as follows. In §2, the fundamentals of HMM theory are introduced. The methodology of converting 2D spatial information into one-dimension HMM is detailed in §3. Two types of approaches, namely observation-sequence-based methods and observation-density-based methods, are proposed. In §4, several methods are compared and the encouraging experimental results along with discussions are described. Finally, concluding remarks and the suggestions for future researches are given in §5.

2. Hidden Markov model

A hidden Markov model (HMM) is distinguished from a general Markov model in that the states in an HMM cannot be observed directly (i.e. hidden) and can only be estimated through a sequence of observations generated along a time series. Assume the total number of states being N, and let q_t and o_t denote the system state and the observation at time t. An HMM, λ, can be formally characterized by $\lambda=(A, B, \pi)$, where A is a matrix of probability transition between states, B is a matrix of observation probability densities relating to states, and π is a matrix of initial state probabilities, respectively. Specifically, A, B, and π are each further represented as:

$$A=\left[a_{ij}\right],\ a_{ij}=P(q_{t+1}=j|q_t=i),\ 1\le i,j\le N, \tag{1}$$

where

$$a_{ij}\ge 0,\ \sum_{j=1}^{N} a_{ij}=1,\ \text{for } i=1, 2, \ldots, N \tag{2}$$

$$B=\left[b_j(o_t)\right],\ b_j(o_t)=P(o_t|q_t=j),\ 1\le j\le N \tag{3}$$

$$\pi=[\pi_i],\ \pi_i=P(q_1=i),\ 1\le i\le N, \tag{4}$$

where

$$\sum_{i=1}^{N}\pi_i=1. \tag{5}$$

For illustration purposes, an HMM model and related parameters, namely, A, B and π, are shown in figure 1. The observation probability density $b_j(o_t)$ for state j given observation o_t is generally modelled as Gaussian distribution (Liporace 1982, Juang 1985, Linde *et al*. 1980, He and Kundu 1991) as:

$$b_j(o_t)=\frac{1}{\sqrt{(2\pi)^k\left|\Sigma_j\right|}}\exp\left[-\frac{1}{2}\left(o_t-\mu_j\right)\sum\nolimits_j^{-1}\left(o_t-\mu_j\right)'\right], \tag{6}$$

where prime denotes vector transpose and k is the dimension of observation vector o_t.

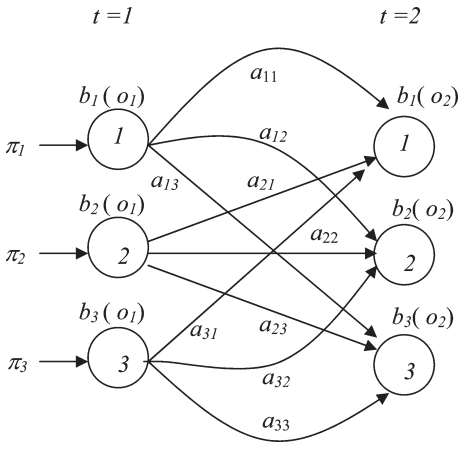


Figure 1. Illustration of HMM parameters A, B and π in the case of $t=1-2$.

Given an HMM λ and observation sequence $O=\{o_1, o_2, \ldots, o_T\}$, one may estimate the best state sequence $Q^*=\{q_1, q_2, \ldots, q_T\}$ based on a dynamic programming approach so as to maximize $P(Q^*|O, \lambda)$ (Rabiner 1989, Forney 1973). In order to make Q^* meaningful, one has to well set up the model parameters A, B and π. The Baum-Welch algorithm (Baum *et al.* 1970) is the most widely adopted methodology for model parameters estimation. The model parameters π_i, a_{ij}, mean μ_i and covariance Σ_i are each characterized as:

$$\pi=\gamma_1(i) \tag{7}$$

$$a=\frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)} \tag{8}$$

$$\mu_i=\frac{\sum_{t-1}^{T}\gamma_t(i)o_t}{\sum_{t=1}^{T}\gamma_t(i)} \tag{9}$$

$$\Sigma_i=\frac{\sum_{t=1}^{T}\gamma_t(i)(o_t-\mu_t)'(o_t-\mu_t)}{\sum_{t=1}^{T}\gamma_t(i)}, \tag{10}$$

where $\gamma_t(i)$ denotes the conditional probability of being state i at time t, given the observations, and $\xi_t(i, j)$ is the conditional probability of a transition from state i at time t to state j at time $t+1$, given the observations.

Both $\gamma_t(i)$ and $\xi_t(i, j)$ can be solved in terms of a well-known *forward-backward* algorithm (Baum and Egon 1967). Define the forward probability $\alpha_t(i)$ as the joint probability of observing the first t observation sequence $O_{1\text{ to }t}=\{o_1, o_2, \ldots, o_t\}$ and being in state i at time t. The $\alpha_t(i)$ can be solved inductively by following formulae:

$$\alpha_1(i)=\pi_i b_i(o_1),\ 1\le i\le N \tag{11}$$

$$\alpha_{t+1}(i)=b_i(o_{t+1})\sum_{j=1}^{N}\left[\alpha_t(i)a_{ij}\right],\ \text{For } 1\le t\le T,\ \text{For } 1\le i\le N \tag{12}$$

Let the backward probability $\beta_t(i)$ be the conditional probability of observing the

observation sequence $O_{t+1 \text{ to } T}=\{o_{t+1}, o_{t+2}, \ldots, o_T\}$ after time t given that the state at time t is i. As with the forward probability, the $\beta_t(i)$ can be solved inductively as:

$$\beta_T(i)=1, 1 \le i \le N \tag{13}$$

$$\beta_t(i)=\sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j),\ t=T-1,\ T-2,\ \ldots,\ 1,\ 1 \le i \le N \tag{14}$$

The probabilities $\gamma_t(i)$ and $\xi_t(i, j)$ are then solved by:

$$\gamma_t(i)=\frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)} \tag{15}$$

$$\xi_t(i,j)=\frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)} \tag{16}$$

By analysing the structure and major parameters in the HMM training algorithm as described in equations (7) to (16), it is clear that the resulting hidden state sequence is strongly dependent on the distribution of observations $b_i(o_t)$, and observation sequence $O=\{o_1, o_2, \ldots, o_T\}$. Specifically, one may sequentially determine the effectiveness of parameters through the following trace:

1. $b_i(o_t)$ and O both affect $\alpha_t(i)$ and $\beta_t(i)$ as shown in equations (11) to (14);
2. $\alpha_t(i)$ and $\beta_t(i)$ make up $\gamma_t(i)$ and $\xi_t(i, j)$ as shown in equations (15) and (16);
3. $\gamma_t(i)$ and $\xi_t(i, j)$ determine π_i, a_{ij}, and $b_i(o_t)$ according to equations (7) to (10), and eventually generate hidden states estimation Q^*.

As a result, if both the observation density $b_i(o_t)$ and observation sequence $O=\{o_1, o_2, \ldots, o_T\}$ are well managed, the revealed hidden state sequence will be closer to the ideal situation (e.g. higher classification accuracy).

When implementing HMM for unsupervised image classification, the pixel values (or vectors) correspond to the observations, and after the estimation for the model parameter is completed, the hidden state then corresponds to the cluster to which the pixel belongs. For incorporating imagery into an HMM, the most straightforward manner is sweeping the image line by line to fit the pixels into HMM. Figure 2(*a*) shows an example of 4×4 image. The sweeping process visits each pixel through left-right, strip-like, direction as shown in figure 2(*b*). It is observed that such an arrangement uses only one-dimensional spatial dependencies. Each pixel gains contextual effectiveness only from its preceding left-hand side pixel. According to the characteristics of HMM, such an arrangement is likely to generate horizontal strip patterns (as will be demonstrated in a later section). There is another method of pixel sequencing known as the Hilbert-Peano sequence (Abend *et al*. 1965), which is commonly seen in the Markovian-based image processing (for example, see Skarbek 1992, Giordana and Pieczynski 1997, Fjørtoft *et al*. 2003). An illustration of the Hilbert-Peano sequence is shown in figure 2(c). In comparison to the strip-like sequence fitting, Hilbert-Peano may provide more flexibility in involving multi-direction neighbouring information.

It is noted that the Hilbert-Peano sequence is a kind of observation-sequence adjustment method to fit the pixels into HMM. Other ways of manipulating the observation sequence are certainly possible, and it will be worthwhile to investigate

how those different observation adjustments affect the presentation of the hidden states (i.e. clustering results). On the other hand, as demonstrated earlier, the observation density can contribute to the change of hidden states. Thus, it is also of interest to ascertain how the classification performs relating to different observation densities. Triggered by the above concerns, two types of strategies of observation adjustments, namely observation-sequence-based methods and observation-density-based methods, are proposed in order to pursue the possibility of unsupervised classification accuracy improvements to the remotely sensed imagery. The methodology is stated as follows.

3. Methodology

3.1 *Observation-sequence-based methods*

Five approaches to observation sequence manipulations to fit HMM are proposed. These methods can be further divided into two categories called pixel non-redundant and pixel redundant schemes, respectively. For the pixel non-redundant scheme, each pixel within an image presents only once in an HMM, while for the redundant scheme, each pixel within an image may present multiple-times depending on the fitting method being applied. Figures 2(*d*) and (*e*) show two ways of incorporating neighbourhood information, a 'V'-like and 'U'-like sequencing approach for a 'pixels non-redundant' sense. Figures 2(*f*) and (*g*) show the same 'V'-like and 'U'-like sequencing approach in terms of the pixel redundant scheme. Figure 2(*h*) shows '◇'-like sequencing, which is also pixel redundant. These methods will be detailed below. It is worthwhile to note that the pixels taken into concern are within the scope of the first- and second-order neighbourhood systems as commonly seen in Markov random fields theory (Tso and Mather 2001).

In following, we define $p_{i,j}$ as a pixel at location row i and column j. An image is assumed to be of size $I \times J$.

- **Pixel non-redundant 'V'-like observation sequencing (figure 2(*d*)).** This incorporates pixels into HMM through a repeated 'V'-shaped scan. For instance,

(*a*)

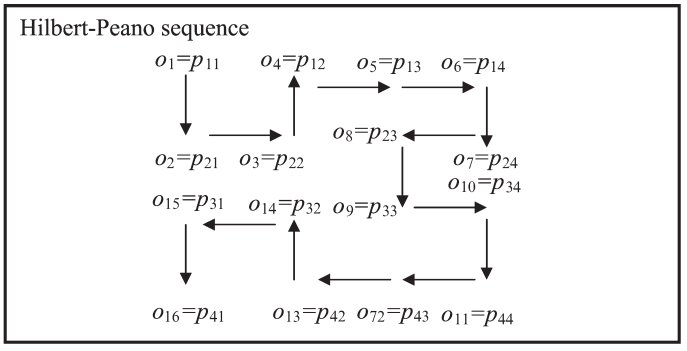
p_{11}	p_{12}	p_{13}	p_{14}
p_{21}	p_{22}	p_{23}	p_{24}
p_{31}	p_{32}	p_{33}	p_{34}
p_{41}	p_{42}	p_{43}	p_{44}

(*b*)

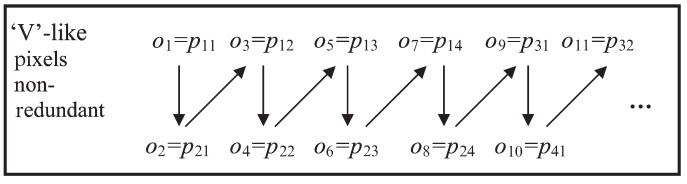
Strip-like $o_1=p_{11}$ $o_2=p_{12}$ $o_3=p_{13}$ $o_4=p_{14}$ $o_5=p_{21}$ $o_6=p_{22}$ …

Figure 2. (*a*) A 4×4 image; (*b*) strip-like sequencing; (*c*) Hilbert-Peano sequencing; (*d*) proposed pixel non-redundant 'V'-like sequencing; (*e*) pixel non-redundant 'U'-like sequencing; (*f*) pixel redundant 'V'-like sequencing; (*g*) pixel redundant 'U'-like sequencing; and (*h*) pixel redundant '◇'-like sequencing.

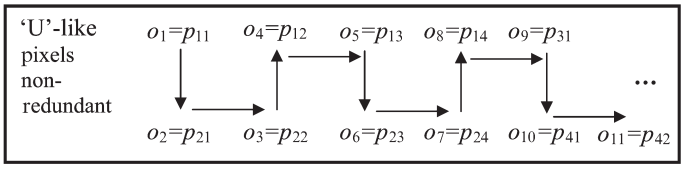
(*c*)



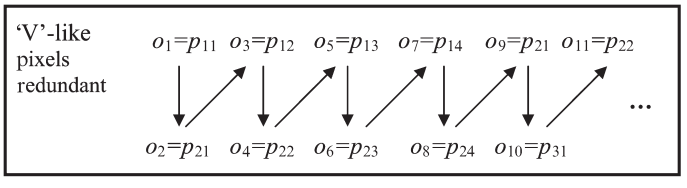
(*d*)



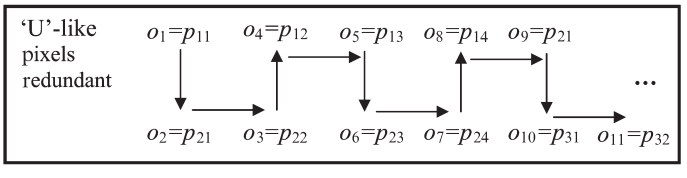
(*e*)



(*f*)



(*g*)



(*h*)

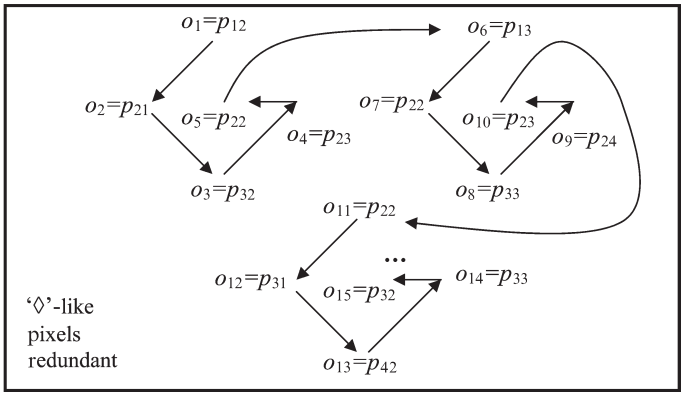


Figure 2 (Continued).

according to the sample image as shown in figure 2(*a*), this method arranges pixels in terms of a 'V' shape sequence as setting $o_1=p_{1,1}$, $o_2=p_{2,1}$, $o_3=p_{1,2}$, $o_4=p_{2,2}$, … to form a hidden Markovian chain. Figure 2(*d*) presents the illustration to this proposed pixel fitting method. After pixel arrangement within row 1 and 2 is completed, one then switches to rows 3 and 4, and repeats the process until all the rows within an image have been visited. In such a way, all the pixels are uniquely introduced to the HMM, and we term this approach pixel non-redundant. The resulting HMM will be of Markov chain $I \times J$ in length.

- **Pixel non-redundant 'U'-like observation sequencing (figure 2(*e*)).** Here, the pixel fitting process follows the 'U'-shaped scan by setting observation $o_1=p_{1,1}$, $o_2=p_{2,1}$, $o_3=p_{2,2}$, $o_4=p_{1,2}$, $o_5=p_{1,3}$, … to form a hidden Markovian chain as shown in figure 2(*e*). After pixel arrangement within rows 1 and 2 are completed, one then switches to rows 3 and 4, and repeats the process until all the rows within an image have been visited. All the pixels are uniquely introduced to the HMM. The resulting length of the Markov chain is again $I \times J$.
- **Pixel redundant 'V'-like observation sequencing (figure 2(*f*)).** This approach is similar to the non-redundant 'V'-like method introduced above. Here, however, after processing on row i and $i+1$ are completed, the process is switched to row $i+1$ and $i+2$, and the same fitting process is repeated. For instance, in the case of $i=1$, after pixels in rows 1 and 2 have been incorporated into the HMM, the fitting process is then switched to rows 2 and 3. The pixels incorporated into an HMM will be redundantly included, and the process is therefore called pixel redundant. This kind of pixel sequence arrangement (and with the other pixel redundant schemes) is intended to increase the length of the Markov chain so as to pursue a more valid parameter estimate within an HMM (Rabiner 1989). The resulting Markov chain will be of $2(I-1)J$ in length. It should be noted that, as in this 'V'-like and later 'U'-like pixel redundant schemes, pixels may present more than once within the HMM. It could result in one pixel holding various hidden states (i.e. belonging to different clusters). In order to resolve such confusion, for each row pair, i and $i+1$, we will treat row i as the primary row and row $i+1$ as the secondary row. Only hidden states obtained from pixels within primary rows are recognized. The reason is that, based on the spirit of our methodology design, the secondary rows are just performing in the role of assistance for taking contextual information into account.
- **Pixel redundant 'U'-like observation sequencing (figure 2(*g*)).** This type of sequencing is similar to the pixel non-redundant 'U'-like observation sequencing, except that, after processing on rows i and $i+1$ are completed, the process is switched to rows $i+1$ and $i+2$, and the same fitting process is repeated. The resulting length of Markov chain is $2(I-1)J$. As addressed above, only hidden states belong to primary rows are concerned.
- **'◇'-like observation sequence arrangement (figure 2(*h*)).** This '◇'-like observation sequence arrangement attempts to incorporate all the first-order neighbours of a pixel of interest into consideration. For a pixel $p_{i,j}$ of interest, it starts from the upper neighbouring pixel by setting $o_t=p_{i-1,j}$, then the left neighbouring pixel $o_{t+1}=p_{i,j-1}$, followed by lower neighbouring pixel $o_{t+2}=p_{i+1,j}$ and right neighbouring pixel $o_{t+3}=p_{i,j+1}$, and finally set $o_{t+4}=p_{i,j}$. The process

then resumes from pixel $o_{t+5}=p_{i-1,j+1}$ and repeats the '◇'-like tracking process to $o_{t+9}=p_{i,j+1}$. Figure 2(*h*) shows such pixel fitting process. The resulting Markov chain will have a length of $5(I-2)(J-2)$. Please note that, for this '◇'-like observation sequence, only hidden states obtained at the locations where the observations counts modulo 5 is zero (i.e. q_5, q_{10}, q_{15}, q_{20}, …) are recognized, since according to such observations fitting methodology design, the other observations are just playing assistance role for bringing contextual information into account.

Once an HMM is formed and the number of states (i.e. clusters) are chosen, the hidden states sequence is traced. Each pixel is then assigned to its corresponding state. Figure 3 shows corresponding examples for pixels state assignments regarding the pixel redundant 'V'-like, 'U'-like, and '◇'-like fitting approaches. The recognized states are shown in bold and marked with '*' sign.

It can be seen that, for the methods described above, we alternatively arrange the information relating to vertical and diagonal contextual dependencies to build up an HMM. Although the HMM is still one-dimensional, the information contained inside is, in effect, 2D. One may treat such HMMs as a kind of pseudo 2D HMM.

3.2 *Observation-density-based methods*

As the descriptions in previous sections have shown, in addition to the observation sequence, there is another parameter, $b_i(o_t)$, i.e. the density of the observations that can bring the effect of context to the hidden states estimation. Normally, in constructing an HMM for image classification in terms of strip-like arrangement as

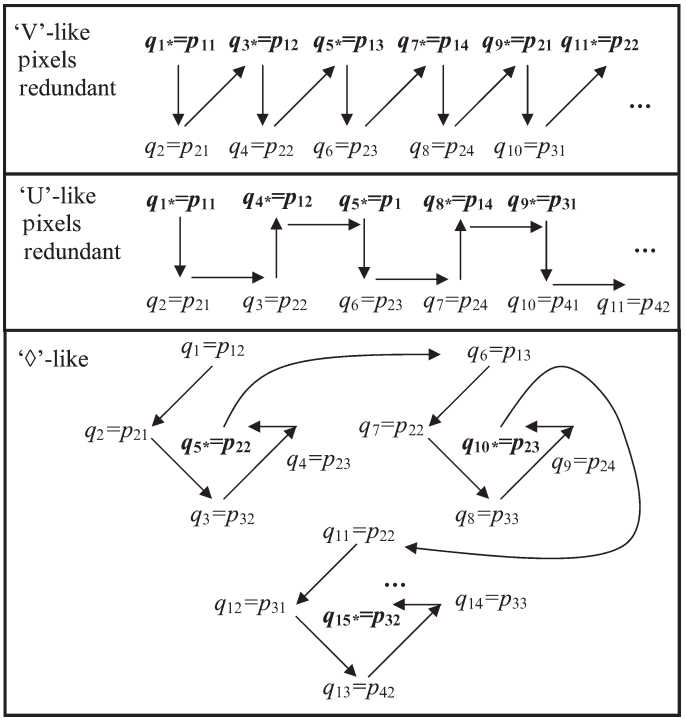


Figure 3. Illustrations of the selection of recognized states. See text for details.

shown, the observation at each step of HMM is only based on the single pixel alone. The estimated state (i.e. cluster) for a pixel of interest is thus highly correlated to the pixel value observed and the spatial direction on which we use HMM to model the contextual dependencies, respectively. Let the spatial direction applied to build up an HMM remain as strip-like in direction. If, to each original observation o_{t0}, two more observations o_{t1} and o_{t2} are added to expand the observation vector into $\{o_{t0}, o_{t1}, o_{t2}\}$, the estimated hidden state is likely to be affected dependent on how the new density being measured in terms of the observation $\{o_{t0}, o_{t1}, o_{t2}\}$. This triggers the following consideration.

If, to each observation within an HMM, one extends the scope of the observation in terms of combining the pixel and its neighbouring pixels to form a new observation vector, the contextual information can thus be incorporated according to the neighbouring directions being involved. This newly formed contextual information will affect the hidden states estimation.

Following such a consideration, to build up an HMM, we maintain the pixel fitting direction aligned to the row direction (i.e. strip-like), but for each pixel of interest, the vertical neighbouring pixel(s) are added in order to take the vertical spatial dependencies into account. We have designed the HMMs with the observation vectors formed in terms of one-side neighbour and two-side neighbours, respectively (figure 4). The experiments are conducted and evaluated based on such observation density adjustment schemes.

4. Experimental results and discussion

4.1 *Study area*

The study area known as Elkhorn Slough is located in the central California coast about 160 km south of San Francisco, California (Silberstein and Campbell 1989). The IKONOS satellite captured the study imagery for this scene on 23 October 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. More particularly, if the developed unsupervised classification schemes can achieve higher accuracy in comparison to the k-means method, it will be a benefit to facilitate the monitoring of the whole area. For experimental purposes, a test area with 1024×1024 pixels was extracted from the IKONOS multispectral imagery (containing 10,000 samples by 20 000 lines), and further re-sampled to 256×256 pixels so as to facilitate the analysis of the classification methodology. Each classified image was then evaluated in terms of its visual quality and corresponding confusion matrix.

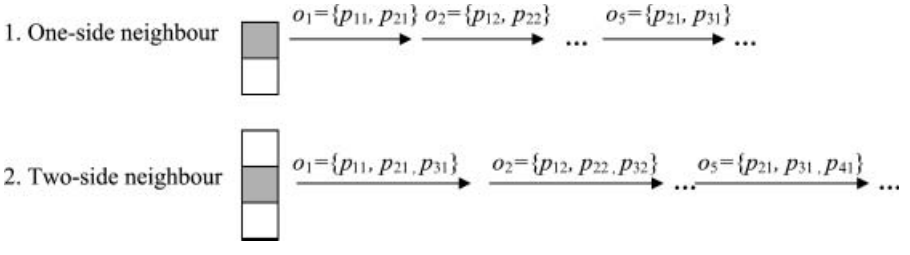


Figure 4. Observation-density-based manipulations according to figure 2(a).

Figure 5 displays the test area in the form of false colour (IR) composite and 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 ploughed in preparation for planting. The planted fields have colour spectra similar to that of the grassland and forest area in the lower portion of the scene and along the river marsh. The region also contains some areas with plastic cover, as part of the field preparation process, or as part of the normal growing process.

Due to the fact that, when dealing with unsupervised real scene classification, each class may contain more than one cluster, we thus conducted the extensive pre-experimental k-means clustering experiments for detecting the relationship between the clusters and the ground data so as to determine the suitable number of clusters and the information classes, to facilitate further the classification accuracy comparisons among the different methods. After tests, it was found that choosing clusters of 10 were easily mapped into six information classes while accuracy was also be achieved in the higher level. The total number of 10 clusters and six information classes were then determined to serve the classification experiments. The information classes are shown in table 1. Those information classes were

Figure 5. Study imagery of the Elkhorn Slough area in terms of false colour composite of IKONOS multispectral data.

Table 1. Selected information classes for unsupervised classification clustering mapping.

No.	Class name
1	Vegetation
2	Water
3	Marsh
4	Forest
5	Bare soil and living area
6	Dry grass and low sprout

determined based on the land cover/land use class 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.

For each information class, the ground data map was referred, and the test samples are then randomly drawn from the image (i.e. in the way of stratified random sampling) (Hammond and Verbyla 1996). Somehow, in order to reduce the bias in the accuracy analysis, the number of samples drawn in this study was related to, though not critically proportional to, the areal extent of each information class. For instance, the sample size was set to around 550–600 pixels for information class 1 to 3. Around 700 samples were obtained for information class 6 (dry grass & low sprout), due to its larger extent relative to the other classes. As information classes 4 and 5 both hold smaller area comparing to the rest classes, fewer samples (around 500 and 400, respectively) were drawn. In total 3442 ground data pixels were obtained for classification accuracy evaluation.

4.2 *Results and discussion*

The schemes for using HMMs to pursue higher classification accuracy in unsupervised sense as described above have been implemented for the test imagery. The HMM constructed by each method went through the parameter estimation process and then the state for each corresponding pixel was extracted to form classified imagery. As described earlier, a total of 10 states (clusters) were used for an HMM to conduct classification. For all the HMMs, the original model parameters, namely, A, B, and π, are randomly assigned as shown in Rabiner and Juang (1986) and Rabiner (1989). The iterations of the model parameters estimations are converged within six or seven runs. One exemplar graph of HMM training taken within our experiments is shown in figure 6. The hidden state (i.e. cluster) sequence Q^* is then estimated using dynamic programming (Rabiner 1989, Forney, 1973) so as to maximize $P(Q^*|O, \lambda)$. The resulting clusters are then mapped to the information classes according to the knowledge of ground data. It is worthwhile to note that, as the resulting classified patterns generated by HMM model are patch-like, the mapping from clusters to information classes can be quite

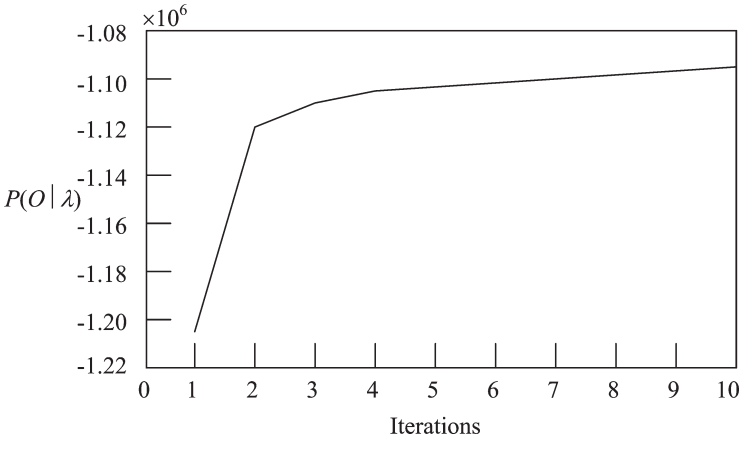


Figure 6. An exemplar HMM training graph.

straightforward according to the ground data map. However, in the case of k-means clustering, since the output clusters are normally in 'busy' look when mapped back to the thematic map, one has to conduct several trials to assign the resulting clusters to suitable information classes so as to make the overall classification accuracy as high as possible. In this study, for each method, the classification experiments were conducted five times. Among those, the results with higher classification accuracy were then chosen for evaluation.

The resulting images for unsupervised classifications using HMM based on the strip-like, 'V'-like, 'U'-like and '◇'-like observation sequence manipulation schemes are shown in figures 7(*a*) to (*f*). The corresponding confusion matrices are displayed in tables 2(*a*) to (*f*), respectively. By observing figure 7(*a*), it can be seen that the simple strip-like observation sequence results in clear horizontal spatial dependency patterns. In particular, on the right part of the image, the class known as 'water' area has been broadly swiped by the class 'marsh'. The upper part of the 'water' shape is hardly recognizable. Also within the image, some horizontal lines are quite apparent, lowering the quality of the classification result. The proposed observation sequence adjustments reveal improvement in both classification accuracy and visual quality. It can be seen from figures 7(*b*) to (*f*) that the classification patterns are more clearly in patch-like distributions. For those HMMs with pixel redundant fitting schemes, some confusion appears within the 'marsh' so as to make the area appear 'busy' in visual appearance (figures 7(*c*), (*e*) and (*f*)). The pixel non-redundant 'V'-like and 'U'-like schemes provide more patch-like patterns as shown in figures 7(*b*) and (*d*). One finds that the resulting 'marsh' area is much clearer than the output obtained by pixels redundant sequencing.

Confusion matrices reinforce the conclusions obtained by visual inspection. The matrices use output data as the rows and reference data as the columns. Here, in addition to the overall accuracy, both producer's accuracies (PA) and user's accuracies (UA) are also calculated to perform the omission and commission error measure. Other statistical indices such as kappa or tau index (Ma and Redmond 1995) can be used to support accuracy measurement. However, those indices are not of particular interest here. The simple strip-like approach with the HMM achieves 71.58% in overall accuracy. Among its corresponding PA and UA, the worst case can be found to be as low as 0.48 and 0.41, respectively. An overall accuracy of around 83% was achieved for the 'U'-like and '◇'-like sequencing approaches (tables 2(*d*) to (*f*)), around a 10% enhancement compared to the strip-like method. In addition, for the 'U'-like scheme, the PA obtained was at least 0.66 (for both pixel non-redundant and redundant cases), while UA achieved to at least 0.69 (for pixel non-redundant) and 0.72 (for pixel redundant case). The '◇'-like scheme holds at least the value of 0.72 in both PA and UA. Compared to the 'U'-like and '◇'-like schemes, the 'V'-like scheme performed less ideally, with 81% in overall accuracy for the pixel non-redundant case, with which the corresponding PA and UA can be as low as 0.63 (table 2(*b*)); an overall accuracy of 79% was achieved in the pixel redundant case and its corresponding PA can be as low as 0.62 and 0.67 for the UA.

For comparison purposes, results output by the traditional k-means clustering and Hilbert-Peano scan are presented in figures 7(*g*) and (*h*), and confusion matrices in tables 2(*g*) and (*h*). The image quality obtained by k-means clustering is poor, with no patch-like patterns formed (see figure 7(*h*)). This again demonstrates the drawback of failing to incorporate spatial dependencies information into classification scheme. An overall accuracy of only 58.57% was obtained with k-means

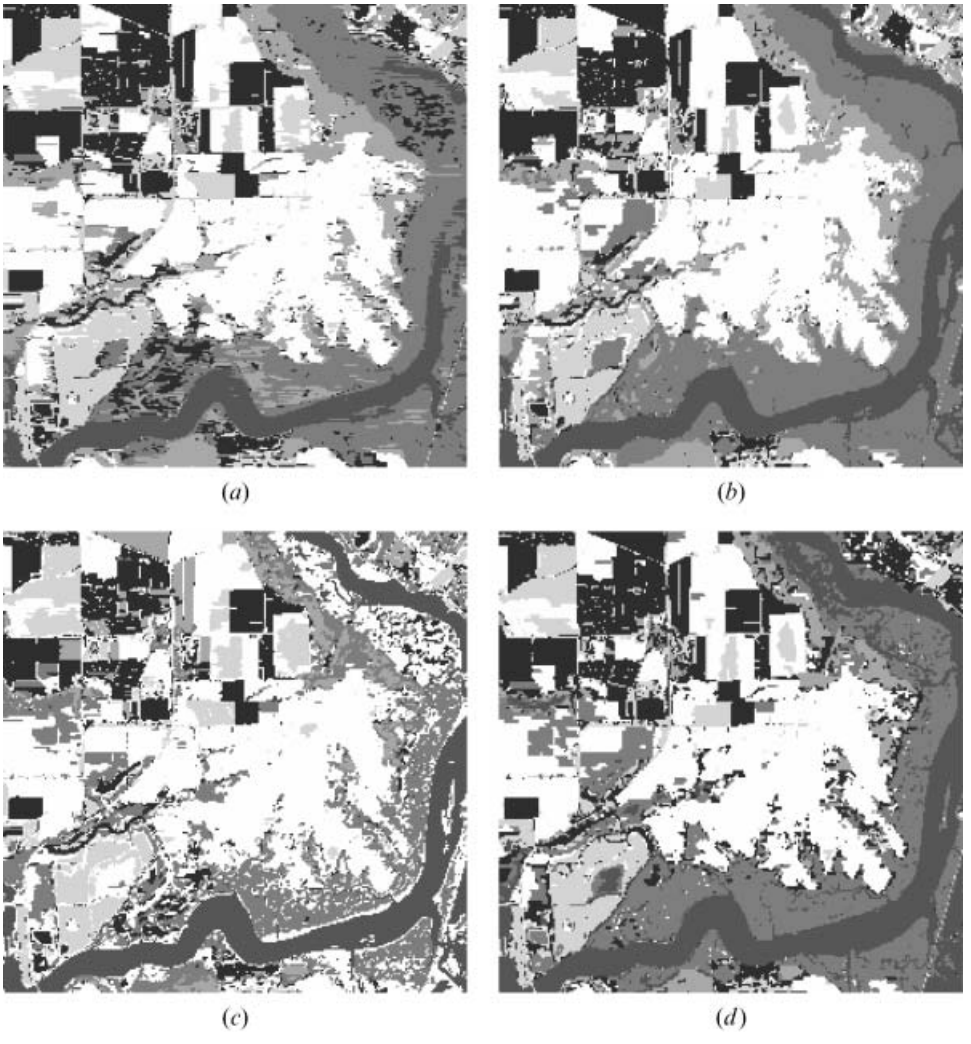


Figure 7. Classification images output by (*a*) HMM with strip-like observation fitting; (*b*) HMM with pixel non-redundant 'V'-like observation fitting; (*c*) HMM with pixel redundant 'V'-like observation fitting; (*d*) HMM with pixel non-redundant 'U'-like observation fitting; (*e*) HMM with pixel redundant 'U'-like observation fitting; (*f*) HMM with '◇'-like observation fitting; (*g*) HMM with Hilbert-Peano scan fitting; and (*h*) traditional unsupervised k-means clustering methods.

clustering. The HMM using Hilbert-Peano sequencing, in our experiments, achieves around 79% in overall accuracy (table 2(*h*)). Such a value is lower than the accuracy obtained by 'U'-like (for both pixel non-redundant and redundant cases), '◇'-like, and pixel non-redundant 'V'-like methods. In table 2(*h*), the corresponding PA and UA values, in the worst case, appear to be as low as 0.66 and 0.52, respectively, which are worse than the proposed U'-like and '◇'-like schemes.

The proposed observation density adjustments for HMM produced even better accuracy achievements compared to the methods described above. The image outputs by unsupervised classification using HMM based on observation density adjustments are shown in figures 8(*a*) and (*b*), and confusion matrices in tables 3(*a*)

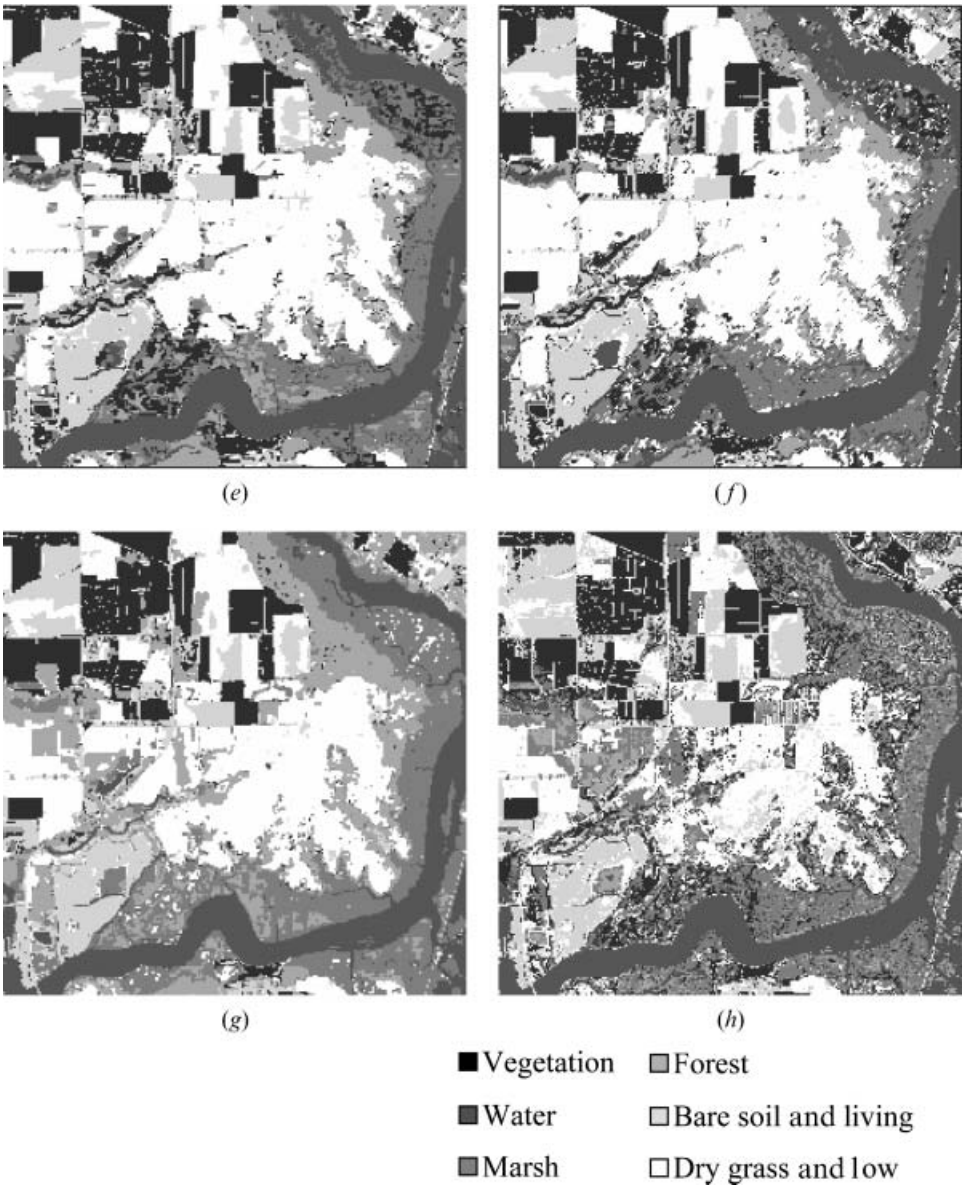


Figure 7 (Continued).

and (*b*), respectively. The overall classification accuracies for both one-neighbour and two-neighbour observation density methods are 88% and 87%. The PA and UA also achieved at least 0.77 for the one-neighbour case. In the two-neighbour scheme, the UA obtained a value of at least 0.77, while at least 0.74 was achieved among PA. Both methods reveal improvements in comparison to the k-mean clustering, Hilbert-Peano scan, and strip-like fitting methods, respectively. There is also an enhancement of about 10% of overall classification accuracy relative to the 'V'-like scheme, and 5% classification accuracy improvement relative to 'U'-like and '◇'-like schemes. The images generated by this method also reveal 'clean' (especially within class 'marsh' area) and patch-like visual appearance.

Table 2. Classification confusion matrices for observation sequence manipulations and traditional k-means clustering.

(*a*) Strip-like.

Class no.	1	2	3	4	5	6	UA
1	463	2	75	24	0	35	0.77
2	0	422	11	0	0	0	0.97
3	77	108	401	225	4	152	0.41
4	0	1	0	257	0	0	0.99
5	0	0	0	1	360	0	0.99
6	44	20	112	28	59	561	0.68
PA	0.79	0.76	0.67	0.48	0.85	0.75	0.7158

(*b*) Pixel non-redundant 'V'-like.

Class no.	1	2	3	4	5	6	UA
1	501	3	8	121	20	1	0.76
2	0	505	33	8	2	17	0.89
3	29	30	554	45	6	209	0.63
4	0	15	0	342	0	0	0.95
5	6	0	0	3	380	4	0.96
6	48	0	4	16	15	517	0.86
PA	0.85	0.91	0.92	0.63	0.89	0.69	0.8131

(*c*) Pixel redundant 'V'-like.

Class no.	1	2	3	4	5	6	UA
1	464	6	78	26	9	30	0.75
2	0	347	0	0	0	0	1
3	0	107	452	8	4	65	0.71
4	12	93	69	474	2	53	0.67
5	10	0	0	4	394	13	0.93
6	98	0	0	23	14	587	0.81
PA	0.79	0.62	0.75	0.88	0.93	0.78	0.7896

(*d*) Pixel non-redundant 'U'-like.

Class no.	1	2	3	4	5	6	UA
1	391	2	0	14	10	31	0.87
2	0	416	5	0	0	0	0.98
3	0	62	581	16	7	95	0.76
4	118	73	13	479	6	0	0.69
5	3	0	0	1	372	2	0.98
6	72	0	0	25	28	620	0.83
PA	0.66	0.75	0.96	0.89	0.87	0.82	0.8306

(*e*) Pixel redundant 'U'-like.

Class no.	1	2	3	4	5	6	UA
1	461	5	106	23	8	36	0.72
2	0	495	29	2	2	10	0.92
3	0	20	400	4	0	63	0.82
4	10	33	64	471	1	0	0.81
5	9	0	0	4	400	7	0.95
6	104	0	0	31	12	632	0.81
PA	0.78	0.89	0.66	0.88	0.94	0.84	0.8306

Table 2 (Continued).

(*f*) '◇'-like.

Class no.	1	2	3	4	5	6	UA
1	458	6	96	26	14	36	0.72
2	0	515	43	7	1	24	0.87
3	4	7	433	65	0	2	0.84
4	7	25	2	396	0	0	0.92
5	13	0	0	2	386	3	0.95
6	102	0	25	39	22	683	0.78
PA	0.78	0.93	0.72	0.74	0.91	0.91	0.8341

(*g*) k-means clustering.

Class no.	1	2	3	4	5	6	UA
1	350	13	114	242	9	32	0.46
2	0	491	32	23	2	11	0.87
3	26	18	374	176	10	160	0.48
4	0	31	56	72	2	133	0.24
5	143	0	8	4	342	25	0.65
6	65	0	15	18	58	387	0.71
PA	0.59	0.88	0.62	0.13	0.80	0.51	0.5857

(*h*) Hilbert-Peano scan.

Class no.	1	2	3	4	5	6	UA
1	450	0	0	12	0	0	0.97
2	0	424	7	0	2	27	0.92
3	8	108	471	24	11	60	0.69
4	95	19	103	484	9	153	0.52
5	9	0	0	4	391	7	0.95
6	22	2	18	11	10	501	0.88
PA	0.77	0.76	0.78	0.90	0.92	0.66	0.7905

Through the above demonstration of the proposed methods, namely observation-sequence-based and observation-density based approaches, higher classification accuracies are obtained. The accuracy and the output imagery performed by each method have been shown. The proposed schemes for incorporating neighbouring pixels into an HMM unsupervised classification process invoke a more conceptual 2D sense, in contrast to attempting to modify the linear HMM structure into 2D style (or so-called Markov meshes). The latter approach complicates the computational algorithm and makes the search for optimum states more bothersome.

One final issue worth our concern is the computational loading related to an HMM. So long as the calculation for the observation density $b_i(o_t)$ remains a constant, Rabiner (1989) showed that the computation involved in an HMM is of order N^2T where N denotes the number of states and T the length of an HMM. The main elements here are the forward and backward calculations for α and β as shown in equations (11) to (14). More precisely, for either forward or backward calculation, it requires $N(N+1)(T-1)+N$ multiplications and $N(N-1)(T-1))$ additions. Table 4 shows the computational loading (running time in minutes calculated by seven training iterations) for the proposed methodology, written in MATLAB 6.5 running on a PC with Intel Pentium III CPU.

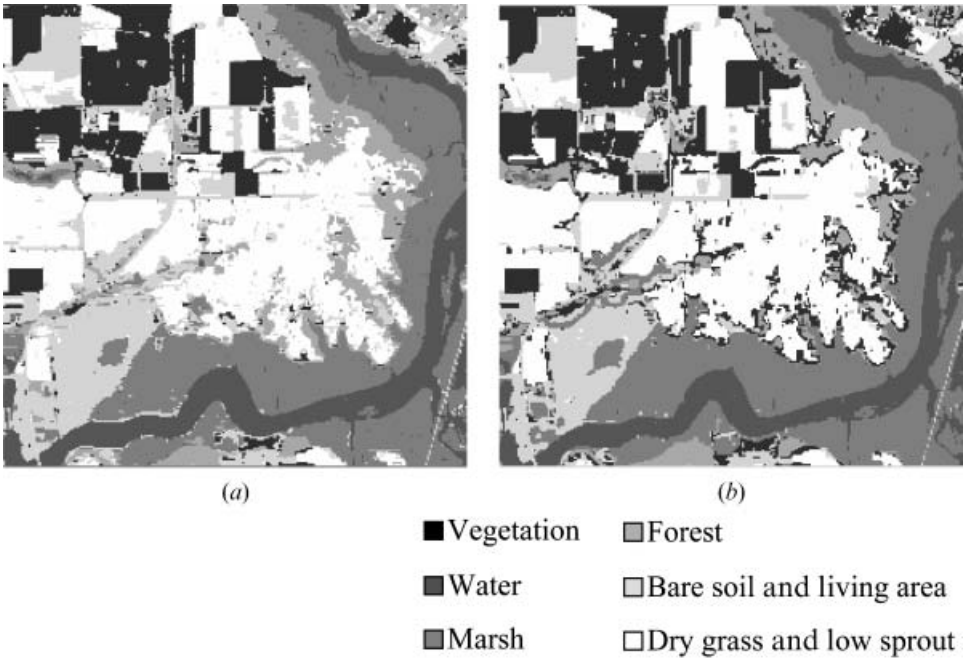


Figure 8. Classification images output by (a) HMM using observation-density based method with one-neighbour added; and (b) HMM using observation-density based method with two-neighbour added.

If one would like to measure the performance of the HMMs based on the overall accuracy alone, it is determined that the proposed one-neighbour observation-density method is relatively ideal scheme (a classification accuracy of 88% was

Table 3. Classification confusion matrices for observation density manipulations.
(*a*) One-side neighbourhood.

Class no.	1	2	3	4	5	6	UA
1	475	0	0	5	0	0	0.98
2	0	427	1	0	0	0	0.99
3	0	103	598	69	0	1	0.77
4	20	23	0	435	0	0	0.91
5	4	0	0	8	421	71	0.83
6	85	0	0	18	2	676	0.86
PA	0.81	0.77	0.99	0.81	0.99	0.90	0.8808

(*b*) Two-side neighbourhood.

Class no.	1	2	3	4	5	6	UA
1	487	0	0	65	0	8	0.86
2	0	410	0	0	0	0	1
3	0	95	599	40	1	38	0.77
4	3	48	0	412	0	0	0.88
5	1	0	0	3	419	28	0.92
6	93	0	0	15	3	674	0.85
PA	0.83	0.74	1	0.77	0.99	0.90	0.8718

Table 4. The running time resulting from different pixel arrangement schemes with HMM.

No.	Pixel fitting scheme	Computational time approximated in minutes
1	Strip-like	8
2	Hilbert-Peano sequencing	8
3	Pixels non-redundant 'V'-like	8
4	Pixels redundant 'V'-like	19
5	Pixels non-redundant 'U'-like	8
6	Pixels redundant 'U'-like	19
7	'◇'-like	26
8	Observation-density-based method One-side neighbour	14
9	Observation-density-based method two-side neighbour	17

*Please note that the running time is obtained by using MATLAB 6.5 program running on Intel Pentium III CPU with training iterations set to 7.

achieved and the method also results in better visual quality) to be applied to remotely sensed imagery classification. However, when the computational loading specified in table 4 is also taken into consideration, where the dimension of observation vector is large, one may make some compromise, and choose the pixel non-redundant 'U'-like HMM as an alternative.

5. Conclusions

New schemes of incorporating 2D spatial information into a one-dimensional linear HMM have been proposed and demonstrated in terms of accuracy analysis and visual quality through unsupervised classifications to the remotely sensed imagery. The traditional k-means clustering approach using spectral data alone is not sufficient so long as higher classification accuracy and clear visual interpretation are pursued. However, by adopting a one-dimensional linear HMM accompanied with proposed observation-sequence-based and observation-density-based methods, both spectral and 2D spatial information can be combined. The results as shown in this study have revealed success in achieving higher accuracy and more patch-like patterns compared to spectrally k-mean clustering and, to a certain extent, are better than another common observation sequence arrangement known as Hilbert-Peano sequencing. The proposed approaches may simplify the efforts of quoting 2D spatial information in the case of applying an HMM to scene classification in remotely sensed imagery. Among the proposed approaches, it is found that the observation density based methods perform better as far as the classification accuracy is concerned. However, one may choose pixel non-redundant 'U'-like sequencing as an alternative if computational burden is also of importance. Future works may focus on the issue of incorporating hyper-spectral data and including multiscale information into HMM.

Acknowledgements
The authors would like to convey their gratitude to the Remote Sensing Laboratory, Naval Postgraduate School, USA, for providing facilities in supporting this research project. Specific gratitude is sent to the reviewers and editorial for their valuable comments to this work.

References

Abend, K., Jarley, T.J. and Kanal, L.N., 1965, Classification of binary random patterns. *IEEE Transactions on Information Theory*, **11**, pp. 538–544.

Baum, L.E. and Egon, J.A., 1967, An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model for ecology. *Bulletin of the American Meteorological Society*, **73**, pp. 360–363.

Baum, L.E. and Petrie, T., 1966, Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*, **37**, pp. 1554–1563.

Baum, L.E., Petrie, T., Soules, G. and Weiss, N., 1970, A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, **41**, pp. 164–171.

Cole, R., Hirschman, L., Atlas, L. and Beckman, M., 1995, The challenge of spoken language systems: Research directions for the nineties. *IEEE Transactions on Speech Audio Processing*, **3**, pp. 1–21.

Devijver, P.A., 1985, Probabilistic labeling in a hidden second order Markov mesh. *Pattern Recognition in Practice*, **II**, pp. 113–123.

Fjørtoft, R., Delignon, Y., Pieczynski, W., Sigelle, M. and Tupin, F., 2003, Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, pp. 675–686.

Forney, G.D., 1973, The Viterbi algorithm. *Proceedings of IEEE*, **61**, pp. 263–278.

Gader, P., Mystkowski, M. and Zhao, Y., 2001, Landmine detection with ground penetrating radar using hidden Markov models. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, pp. 1231–1244.

Giordana, N. and Pieczynski, W., 1997, Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**, pp. 465–475.

Hammond, T.O. and Verbyla, D.L., 1996, Optimistic bias in classification accuracy assessment. *International Journal of Remote Sensing*, **17**, pp. 1261–1266.

He, Y. and Kundu, A., 1991, 2-D shape classification using hidden Markov model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**, pp. 1172–1184.

Huang, X.D., Ariki, Y. and Jack, M.A., 1990, *Hidden Markov Models for Speech Recognition* (Edinburgh, UK: Edinburgh University Press).

Juang, B.H., 1985, Maximum likelihood estimation for mixture multivariate stochastic observartions of Markov chains. *AT&T Technical Journal*, **64**, pp. 1235–1249.

Li, J., Najmi, A. and Gray, R.M., 2000a, Image classification by a two-dimensional hidden Markov Model. *IEEE Transactions on Signal Processing*, **48**, pp. 517–533.

Li, J., Gray, R.M. and Olshen, R.A., 2000b, Multiresolution image classification by hierarchical modeling with two-dimensional hidden Markov models. *IEEE Transactions on Information Theory*, **46**, pp. 1826–1841.

Linde, Y., Buzo, A. and Gray, R.M., 1980, An algorithm for vector quantizer design. *IEEE Transactions on Communications*, **COM-28**, pp. 84–95.

Liporace, L.A., 1982, Maximum likelihood estimation for multivariate observations of Markov sources. *IEEE Transactions on Information Theory*, **IT-28**, pp. 728–734.

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

Mather, P.M., 1999, *Computer Processing of Remotely Sensed Images: An Introduction*, 2nd edn (Chichester, UK: John Wiley & Sons).

Rabiner, L.R., 1989, A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of IEEE*, **77**, pp. 257–286.

Rabiner, L.R. and Juang, B.H., 1986, An introduction to hidden Markov models. *IEEE ASSP Magazine*, **3**, pp. 4–16.

RUNKLE, P., NGUYEN, L.H., MCCLELLAN, J.H. and VARIN, L., 2001, Multi-aspect target detection for SAR imagery using hidden Markov models. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, pp. 46–55.

SILBERSTEIN, M. and CAMPBELL, E., 1989, *Elkhorn Slough* (CA: Monterey Bay Aquarium Foundation).

SKARBEK, W., 1992, Generalized Hilbert scan in image printing. In *Theoretical Foundations of Computer Vision*, R. Klette and W.G. Kropetsh (Eds), pp. 433–454 (Berlin Germany: Akademie Berlag).

TSO, B. and 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**, pp. 1255–1260.

TSO, B. and MATHER, P.M., 2001, *Classification Methods for Remotely Sensed Data* (London: Taylor and Francis).

VIOVY, N. and SAINT, G., 1994, Hidden Markov models applied to vegetation dynamics analysis using satellite remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, **32**, pp. 906–917.