1-1-1976

Processing Multispectral Scanner Data Using Correlation Clustering and Nonparametric Classification Techniques

Richard E. Haskell

David E. Boddy

Reprinted from

Symposium on

Machine Processing of

Remotely Sensed Data

June 29 - July 1, 1976

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

IEEE Catalog No.
76CH1103-1 MPRSD

PROCESSING MULTISPECTRAL SCANNER DATA USING CORRELATION CLUSTERING AND NONPARAMETRIC CLASSIFICATION TECHNIQUES*

Richard E. Haskell and David E. Boddy
Oakland University
Rochester, Michigan 48063

ABSTRACT

A two-step classification algorithm for processing multispectral scanner data has been developed and tested. The algorithm is carried out by two separate programs called CLUSTX and GROUPX. The program CLUSTX is a single pass clustering algorithm that assigns each pixel, based on its spectral signature, to one of NCLUST clusters. The output of the program CLUSTX is a cluster tape in which a single integer is associated with each pixel. This integer is the cluster number to which the pixel has been assigned by the program.

The clustering program CLUSTX can be considered to be a data reduction and preprocessing step in the classification algorithm. The assumption is that enough clusters are chosen so that all pixels assigned to a particular cluster have very similar spectral signatures and thus belong to the same class. The cluster tape which is the output of the program CLUSTX is the input to the classification program GROUPX. Ground truth information is used in GROUPX to classify each cluster as one of a small set of classes. Once the clusters have been assigned to classes the cluster tape is read pixel by pixel and an output tape is produced in which each pixel is assigned to its proper class. This output classification tape can then be used directly to produce classification maps, compute acreage of different classes, or test the accuracy of the classification method by comparing the results with additional ground truth.

Ground truth information is used to train the classifier that will classify each pixel. This classifier creates nonlinear decision surfaces based on the method of potentials. Two types of training are possible. If the ground truth is limited then the spectral signatures from each pixel are used to construct the decision surfaces. On the other hand, if a large quantity of ground truth is available, then it can be used to produce a cost matrix giving the number of pixels in each cluster that belongs to each of the various classes. These numbers are used to estimate the a posteriori probabilities of a particular cluster belonging to a particular class. The cluster is then assigned to the class for which this a posteriori probability is a maximum. The clusters classified in this manner serve as the training data for constructing the decision surfaces using the potential functions. The remaining clusters are then classified using the method of potentials.

The classification algorithm can be operated in a hierarchical manner. A classification tree, or taxonomy, is defined for the data. Each ground truth datum is classified at various levels in the tree, depending on how much detail is known for that datum. The ground truth data can then be used to classify each unknown pixel, or cluster, to any desired level of detail in the classification tree.

The two major advantages of this classification algorithm are: 1) The classification method is entirely nonparametric and thus avoids the errors that are inherent in estimating parameter vectors in parametric methods. 2) Changes in the spectral signature of a particular class along the flight line cause no problem as long as representative ground truth is available, since the result will simply be the generation of new clusters.

*Work supported by NASA under Contract NAS-9-14195.