Abstract: A new multispectral image context classification, which is based on a recursive algorithm for optimal estimation of the state of a two-dimensional discrete Markov Random Field, is presented. The implementation of the recursive algorithm is a form of dynamic programming. Finally, experimental results with remote sensed multispectral scanner data using the recursive context classification are presented and contrasted with results from context free classification.

(1.0) Introduction

Because of the high demand for classification efficiency and accuracy in remote sensing applications, making full use of spectral and spatial context in MSS (Multispectral Scanner) or TM (Thematic Mapper) data is very important. In recent years, this realization is becoming increasingly prevalent, and progress is being made. A spatial stochastic recursive contextual classification was proposed by Yu and Fu [4]; an estimation method of the context function was discussed by Tilton, Vardeman and P.H. Swain [5]; Haralick [1] gave a survey of decision making in context; and Gurney and Townshend [3] used a contextual model based on frequency distributions in the classification of Remotely sensed data.

The conventional automatic classification techniques for remote sensing data classify each pixel (picture element) independently. This type of classification can only exploit spectral or, in some cases, spectral and temporal information. There is no provision for using coherent spatial information in the classification of MSS or TM data. In order to improve classification accuracy, it is natural therefore to use information not only from an individual pixel but also from its neighbors within the image. Such spatial information is usefully subdivided into two types: textural, which is simply spatial variation in tone (spectral response) over relatively small areas, and contextual, which is spatial coherence within the image. This paper presents a new approach to multispectral image context classification based on a recursive algorithm for optimal estimation of the state of a two-dimensional discrete Markov Random Field. The first task of this approach is to define the two-dimensional discrete Markov Random Field model. A recursive algorithm is then obtained from local measurements and from the noise independence assumption of the Markov Random Field model, and finally probabilities for each category are estimated.

Implementation of the recursive algorithm is a form of dynamic programming. A two pass and four pass algorithm are discussed in this paper. Because the estimation equations of the recursive algorithm are quite simple, the computation complexity of the approach is low. We show that recursive contextual classifications can improve classification performance, as compared to noncontextual classification. In addition, this algorithm has the advantage over other techniques in that it handles multiple observation data naturally and simultaneously.

(2.0) Statement of problem

Consistent with the two-dimensional (2-D) discrete Markov random field for multispectral image processing applications we assume a random observation vector \( X_i \), in which each component is a two dimensional \( I \times J \) array, and each \( I \times J \) array corresponds to one band of \( I \times J \) pixels in MSS image.

Unlike 1-D discrete time series, where the existence of a preferred direction is inherently assumed, no such preferred ordering of the discrete lattice is appropriate. In other words, the notion of "past" and "future" as understood in a unilateral 1-D Markov processes is restrictive in 2-D as it implies a particular ordering in which the observations are scanned down top and left to right. It is quite possible that an observation at a pixel \( p \) may be dependent on surrounding observations in all directions. More important dependence in 2-D Markov field is nearest neighbors, which are defined as the nearest north, south, east and west neighbors. The Markov random field models may be defined as below: Let

\[ \{ x(s), s \in \mathcal{R} \}, r = \{ s = (i,j); f \leq i,j \leq M \} \]

be an observation from an image, where \( M \) designates row and column size of an image; \( (i,j) \) designates a 2-D image position. It is postulated that this data is generated by an appropriate 2-D (non causal) Markov Random Field model. The models characterize the statistical dependency among pixels by requiring that \( P(X(r)|allX(r), r \neq s) = P(X(s)|allX(s+r), r \in N) \) Where \( N \) is the appropriate symmetric neighbor set. For instance \( N = \{ (0,1), (0,-1), (-1,0), (1,0) \} \) corresponds to take simplest Markov model and by including more neighbors we can construct higher order of Markov model. Since the model is
defined only for symmetric neighbor sets, often \( N \) is equivalently characterized using an asymmetrical neighbor set \( N^* \); i.e. if \( r \in N \) then \(-r \notin N^* \) and \( N = \{ r = -r \mid r \in N \} \). The nearest neighbors will be used in this paper. Let the observation vector \( \mathbf{X}_i \) have fixed but unknown classification \( Q_i \), as shown in Figure 1. The digital multiband image is divided up into discrete, recognizable, measurable pieces which we call units (i.e., pixels). Let \( d \) be an observed measurement vector from a unit whose true but unknown category label is \( C \). Assignment of the best category label \( C \) to the unit is based on the conditional probability \( P(C|d) \) when using pixel independent processing. In contrast, context dependent processing does not use only the locally observed \( d \) in order to assign the best category label to each pixel, but rather is based upon all of the observed data from all units to decide the best label. Letting \( D \) be the collection of all observed measurements, context dependent processing makes the assignment of the best category label based on the conditional probability \( P(C|D) \).

In the next section we will show how to compute \( P(C|D) \) using recursive neighborhood operators under conditions of a Markov random field.

(3.6) Recursive Algorithm Deviation

The multiband image considered in this paper posses a finite gray tone value \( D = \{0, \ldots, 255\} \) and is defined on the two-dimensional integer of finite size \( I \times J \). Before presenting this algorithm, we must first give the notation and assumptions for the two-dimensional Markov Field under which the algorithm can be derived.

(3.1) Notation:

(i, j) = designates a position (i-row, j-column)

\( d_{ij} \) = an observed measurement vector from pixel \( i, j \)

\( C_{ij} \) = an assigned category label \( C \) to pixel \( i, j \)

\( D_{ij} \) = the collection of all observed measurements from the set of \( I \times J \) pixels.

\( D_{ij}^1 \) = set of all observed data values \( d_{i,k} \) for \( (i,k) \) to the left or above \( (i,j) \), including \( (i,j) \)

\( E_{ij}^1 \) = set of all observed data values \( d_{i,k} \) for \( (i,k) \) to the right or below \( (i,j) \), including \( (i,j) \)

\( D_{ij}^2 \) = set of all observed data value \( d_{i,k} \) for \( (i,k) \) to the left or above \( (i,j) \), excluding \( (i,j) \)

\( E_{ij}^2 \) = set of all observed data value \( d_{i,k} \) for \( (i,k) \) to the right or below \( (i,j) \), excluding \( (i,j) \)

See figures 1 and 2 for the geometric picture of the above sets.

(3.2) assumption:

a) We assume that an n-tuple of measurements is determined by some local measuring process. So give the true category of a pixel all measurements of units are independent of each other. In effect this says that given the true state of affairs, we observed measurement variations are independent.

\[
P(D_{ij}|C_{ij}) = P(D_{ij}^1|C_{ij})P(E_{ij}^1|C_{ij})P(d_{ij}|C_{ij})
\]

\[
P(D_{ij}^1|C_{ij}) = P(D_{ij}^1|C_{ij})P(d_{ij}^1|C_{ij})
\]

\[
P(E_{ij}^1|C_{ij}) = P(E_{ij}^1|C_{ij})P(d_{ij}^1|C_{ij})
\]

b) the n-tuple measurement of unit \( i \) depends only upon the true interpretation associated with unit \( i \) and does not depend upon any relationship unit \( i \) may have with other units or upon the interpretation associated with any other unit.

\[
P(d_{ij}|C_{ij}, D_{ij}^1) = P(d_{ij}|C_{ij})
\]

\[
P(d_{ij+1}|C_{ij+1}, C_{ij}, E_{ij+1}^1) = P(d_{ij+1}|C_{ij+1})
\]

c) given neighboring categories, the observed data for pixels other than the current pixel tell nothing more about the current pixel's category.

\[
P(C_{ij}|C_{ij-1}, C_{ij-1}, E_{ij-1}^1) = P(C_{ij}|C_{ij-1}, C_{ij-1})
\]

\[
P(C_{ij+1}|C_{ij+1}, E_{ij+1}^1) = P(C_{ij+1}|E_{ij+1}^1)
\]

d) measurements of the nearest neighboring pixels tell everything that the surrounding categories (i.e. not nearest neighboring pixel) tells.

\[
P(C_{ij-1}|C_{ij-1}, E_{ij-1}^1) = P(C_{ij-1}|D_{ij-1}^1)
\]

\[
P(C_{ij+1}|C_{ij+1}, E_{ij+1}^1) = P(C_{ij+1}|E_{ij+1}^1)
\]

e) approximation: (omitting row data values should not make much of a difference)

\[
P(C_{ij-1}|D_{ij-1}^1) = P(C_{ij-1}|D_{ij-1}^1)
\]

\[
P(C_{ij+1}|E_{ij+1}^1) = P(C_{ij+1}|E_{ij+1}^1)
\]

e) A weaker assumption than independence of neighboring categories is the Markov Random Field assumption, which is used when the true interpretation of any unit given the true interpretations of all the surrounding depends only upon the interpretation of the nearest neighboring pixel.

\[
P(C_{ij-1}, C_{ij-1}|C_{ij}) = P(C_{ij-1}|C_{ij})P(C_{ij-1}|C_{ij})
\]

\[
P(C_{ij+1}, C_{ij+1}|C_{ij}) = P(C_{ij+1}|C_{ij})P(C_{ij+1}|C_{ij})
\]

(3.3) implementation of the recursive algorithm:

The recursive algorithm determines the Bayes labeling \( C_1, C_2, \ldots, C_M \) under Bayes decision rule and Markov field assumption by determining the maximizing \( C_1, C_2, \ldots, C_M \) of \( P(C_{ij}|D_{ij}) \). Both two pass and four pass algorithms are discussed in this section.

a) two pass algorithm:

To solve the conditional probability \( P(C_{ij}|D_{ij}) \) we can argue as follows.

From the definition of conditional probability:
\[ P(C_{ij}|D_{ij}) = \frac{P(D_{ij}|C_{ij})P(C_{ij})}{P(D_{ij})} \]

Then,

\[
P(C_{ij}|D_{ij}) = \frac{P(D_{ij}^j, E_{ij}^j, d_{ij}|C_{ij})P(C_{ij})}{P(D_{ij})} \]
\[
= \frac{P(D_{ij}^j|C_{ij})P(E_{ij}^j|C_{ij})P(d_{ij}|C_{ij})P(C_{ij})}{P(D_{ij})} \]
\[
= \frac{P(C_{ij})[P(D_{ij}^j|C_{ij})P(d_{ij}|C_{ij})][P(E_{ij}^j|C_{ij})P(d_{ij}|C_{ij})]}{P(D_{ij})} \]
\[
= \frac{P(D_{ij}^j|C_{ij})P(E_{ij}^j|C_{ij})P(d_{ij}|C_{ij})}{P(D_{ij})} \]
\[
= \frac{P(C_{ij}|D_{ij}^j)P(D_{ij}^j|E_{ij}^j)/P(C_{ij})P(d_{ij}|C_{ij})}{P(D_{ij})} \]
\[
= \frac{P(C_{ij}|D_{ij}^j)P(D_{ij}^j|E_{ij}^j)/(P(C_{ij})P(d_{ij}|C_{ij}))}{P(D_{ij})} \] ...

(1.1)

In equation (1.1) probability distributions \( P(C_{ij}) \) and \( P(d_{ij}|C_{ij}) \) are either known a priori, or estimated in the case of supervised remote sensing classification, and the denominator is some constant not dependent upon the true category. Then to solve (1.1) the problem becomes to calculate: \( P(C_{ij}|D_{ij}^j) \) and \( P(C_{ij}|E_{ij}^j) \).

\[
P(C_{ij}|D_{ij}^j) = P(C_{ij}, D_{ij}^{j-1}, d_{ij}) \]
\[
= \frac{P(C_{ij}, d_{ij}|D_{ij}^{j-1})P(D_{ij}^{j-1})}{P(D_{ij}^j)} \]
\[
= \frac{P(d_{ij}|C_{ij}, D_{ij}^{j-1})P(C_{ij}, D_{ij}^{j-1})}{P(D_{ij}^j)} \]
\[
= \frac{P(d_{ij}|C_{ij})P(C_{ij}, D_{ij}^{j-1})}{P(D_{ij}^j)} \] (by assumption 3)
\[
= \frac{P(d_{ij}|C_{ij}) \left( \sum_{C_{i-1,j-1}} \sum_{C_{i-1,j}} P(C_{i-1,j-1}, C_{i-1,j}, C_{ij}, D_{ij}^{j-1}) \right)}{P(D_{ij}^j)} \]
\[
= \frac{P(d_{ij}|C_{ij}) \left( \sum_{C_{i-1,j-1}} \sum_{C_{i-1,j}} P(C_{i-1,j-1}, C_{i-1,j}, D_{ij}^{j-1})P(C_{i-1,j-1}|C_{ij}, C_{i-1,j}, D_{ij}^{j-1}) \right)}{P(D_{ij}^j)} \]
\[
= \frac{P(d_{ij}|C_{ij}) \left( \sum_{C_{i-1,j-1}} \sum_{C_{i-1,j}} \left[ P(C_{i-1,j-1}, C_{i-1,j}, D_{ij}^{j-1}) \right] \right)}{P(D_{ij}^j)} \] (by assumption 4)
\[
= \frac{P(d_{ij}|C_{ij}) \left( \sum_{C_{i-1,j-1}} \sum_{C_{i-1,j}} \left[ P(C_{i-1,j-1}, C_{i-1,j}, D_{ij}^{j-1}) \right] \right)}{P(D_{ij}^j)} \]
\[
= \frac{P(d_{ij}|C_{ij}) \left( \sum_{C_{i-1,j-1}} \sum_{C_{i-1,j}} \left[ P(C_{i-1,j-1}, C_{i-1,j}, D_{ij}^{j-1}) \right] \right)}{P(D_{ij}^j)} \]
\[
= \frac{P(C_{ij}|D_{ij}^{j-1}, d_{i-1,j-t}, \ldots, d_{i-1,j-1}, P(D_{ij}^{j-1}))}{P(D_{ij}^j)} \] Using approximation (5),

\[
= \frac{P(C_{ij}|D_{ij}^{j-1})}{P(D_{ij}^j)/P(D_{ij}^{j-1})} \]
\[
= \frac{P(C_{ij}|D_{ij}^{j-1})}{P(D_{ij}^j)/P(D_{ij}^{j-1})} \]

... (1.2)
so finally we have the formula (1.3)

\[ P(C_{i,j}|D_{i,j}) = \frac{P(d_{i,j}|C_{i,j}) \sum_{C_{i-1,j}} P(C_{i-1,j}|C_{i-1,i}) P(C_{i-1}|D_{i,j}^{-1}) P(C_{i-1}|D_{i,j}^{-1})}{\Delta} \]

\[ = P(d_{i,j}|C_{i,j}) \sum_{C_{i-1,j}} P(C_{i-1,j}|C_{i-1,i}) P(C_{i-1}|D_{i,j}^{-1}) P(C_{i-1}|D_{i,j}^{-1}) \]

where \( \Delta = \sum_{k_{i,j}} P(d_{i,j}|k_{i,j}) \sum_{k_{i-1,j}} P(k_{i,j}|k_{i-1,j}) P(k_{i-1}|D_{i,j}^{-1}) P(k_{i-1}|D_{i,j}^{-1}) \)

which is a normalizing constant, not dependent on the true category.

(1.3)

The recursive formula (1.3) is from top-left to bottom-right. Similarly, if scanning from bottom-right to top-left, we also have a recursive formula (1.4):

\[ P(C_{i,j}|E_{i,j}^{-1}) = \frac{P(d_{i,j}|C_{i,j}) \sum_{C_{i+1,j}} P(C_{i+1,j}|C_{i+1,i}) P(C_{i+1}|E_{i,j}^{-1}) P(C_{i+1}|E_{i,j}^{-1})}{\Delta} \]

(1.4)

In (1.3) and (1.4) \( P(d_{i,j}|C_{i,j}) \) is the same as in (1.1), and the denominator \( \Delta \) is some constant not depending on the true category. In fact, it is a normalizing constant, which makes \( \sum_{C_{i,j}} P(C_{i,j}|E_{i,j}^{-1}) = 1 \). The transition probabilities \( P(C_{i,j}|C_{i-1,j}, C_{i-1,i}), P(C_{i,j}|C_{i+1,j}, C_{i+1,i}) \) in (1.3) and (1.4) are estimated from the conventional non-contextual preclassification results. The initial values are evaluated from \( P(d_{i,j}|C_{i,j}) \), because we have

\[ \sum_{C_{i,j}} P(C_{i,j}|C_{i-1,j}, C_{i-1,i}) P(C_{i-1,j}|D_{i,j}^{-1}) P(C_{i-1,j}|D_{i,j}^{-1}) = 1 \]

when \( i = 1 \) or \( j = 1 \)

So finally, application of the Bayes decision rule requires calculation of the following formulas and finding the class \( C_{i,j} \) which produces the maximum probability \( P(C_{i,j}|D_{i,j}) \).

\[ P(C_{i,j}|D_{i,j}) = \frac{P(C_{i,j}|D_{i,j}) P(C_{i,j}|E_{i,j}^{-1})}{P(D_{i,j}) P(E_{i,j}^{-1})} \]

where

\[ P(C_{i,j}|D_{i,j}) = \begin{cases} P(d_{i,j}|C_{i,j}) \left( \sum_{C_{i-1,j}} P(C_{i-1,j}|C_{i-1,i}) P(C_{i-1}|D_{i,j}^{-1}) P(C_{i-1}|D_{i,j}^{-1}) \right) & (i = 1 \text{ or } j = 1) \\ P(D_{i,j}|C_{i,j}) \left( \sum_{C_{i+1,j}} P(C_{i+1,j}|C_{i+1,i}) P(C_{i+1}|E_{i,j}^{-1}) P(C_{i+1}|E_{i,j}^{-1}) \right) & \text{otherwise} \end{cases} \]

\[ P(C_{i,j}|E_{i,j}^{-1}) = \begin{cases} P(d_{i,j}|C_{i,j}) \left( \sum_{C_{i+1,j}} P(C_{i+1,j}|C_{i+1,i}) P(C_{i+1}|D_{i,j}^{-1}) P(C_{i+1}|D_{i,j}^{-1}) \right) & (i = j \text{ or } j = 1) \\ P(d_{i,j}|C_{i,j}) \left( \sum_{C_{i+1,j}} P(C_{i+1,j}|C_{i+1,i}) P(C_{i+1}|E_{i,j}^{-1}) P(C_{i+1}|E_{i,j}^{-1}) \right) & \text{otherwise} \end{cases} \]

193
b) Four pass algorithm:

A more symmetric and accurate model to solve the conditional probability $P(C_{i,j} | D_{ij})$ is the four blocks model (i.e., we consider propagation of conditional probability in four directions rather than in two directions only. The notation for the recursive algorithm in the four blocks case is shown in Figure 3. Note that some of the symbols we use in this section have a different meaning than in the previous section. We have

$$P(C_{i,j} | D_{ij}) = \frac{P(D_{ij}^i, E_{ij}^i, F_{ij}^i, G_{ij}^i, d_{ij} | C_{i,j}) P(C_{i,j})}{P(D_{ij})}$$

$$= \frac{P(D_{ij}^i | C_{i,j}) P(E_{ij}^i | C_{i,j}) P(F_{ij}^i | C_{i,j}) P(G_{ij}^i | C_{i,j}) P(d_{ij} | C_{i,j}) P(C_{i,j})}{P(D_{ij})}$$

$$= \frac{[P(D_{ij}^i | C_{i,j}) P(d_{ij} | C_{i,j})] [P(E_{ij}^i | C_{i,j}) P(E_{ij}^i | C_{i,j})] [P(F_{ij}^i | C_{i,j}) P(d_{ij} | C_{i,j})]}{P(D_{ij})}$$

$$\frac{1}{[P(G_{ij}^i | C_{i,j}) P(d_{ij} | C_{i,j})] P(C_{i,j})}$$

$$= \frac{P(D_{ij}^i | C_{i,j}) P(E_{ij}^i | C_{i,j}) P(F_{ij}^i | C_{i,j}) P(G_{ij}^i | C_{i,j}) P(C_{i,j})}{P(D_{ij})}$$

$$= \frac{[P(C_{i,j} | D_{ij}^i) P(D_{ij}^i) / P(C_{i,j})] [P(C_{i,j} | E_{ij}^i) P(E_{ij}^i) / P(C_{i,j})] [P(C_{i,j} | F_{ij}^i) P(F_{ij}^i) / P(C_{i,j})]}{[P(C_{i,j} | G_{ij}^i) P(G_{ij}^i) / P(C_{i,j})]}$$

$$\Delta [P(d_{ij} | C_{i,j})]^{-3}$$

where $\Delta = P(D_{ij}) / P(E_{ij}^i) P(F_{ij}^i) P(G_{ij}^i) P(C_{i,j})$ \hfill ...(2.1)

In (2.1) $P(C_{i,j})$ and $P(d_{ij} | C_{i,j})$ are known a priori information for supervised remote sensing classification, and $\Delta$ is independent from categories $C$. Then to solve (2.1) the problem becomes to calculate $P(C_{i,j} | D_{ij}^i), P(C_{i,j} | E_{ij}^i), P(C_{i,j} | F_{ij}^i)$ and $P(C_{i,j} | G_{ij}^i)$.

From the derivation which is the same as the two block cases, we have:

$$P(C_{i,j} | D_{ij}^i) = \frac{\sum C_{i,j-1} \left[ \sum C_{i-1,j-1} P(C_{i,j-1} | C_{i-1,j-1}) P(C_{i-1,j-1} | D_{ij}^i) \right] P(C_{i,j-1} | D_{ij}^i)}{\Delta}$$ \hfill ...(2.2)

$$P(C_{i,j} | E_{ij}^i) = \frac{\sum C_{i,j+1} \left[ \sum C_{i+1,j+1} P(C_{i,j+1} | C_{i+1,j+1}) P(C_{i,j+1} | E_{ij}^i) \right] P(C_{i,j+1} | E_{ij}^i)}{\Delta}$$ \hfill ...(2.3)

$$P(C_{i,j} | F_{ij}^i) = \frac{\sum C_{i+1,j-1} \left[ \sum C_{i+1,j-1} P(C_{i+1,j-1} | C_{i+1,j-1}) P(C_{i+1,j-1} | F_{ij}^i) \right] P(C_{i+1,j-1} | F_{ij}^i)}{\Delta}$$ \hfill ...(2.4)

$$P(C_{i,j} | G_{ij}^i) = \frac{\sum C_{i+1,j} \left[ \sum C_{i+1,j} P(C_{i+1,j} | C_{i+1,j}) P(C_{i+1,j} | G_{ij}^i) \right] P(C_{i+1,j} | G_{ij}^i)}{\Delta}$$ \hfill ...(2.5)
The recursive formulas (2.5), (2.3), (2.4) and (2.5) are from top-left to bottom-right, from top-right to bottom-left, from bottom-left to top-right and from bottom-right to top-left scanings, respectively. The procedure is exactly the same as the previous section except there is a four pass scanning instead of a two pass scanning.

(3.4) Computation requirement:

Because the estimation equations of the recursive algorithm are quite simple, the computation required increases linearly with the number of picture points. Let C be the number of categories and N the number of pixels. In the two pass algorithm, evaluation of \( P(C_i|D_1^{12}) \) and \( P(C_i|E_1^{12}) \) required \( 2C^2N \) multiplications, and \( P(C_i|D_2^{12}) \) required \( CN \) multiplications. So, the total multiplications are \( (4C + 1)CN \). Similarly, in the four pass algorithm, the computation complexity is \( (8C + 1)CN \).

When we implement recursive formula for each pixel, it requires only adjacent pixels previously processed, and the measurement dependency is entirely included in \( P(d_i|C_i) \). So in this algorithm the internal memory required grows linearly with the number of columns in the image. Because of the above two facts, the method is particularly suitable for large sized images.

(4.0) Implementation details and experimental results

In this section we summarize the procedure of the recursive contextual classification, and then present experimental results in an effort to demonstrate the feasibility of the recursive contextual classification algorithm. Finally, we compare this technique with some of the techniques discussed in the literature.

(4.1) Summary of the recursive contextual classification procedure.

Before presenting some experimental results with the proposed classification procedure, we shall first summarise the steps necessary to perform such a classification where the conditional distributions as assumed normal.

1. Evaluate training statistics, including mean vector and covariance matrices for each class, from ground data.

2. Preclassify the image using a pixel independent or context free Bayes classification technique.

3. Evaluate transition probability \( P(C_i|C_{i-1}, C_{i-1,3}) \) from the preclassification results.

4. Use recursive formula (1.2-1.3) to compute \( P(C_i|D_1^{12}) \) and \( P(C_i|E_1^{12}) \) (the two pass algorithm) or \( P(C_i|D_1^{12}), P(C_i|E_1^{12}), P(C_i|D_2^{12}) \) and \( P(C_i|E_2^{12}) \) (for the four pass algorithm), respectively.

5. Use formula (1.1) to evaluate \( P(C_i|D_1^{12}) \) for each category.

6. Use the classification rule to select that class \( C_i \) which produce the maximum probability \( P(C_i|D_1^{12}) \).

(4.1) Experimental results:

The technique is illustrated using digital remote sensing data collected by the Landsat MSS. Several experimental results, which collected from different areas with different categories and different preclassification accuracy. The first experimental data, which was a subset of the 13 April 1976 MSS scene of Roanoke, VA, was selected as the first study area. Four bands of digital, multispectral data were classified by the conventional contextual method in order to compare the results. The classification method was used to identify different classes in the LANDSAT MSS data. We use the following ground cover classes, which are the USGS Land use and land cover classification: (1) Urban or Built-up Land; (2) Agricultural Land; (3) Rangeland; (4) Forest land; (5) Water; (6) Wetland; (7) Barren land; (8) Tundra; (9) Perennial Snow or ice.

In Formula (1.1) \( P(d_i|C_i) \) is the class conditional distribution, which is estimated from the training sets and ground truth. In this experiment means and covariance matrices of each category were calculated from ground truth data. The class conditional probability \( P(d_i|C_i) \) are assumed multi-dimensional normal:

\[
P(d_i|C_i) = \frac{1}{(2\pi)^{j/2} \sum |\Sigma |^{1/2}} \exp \left( -\frac{1}{2} (d - m)^T \Sigma^{-1} (d - m) \right)
\]

The study area was selected form the Roanoke, VA region (longitude from 79°55' to 80°00' W; latitude from 37°15' to 37°23' N). The land cover of this mountainous region is a complex pattern of diverse spectral classes presented in small parcels. The more easily classified of this land cover class-open water-is not represented in this test area. Thus, this area is a difficult area for conventional classification. The accuracy of context free classification algorithms, including Bayesian classifiers, 60%; AMOEBA (Bryant, 1979), and ISODATA (Duda and Hart, 1973), are not unusual for scenes of this complexity.

The context free and contextual classification results are shown in Figure 4. A comparison of several methods with the Markov context model is illustrated in Table 1. By visually examining these results, one can easily tell how good the performances are within each class, and also along the boundaries between classes. At first sight, we see that the Bayesian classifiers results are quite noisy. The Markov context classification results served to "clean up" the picture significantly. It is clear that many small isolated pixels were eliminated, and each area was much more homogeneous in the contextual classification results. Boundaries remained accurately placed. The MSS four band image, ground truth map, which had been classified by professional analysts, is given in Figure 4. The above comparison and TAB 1 indicated that a 5 to 10% improvement of accuracy was obtained by the context classification method. So in addition to the visual improvement, the context classification scheme improves the classification accuracy as well.

The second study area is California. Three MSS classification results where sizes of subsets are 130 x 90, 101 x 70, 130 x 60 respectively, are shown in Figure 5. These results show that the algorithm was effective in several different areas with varied categories and preclassification accuracies (these areas had about 90% preclassification accuracy).
(4.2) Related literature:

The approaches to the context classification scheme reviewed here are the presented by Tilton [5] and Yu [4]. Tilton uses an estimation method of the context function, which had 2-6% improvement in classification accuracy; Yu uses a coding technique, by which the spatial correlation parameter can be estimated. His contextual classification provided about 5% improvement in the first-stage, and 2% additional improvement in the second-stage. So a comparison of our method with these two methods shows the classification accuracy improvement of our approach to be about the same. (5%-6%). The decision rule in Tilton is \( d(X_i) = \text{action (classification)} \) a which maximizes

\[
\sum_{\theta^p \in \Theta^p} \frac{G(\theta^p)}{\sum_{\theta^p} G(\theta^p)} \prod_{k=1}^{p} f(X_K|\theta_K)
\]

where \( G(\theta^p) \), the “context function”, is the relative frequencies with which \( \theta^p \) occurs in scene being analyzed. \( X_i \) is a vector of observation \( X_{ij} = (x_1, x_2, \ldots, x_p)^T \), and \( \theta_i \) be the vector of true but unknown classifications associated with the observations in \( X_i \). \( \theta^p \in \Theta^p \) is a vector of possible classifications for the elements of any P-context array. So the computation required grows p-the power with the number of picture points. (p = 3 or p = 5). In the paper only 50 x 50 size images were tested.

The classification rule in Yu paper [4] is to select that class \( \theta_K \) which, after normalizing cell K, will produce the maximum joint probability \( P(E_k, E_{k1}, E_{k2}, E_{k3}, E_{k4}|\theta_k) \). In this method use the coding technique to estimate the spatial correlation for each paired, and for each iteration it had to evaluate the joint normal density function and use maximum likelihood decision rule for each pixel once, so the computation cost is relatively high.

Compared with the above two papers, the main advantage of our approach is its low cost in computation, so it is particularly suitable for larger sized images. In addition, this algorithm has the advantage over other techniques in that it handles multiple observation data naturally and simultaneously.

Summary

We have developed a new multispectral image context classification with Markov Random Field, where remotely sensed data are more efficiently and more accurately classified compared to traditional context free classifiers. This new approach of multispectral image context classification is based on a recursive algorithm for optimal estimation of the state of a two-dimensional discrete Markov Random Field.

The first task of this approach is to develop a two-dimensional discrete Markov Random Field model. A recursive algorithm is then obtained from local measurement and noise independence assumptions of the Markov Random Field model, and finally the algorithm determines the conditional probabilities of categories given all the data.

Because of the quite simple algorithm, the computation complexity and memory space required by the approach is low. So the method is particularly suitable for large images.

References


**Figure 1:** The 2-D discrete Markov random field for multispectral image processing application

**Figure 2:** Notation for a two-pass recursive algorithm

\[ d_{ij} = d_{i+1,j} \]
\[ e_{i,j} = e_{i,j+1} \]

**Figure 3:** Notation for a four-pass recursive algorithm

\[ d_{i,j} \]
\[ e_{i,j} \]
\[ f_{i,j} \]
\[ g_{i,j} \]
Fig 4 13 April 1976 MSS scene of Rosnake VA

Ground truth of same area

Bayes' preclassification result

Markov contextual classification result

MSS image of California (J)

Bayes preclassification result
## CMI 1: Comparison of classification accuracy using Markov context

<table>
<thead>
<tr>
<th>Location</th>
<th>Size</th>
<th>Noncontext Classification</th>
<th>Markov Context Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>3 of error</td>
<td>error%</td>
</tr>
<tr>
<td>Penns.-blue (VTL)</td>
<td>151X151</td>
<td>342</td>
<td>37.7%</td>
</tr>
<tr>
<td>Cali.-formic (F)</td>
<td>130X90</td>
<td>10.37%</td>
<td>four</td>
</tr>
<tr>
<td>Cali.-formic (H)</td>
<td>101X70</td>
<td>12.8%</td>
<td>four</td>
</tr>
<tr>
<td>Cali.-formic (M)</td>
<td>120X60</td>
<td>13.2%</td>
<td>four</td>
</tr>
</tbody>
</table>

* Two is a two-pass algorithm and four is a four-pass algorithm in the TAMI.