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

CLASSIFICATION BY CLUSTERING*

Alex Pentland
Environmental Research Institute of Michigan
Ann Arbor, Michigan

ABSTRACT

Conventional classification procedures have several difficulties which sometimes limit the usefulness of computer aided analysis techniques on multispectral scanner data. In order to minimize some of these problems, the clustering algorithm used at ERIM (called CLUSTR) was adapted for use as a classifier. Briefly, the technique devised is to cluster the scene, assigning each pixel to a cluster, and then to identify the crop type of the clusters by examining training areas to determine the crop type of pixels assigned to each cluster. In this manner, the classification of each pixel to a particular crop class is accomplished.

This approach to classification has several advantages over more conventional classification techniques. Among these advantages are:

1) CLUSTR is designed to use several small normal distributions (clusters) to approximate the non-gaussian spectral distributions of the various ground classes -- thus minimizing problems with non-gaussian distributions.

2) CLUSTR continually updates its estimate of the various spectral distributions, including modifying the means, variances and even the number of clusters as the distributions in the data change. This minimizes the effects of most variations in the data.

3) Problems stemming from the inability to obtain representative training data are reduced, because all of the data is used in constructing the signatures, instead of just the data from the training areas.

4) Inaccuracies in the ground truth for training areas are less important than in conventional techniques, e.g., you do not cluster the wheat training regions and call all resulting clusters "wheat", even if they look like corn, instead you cluster the entire scene and only those clusters which have more "wheat" pixels assigned to them than "other" pixels are identified as "wheat". With conventional techniques, all pixels must be correctly identified.

5) Human participation in the signature extraction and classification procedures is reduced, because they are combined into one step.

From preliminary tests it appears that the CLUSTR classifier is as accurate as the Bayes maximum likelihood decision rule and may be useful for proportion estimation, especially in cases where ground truth is limited, or where there are variations in the data, or where conventional signature extraction is difficult.

*The effort described herein was supported by the Earth Observations Division of the NASA/Johnson Space Center under contract NAS9-14123.