EVALUATION AND ANALYSIS OF TRAINING SETS AND TRAINING STATISTICS OF MULTISPECTRAL SCANNER DATA RELATED TO AGRICULTURAL CROP COVER TYPES

R.S. Ayyangar, P.M. Sarma, P.P. Nageswara Rao and S.S. Budwal

National Remote Sensing Agency, Secunderabad

ABSTRACT

The size and reliability of the training sets or sample area for the classification of airborne multispectral scanner data obtained over an agricultural area with the help of an interactive computer system have been examined in this study. The experiment reported herein suggests that a training set of not less than 50 pixels would adequately represent all the likely variations in any particular field. The evaluation of the results further reveals that if the training sets can adequately represent the field variations characteristic of the region, the corresponding training statistics can be utilized both on scanline and pixel directions.

INTRODUCTION

Until recently, for extracting useful information from remotely sensed data, interpreters have been heavily dependent on manual interpretation. With the development of efficient data acquisition system and capability to collect large volume of data in short time, it is becoming exceedingly difficult for trained interpreters to derive useful information quickly from remotely sensed data unless some other methods, such as computer processing, are adopted. Since the remotely sensed data is obtained in digital format, it offers scope to utilize the computer techniques for rapid processing of the data. The computer analysis of remotely sensed data, essentially a numerical analysis, utilizes the spectral variations exhibited by the objects of interest, besides, spatial and temporal information, using certain statistical principles.

Among the several algorithms (calculating procedures) available for digital processing of remotely sensed data, pattern recognition involving a Gaussian maximum likelihood decision rule is the most commonly used. Using the above statistical procedures, two methods of classification procedures are recognised i.e. supervised classification and unsupervised classification. The supervised classification involves the use of training samples (training sets) to train the classifier to recognise informational classes and often involves the interaction of the interpreter and the computer. Unsupervised classification utilizes the spectral variations exhibited by the objects and separate them into spectral classes using the clustering technique, based on which informational
classes are derived. However, when there is an access to an interactive computer system, supervised classification is usually adopted. Supervised classification of multispectral data with the aid of a computer system often poses two basic questions.

i. the size of the training set and its adequacy as a representative sample for the training statistics.

ii. the reliability of the training statistics in the classification performance.

These are especially important in the analysis of airborne multispectral scanner data where one would be interested, not only in determining the adequate size of the sample but also in knowing the reliability of the training statistics in the classification of the data along the entire run and data swath.

To ascertain and examine the above points, an attempt that was made in this direction using the airborne multispectral scanner data obtained over an agricultural area is described in this paper.

**METHODOLOGY**

The experiment was conducted in one of the intensively cultivated agricultural regions of Karnataka state. The agriculture included various crops under different management practices. Among the several crops grown in this region, rice and sugarcane were the major crops. The multispectral data was collected during the Kharif cropping season (September, 1978) with the aid of an eleven channel Modular Multispectral Scanner (M2S) on board from an aircraft flown at an altitude of about 1000 meters above ground level. The ground resolution at that altitude was 6.25 sq. m. The airborne multispectral data had the concurrent support of the ground truth data.

The spectral data recorded on high density digital tape (HDDT) was converted into computer compatible tapes for data analysis. The data was analysed with the aid of an interactive computer system, the multispectral data analysis system (M-DAS). For determining the appropriate size of the training set, three sample sizes, viz. 100 ± 5 pixel, 50 ± 5 pixel and 25 ± 5 pixels were considered. Several training sets representing various agricultural categories were identified with the help of the computer and their respective spectral statistics extracted. The spectral statistics consisted of the mean, standard deviation etc. for ten spectral bands of the M2S (the eleventh band was not functioning). Using the above spectral statistics, polynomial curves for the mean values of the spectral bands were drawn to evaluate and understand the variation exhibited by the sample size.