Cluster Method Analysis of K.S.C. Image

Joe Rodriguez Jr. and M. Desai
Division of Engineering.

University of Texas at San Antonio
6900 Northloop 1604 West
San Antonio, TX 78249
joe@runner.utsa.edu

Abstract:

Information obtained from satellite-based systems has moved to the forefront as a method in the identification of many land cover types. Identification of different land features through remote sensing is an effective tool for regional and global assessment of geometric characteristics. Classification data acquired from remote sensing images have a wide variety of applications. In particular, analysis of remote sensing images have special applications in the classification of various types of vegetation. Results obtained from classification studies of a particular area or region serve towards a greater understanding of what parameters (ecological, temporal, etc.) affect the region being analyzed. In this paper, we make a distinction between both types of classification approaches although, focus is given to the unsupervised classification method using 1987 Thematic Mapped (TM) images of Kennedy Space Center.

I. Introduction:

The primary objective of image classification is to identify, as a unique gray level, the features occurring in an images in terms of the type of land cover these features actually represent. Picture elements, pixels, within an image represent the smallest unit of spatial area on the ground for which data is collected. Image analysis is done to provide a quantitative analysis of pixels for which, using computer based algorithms, they are counted for area estimates and identified based on their numerical properties.

Basic to the understanding of multispectral classification is the concept of the spectral signature or spectral response of an object on the ground. The spectral response for a given object is a measure of the amount of electromagnetic radiation it reflects as a function of wavelength. This quantitative measure of the reflected electromagnetic radiation sampled in a series of different wavelength bands produces a unique response called a signature. Therefore, the objective of classification becomes recognition of unique pixel signatures [2]. The governing idea is to automatically categorize all signatures in an image into special land cover classes, more commonly know as themes.

Remote sensing data sets exits as a or way to integrate spatially heterogeneous responses into a more easily measurable format by quantifying them at a specific scale (e.g. 10, 20, or 30m). For instance, Landsat Thematic Mapper (TM) sensors have a spatial resolution or pixel size of 30m, which represents a 30m x 30m area on the ground. In an area of heterogeneous land covers, spectral responses for different objects within a pixel will be averaged or aggregated into a composite spectral response for any particular pixel that falls over a specific area on the ground.

Thus the multispectural domain exists as both an additional axis of information available for analysis and as any integrating factor of scale-related phenomena [1,3].

H. Classification Analysis:

Multispectral classification is an information-extraction process that analyzes the spectral signatures determined in a region and then assigns pixels into categories based on similar signatures obtained in the entire image. There are generally two types of classification approaches used, supervised and unsupervised.

Supervised classification procedures are part of the essential tools used to extract quantitative information from remotely sensed data. In this type of approach, the analyst defines on the image a small area, called a training site, which is representative of each terrain category, or class. Then spectral values for each pixel in a training site are used to define the decision space for that class. After each training site is defined, the computer algorithm then classifies all the remaining pixels in the scene accordingly.

The second classification approach, for which we focus our analysis, is called unsupervised. Unsupervised classification is a method which examines a large number of pixels and divides them into a number of classes based on natural groupings present in the image. Unsupervised Classification is performed most often by using clustering methods to assign each pixel in an image to spectral classes, of which a user has no previous knowledge. Unlike supervised classification, unsupervised classification does not require analyst-specified training data (previously acquired data of the scene being analyzed). This procedure can be used to determine the number and location of the spectral classes into which the pixels are assigned. Finally, using the existing information from site photos, visits, and maps the resulting classification can then be identified [2].

The unsupervised approach to image classification always requires the classifier, the algorithm used to carry forth the pixel analysis, to learn or cluster. Clustering techniques are useful for image segmentation and for classification of raw data, for which there is no previous knowledge, to establish classes. Statistical techniques can be utilized to automatically group an n-dimensional set of observations in their natural spectral classes. Therefore we use clustering techniques to define a set of feature points, pixels, in the region being analyzed for which their is a large density compared to the density of features points in the surrounding region.

For our analysis we adopted a computer based clustering method, ISOCLUS [8], which is an iterative statistical method for clustering of feature points. This clustering method is based on the Isodata algorithm. In the analysis, we first assumed the number of clusters, K. Next, the partitioning of the data is done such that the average spread or variance of the partition is minimized. Let $\mu_k(n)$ denote the kth cluster center and the nth iteration and $\mathbf{R}_k$ denote the region of the kth cluster at a given iteration. Initially, we assigned arbitrary values of $\mu_k(0)$. At the nth iteration we took one of the data point $\mathbf{x}_i$ and assigned it to the cluster whose center is closest to it, that is,

$$\mathbf{x}_i \in \mathbf{R}_k \iff d(\mathbf{x}_i, \mu_k(n)) = \min [d(\mathbf{x}_i, \mu_k(n)]$$

where $d(\mathbf{x},\mathbf{y})$ is the distance measure used. Then we recompute the cluster centers by finding the point that minimizes the distance for elements with each cluster. Thus,

$$\mu_k(n + 1): \Sigma\, d(\mathbf{x_i}, \mu_k\,(n+1)) = \min \Sigma\; d(\mathbf{x_i},\mathbf{y}), \qquad k = 1, \ldots, K.$$

The procedure is repeated for each $\mathbf{x_i}$, one at a time, until the clusters and their centers remain unchanged. If $d(\mathbf{x},\mathbf{y})$ *is* the Euclidean distance, then a cluster center is simply the mean location of its elements. If K is not known, we start with large values of K and then merge to K-1, K-2, ... clusters by a suitable cluster-distance measure. [3,4].

Figure 1, Original 1987 TM Image of KSC

An original portion of a 1987 TM image (using BANDS 1, 2, 3) taken of Kennedy Space Center for which we used in our analysis is represented in *Figure 1.* The objective is to identify major land cover classes, primarily vegetation, using the ISOCLUS algorithm. We suspect that using this type of classification scheme will yield at least three different distinctions: manmade structures, water, and vegetation.

III Classification Results:

Figure 2. shows the image after the classification analysis has been done using the ISOCLUS algorithm. *Figure 3* represents color composite of *Figure 2,* this was done to highlight the differences in contrast in order to more easily analyze and review the results. Careful examination of the results shows that the ISOCLUS algorithm successfully identified the those regions corresponding to water, manmade structures, and vegetation. The signature given for water is clearly given by the major black regions in *Figure 2 &3.* Manmade structures are easily depicted from the *Figure 2* and are given a signature in the white to light-level gray areas. The mid-level to somewhat dark shades of gray in *Figure 2* and blue-green regions in *Figure 3* correspond to vegetation. It is important to point out that classification of vegetation occurred in two different types of regions, meaning that the ISODATA algorithm recognized healthy mainland vegetation

637

as well as vegetation growing in lower wetland areas. Close observation of the upper left-hand portion of *Figure 2* illustrate this

Figure 2. Image after Clustering Analysis

Figure 3. Color Composite of Clustered Image

IV. **Conclusion:**

For the analysis of these images using the unsupervised technique vegetation signatures identified were only put into general class and were not discretely defined exclusively by vegetation type. Results obtained form these K.S.C. image primarily serve as a stepping stone for more extensive analysis using more complicated techniques. We will use the these results and incorporate them with fuzzy logic analysis to obtain an exclusive distinctions between vegetation types.

V. **Acknowledgments:**

VI. **References:**

[1]   D. Quattrochi and R. Pelletier *Remote Sensing for analysis of landscapes: An Introduction,* 1991.

[2]   F.F. Sabin, Jr.. *Remote Sensing Principles and Interpretation.* New York: W.H. Feeman & Co. 1987.

[3]   A.K. Jain. *Fundamentals of Digital Image Processing.* New York: Prentice f lal 1. Inc.1989.

[4]   T.M. Lillesan and R.W Kiefer. *Remote Sensing and Image Interpretation.* New York: Wiley 1979.

[5]   B. Wood, L. Beck, S. Dister, B. Lobitz, and V. Ambrosia. *"Readings in Remote Sensing, Geographic Information Systems and Related Technologies."* NASA Ames Research Center, 1996.

[6]   A. H. S. Sloberg, A. K. Jain, and T. Taxt, *"Multisource classification of Remotely Sensed Data: Fusion of Landsat TM and SAR Images".* IEEE Trans. Geosci., Remote Sensing, vol.   32 no. 4, p 768-777

[7]   L. E. Pierce, F.T. Ulaby, K. Sarabani, and M.C. Dobson *"Knowledge-Based Classification of Polarimetric SAR Images".* IEEE Trans. Geosci., Remote Sensing, vol.32 no.5, p 081-1086.

[8]   *Using PCI Software Vol. II.* PCI Inc. Version 5.3. Copyright 1 994.