081278

AN APPROACH TO THE USE OF STATISTICAL CONTEXT IN REMOTE SENSING DATA ANALYSIS

E.F. Kit and P.H. Swain
School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907

ABSTRACT

A statistical model is developed for using image context in maximum likelihood classification. Experimental results using both simulated and real multispectral remote sensing data demonstrate the utility of the model. Some practical problems associated with the use of the model are discussed.

RESUME

Les auteurs développèrent un modèle statistique afin d'utiliser un "contexte" de l'image en matière de classification des probabilités maximales. Ils démontrent l'utilité du modèle par des résultats d'essais qui utilisent des données simulées et réelles de télédétection multispectrale. Ils discutent de quelques problèmes pratiques reliés à l'utilisation du modèle.

INTRODUCTION

It does not require a highly skilled photo interpreter to appreciate the fact that much remote sensing imagery is rich in spatial information content; i.e., information inherent in the image nature of the data which improves our knowledge about the ground scene. Historically, however, efforts to analyze multispectral remote sensing data by automatic or computer-assisted methods have focused largely on the spectral information contained in individual pixels ("picture elements")[1], thereby ignoring the spatial information. Only fairly recently has attention turned to making use of spatial information in the data.

The most effective approaches to utilizing spatial information in multispectral imagery have drawn on special characteristics of the imagery or intuitive notions of features which seem likely to be information-bearing. A notable example is the ECHO (Extraction and Classification of Homogeneous Objects) process which segments a scene into "objects," and then uses sample classification to assign each object, as a whole, rather than its individual pixels, to an appropriate ground cover class[2]. Provided the average size of the objects is large relative to the resolution of the sensor system, this method offers both improved classification accuracy and speed.

Another example is the use of features based on gray-tone spatial-dependence matrices, as described by Haralick et al[3], to characterize local scene texture. These matrices are features which seem to be comparable to intuitive notions of texture. Experiments have shown them to be useful for classification purposes.

In this paper, we consider the "context" of a pixel in a still more general way. It has long been known that in written language one finds certain letters occurring frequently in the company of others. Examples include "qu," "ee," "ing." This phenomenon can be used to increase the probability of correct recognition of letters in the machine analysis of hand written or printed text. The same principle can be used effectively in the analysis of multispectral remote sensing image data. The method is an extension of the statistical decision theory approach used for pixel-by-pixel classification. Its advantages over the spatially oriented methods cited above are at least two-fold: (i) the mathematical foundations of the underlying model are firmly rooted in the theory of optimal decision making, and (ii) the context classifier does not require a number of user-specified and scene-dependent parameters.

THE MODEL

We assume a two-dimensional array of pixels of fixed but unknown classification, as shown in Figure 1. Associated with the pixel having coordinates (i,j) is its true classification θ_{ij}, where $\theta_{ij} \varepsilon \Omega = \{1,2,\ldots,k\}$, and a random vector (observation) X_{ij} having class-conditional distribution $f_{\theta_{ij}}$, where $f_{\theta_{ij}} \varepsilon \{f_1,f_2,\ldots,f_k\}$ is a member of the set of density functions associated with the classes. The observations are assumed to be class-conditionally independent. We wish to classify the $N = N_1 \times N_2$ observations in the array.

Let the action (classification) taken with respect to pixel (i,j) be denoted by $a_{ij} \varepsilon \Omega$. Let the loss suffered by taking action a_{ij} when the true class is θ_{ij} be denoted by

Presented to the 5th Canadian Symposium on Remote Sensing, Victoria, August 1978.

$L(\theta_{ij}, a_{ij})$, for some fixed non-negative function $L(\cdot,\cdot)$. Then the average loss suffered over the N classifications in the array is

$$\frac{1}{N}\sum_{i,j} L(\theta_{ij}, a_{ij}) \qquad (1)$$

If we make the action a_{ij} a function of the observations, then the expected loss is

$$R = E\left[\frac{1}{N}\sum_{i,j} L(\theta_{ij}, a_{ij}(\underline{X}))\right] \qquad (2)$$

where $\underline{X}$ is the set of N observations in the array. When context is ignored, the action (classification) depends only on the cell to be classified. In that case,

$$R = E\left[\frac{1}{N}\sum_{i,j} L(\theta_{ij}, a(X_{ij}))\right]$$

$$= \sum_{\theta\varepsilon\Omega} G(\theta)E_\theta\left[L(\theta,a(X))\right] \qquad (3)$$

which is the Bayes risk given the frequency distribution $G(\theta)$ of the classes (prior probabilities). The decision rule $a(X)$, which maps the observation space into Ω, should be chosen to minimize R.

To introduce the context, we focus on some fixed arrangement of p pixels in the frame which we wish to incorporate in each decision. For example, see Figure 2. The arrangement actually chosen will be dictated by practical considerations (recall we are classifying pixel (i,j)). Denote the p-vector of states as $\underline{\theta}_{ij}$ and the p-vector of observations $\underline{X}_{ij}$. Now a function $a(\underline{X}_{ij})$ will map p-vectors of observations into actions. In this case, (2) and (3) become

$$R = E\left[\frac{1}{N}\sum_{i,j} L(\theta_{ij},a_{ij}(\underline{X}))\right]$$

$$= \frac{1}{N}\sum_{i,j} E\left[L(\theta_{ij},a(\underline{X}_{ij}))\right]$$

$$= \sum_{\underline{\theta}} G^P(\underline{\theta})E_{\underline{\theta}}\left[L(\theta,a(\underline{X})\right] \qquad (4)$$

where now $G^P(\underline{\theta})$ is the frequency distribution of the p-vectors $\underline{\theta}$. Again, this expression gives the Bayes risk associated with the decision function $a(\underline{X})$ which should be chosen to minimize R.

What form should the decision function take? To minimize (4), it suffices to choose $a(\underline{X})$ to be an action which minimizes

$$\sum_{\underline{\theta}} L(\theta,a)\left[\prod_{i=1}^{p} f_{\theta_i}(X_i)\right]G^P(\underline{\theta})$$

or, taking $L(\theta,a)$ to be the usual 0-1 loss function, choose $a(\underline{X})$ so that it maximizes

$$\sum_{\substack{\underline{\theta},\\ \theta_1=a}} \left[\prod_{i=1}^{p} f_{\theta_i}(X_i)\right]G^P(\underline{\theta})$$

where θ_i and X_i are the state and measurement of the ith pixel in the p-array and θ_1 and X_1 refer to the pixel being classified. The sum is over all p-vectors for which $\theta_1 = a$. This strategy parallels closely that followed for pointwise classification, but incorporates the p-array context.

EXPERIMENTAL RESULTS

A data analysis experiment was designed to explore the effectiveness of this approach for classification of earth resources data. More specifically, it was desired to determine the degree to which the use of context characterized as described above would improve classifications as compared to results achieved without context.

In order to avoid confounding other effects with the impact of context, it was decided to use a simulated data set generated as follows. A classification of multispectral remote sensing data was selected which had been judged to be very accurate (typically, produced by careful analysis and refinement of multitemporal data). Such a classification could be expected to embody the contextual content of an actual ground scene. Using the classification map and the associated statistics of the classes (developed in producing the classification), data vectors were produced by a Gaussian random number generator and composed into a new data set. Thus the new data set had the following characteristics:

1. Each pixel in the "simulated" data set represented the same class as in the "template" classification. The template could be considered the "ground truth" for the new data set.

2. All classes in the data set were known and represented.

3. All classes had multivariate Gaussian distributions with statistics typical of those found in real data.

4. All pixels were class-conditionally independent of adjacent pixels.

5. There were no mixture pixels.

Although the simulated data are somewhat of an idealization of "real" remote sensing, its spatial organization is consistent with a real world scene and its overall characteristics are consistent with the context model set out above. In essence, then, what the experimental results based on the simulated data show is the effectiveness of the context classifier, given that the underlying assumptions (approximations) are reasonable. Further experiments are required to generalize the conclusions of these results to real data.

Three data sets were selected to represent a variety of ground cover types and textures. Data set 1 is agricultural (Williston, North Dakota), with ground resolution and spectral bands approximating those of the projected Landsat D Thematic Mapper. Data set 2a is Landsat 1 data from an urban area (Grand Rapids, Michigan). Data set 2b is from the same Landsat frame as 2a, but from a locale having significantly different spatial organization. Each data set is square, 50 pixels on a side.

Figure 3 shows the achieved classification results. The "no context" classification accuracy is plotted on the vertical axis of each graph. Data set 1 was classified using successively 0, 2, 4, 6 and 8 neighboring pixels; data sets 2a and 2b were classified using 0, 2, 4 and 8 neighboring pixels. The results speak for themselves. The accuracy improvement resulting from the use of contextual information is quite significant.

To accomplish the context classification using this approach, it is necessary to have available a set of class-conditional density functions (f_θ) for the classes to be recognized and the frequency distribution for the p-vectors ($G^p(\underline{\theta})$). In remote sensing applications, the class-conditional density functions are typically learned from training samples. For the experiments described above, the Gaussian class statistics on which the data simulation was based were used for the classification (these were originally the training statistics used to produce the "template" classification). An important question is how in practice to determine the p-vector frequency distribution. In the foregoing experiment, this distribution was simply tabulated from the "template" classification. But in a real data situation, such a template is not available (else there would be no need to perform any further classification).

One can envision a number of ways in which the p-vector distribution might be estimated for a remote sensing application. For example, it could be extracted from a classification of the same area obtained previously. This would require that the area not have changed much in its class make-up since the earlier data were collected and that the earlier classification was reasonably accurate. Or, the distribution might be obtained from a classification of any similarly constituted area. Still another possibility would be to estimate the p-vector distribution for the context classification from a "conventional" classification with "reasonably good" accuracy. All of these methods produce an estimate of the p-vector distribution, and a crucial question on which hinges the utility of this approach is how sensitive the contextual algorithm is likely to be to the "goodness" of the estimate. This question is the subject of ongoing research.

An experiment was formulated to obtain some evidence concerning the feasibility of applying the context classifier to a real data situation. The data set used covered a somewhat larger area of Grand Rapids, Michigan, containing both data sets 2a and 2b. Data from small areas of known ground cover were used to estimate the training class statistics, and data from a disjoint set of areas of known ground cover were used as "test samples" to evaluate the classifier accuracy (unfortunately, the set used for this test was rather small, consisting of only 136 pixels distributed among 4 urban classes).

A non-contextual classification was performed and found, based on the test set, to be 81.6 percent accurate. The p-vector distributions were estimated from this classification and used to perform context classifications using first four and then eight nearest neighbors. The four-neighbor classification was 83.1 percent accurate; the eight-neighbor classification was 84.6 percent accurate. For this case, then, some improvement in classification accuracy was again achieved by incorporating context in the decision process.

One might ask what would happen if the contextual classification were iterated on each interaction, basing the p-vector distribution on the results of the previous iteration. Certainly the use of such a procedure would involve questions of stability.

CONCLUSIONS

We have formulated an approach for maximum likelihood classification of multispectral image data using the context of each pixel to be classified. Experimental results using simulated data have demonstrated that the context classifier is indeed capable of improving classification accuracy over that obtainable by means of no-context classification. Very limited results suggest the feasibility of the approach for real data classification.

The price paid for incorporating context into the classification process by the approach suggested here is considerable in terms of the amount of computation required and the amount of prior knowledge about the data which is used in the classification. However, these problems appear to be resolvable, and the general approach to contextual classification should prove valuable where context is an important information-bearing characteristic of the scene.

ACKNOWLEDGEMENTS

This research was funded in part by the Defense Advanced Research Projects Agency, U.S. Department of Defense, under Contract No. MDA903-77-G-1.

REFERENCES

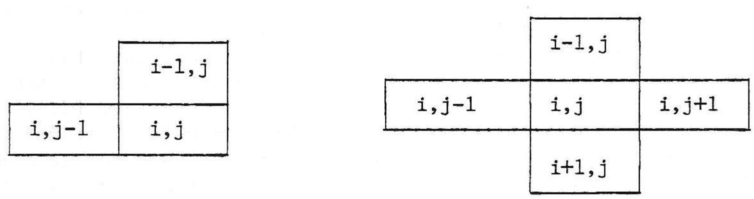
1. P. H. Swain and S. M. Davis, eds., Remote Sensing: The Quantitative Approach, McGraw-Hill International Book Co., 1978.

2. R. L. Kettig and D. A. Landgrebe, "Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects," IEEE Trans. Geoscience Electronics, Vol. GE-14, No. 1, pp. 19-26, January 1976.

3. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," IEEE Trans. Systems, Man and Cybernetics, Vol. SMC-3, pp. 610-621, November 1973.

$$
\begin{array}{|cccc|}
\hline
\theta_{11} & \theta_{12} & \cdots & \theta_{1N_2} \\
\theta_{21} & \theta_{22} & \cdots & \theta_{2N_2} \\
\vdots & & & \\
\theta_{N_1 1} & & \cdots & \theta_{N_1 N_2} \\
\hline
\end{array}
$$

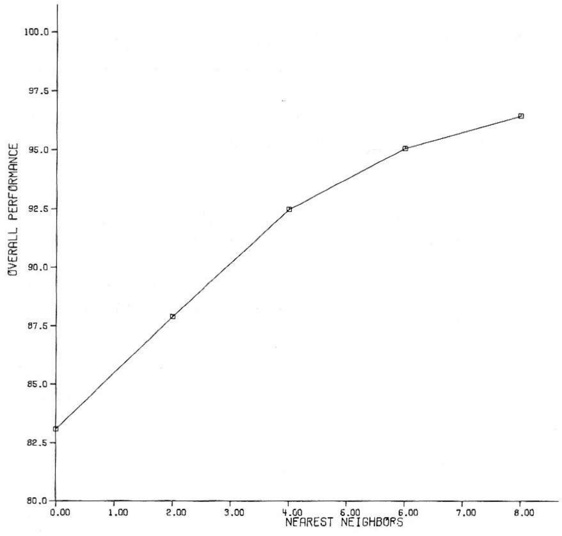
Figure 1. Image Array.



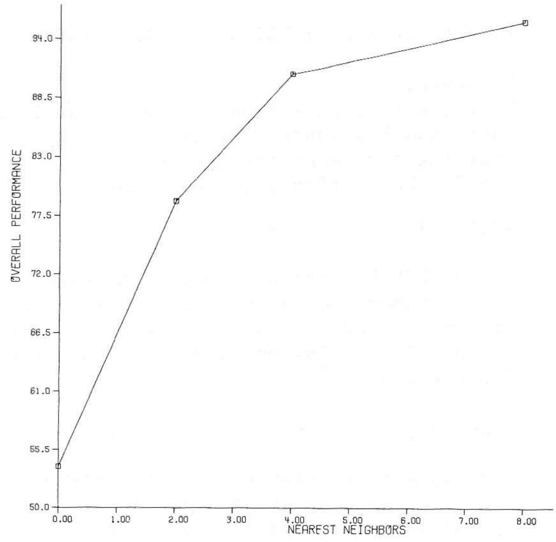
a p=3 choice

a p=5 choice

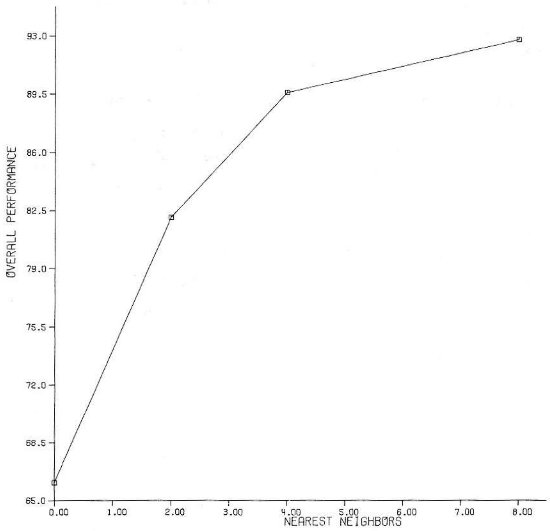
Fig. 2. Contextual p-arrays.



(a)



(b)



(c)

Figure 3. Context classifier results. (a) Data set 1 (agricultural, Landsat D resolution); (b) Data set 2a (urban, Landsat 1 resolution); (c) Data set 2b (urban, Landsat 1 resolution). Vertical scale is percent correct recognition.