E77-10214
CR-152574

UCD

ORIGINAL PAGE IS
OF POOR QUALITY

(E77-10214) A STUDY OF THE USE OF REMOTE SENSING DATA IN HYDROLOGIC ENGINEERING MODELS Progress Report (California Univ.) 41 p HC A03/MF A01 CSCL 08H

N77-31568

Unclas
G3/43 00214

Progress Report on NASA Grant NSG 5092

A Study of the Use of Remote Sensing Data in Hydrologic Engineering Models

Department of Electrical Engineering
University of California, Davis

January 1977

Progress Report on NASA Grant NSG 5092

"A Study of the Use of Remote Sensing Data in Hydrologic Engineering Models"

Department of Electrical Engineering
University of California, Davis

Principal Investigator: V. R. Algazi

Co-Investigator: M. Suk

Contributors: G. Ford, K. Ellis, J. Stewart, R. Eid
M. Duncan

CONTENTS

I. Introduction

This research was started in January 1976 and has as a goal to develop and examine the use of remote sensing data (LANDSAT and others) in the acquisition of data, and the calibration of widely used hydrologic computer models of the Corps of Engineers. Objectives of the current phase of the research are to incorporate existing data processing methodologies used in remote sensing to the specific data needs of the hydrologic models, and to determine the limitations in the use of remote sensing data currently available for such hydrologic models. We have specifically focused on the well developed methodologies for the quantification of land use by remote sensing.

A general underlying concern is to consider the suitability of the procedures used to the equipment and capabilities available to the Corps of Engineers. On these grounds, highly interactive methods requiring high quality image displays are not suitable. Thus, since our last interim report, the emphasis in our work shifted from maximum likelihood classification, a supervised technique, to clustering, an unsupervised technique, which seems more likely to provide accurate land use classification results with a limited amount of interaction. This clustering approach has been applied to the Trail Creek Watershed with a substantial resulting improvement on land use classification accuracy as compared to maximum likelihood. The clustering approach has also been applied to the Castro Valley Watershed providing us with a preliminary land use classification. This annual report concentrates on follow up results since our July 1976 report and considers successively:

1. Previous work on the Trial Creek Watershed.
2. Use of clustering for land use classification.
3. Karhunen-Loeve transformation or principal component analysis of the data.
4. Incorporation of spatial information or texture into classification algorithms.
5. Outline of a tentative operational procedure for land use classification.
6. Classification results for the Trial Creek Watershed.
7. Ground truth for the Castro Valley Watershed.
8. Preliminary classification result for the Castro Valley Watershed.
9. Discussion of the work done and work proposed for the following phase.

II. The Trail Creek Watershed

The first watershed which has been under study is the Trail Creek Watershed located in Athens and Clarke County, Georgia. The watershed is relatively small, approximately 12 square miles, and has been further subdivided into 21 subbasins in an HEC study.

HEC has recently completed a comprehensive Flood Plain Information (FPI) study for the watershed. In the study, data management and analytical techniques, emphasizing consistent comprehensive assessments of the effects of alternative land use patterns on the flood hazard, general damage potential, and environmental status of the study area, have been developed.

In our work we have applied several well developed machine classification algorithms and have examined the classification results for the quantification of land use by remote sensing. As indicated earlier, emphasis has

)een given to the development of a procedure which appears operationally
isable by the Corps of Engineers.

In this section, we shall first briefly review some previous work
reported in the July interim report. The work had been mainly concerned
with the acquisition of remote sensing data and of ground truth, and with
:he application of the maximum likelihood classification algorithm. We
:hen discuss our shift in emphasis to classification using a clustering
ipproach, which is an unsupervised classification scheme. We then
)resent new results obtained by this approach, and compared them to
jround truth as well as to the results obtained using maximum likelihood
:lassification.

A. Ground Truth

After giving consideration to several watersheds, we chose, in
:onsultation with HEC, the Trail Creek Watershed as the first test watershed.
This choice was based on two facts: the availability of land use informa-
:ion to serve as ground truth and an extensive Flood Plain Information
study done by HEC. As we proceeded in our work we reached the conclusion
that the land use pattern published in the HEC report was inadequate as
jround truth because of obvious and significant discrepancies with the
information available in recent detailed aerial photographs of the water-
shed to which the satellite remote sensing information should logically be
:ompared. We are uncertain as to all the causes for the poor quality of
the land use information published by HEC, but this might be partially
explained by the fact that the emphasis of the FPI study was on techniques
and their applications rather than on the specific development of the flood
)lain information.

Because of the need to establish a basis of comparison for our work using satellite data we had to undertake the substantial additional task of obtaining ground truth land use information. We proceeded to do a manual classification of the land use in the watershed using high flight photographs. This is a time consuming and costly procedure which was not planned at the onset and which was required by the lack of suitable ground truth. Since our manual classification appears to be consistent and based on recent photographs close in time to available LANDSAT imagery, it will serve as a principal basis for the verification of remote sensing classification results throughout the study. The result of the manual classification is displayed as a color image in Figure II-1. We also tabulate the percentages of each land use class in Table II-1.

Land Use	Percent of Areas
Natural Vegetation	50.17
Developed Open Space	.49
Low Density Residential	2.45
Medium Density Residential	6.79
High Density Residential	.11
Agricultural	28.73
Industrial	2.59
Commercial	1.55
Pasture	3.04
Water Bodies	.57
Trailer Parks	2.47
Highways	1.06

Table II-1. Areal percentages of land use classes from manual classification of high flight photograph. Trail Creek, Georgia.

Figure II-1. Trail Creek Watershed existing land use - ground truth.

B. Maximum Likelihood Classification

In a first attempt to machine classification of land use from remote sensing data, we made use of the CALSCAN program.

CALSCAN is an RSRP* version of LARSYSAA, the LARS- Purdue Image classification Program rewritten for the CDC 66-7600 computer system at the Lawrence Berkeley Laboratory (LBL). The program is basically a maximum likelihood classifier, which is a supervised classification algorithm. The need for accurate training fields is one of the major difficulties encountered in the use of the CALSCAN program. The CALSCAN program was chosen as our first classification algorithm because of the availability of LBL computers to HEC via a telephone line as ~~well as because of~~ the well established application of maximum likelihood classifiers in agricultural land use classification.

We have attempted to classify an October scene of LANDSAT image using the CALSCAN program. The steps in the classification procedure are: (1) define a reasonable set of land use categories, (2) locate one or several training fields for each class on LANDSAT imagery and identify the coordinates of each training field, (3) run the CALSCAN program, and (4) process the result for display and for tabulation of results. The classification results are displayed in Figure II-2 and summarized in Table II-2.

---

* RSRP: Remote Sensing Research Project, University of California at Berkeley.

Figure II-2. Machine classification of land use pattern using a maximum likelihood classification algorithm - the Trail Creek Watershed.

Land Use	Percent of Areas	
	Ground Truth	Remote Sensing Data
Natural Vegetation	50.17	36.19
Dev. Open Space	.49	8.82
Agricultural	28.73	19.17
Pasture	3.04	
Residential (low density)	2.45	23.78
Residential (medium density)	6.79	
Residential (high density)	.11	
Commercial	1.55	1.97
Industrial	2.59	6.08
Water Bodies	.57	1.02
Trailer Parks	2.47	2.98
Highways	1.06	

Table II-2. Areal percentages of land use classes as determined using a maximum likelihood classifier and comparison with ground truth. T. C. Georgia

C. The Clustering Approach to Machine Classification of Land Use

In the July report, we pointed out a number of difficulties encountered in the use of a maximum likelihood classifier. Among these are:

(1) Choice of land use categories.

A maximum likelihood classifier is a supervised algorithm and it requires a predetermined set of land use categories on which the classifier will be trained (acquire statistics). The set of land use categories should be complete in the sense that every portion of the watershed should belong to one of these categories. There are several difficulties associated with this approach:

(a) It is not always easy to predetermine a complete set of land use categories which encompasses all of the actual land uses of the watershed.

(b) What the machine measures may be significantly different from what a person perceives or interprets. Human beings tend to lump things together spatially, which is difficult to do by machine. The maximum likelihood classifier, trained on a predetermined set of land use categories based on human perception of classes may be forced to make serious mistakes. For example, a machine cannot provide the intended or spatially composite use of land (functional land use) such as a school which is a mixture of buildings and of open fields.

(c) Most supervised classifiers including the CALSCAN program are based on the statistics of the training areas and assume unimodal distributions for each land use category. This requires that each land use category should be reasonably homogeneous in nature, and thus the differentiation of all possible subclasses within a land use class. It is often difficult, for example, to predetermine all subclasses within an agricultural land use class composed of several crop types.

(2) Training areas.

As we already mentioned, the CALSCAN program requires the estimation of the statistics of training areas. It is well known that to have a reliable estimate of statistics, a large number of sample points (corresponding to a large size training area) is required. In actual applications, it is not easy to find training areas of large size for certain land use categories. Further, the

determination of the exact outlines and coordinates on the LANDSAT data for each training field is also difficult, and is commonly done interactively, by examination of partial results presented as color images.

Considering these difficulties and our objective to develop an operational procedure with a minimum amount of interaction, we shifted emphasis from supervised to unsupervised classifiers. The clustering approach is a well known unsupervised classification algorithm, which does not require a priori knowledge of land use categories nor locating training areas. It also appears relatively easy to implement as an operational procedure with limited interaction. The clustering approach to land use classification is based on classifying first the data into machine classes or clusters according to machine measure of homogeneity without injecting into the process the human preconception of what the land use categories should be. Then human being interacts with the machine to interpret and refine the result of the machine classification. At this second stage, the prior knowledge of land use and the relative importance of achieving accurate classification results for each land use category play an important role.

There are several well known clustering algorithms available, but we chose to use the ISOCLAS program, mainly because it has already been implemented on CDC 66-6700 computer system at the Lawrence Berkeley Laboratory by the RSRP group. To describe the basic ideas of ISOCLAS program, we directly quote the Introduction part of the ISOCLAS USER'S MANUAL published by RSRP.

ISOCLAS

"This program performs a modified version of the clustering algorithm known as ISODATA to multispectral scanner data. The acronym ISODATA stands for Iterative Self-Organizing Data Analysis Technique (A). As its name implies, the algorithm is an iterative procedure which groups similar 'objects' into sets called clusters. The algorithm was originally developed by Ball and Hall of Stanford Research Institute and used in their PROMENADE system. (See References 1 and 2 for articles written by Ball and Hall on this subject.) A clustering technique based on ISODATA and suitable for JSC's* use in processing multispectral scanner data, was developed by E. Kan and A. Holly (LEC)*. To distinguish between the original and revised programs it was decided to call JSC's version of the clustering program ISOCLS (Iterative Self-Organizing Clustering Program). At RSRP the program is called ISOCLAS.

The procedure will, ideally, separate all of the data into distinct groups or clusters, the center of each cluster being represented by its mean. The process is initialized by assigning each data point to the nearest estimated cluster center (absolute distance is calculated to each cluster mean). After assigning all of the data to clusters, new means are calculated and tests are made to see if clusters should be split or combined. A cluster is split if the standard deviation of the cluster exceeds a specific threshold value. Two clusters are combined if the distance between the cluster centers is smaller than the specified threshold. A cluster is deleted if it has fewer than some specified number of points. The data is reassigned after each split or combine iteration to the new clusters and the process continues until the desired number of iterations

---

* JSC: NASA Johnson Space Center.
LEC: Lockheed Electronics Company, Inc.

has been obtained."

There are five parameters of importance in ISOCLAS program: (1) STDMAX - threshold for splitting clusters, (2) DLMIN - threshold value for combining clusters, (3) MNIN - minimum number of data points allowed per cluster, (4) MAXCLS - maximum number of clusters allowed, and (5) ISTOP - maximum number of iterations.

D. Importance of Spatial Information in Classification.

It is known that spatial information plays an important role in the human perception and pattern recognition. Recently, a great deal of attention has been given to the utilization of the spatial or textural information in remote sensing data to improve the accuracy of machine classification. We briefly describe some approaches used to incorporate spatial information into classification procedures.

(1) ECHO (Extraction and Classification of Homogeneous Objects)[3].
The procedure is based on partitioning the data in the spatial domain into "objects" which are groups of spatially contiguous data points belonging to a homogeneous class. These "objects" are then classified using one of well known classification algorithms, e.g. maximum likelihood classification.

(2) Spatial Clustering [4].
The procedure is based on clustering the thresholded gradient images. Several gradient algorithms have been used to generate gradient images*.

(3) Utilization of textural information [5].
Textural information is based on the spatial distribution of spectral responses of remote sensing data. Measures of textural information such as angular second moment, contrast, correlation, entropy and many more have been defined.

Recently, Wiersma and Landgrebe [6] have shown that the utilization of spatial information improves the classification accuracy significantly over conventional classification algorithms such as maximum likelihood classification without using spatial information. We have noted the importance of spatial information as well, and we are currently investigating the possibility of incorporating spatial information into our proposed operational procedure. Texture may be used before clustering as an additional feature or after classification as a means to consolidate and refine the results.

E. Karhunen - Loeve Transformation (Principal Component Analysis) of Remote Sensing Data

The LANDSAT images are composed of four different spectral responses. Thus, each point in the digitized LANDSAT image can be considered as a four dimensional vector. In the machine classification of remote sensing data, the speed and accuracy of the algorithm are largely dependent on the dimensionality of the data. Furthermore, when we decide to use spatial information or textural information as additional features, we increase the dimensionality of the data. This results in a large volume of data to be processed, especially for large watersheds. It is obviously desirable to reduce the amount of data to be processed while retaining most of the information content. This dimensionality reduction problem is solved by an approach called the Karhunen-Loeve (KL) transformation analysis in Electrical Engineering and as principal component analysis in statistics.

The KL transformation can be incorporated into our classification procedure as a means to reduce the amount of data. The original LANDSAT image composed of 4 Bands, B4, B5, B6, and B7 is transformed into four new components KL1, KL2, KL3, and KL4 by a linear transformation. The first

two components KL1 and KL2 can be used for classification without a significant loss of information. Furthermore, most data errors in LANDSAT image, such as random noise and periodic banding effect due to sensor nonuniformity are transformed into the last two KL components. Thus, by using only the first two KL components, the effect of data error can be reduced significantly.

F. Towards an Operational Procedure

An outline of a tentative operational procedure for land use classification is given here. The steps and sources of data for such a procedure will probably be as follows:

a. Digital Specification of the Watershed. Information on the watershed obtainable from maps, such as the watershed boundary and major roads, is entered on a grid oriented data base using an x-y digitization tablet or by key punching the data on cards. A major portion of this important step is being implemented at the UC Davis Image Processing Facility.

b. Preprocessing of the Data. The original LANDSAT data is transformed using a KL transformation. This step is optional and can be bypassed for a small watershed. KL transformation algorithms have been implemented on the LBL computers as well as at UC Davis.

c. Clustering. Currently the ISOCLAS program is used for clustering of data. We are contemplating at this time to implement some other clustering algorithms or to develop one of our own. The ISOCLAS program is implemented on the LBL computers by the RSRP group of UC Berkeley.

d. Classification. The data is classified after clustering by labelling each cluster as belonging to one of the land use categories. This requires ground truth information in the form of maps and

(high-flight) photographs. After examining all the available information such as the display of the centers of resulting clusters, maps, and aerial photographs, one of the following decisions is made. (1) The cluster belongs to a specific land use category. (2) It is a mixture of two or more land use categories or the information at hand is not sufficient to label the cluster, i.e., the cluster is either in conflict or inconclusive in nature. (3) The cluster is of no importance or not valid. We then assign that cluster to an "other" category (none of the desired land use classes). The ground truth information such as maps and aerial photographs is used to label the clusters in the following manner. From the examination of the computer printout of clustering results, several spatially contiguous areas (each having more than M points) within each cluster are chosen and the corresponding LANDSAT data is brought in registration with maps and aerial photographs. By studying corresponding areas on all available data we make one of three decisions for each cluster as outlined above. We can also use the reverse process, i.e., define some ground truth points or areas on maps and photographs, and transform those points or areas to LANDSAT image coordinates. We can then label the data clusters. The registration procedure of maps, aerial photographs and LANDSAT data will be discussed later in this section.

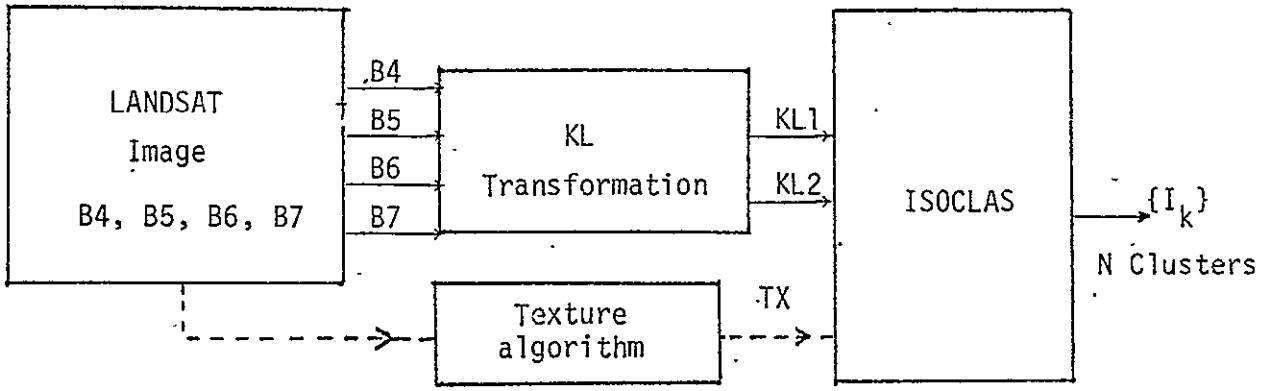
e. Reclustering. For the points belonging to the second group of clusters, i.e., clusters in conflict or inconclusive, a reclustering step is applied. First points belonging to this group are selected from the original LANDSAT data. Then the ISOCLAS

program or a similar clustering algorithm is applied again to those points. The purpose of this reclustering is to more finely subdivide the data in the difficult areas to allow unequivocal labeling of clusters. After reclustering, all the resulting clusters are labeled using the procedures described in Step d, with the only difference that we now try to label all clusters. Clusters which cannot be labeled properly are assigned to the "other" class. However, we expect that very few points will belong to this group. A program to mask out and select part of the original LANDSAT data is implemented on the LBL computers. A schematic diagram showing the steps above is given in Figure II-3.

f. Textural Information: The textural information of the remote sensing data can be incorporated into our procedure by modifying the clustering and reclustering steps above. This modification is shown by dotted lines in Figure II-2. A program to evaluate textural information has been implemented on the LBL computers.

g. Geometric Correction and Registration of Maps, Aerial Photography and LANDSAT Data. This geometric correction, using principally a least square geometric correction program, will require that a number of control points be obtained from all the sources of data and entered in numerical form into a program. Obtaining such ground control points for LANDSAT data is an important problem which remains to be solved for the case in which no high quality, high resolution display of LANDSAT data is available. Most of this step is currently implemented at the UC Davis Image Processing Facility.

Preprocessing and Clustering



Classification

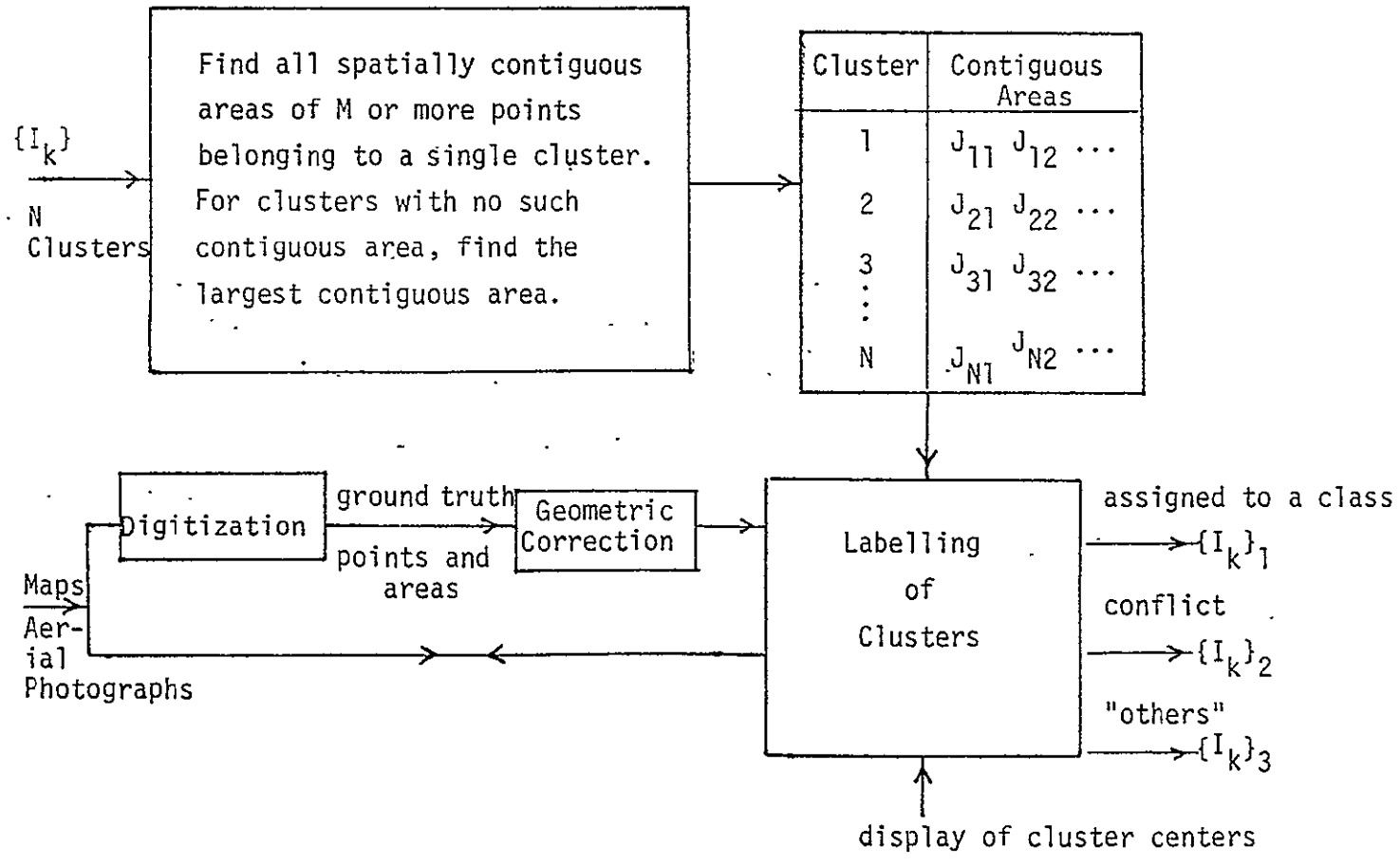


Figure III-2. A proposed operational procedure for land use classification.

Reclustering

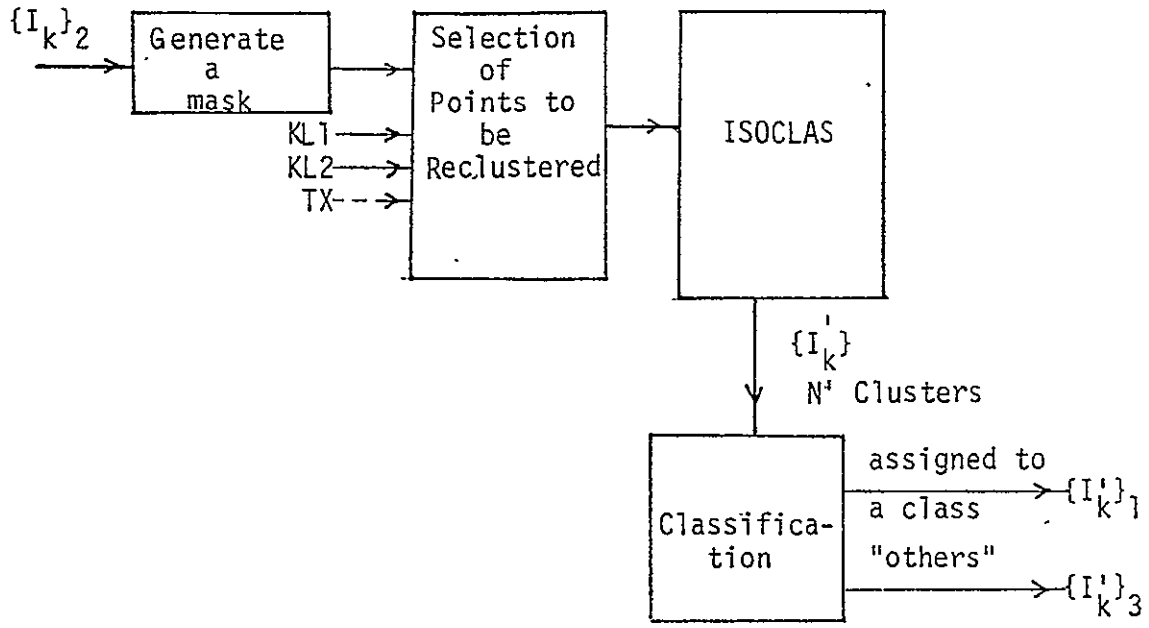


Figure II-3. (Cont.)

G. New Results on Land Use Classification

We have attempted to classify the remote sensing data of the Trail Creek Watershed using the clustering approach. The same October scene used in maximum likelihood classification was used here again. In the following, we describe the steps used and compare the results with the ground truth and the results of maximum likelihood classification.

(1) KL transformation of the data.

The original LANDSAT image is transformed for data compression using programs implemented on the LBL computers.

(2) Clustering of the data.

For parameter values STDMAX = 2.5, DLMIN = 2.5, NMIN = 60 MXCLS = 40 and ISTOP = 30, the first two components of transformed data KL1 and KL2 are clustered using the ISOCLAS program. A display of the centers of resulting 34 clusters are given in Figure II-4.

For the purpose of comparison, we tried to label all clusters with land use categories as best as we can without using reclustering. The result of the classification is given in Figure II-5 and Table II-3(a).

(3) Initial Classification

As described in the step d of proposed operational procedure, we divide the 34 clusters into three groups shown also in Figure II-4.

(4) Reclustering

For the clusters marked "reclustering" in Figure II-4, we applied the ISOCLAS program again using as parameters STDMAX = 1.5, DLMIN = 2.5, NMIN = 10, MAXCLS = 20 and ISTOP = 30.

The final result of steps (3) and (4) are shown in Figure II-6 and Table II-3(b).

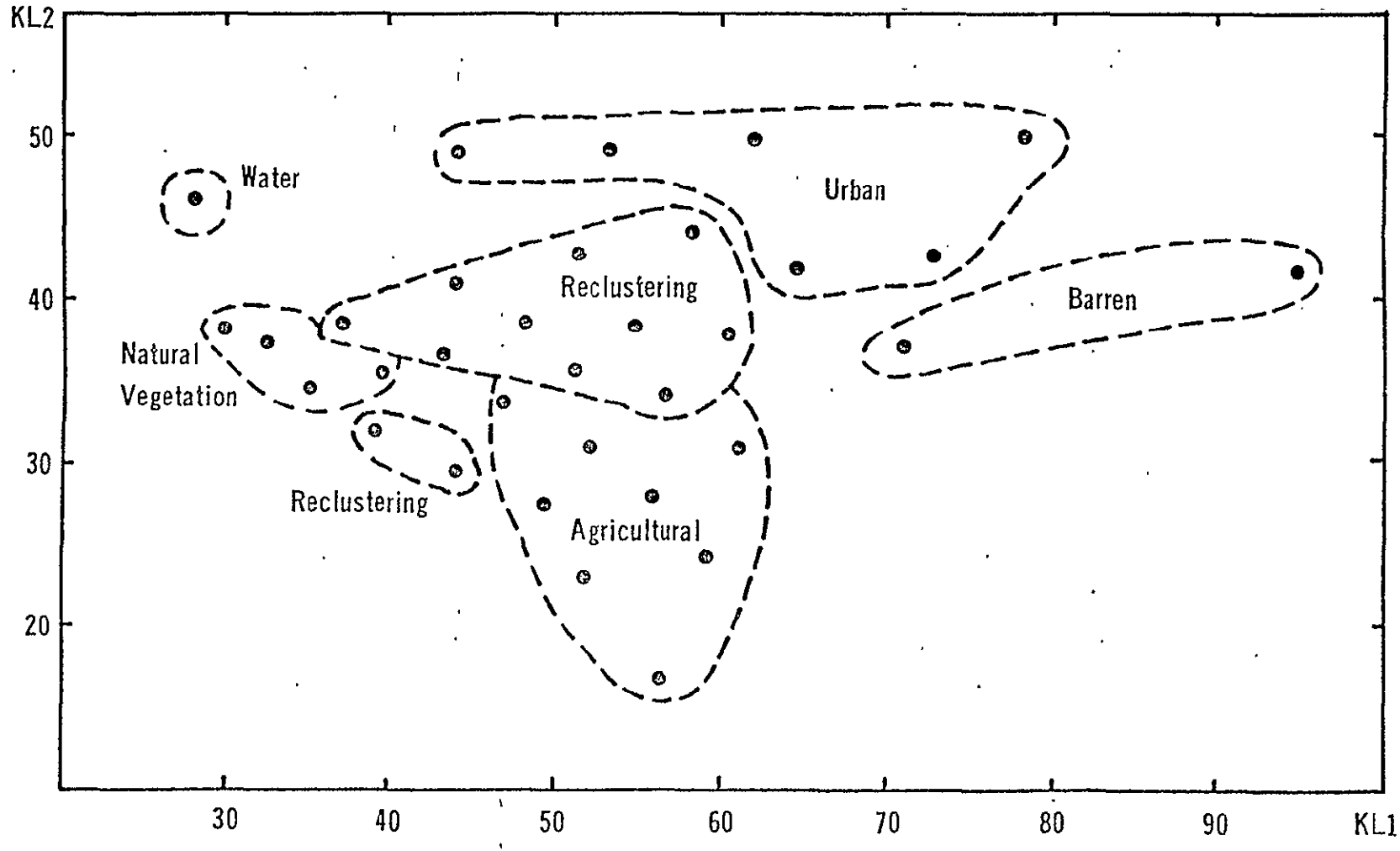


Figure II-4 The 34 Cluster Centers and the Initial Decision Made on Each Cluster

Figure II-5. Machine classification of land use pattern using one step clustering approach - the Trail Creek Watershed.

Figure II-6. Machine classification of land use pattern using the iterative clustering approach - the Trail Creek Watershed.

(5) Textural Information.

There are several possible measures of textural information which have been used in remote sensing data. In fact, Haralick [5] has proposed a set of 28 texture features. Textural information measures the spatial relationship for each spectral responses, and can be tabulated in a set of matrices called group-tone spatial-dependence matrices by Haralick. All the 28 texture features can be evaluated from these matrices.

As an experiment, we chose the angular second moment which is a measure of the homogeneity of the image. In that process, first band 5 of the LANDSAT data was re-quantized into eight gray levels. Then the gray-tone spatial-dependence matrices, and in turn the angular second moment, were determined based on the requantized image. A 5 by 5 spatial block size was used to determine the measure of texture assigned to the central point of the block. We examined the effect of the texture on classification by clustering KL1, KL2 and TX, and labeling the resulting clusters. We did not apply the reclustering step. The results of the classification are shown in Figure II-7 and Table II-3(c)

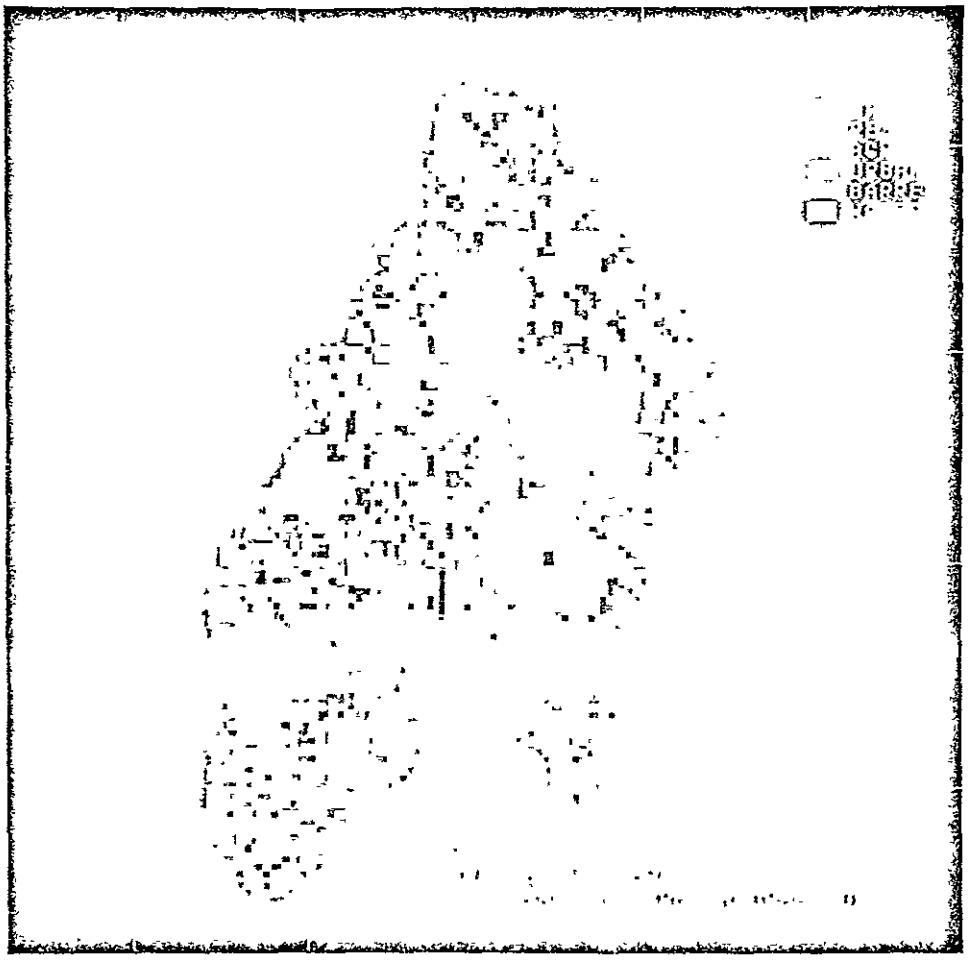


Figure II-7. Machine classification of land use pattern using one step clustering approach with textural information - the Trail Creek Watershed.

Land Use	Percent of Areas			
	Ground Truth	One Step Clustering (a)	Clustering and Reclustering (b)	With Texture (c)
Natural Vegetation	50.17	48.69	45.06	52.27
Residential	(low density) 2.45 (medium density) 6.79 (high density) 0.11	16.60	15.31	14.96
Dev. Open Space Agricultural Pasture	0.49 28.73 3.04	26.69	32.24	26.52
Urban	(Industrial) 2.59 (Commercial) 1.55	5.84	5.21	3.75
Water Bodies	0.57	0.97	0.97	0.49
Trailer Parks Highways Barren	2.47 1.06	1.21	1.21	2.01

Table II-3. Aerial percentages of land use classes as determined from remote sensing data using the clustering approach. [C. G.

Discussions

The new results can be compared to the ground truth and to the maximum likelihood classification results. Even though, our new results are significantly better than the maximum likelihood classification results, improvements might not be as apparent in direct numerical comparison of percentage of land use in each class, since generally numerically compiled results are the average of many detailed effects. However, the following conclusions seems to be justified.

(a) Our new results using clustering show significant improvement over maximum likelihood classification when we examine the detailed results point by point on an image.

(b) The clustering approach is much more flexible in the sense that the classes are assigned after the fact.

(c) Reclustering results in a significant improvement over the one step clustering classification. This conclusion can be drawn from the examination of Figures II-5 and II-6.

(d) It is difficult from the results obtained to assess the effect of textural information.

(e) It still appears to be necessary to devise some kind of consolidation program to remove extraneous misclassified points.

Note also that we have finally only 6 land use categories instead of the 10 catagories used by HEC. The following comments are pertinent:

(a) The separation of industrial and commercial classes: These two categories may not be differentiated accurately from remote sensing data. By applying a spatial consolidation algorithm, a partial success appears possible. For example, in clustering, we have a good indication that downtown commercial area and large size parking spaces around shopping centers might be distinguished. Further work is needed.

(b) Density of residential areas: That depends largely on the definition of low, medium and high density residential areas. Residential areas, generally, tend to be clustered as newly developed or old residential areas on the basis of surroundings rather than density. More work on fairly large urban areas is needed to determine whether density of residential areas can be determined by using Remote Sensing.

(c) Separation among agricultural, pasture and developed open space: We have not payed too much attention to this problem as yet. Even with lots of care and attention, it seems difficult to separate these classes even from high-flight image. A careful study is needed.

From a user's point of view, personnel of Savannah District of the Corps of Engineers through HEC provided us with the following comments on the importance of confusion between classes:

(a) Confusion between agricultural and developed open space

Economics - very minor problem

Hydrology - significant problem

(b) Highways classified as residential*

Economics - problem in flood plain

Hydrology - significant problem

(c) Confusion between industrial and commercial

Economics - significant problem in flood plain

Hydrology - no problem

---

* This problem can be simply eliminated by entering major highways from maps to LANDSAT image.

(d) Classification of highways and trailer parks as industrial

Economics - problems in flood plain

Hydrology - no problem

(e) Lumping of all residential categories together

Economics - problem in flood plain

Hydrology - significant problem

From our experience, it is likely that the proposed operational procedure can be adopted with minor modification and improvements. However, we cannot, as yet, draw any conclusions on the adequacy and computational efficiency of the ISOCLAS algorithm and on the usefulness of textural information. Our plans are as follows in this regard:

(a) As we have mentioned, there are five parameters of importance in the ISOCLAS program, and the performance of the program, by and large, depends on the choice of these parameters. Experiments to devise a rational procedure for choosing these parameters will be performed. Further, based on the results of these experiments, we shall decide whether the ISOCLAS program is adequate.

(b) We shall consider the implementation of some other clustering algorithms or develop a new clustering algorithm of our own.

(c) Although we note the importance of spatial information, it is not clear at this time how to utilize that information or what kind of texture measure should be used. A systematic and careful study on this problem will be conducted.

III. The Castro Valley Watershed

The second watershed of our study is the Castro Valley Watershed in California. A portion of a 1: 2400 USGS map of the Castro Valley Watershed

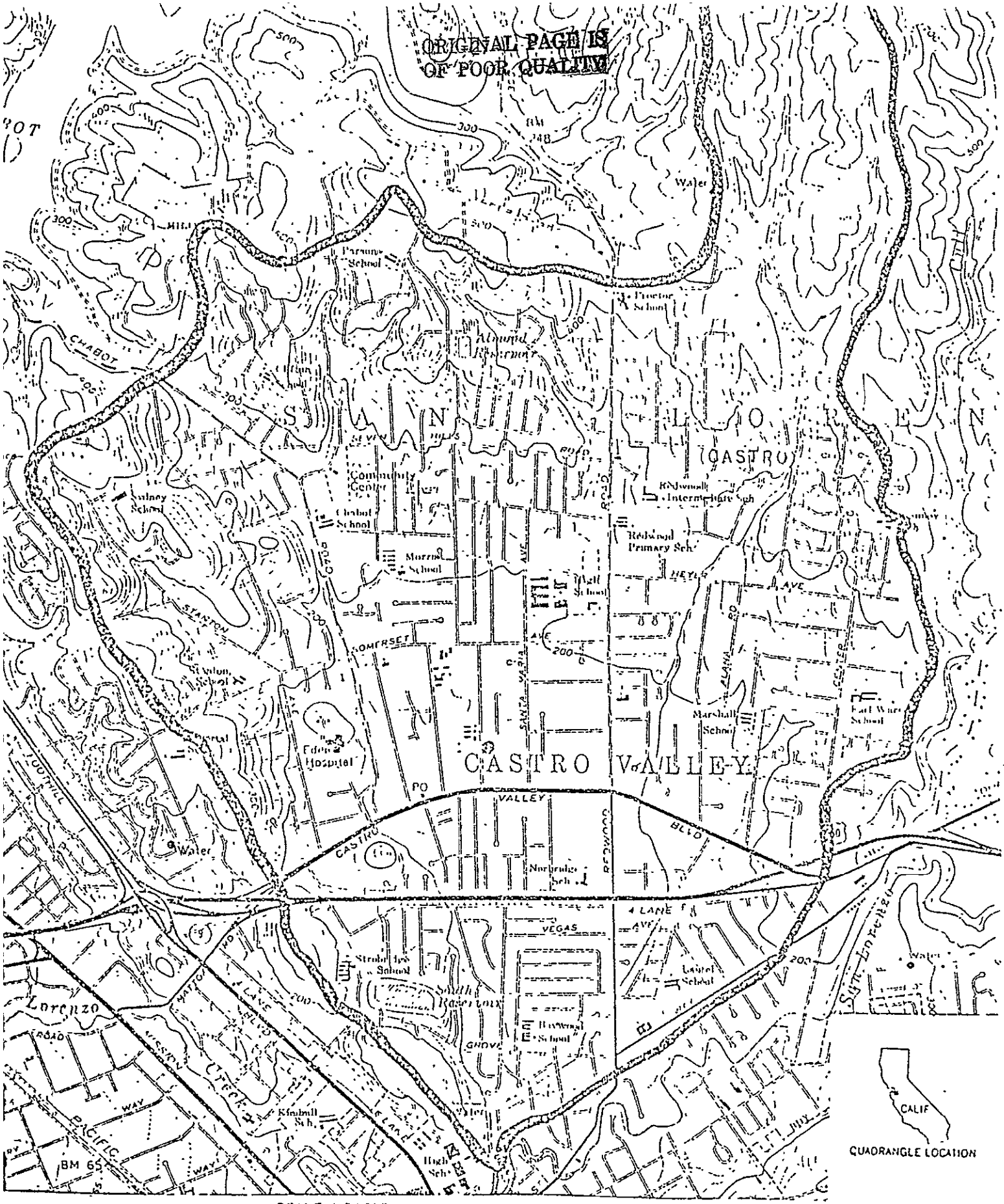


Figure III-1. The Castro Valley Watershed

is shown in Figure III-1. This part of our study has just begun, and will be reported briefly in this section.

A. Ground Truth

From our work on the Trail Creek Watershed, one important question emerged: what is the "ground truth" which can be reasonably compared to the results of machine classification of land use. Shall it follow closely to human perception and interpretation of images, or shall it be based on detailed machine measurements of land cover?

Examples which illustrate the difficulties are:

(1) A school area with buildings and playing fields: Is it an institutional land use collectively, or do we subdivide this area into two land uses of buildings and open field?

(2) A church: Is it commercial land use (a large building with parking lots) or institutional?

(3) A wooded area in the middle of a residential area: Can it be lumped into the residential area, or do we classify it as natural vegetation or even as developed open space?

(4) Some major highways: Are they transportation land use, or just residential (with an appropriate mixture of concrete and vegetation)?

Some of these questions will never be answered satisfactorily for all users and different applications. Thus, when we compare the results of machine classification of remote sensing data to the "ground truth" and estimate the accuracy of classification, we have to keep the above questions in mind. The "ground truth" for the same area may be different things for different objectives and applications.

We are currently preparing explicit rules for manual classification of high flight images to be used as the "ground truth" suitable for our

specific objectives; namely, applications to hydrology and possibly to economic analysis. This requires substantial interaction with HEC personnel. In the following, we list tentative rules used in the manual classification of the high-flight image of the Castro Valley Watershed. These rules will be subject to modification and improvement after consultation with the HEC. Figure III-2 shows the result of our manual classification.

Classification Guidelines

The classification scheme lies somewhere between a strict land use classification and that required for machine classification. Hence, some consolidation of classes may be necessary at the machine level.

(1) In interpreting and classifying, one must constantly be aware of the resolution of the comparison product. With this in mind, a minimum of one-half pixel was determined to be that minimum width for all aerial units mapped.

(2) To facilitate machine classification, buildings were separated from open space in the case of schools and hospitals, subject to the limitation described in (1).

(3) Where two different classes were adjacent to each other with a road or highway separating them, the boundary between the two classes was determined to be the centerline of the road or highway (assuming the width of the road/highway was less than the one-half pixel limitation).

(4) Where a road or highway traversed a given class, that road or highway was included as part of the given class (assuming the width of the road or highway was less than the one-half pixel limitation).

ORIGINAL PAGE IS
OF POOR QUALITY

Figure III-2.

Figure III-2. Manual land use classification of the Castro Valley Watershed.

Land use

1. School
2. Trailer park
3. Reservoir
4. Dedicated open space (park, golf course, etc.)
5. Hospital
6. Highways
7. Improved open space
8. Commercial
9. Residential
10. Multi-family residential
11. Natural vegetation

B. Machine Classification of Land Use in the Castro Valley Watershed.

At this time, we have just obtained preliminary classification results. A July 26, 1972 LANDSAT scene was first clustered using the ISOCLAS program. Then after examination of the clustering results, the map and the high-flight image (our ground truth), a preliminary classification of the watershed was obtained. The operational procedure discussed in the previous section was applied without reclustering. We show classification results in Figure III-3. We do not provide numerical comparisons at this time. However, examining Figure III-3, we note that:

(a) The classification results match well with the urban areas along the Castro Valley Blvd. and Redwood Road.

(b) The classification of the residential areas is fairly accurate. Here, again we have not separated the densities of the residential areas.

(c) There is a confusion on the northern hillside. The area is mainly natural vegetation, but the machine classified the area as a mixture of natural vegetation, open space and residential areas.

IV. Discussions and Continuing Work

We have recently proposed a continuation of this project. Work reported here has been motivated in part by the longer term needs and objectives of this project. Our activities during the past year which meet these long term needs and objectives can be summarized as follows:

1. Supporting work of general interest.

We are actively engaged in the development of a digitization algorithm, and in the geometric correction and compilation of data available from various sources. It is self evident that all useful and readily accessible data, such as maps and photographs, as well as

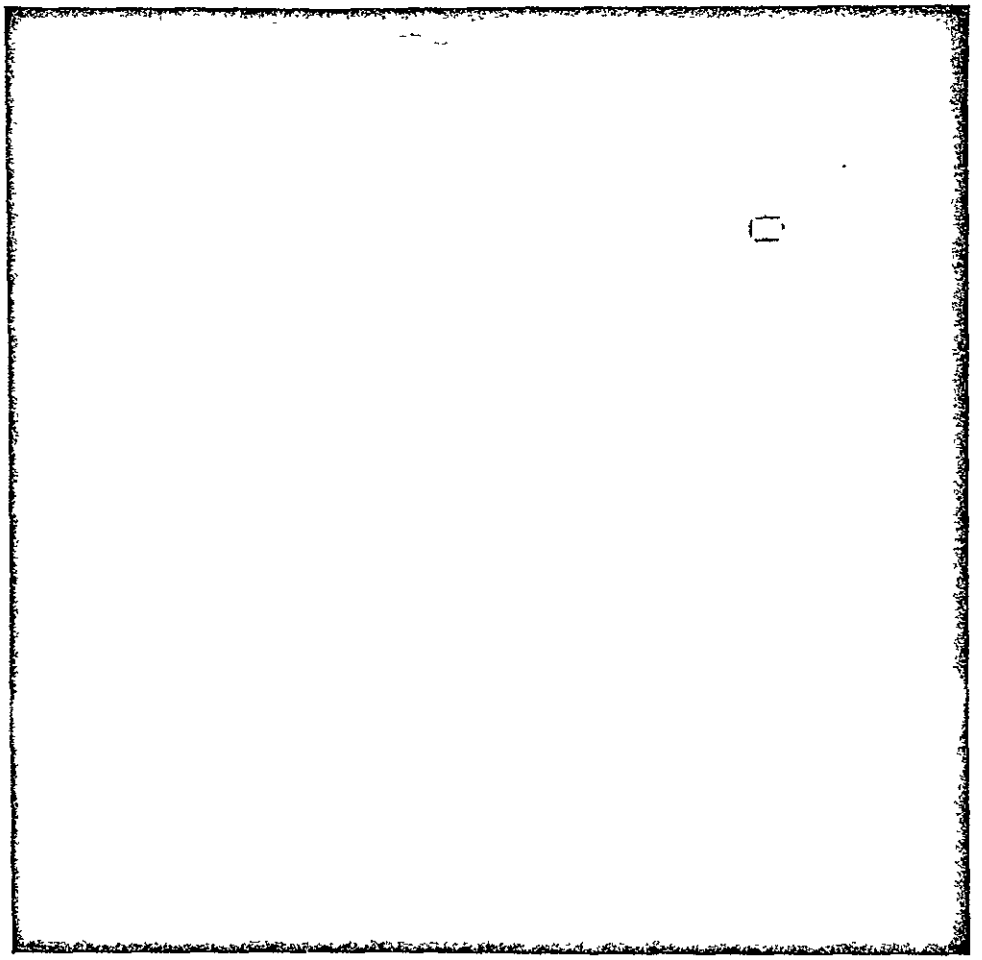


Figure III-3. Machine classification of land use pattern using one step clustering approach - the Castro Valley Watershed.

satellite data, should be exploited in all land use classification work. This creates a substantial need for computer programs and algorithms which compile all the information on a single data base. This is an area of work to which substantial manpower is being assigned within the grant. These programs are being implemented on the UC Davis Image Processing System.

2. Implementation of necessary computer programs.

Toward the development of an operational procedure, we have implemented several computer programs besides digitization and geometric correction on various computers. Among those are:

- ISOCLAS - implemented on the LBL computers by the RSRP group
- Reformatting program for LANDSAT CCT's - the CALSCAN and ISOCLAS programs on the LBL computer assume different tape formats than LANDSAT CCT's. We have implemented reformatting programs on the LBL computers.
- Karhunen-Loeve Transformation - implemented on the LBL computer.
- Masking program used in reclustering. A portion of data should be selected or masked out. This program has been implemented on the LBL computers.
- Texture program - implemented on the LBL computers.

3. Development of an operational procedure.

We proposed a tentative procedure which appears operationally usable by the Corps of Engineers in the determination of land use information from remote sensing data. Note that most of computer programs needed have been already implemented on the LBL computers which can be accessed by the Corps of Engineers.

4. Systematic study of classification algorithms.

We have studied both maximum likelihood classification and clustering. We concluded that classification based on clustering appears to be more suitable to the development of an operational procedure. We applied this approach to both the Trail Creek Watershed and the Castro Valley Watershed.

5. Utilization of textural information.

We recognized the importance of spatial information in classification of remote sensing data. We have tested one measure of texture for the Trail Creek Watershed. However, it is inconclusive, at this time, how much can be gained by using textural information.

As a continuation of our project, we have proposed to consider among others:

1. The completion and the development of an operational procedure.

a) The accuracy and efficiency of the ISOCLAS program depends on the choice of five parameters. We will continue to experiment to choose these parameters more rationally.

b) We are not satisfied with the performance of the ISOCALS program, at this moment. We will look into the possibilities of either adapting or developing a more suitable clustering algorithm.

c) The effect of textural information in classification is inconclusive yet. We will perform a systematic study on this and study other measures of textural information than one used before.

d) Development of spatial consolidation program as discussed in the July report.

e) Most of programs needed are implemented already. We will put these programs together and develop one step procedure which can be processed on the LBL computers.

2. Continuation of supporting work such as digitization, geometric correction, and display of information.
3. The determination of present land use of four additional watersheds chosen in consulation with the Hydrological Engineering Center of the Corps of Engineers.
4. The start of the process of transferring algorithm methods and procedures to HEC for their own use.

REFERENCES

1. G. H. Ball, and D. J. Hall, "A clustering Technique for Summarizing Multivariate Data," Behavioral Science, Vol. 12, 1967.

2. G. H. Ball and D. J. Hall, ISODATA, A Novel Technique for Data Analysis, and Pattern Classification, Technical Report, Stanford Research Institute, Menlo Park, Calif., May 1965.

3. R. L. Kettig and D. A. Landgrebe, "Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects," IEEE Trans. Geoscience Elec., Vol. GE-14, No. 1, pp. 19-26, Jan. 1976.

4. R. M. Haralick and I. Dinstein, "A Spatial Clustering Procedure for Multi-Image Data," IEEE Trans. Circuit and Systems, Vol CAS-22, No. 5, pp. 440-450, May 1975.

5. R. M. Haralick and K. S. Shanmugam, "Combined Spectral and Spatial Processing of ERTS Imagery Data." Symposium on Significant Results Obtained From ERTS-1, Vol. 1, NASA Goddard Space Flight Center, Maryland, pp. 1219-1228, March 1973.

6. D. J. Wiersma and D. Landgrebe, "The Use of Spatial Characteristics for the Improvement of Multispectral Classification of Remotely Sensed Data," Symposium Proceedings, Machine Processing of Remotely Sensed Data, LARS, Purdue University, West Lafayette, Indiana, pp. 2A-18-25, June 1976.