Reprinted from

Symposium on Machine Processing of Remotely Sensed Data

June 27 - 29, 1979

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana 47907 USA

IEEE Catalog No.
79CH1430-8 MPRSD

CALIFORNIA DESERT RESOURCE INVENTORY USING MULTISPECTRAL CLASSIFICATION OF DIGITALLY MOSAICKED LANDSAT FRAMES

NEVIN A. BRYANT, RONALD G. MCLEOD, AND ALBERT L. ZOBRIST
Jet Propulsion Laboratory

HYRUM B. JOHNSON
Bureau of Land Management

ABSTRACT

Time and budget constraints have precluded the use of conventional mapping techniques for the U.S. Bureau of Land Management to produce a comprehensive resource inventory of the California deserts region as mandated by the California Desert Conservation Act. A Landsat mosaic at full resolution for ten scenes over the California deserts region was used to provide a continuous data base for multispectral thematic classification. Procedures for adjustment of brightness values between frames and the digital mosaicking of the Landsat frames to standard map projections were developed for this task and are discussed.

The principle of transect sampling was adopted as a means to obtain a uniform classification throughout the entire desert while generating one set of classification statistics. The transects were selected by the BLM science team to include all variable types in the landscape. The initial set of unsupervised statistical clusters was reduced by the BLM staff to 100 primary statistical clusters and applied to twelve small (512x512) test areas extracted from the Landsat mosaic. Subsequent to verification, classification was performed on the entire desert mosaic in 1°x1° segments. Resource class assignment was aided by including, in a post classification procedure, DMA/NCIC digital terrain elevation data (with derived measures of slope gradient and aspect) registered to the Landsat mosaic. The combination of local terrain variations and a global sampling strategy based on transects provided the framework for an accurate classification throughout the entire desert region.

*This paper presents one phase of research conducted at the Jet Propulsion Laboratory, California Institute of Technology, under NAS 7-100, sponsored by the National Aeronautics and Space Administration.

I. INTRODUCTION

The Bureau of Land Management (BLM) in the State of California has been mandated by Congress to prepare a comprehensive, multiple use management plan for the California Desert Conservation Area (CDCA) by October 1, 1980 (Federal Land Policy and Management Act, Section 601, 1976)[1]. The CDCA encompasses an area of approximately 25 million acres, or about one-fourth of the State of California. Resource information for this sparsely occupied portion of the state is incomplete in many areas and nonexistent in others. Time and budget constraints have precluded the use of conventional mapping and inventorying techniques for obtaining much of the needed resource data. The feasibility of using multispectral classification of Landsat imagery for the CDCA Plan was explored by representatives of the BLM Desert Plan Staff at the EROS Data Center in early 1977. In general the results were highly favorable and the conceived potential was found to greatly exceed the initial expectations. However, different systems and procedures needed to be implemented if an approach to uniform classification over the entire region was to be realized.

The BLM personnel turned to the Jet Propulsion Laboratory's Image Processing Lab to assist in preparing a controlled digital mosaic of Landsat frames for the entire CDCA prior to digital classification of spectral data and for the incorporation of digital terrain data in resource classification strategies. The artificialty of essentially all resource classification schemes (ground based and otherwise) was tacitly accepted at the outset. Thus, during the initial phases of establishing the training statistics little concern was shown for trying to establish spectral

CH1430-8/79/0000-0069$00.75

clusters that corresponded to "known" resource types. Typal classifications of natural biological resources (vegetation, habitats, etc.) are most often only intuitive in substance and seldom adequately quantified. This condition has contributed to difficulties for testing Landsat classification as observed by Maxwell[2] working with vegetation production in Colorado, Bentley et al[3] working with vegetation and soils in Arizona and Montana and Hutchinson[4] striving for an integrated survey method in the Mojave Desert of California. It appears that extreme variability within the ground control class areas commonly leads to improper evaluations of how well Landsat data can define resources.

A basic assumption made here is that spectral reflectance patterns are related in a rational way to resource characteristics and that from this relationship practical classification schemes can be developed. The first goal in proceeding with the digital spectral classification then became one of establishing a set of spectral clusters that would embrace the range of spectral reflectance intensities and patterns found in the data set for the CDCA. There is a need for such an approach to be fine grained enough, i.e. have enough clusters, to separate out essentially all signatures that might correspond to special purpose resource classes desired for practical applications. The underlying principle of this approach is the opposite of the null hypothesis most commonly used in statistics (type and error) but instead relies on minimizing the chances of keeping things together that are different (type 2 error). In terms of the splitter taxonomists, "It is a greater sin to call things that are different the same, than to call things that are the same different." In this case more than one spectral cluster representing a given special purpose resource type can easily be combined. A pixel transect sampling method was developed to assure uniform classification throughout the entire area of analysis while generating one set of classification statistics. An unsupervised, purely statistical classification of the Landsat data could then be employed which would deliver an initial classification of enough discrete classes that all resource cover types, as known to resource specialists, could be delineated. The combination of local terrain variations and a global sampling strategy based on transects was to provide the framework of spectral classification results that could be used more or less independently by vegetation, soils and wildlife specialists in defining their own class aggregation parameters for resource mapping.

II. LANDSAT MOSAIC DATA BASE

Considerable work has been done at JPL's Image Processing Laboratory developing geometric rectification and registration algorithms, and more recently, software has been developed for digital image map reprojection and image mosaicking. Image processing support for JPL's planetary program provides the basic software and procedures necessary to achieve digital image mosaicking[5,6]. Methods have been developed that not only geometrically correct and register images in the x and y directions but also correct and register images in the z domain (i.e. brightness). After corrections are applied to the three axes (x, y, and z) of each Landsat frame in the set of multiple images, a mosaicking algorithm is employed to construct a single larger mosaic image.

Planetary mission images, and the associated image rectification software differs from the Landsat MSS digital imagery in that reprojection of vidicon images to specific projections could be achieved by using the pointing statistics of the spacecraft and then fine-tuning the local misregistration between frames by applying a rubber-sheet geometric distortion correction algorithm a second time[7]. In the case of Landsat mosaicking, it was found that it was necessary to incorporate ground control points of known position in the image, as well as relative registration points to fine-tune the local misregistration between frames[8]. The reasons for incorporating known ground control points are twofold: a) the MSS multispectral scanner on Landsat is not a framing imaging system, so that continuous changes in pointing perspective geometry make it virtually impossible to reconstruct a perfect orthophoto image from calibration data; and b) the relative position of points on the earth's surface is precisely known, with the result that geodetic control points must be used as input to the geometric correction of the satellite image data if any satellite mosaic is to be expected to conform to the planimetry of existing maps. All the imagery used to construct the CDCA mosaic data base was, therefore, resampled in the x and y directions to accommodate Landsat sensor and orbital characteristics, local topographic offset effects, and to conform to the Lambert Conic Conformal Map Projection[8]. The pixel resolution selected was 80 meters square. The data were also resampled in the z domain to present a smooth surface that no longer reflected abrupt changes in sensor calibration or day-to-day side-lap differences associated with atmospheric effects.

The images used to generate the CDCA regional mosaic were ten Landsat scenes taken on four sequential cloud-free days in August, 1976. A detailed analysis of image products generated has shown that the brightness value adjustment (z correction) has successfully smoothed out any sharp edge effects between frames due to atmospheric differences at different days, thereby permitting the extension of thematic classification across Landsat frames[9]. Converting the image data to a specific map projection allows for easy, repeatable access to any portion of the mosaic. Coordinates of the vertices of an area requiring access are specified in latitude and longitude (degrees, minutes, seconds) which are in turn converted into the x and y positions of the image raster grid. This capability is useful for detailed inventory and analysis of sub areas, since the CDCA mosaic comprises some 50 million picture elements (pixels) per Landsat band (Figure 1). Such a corrected image grid permitted the overlaying of the National Cartographic Information Center/Defense Mapping Agency (NCIC/DMA) digital terrain files. The registration accuracy of both multispectral and terrain elevation files was sufficient to effectively utilize the elevation data as an aid to thematic classification.

III. OBTAINING CLASSIFICATION STATISTICS

Despite the fact that the mosaicking procedure's brightness value adjustment (a correction) had successfully smoothed out any sharp edge effects between frames due to atmospheric differences on different days, it was felt that assumptions of signature extension from selected ground plots across Landsat scene boundaries should probably not be made. Even though desert regions are characterized by stable atmospheric conditions for long periods of time, changes in plant species composition for similar habitats are known to develop over several degrees of latitude, as are reflectivity signatures for similar areas. A system of linear transect sampling was adopted as a means to assure uniform classification statistics with sufficient discrete classes to accommodate essentially all resource cover types defined by the BLM resource specialist. The initial pixel brightness values for all four MSS bands were extracted in two pixel wide swaths along approximately 3,000 miles of transects specified by the Bureau of Land Management science team to include all variable types in the landscape (see Figure 2). The transects traversed all types of terrain elevation, slope gradient, slope aspect, and the various cultivated and populated areas. The transect data were then aggregated into one small image per Landsat band which represented one-half of one percent (0.5%) of the entire CDCA. An unsupervised clustering algorithm was applied, which initially separated out 1993 clusters.

The unsupervised clustering algorithm, which builds up clusters as it passes through the training data was structured to maximize the number of initial clusters. The initial cluster data set displayed the very broad range in reflectivity values found in the California deserts regions, which range from highly reflective white sands and playas to dark basalts, and irrigated agriculture to forested mountain areas, as well as the diverse variety of rangeland and dry cropland conditions. Much of the CDCA is sparsely vegetated, i.e. has less than 10% plant ground cover, and maximum ground cover seldom exceeds 50% except in cultivated lands which were of little concern to BLM scientists. Wildland areas, with the higher percentage ground cover values, though comparatively small in size, are of great significance in terms of total plant productivity and related resource values. The task of reducing the 1993 clusters to a manageable number became a problem of devising a way to retain significant but small clusters having strong vegetation components in their signatures. It is generally known that such clusters should be those with comparatively low reflectance in the red band (MSS 5) and high reflectance in the infrared bands (MSS 6&7) (Maxwell[2], Bently et al[3]). The BLM science team developed a systematic procedure to reduce the initial 1993 clusters into 100 principal clusters while maintaining the critical elements needed to define landscape diversity. To that end, the 1993 initial clusters were first separated into two partitions based upon the strength of the vegetation component of the spectral signature as indicated by the brightness ratio of MSS band 6 to MSS band 5. If the ratio of MSS band 6 to MSS band 5 was greater than or equal to one, the clusters were determined to signify a chlorophyll type signature and if the ratio was less than one, the clusters determined a nonchlorophyll signature. This basically broke the clusters into vegetation reflecting (684 clusters) and non-vegetation reflecting signatures (1309 clusters). To further reduce the amount of clusters each resultant group was broken down into two partitions. Of the vegetation reflecting signatures where MSS band 6 was greater than MSS band 5, these signatures were divided into groups where MSS band 7 was greater than or equal to MSS band 5 (very high vegetation signature content; 363 clusters) and where MSS band 7 was less than MSS band 5 (high vegetation signature; 321 clusters). The non- or low vegetation clusters were divided on the basis of whether MSS band 5 brightness was greater than 130 (on a pixel reflectance scale of 0-255), (very bright surface conditions, e.g. playas and sand; 690 clusters) and whether MSS band 5 brightness was less than or equal to 130 (dark surface features, e.g. basalt flows, shadows, 619 clusters). This resulted in partitioning the 1993 initial clusters in four groups representative of four major signatures types in the CDCA.

The reduction of the clusters to approximately 25 classes within each of the four partitions was performed by first ranking clusters in accordance with their brightness coefficient (Figure 3) and then by merging overlapping clusters and deleting the population and associated statistics of clusters where the mean value occurred within the one standard deviation ellipse of a cluster with a larger population. In order to assure that population size and standard deviation about the mean value of a cluster did not predominate in the cluster merging process, each partitioned group was treated individually. Important low frequency pixel clusters representing strong vegetation signatures were thus retained to be used in the classification process (Table I).

IV. CLASSIFICATION

The final one hundred clusters selected for the CDCA were then used as input to a hybrid Bayesian/Parallelepiped type classifier and applied to four primary test areas comrpising over 400,000 acres each[11]. These areas were 25.4 miles (40.9 km) on a side or 512x512 pixels. Eight secondary test areas were also used for classification verification. The test areas were selected by the BLM Desert Plan science team for their diversity and representativeness of the varieties of landscape found throughout the CDCA. The four primary test areas were intensively studied and analyzed by the BLM staff in their office by using black and white film overlays of each class on color transparencies of the raw Landsat data. As general patterns of class distributions emerged names were given to the classes. Several class combinations for both primary and secondary test areas were then displayed on an interactive display system at JPL and analyzed by the BLM science team to permit the assignment of each class or class combination to either a plant community, a soil series, or wildlife habitat type. Class assignment was aided by extensive supplemental ground truth collected by the BLM staff through field surveys and low altitude aerial photography. The ground truth information helped spot inconsistencies in classes from test area to test area. It was found, for instance, that with the aid of registered NCDC/DMA elevation data, and derived slope aspect data, many of the inconsistencies in class assignments were resolved. Changes in elevation stratification of communities with similar reflectance characteristics often occurred between those test areas in the high desert and those in low desert areas. In a similar manner, problems such as dark desert pavement and shaded north slopes in mountains could be differentiated.

Once the BLM science team was assured of the proper class assignments to soils, vegetation, and wildlife habitat types, the respective maps were generated for each of the 22 quadrangular areas encompassing one degree of latitude and longitude in the CDCA (Figure 4). As the results of the 12 test areas analyses showed a systematic drift in signature characteristics from north to south, the quadrangles were each assigned to one of four quadrangle groups. Each quadrangle is considered as an individual unit for further analysis and resource inventorying. The resouce data for the area had now been "packaged" into manageable sizes for more detailed investigation during the comparative study phase needed to draw up a comprehensive plan for use of the CDCA. Furthermore, the 1°x1° quadrangle data sets can be easily accessed on BLM minicomputer systems in the future.

V. CONTINUING WORK

Two further phases of the project are underway at the time this report has been written. One involves a verification procedure that goes beyond that already performed at the 12 test areas, and the other involves the derivation of resource inventory statistics by administrative areas. The verification procedure involves the integration of approximately 500 large scale (1:2000) low altitude color photograph strips approximately 750x2000 meters each. The air photo strips were taken throughout the CDCA, and will be used to both verify further the classification procedure used with the Landsat mosaic and derive more specific estimates of biomass concentration and plant species distribution. The latter measures will help the decision making process for both wildlife habitat and range carrying capacity assessments. The resource inventory statistics will be aggregated according to land ownership and areas leased for grazing range allotments for each 1°x1° quadrangle. The procedure used involves the application of the JPL created Image Based Information System[12] to generate georeference planes for the administrative areas requested and then through the application of image overlay software derive the resource inventory acreage statistics and related coefficients of range carrying capacity obtained from the 500 low altitude photo test areas.

VI. REFERENCES CITED

1. Federal Land Policy and Management Act, Section 601, Public Law 94-579, The 94th U.S. Congress, 1976.

2. Maxwell, Eugene L., "A Remote Rangeland Analysis System," Journal of Range Management, 1976, Vol. 29, pp. 66-77.

3. Bentley Jr., Gordon, Bette C. Salmon-Drexler, William J. Bonner and Robert K. Vincent, Landsat study of ephemoral and perennial rangeland vegetation and soils, United States Department of Interior, Bureau of Land Management Report sponsored by National Aeronautics and Space Administration under contract number S-53766-A, December, 1976.

4. Hutchinson, Charles F., "Digital Use of Landsat Data for Integrated Land Resource Survey: A Study in the Eastern Mojave Desert," Ph.D. Dissertation, University of California, Riverside, 1978.

5. Elliott, D.A., "Digital Cartographic Projection," Proceedings Caltech/JPL Conference on Image Processing Technology, November 3-5, 1976, pp. 5-10.

6. Johnson, T.V., J.A. Mosher and D.L. Matson, "Lunar Spectral Units: A Northern Hemisphere Mosaic," Proceedings of the Eighth Lunar Science Conference, 1977, pp. 1013-1028.

7. Rofer, C.A., A.A. Schwartz and J.B. Seidman, Computer Cartographic Projections for Planetary Mosaics, 3 volumes, JPL Internal Document 900-639, Jet Propulsion Laboratory, California Institute Technology, Pasadena, California, November 15, 1973.

8. Zobrist, A.L., "Multiple-Frame, Full-Resolution Landsat Mosaicking to Standard Map Projections," Proceedings American Society of Photogrammetry Fall Technical Meeting, Albuquerque, N.M., October 15-20, 1978, pp. 608-616.

9. Zobrist, A.L. and Bryant, N.A., "Map Characteristics of Landsat Mosaics," Proceedings American Society of Photogrammetry Annual Convention, Washington, D.C., March 19-23, 1979, pp. 260-273.

10. Meuller-Domois, Dieter and Heinz Ellenburg, Aims and Methods in Vegetation Ecology, John Wiley & Sons, New York, 1974, pp. 212-217.

11. Addington, J.D., "A Hybrid Classifier Using the Parallelepiped and Bayesian Techniques," Proceedings of the American Society of Photogrammetry, 41st Annual Meeting, March 1975, pp. 772-784.

12. Bryant, N.A. and A.L. Zobrist, "IBIS: A Geographic Information System Based on Digital Image Processing and Image Raster Datatype," Proceedings of the LARS Symposium on Machine Processing of Remotely Sensed Data, Lafayette, Indiana, 29 June 1976, p. 1 A-1.

Table 1

RELATIVE PERCENT OF PIXELS AND CLUSTERS IN THE BAND RATIO-BRIGHTNESS CATEGORIES FROM THE TRANSECT SAMPLE TRAINING SET

	LANDSAT BAND RATIO-BRIGHTNESS CATEGORIES			
	1	2	3	4
	Band 5 ≤ Band 6 Band 5 ≤ Band 7	Band 5 ≤ Band 6 Band 5 > Band 7	Band 5 > Band 6 Band 5 ≤ 130	Band 5 > Band 6 Band 5 > 130
Percent of Pixels	6%	12%	52%	30%
Percent of Clusters	18%	16%	35%	31%

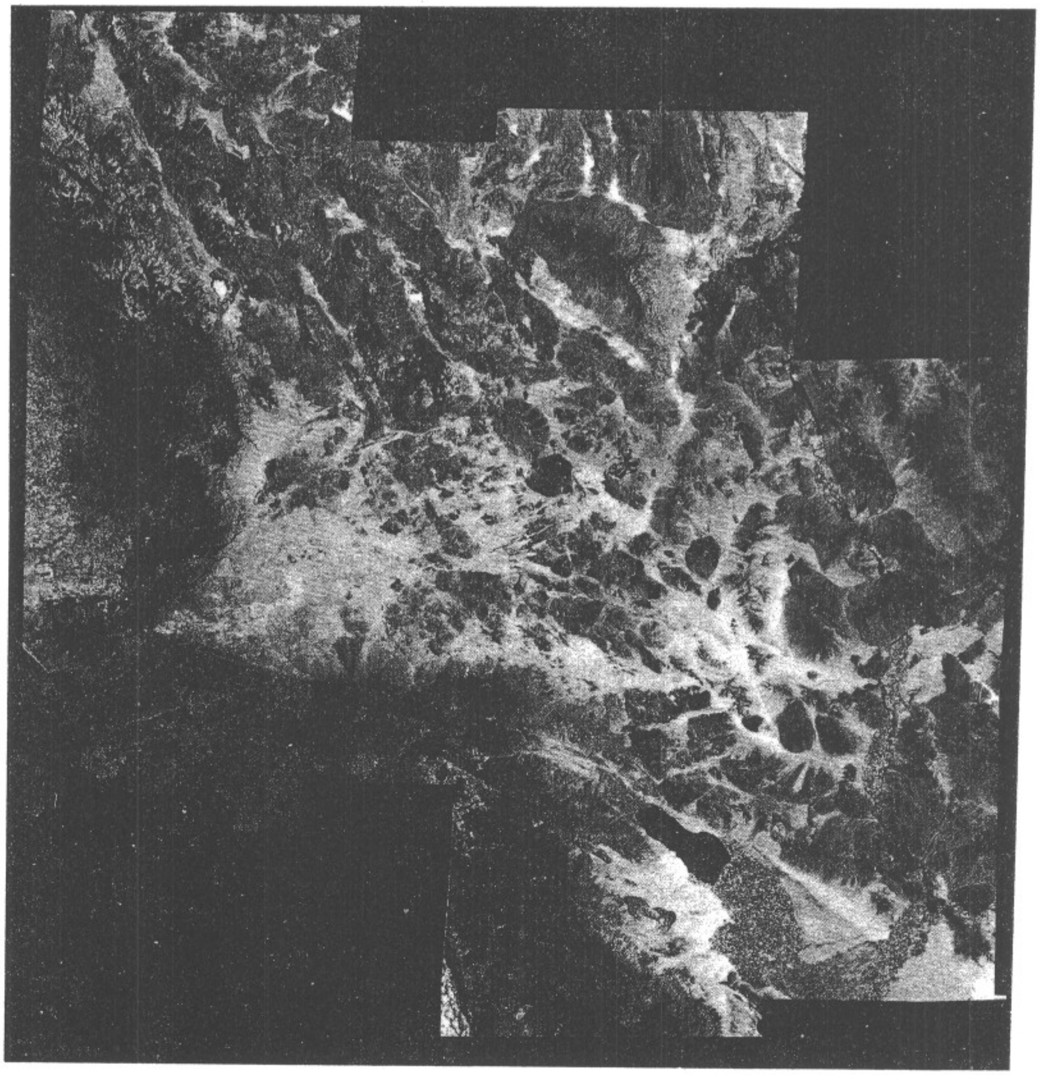

Figure 1. Ten frame Landsat mosaic encompassing the California Desert Conservation Area.

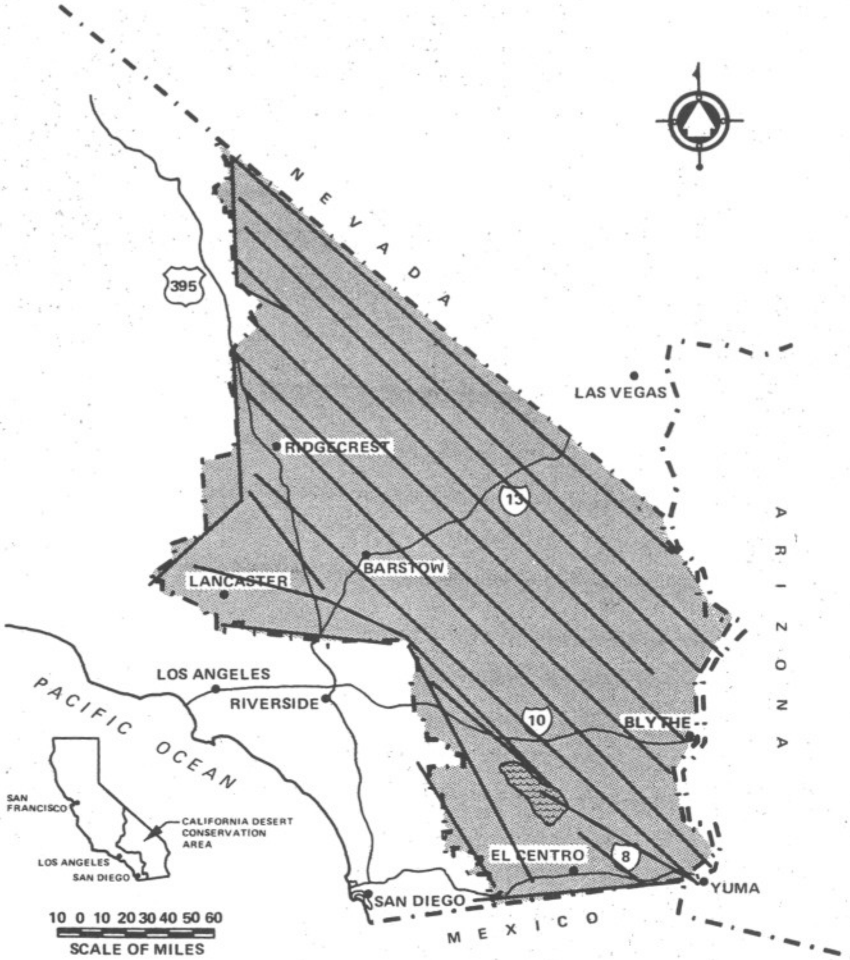

Figure 2. Location of Transect Lines Used to Gather Training Sample Statistics for California Desert Conservation Area Multispectral Classification.

DETERMINING BRIGHTNESS COEFFICIENT

1. For Clusters 1-1993 $\frac{B_4 + B_5 + B_6 + B_7}{4} = X$

2. Sort x in Ascending Order to Find Minimum
   Minimum Aggregate Cluster = $\sum x_{MIN_{1-4}}$

   CL_{MIN} = MIN 1, MIN 2, MIN 3, MIN 4
   B_4 B_5 B_6 B_7

3. For Clusters 1-1993 compute:

$$B_4 - MIN_1 = M\ 1$$
$$B_5 - MIN_2 = M\ 2$$
$$B_6 - MIN_3 = M\ 3$$
$$B_7 - MIN_4 = M\ 4$$
$$\sum MIN_{1-4} \sum M_{1-4}$$

4. Brightness Coefficient For Each Cluster:

$$CL = \frac{\left(\sum M_{1-4}\right) \bullet 2}{\left(\sum M_{1-4}\right) + \left(\sum MIN_{1-4}\right)}$$

Figure 3. Procedure Used to Determine Brightness Coefficient (after Meuller-Dombois & Ellenberg (10))

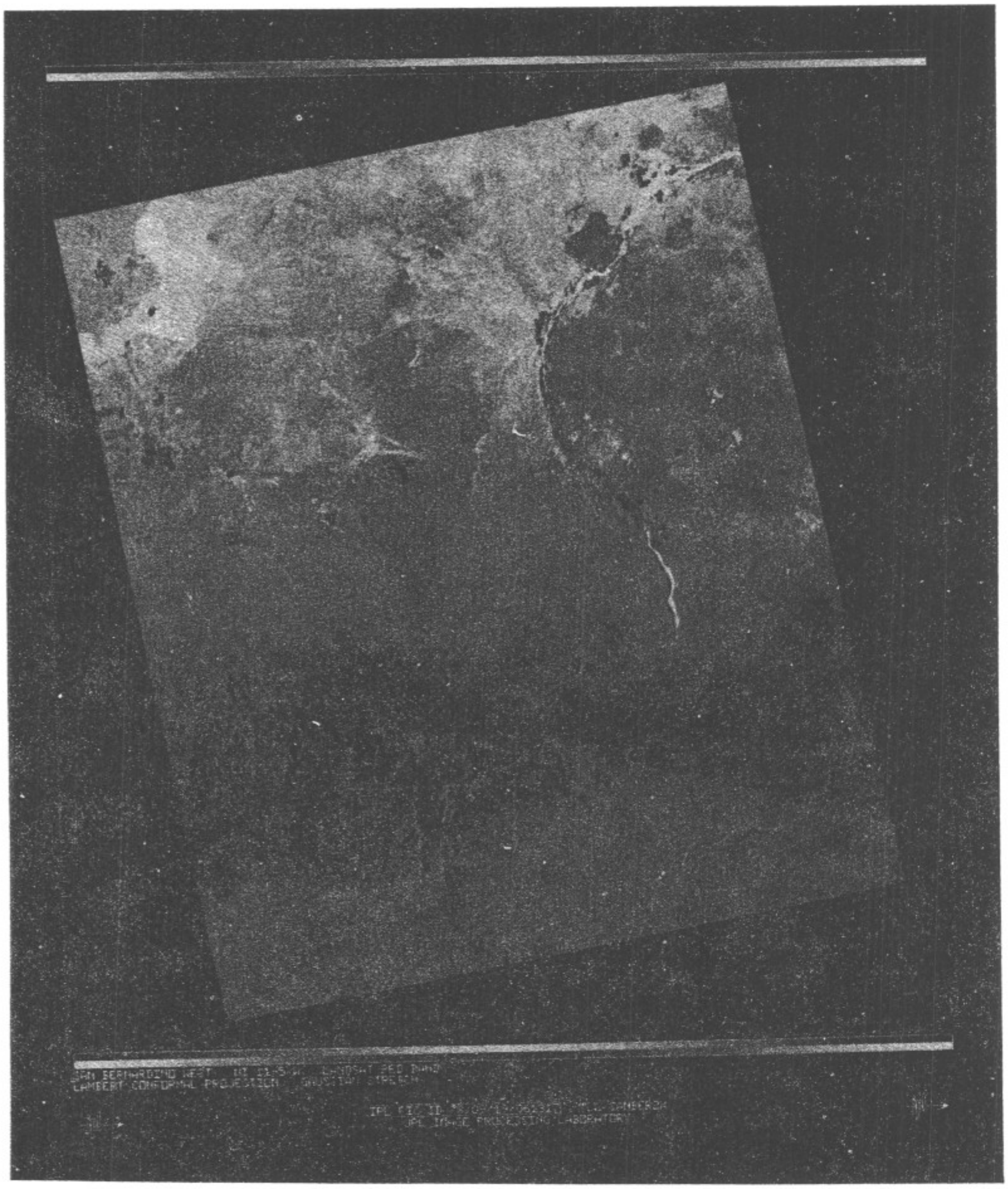

Figure 4. Sample one degree latitude, one degree longitude quadrangle of data extracted from CDCA mosaic. Quadrangle corresponds to U.S. 1:250,000 map series NI 11-5/W, San Bernardino West.

Nevin A. Bryant is responsible for the conceptualization and application of geographic information systems based on digital image processing technology and the use of remotely sensed imagery at the Jet Propulsion Laboratory's Image Processing Laboratory. Dr. Bryant received his B.A. degree from McGill University in Honours Geography; his M.A. degree from the University of Hawaii in Geography, and his Ph.D. from the University of Michigan. Dr. Bryant has conducted field research in Western Samoa and Java, Indonesia related to the interaction between population growth and regional economics and has more recently undertaken projects related to vegetation and land use mapping. His present interest include data reduction and interpretation of satellite imagery, geographic information systems, and regional economic modelling with particular emphasis on metropolitan areas.