Paper L 9

THE UTILITY OF ERTS-1 DATA FOR APPLICATIONS IN LAND USE CLASSIFICATION

John E. Dornbach and Gerald E. McKain, *NASA, Johnson Space Center, Houston, Texas*

ABSTRACT

A comprehensive study has been undertaken at the Johnson Space Center to determine the extent to which conventional image interpretation and computer-aided (spectral pattern recognition) analysis techniques using ERTS-1 data could be used to detect, identify (classify), locate, and measure current land use over large geographic areas.

It can be concluded that most of the Level I and II categories in the USGS Circular #671 can be detected in the Houston-Gulf Coast area using a combination of both techniques for analysis. These capabilities could be exercised over larger geographic areas, however, certain factors such as different vegetative cover, topography, etc. may have to be considered in other geographic regions.

The best results in identification (classification), location, and measurement of Level I and II type categories appear to be obtainable through automatic data processing of Multispectral Scanner computer compatible tapes. To achieve these results, however, a slight modification of the Circular #671 categories and definitions would be required.

INTRODUCTION

The purpose of this paper is to report on one part of the ERTS-1 investigation at the Johnson Space Center (JSC). In particular, this report will focus on the use of the ERTS-1 Multispectral Scanner System (MSS) data as a source of information for compiling land-use inventories, such as those which might be required as part of recent coastal zone and pending national land-use legislation. The work of the Earth Observations Division Investigation Teams and the Land-Use/Urban Analysis Teams in particular are acknowledged.

The objectives of the ERTS-1 land-use investigation were as follows:

- <u>General</u> - Within the Houston Area Test Site (HATS), determine the role of ERTS-1, aircraft, and ground data in the definition and development of the technology for land-use classification over a large geographic (multi-county) region.

- Specific

  1. Using MSS imagery, establish the land-use categories which can be detected, identified, located and measured by conventional image interpretation techniques.

  2. Using MSS computer compatible tapes (CCT's), determine the degree to which automatic data processing (ADP) techniques can be applied to detect, identify, locate, and measure the various categories of a land-use classification system.

  3. Using existing or by developing new manual and automatic techniques, produce land-use classification maps of a portion of the test site.

  4. Using aircraft and ground data, relate results from the above to land use present at the time of the ERTS-1 overflight.

The investigation teams used imagery generated not only at the National Data Processing Facility (NDPF) at the Goddard Space Flight Center (GSFC) but also color-composite imagery generated directly from the MSS CCT's at JSC.

The 18-county Houston Area Test Site (HATS) covers an area about 41,600 square kilometers (16,000 square miles) around the City of Houston. It is an extremely diverse land-use study area possessing a large Central Business District, a rapidly expanding urban fringe, a large petrochemical industrial complex, many square kilometers of extractive industries, extensive agricultural and range lands, coastal zone features, and the Sam Houston National Forest. Figure 1 illustrates the general area covered by the ERTS-1 scenes which were studied. The area covered by the one-fourth scene CCT which was used in automatic data analysis, is the area which is outlined and that which received the most intensive analysis. Therefore, in the remainder of this paper, this area will be referred to as the Prime Study Area.

CLASSIFICATION SCHEME

The classification scheme used in this investigation is that described in the U.S. Department of the Interior Geological Survey, Circular #671 (Reference 1). This scheme (or a planned revision thereto) is being considered for possible application on a national basis, in the event legislation such as contained in the Senate Bill 268, "Land Use Policy and Planning Assistance Act" is enacted. Although not specifically an objective of this task, the results of this investigation illustrate the applicability of the Circular #671 scheme to the Central Gulf Coast Region and what, if any, modifications might be recommended to produce a more meaningful classification, if ERTS MSS data is to be one of the major data sources. These modifications are especially significant if ADP techniques are to be the main source of data analysis, since Circular #671 on Page 6 recognizes that, "As further advances in technology are made, it may be necessary to modify the classification system for use with automatic data analysis."

DATA USED

ERTS-1 MSS data from two passes over HATS were utilized in the investigations on Land-Use Classification. These passes, on August 28 and 29, and October 3 and 4, 1972, occurred in generally clear conditions with only a small percentage of cloud cover. The major analysis activities used the August and October data. During the August pass, high altitude underflights were flown over a portion of the test site.

Both system-corrected imagery and digital data were acquired from the NDPF. The black and white 70 mm and 240 mm (9 inch) imagery from the NDPF were utilized primarily in the form of black and white prints of each band enlarged from the 70 mm transparencies to a scale of approximately 1:250,000. For this land-use study, the color composites from the NDPF were not analyzed, due to the availability of higher resolution color composites directly from the digital tapes. The digital tapes were used in two ways; they were used for computer-aided classification by spectral pattern recognition using supervised and nonsupervised techniques, and were also used to produce two or three band color composites using the film recorder on the JSC multispectral Data Analysis Station, giving essentially a first generation, positive transparency film product at a scale of approximately 1:250,000.

CONVENTIONAL IMAGE INTERPRETATION

- Level I and II Classification of GSFC Black and White Imagery

Conventional image interpretation for a Level I classification was conducted using black and white transparencies and enlarged prints from Bands 5 and 7 over the prime study area. The following Level I categories of Circular #671 were detected and identified:

Urban and Built-Up Land	Forest Land
Agricultural Land	Water
Rangeland	Non-Forested Wetland

There were no Barrenlands identified in the area covered by this data. Also, the categories of Tundra and Permanent Snow and Icefields are not found in Texas.

As individual categories, Agricultural Land and Rangeland were usually easily identifiable from everything else. However, the least reliable of the Level I determinations was delineating the boundary between these two categories, even when field sizes and shapes were used. It was necessary to define all identifiable pastureland as Rangeland, in order to alleviate this problem. The authors are aware that this is probably a problem unique to this part of the United States. In the remainder of this report, Agricultural Land and Rangeland will be treated as a single Level I category.

There was also a definitional problem regarding the Level I classification of oil and gas Extractive areas versus the Rangeland and Non-Forested Wetland categories, since in Texas many oil and gas fields are grazed. Generally on ERTS-1 imagery, clues to the existence of oil or gas fields are far below the resolution of the system. However, close in to the urbanized area, this use probably would be defined as Open and Other; hence, would fall into the urbanized category. As distance from the urban fringe increases, it becomes more difficult to make the Extractive versus Rangeland decision. Near the Gulf of Mexico, much of the Non-Forested Wetland is also grazed and therefore, the decision becomes even more difficult.

Highly reflective areas in the two bands were also often difficult to classify, especially at the urban fringe, since the interpreter did not have enough clues to determine whether it were urban (new subdivision or industry), agriculture (bare soil), or cleared forest (to be reforested). It is significant to note that the definitions in Circular #671 did not adequately cover the commercial timber cutting which occurs each year in different parts of the Sam Houston National Forest or the extensive cutting and clearing which is occuring north of the City of Houston, where new fresh water reservoirs for municipal surface-water supplies are being constructed.

From a standpoint of the minimum size of Level I features which were detected by this procedure, the following results were achieved:

Feature	Performance
Agricultural Fields	4 Hectares
Forest Stands	4 Hectares
Rights-of-Ways in Forest	28 Meters (width)
Water Bodies	1 to 2 Hectares
Rural Settlements	40 Hectares

The performance accuracy in Level I classification using this technique ranged between 80 to 98 percent. A more detailed comparison of the performance of the three techniques used for each category is given in the results section of this paper. Also, as will be seen later, the color composite generated at JSC provided a better data source for attempts at conventional image interpretation for Level II and hence Level II classification was not attempted using black and white imagery.

° Level I and II Classification of JSC Generated Color Composite

One computer compatible tape was used to produce a first generation, positive color transparency on the digital tape to film recorder which is part of the Data Analysis Station equipment at JSC. The composite was produced with Channel 1 (Band 4) portrayed as blue, Channel 2 (Band 5) as green, and Channel 4 (Band 7) as red. The result was a 1:250,000 scale (approximate), false-color infrared ektachrome transparency which was of considerably higher resolution than the color prints received from the NDPF. This is understandable since the NDPF goes through several image reproduction and scanning steps prior to production of a 1:1,000,000 scale system-corrected, color composite paper print.

Conventional image interpretation for Levels I and II was conducted using this transparency (Figure 2). Each category was annotated on a stable base mylar overlay and keyed to the color composite film strip. The following Level I and II categories of Circular #671 were detected and identified:

Level I	Level II
Urban and Build-Up Land	
	Residential and Open
	Commercial and Industrial
	Transportation, Communications and Utilities
	Strip and Clustered Settlements
	Extractive
Forest Land	
Agricultural Land/Rangeland	
Water	
Non-Forested Wetland	

As previously explained in the Section on the conventional interpretation of NDPF black and white imagery; Barren Lands, Tundra, and Permanent Snow and Icefields were not present. Furthermore, other category classifications at the Level II were not attempted by the Land-Use Study Team for the following reasons:

Agricultural Land - All land in this category could only fall into either the Cropland and Pasture or the Other category. It was impossible to determine unused cropland or brushland from vacant or unused land in the Urban category of Mixed or Open and Other, especially in the multi-county, urban fringe area.

Rangeland - According to the Circular #671 definitions, only the Grass category of Level II was present in the Prime Study Area.

Forest Land - Since only single date imagery was used, decisions on Level II were not possible. However, other study teams using growing and dormant season imagery could make this type determination.

Water - Almost all water features in the Prime Study Area would have had to be classified as Reservoirs. Due to the scale involved, the image interpreters did not attempt to identify (classify) water areas smaller than 16 hectares (40 acres) which would have been the Water - Other Category at Level II. No streams were wider than 200 meters

(1/8 mile) as defined in Circular #671 except in one instance, where a classification decision could not be made due to the tidal nature of the mouth of the stream.

Non-Forested Wetland - Only a very small percentage of this category was found in the Study area, and that which existed was vegetated. It is uncertain how well this technique could differentiate the vegetated from the bare in the Non-Forested Wetland category.

As will be seen in the results section, the performance accuracy in Level I and II classification was slightly better than when using the black and white enlargements. The ability to detect minimum size objects was also about the same.

During the first and second phases of this study, i.e., those which used conventional image interpretation techniques, special enhancements were produced using black and white transparencies of two or three bands in various optical and electro-optical additive color viewers and printers available at JSC. It was found that these did not appear to provide better or faster results than use of the black and white enlargements or JSC color composites. In fact, the enhancing process appeared to degrade the ability to measure the acreage of certain types of land use.

° Level I and II Classification Using Computer-Aided Techniques

The two methods of computer-aided processing used to analyze the digital tape data were supervised maximum likelihood classification (LARSYS II, Purdue LARS) and unsupervised classification using clustering techniques (ISOCLS - Iterative Self-Organizing Clustering Program, JSC). The first method of processing requires information from training fields that are defined by the analyst and then all other data is classified into the various defined classes (i.e., into the classes that had been input into the computer). A threshold is usually specified, so that data points that are far from the mean of each class are left unclassified. The second method of processing organizes all of the data into spectrally homogeneous groups (clusters) and produces classification-type clustering maps, in which the clusters require identification and interpretation in a postprocessing analysis. It is also possible to use the ISOCLS clustering algorithm to generate statistics for LARSYS II classification in lieu of training field statistics.

The digital data used for this analysis was a 9 track, computer compatible tape from the August 29, 1972, ERTS-1 overpass of the Houston area.

Using the computer-aided techniques, the following Level I and II categories of Circular #671 were detected and identified:

Level I	Level II
Urban and Built-Up	
	Residential

Level I	Level II
	Commercial, Industrial and Transportation (Open and Other)
Agricultural Land/Rangeland	
Forest Land	
Water	
Non-Forested Wetlands	

When attempts were made to classify Level I categories for the entire Prime Study Area, it was found that there was a marked spectral similarity between highly vegetated Urban categories such as Open and Other and the Cropland and Pasture categories under Agricultural Land (Figure 3). The following performance percentages were determined by computing the area of a predetermined test location which had been correctly classified:

All Level I categories, except Urban	91 - 97%
Level I, Urban	44%

When an effort was made to limit the area to be classified to what might be termed the area inside the urban fringe, the following range of Level II results were obtained:

Residential	67 - 83%
Commercial, Industrial & Transportation	60- 94%
Open and Other	72%

Figure 4 is an example of a supervised classification limited to the highly urbanized area around Houston. Please note that the category Vegetated was used as a class to include urban and non-urban land use which was spectrally similar. It was found that to produce this classification, it was necessary to select more training fields than would be normally required in classifying more homogeneous data sets such as forest or water. Even when many types of residential areas were selected as training fields which could conceivably be considered representative of a class called general residential land use, it appeared that picture elements for all urban categories other than Commercial/Industrial/Transportation were being classified as either Residential or Open and Other (Vegetated). Due to the difficulty in selecting training fields representative of a class, no attempt was made to generate performance statistics on the supervised analysis.

Figure 5 is an example of the use of ISOCLS as a preprocessor to organize certain clusters into meaningful categories of urban land use. In the process, clusters consisting of a few pixels or those with large distances from the mean of the chosen clusters were eliminated. The organizing process used aircraft underflight imagery to correlate the urban land use with each specific cluster. Using this technique, it was shown that not only could some Level I and II categories be mapped, but the Residential category could further subdivided into what might be termed Level III.

A sampling technique was developed for this classification based on the random selection of grid squares superimposed on a base map, which contained land use as interpreted from the 1:120,000 scale high altitude aircraft imagery. By comparison on a picture element by picture element basis within the selected grid squares, it was possible to determine how well the clusters agreed with existing land use. The following table illustrates how well the classifications derived from the clusters agreed with the image interpretations on the base map.

Agreement With Base (%)	Commercial/Industrial/ Transportation	Residential	Mixed Urban	Vegetation (Woody)	Vegetation (Non-Woody)	Water
Commercial/Industrial/ Transportation	94.2	5.5				0.3
Residential	2.6	66.8	23.0	4.5	3.2	
Mixed Urban	1.0	20.8	51.1	3.8	23.5	
Vegetation (Woody)		0.7	0.2	95.1	4.0	
Vegetation (Non-Woody)	1.1	12.1	25.7	4.8	56.2	
Water	3.9	3.0	1.9	2.2	1.5	87.7

As was the case with the supervised classification, the Commercial/Industrial/Transportation, the Woody Vegetation (Forest and Brush), and Water categories were well differentiated. The major problem to be encountered in the use of the Circular #671 categories would be in differentiating picture elements with similar spectral response in all four channels in the urban and non-urban areas, as illustrated by the Mixed Urban and Vegetation (Non-Woody) categories.

This study indicated also that if those clusters combined as Residential were analyzed, it was possible to identify each cluster as a separate category which could be classified as Residential, Residential (New) and Residential Mixed. These categories could be considered to be a Level III classification. They appeared to agree generally well with interpretations made from aerial photographs, but a system was not developed which could provide repeatable results with these categories included with the other Level I and II categories.

A limited experiment was conducted to determine how well the areal extent of certain urban land use categories could be measured. The following table illustrates these results:

Accuracy of Areal Extent of Land Use Categories Delineated by Computerized Classification Technique (ISOCLS)

Residential Study Area	Ratio of Area Acreage $\frac{C^*}{B}$
Cloverleaf (Area L)	1.12
Garden Villas (Area N)	1.16
Center City (Area P)	0.91

* B = Acreage of Area as delineated on an aerial photograph.
C = Acreage of Area as delineated on computer classification map (manual picture element count)

Insofar as location of features is concerned, a capability has been developed at JSC to reformat the ERTS-1 CCT as received from the NDPF to correct for:

1. Differential in line and sample spacing
2. Earth Rotation
3. Scale and aspect ratio of the JSC film recorder

The reformatted (registered) tape can then be used to generate a color composite image or after classification by LARSYS II, a color-coded classification map upon which a 10 square kilometer Universal Transverse Mercator (UTM) grid is superimposed (Figure 6). A cursory analysis has indicated that within a local area, this grid is located with respect to the position of features on the earth's surface to much better than 500 meters. Larger areas, as for example an entire CCT, suffer some degradation in the position of all features, due mainly to the instability inherent in the film recorder.

Another paper at this Symposium from JSC will address the locating of ERTS-1 picture elements classified as water to an accuracy of at least 300 meters (1,000 feet) at a 1:24,000 scale. This was made possible by assigning scan line and sample numbers to control points selected from the USGS 1:24,000 scale topographic map series. On the basis of this work and in recognition the UTM overlay previously referred to, it has been demonstrated that it is possible to locate picture elements which have been classified as some specific category of land use, to an accuracy of at least 300 meters or better. Due to the "mixed" or "unresolved object" picture element problem between the different categories of land use, it is not believed that the boundary could be located more accurately. It should not, however, be more than 300 meters or less than two millimeters on a 1:250,000 scale land use map. Depending upon the manner in which the land use maps were to be used, this may be acceptable to those agencies interested in regional or national land use mapping of Level I and II categories.

RESULTS

1. The results of this land use investigation and those of the other ERTS-1 Teams at JSC indicate that using ERTS MSS data as the only information source, a Level I and II Land Use Map of the Texas-Lousiana Gulf Coast Region could be produced. This would require a slight modification in the categories and definitions contained in Circular #671. Assuming that growing and dormant season data is available and that both conventional image interpretation and computer-aided processing are utilized, the following categories could be classified for areas in excess of 4 hectares (10 acres):

Level I	Level II
Urban and Built-Up Land	Residential
	Commercial and Industrial
	Extractive
	Transportation, Communications, Utilities
	Strip and Clustered Settlements
	Open and Other
Agricultural Land	Cropland
	Other
Rangeland	Open
	Wooded
Forest Land	Deciduous
	Evergreen (Coniferous and Other)
	Mixed
Water	Streams and Waterways
	Lakes
	Reservoirs
	Bays and Estuaries
	Other
Non-Forested Wetland	Swamp
	Marsh
	Non-Vegetated
Barren Land	Beaches
	Sand Other Than Beaches
	Other

2. All three techniques examined for producing Level I land use maps produced acceptable results insofar as classification accuracy was concerned. As indicated in the paper, when much of the open and heavily vegetated areas inside the urban fringe were treated separately from non-urban areas, the classification accuracy was in the order of 67 to 83%. A comparison of the accuracies achieved by comparing the three classification products with a standard base derived from image interpretation of aerial photography is as follows:

Land Use Category	GSFC B&W Images	JSC Color Composite	Computer-Aided
Forest Land	98.4%	99.9%	97.4%
Agriculture Land/ Rangeland	89.9	92.3	92.9
Water	79.6	76.0	91.1
Urban and Built-Up Land	93.7	96.5	44.1(67-83)*

* Accuracy achieved when Urbanized area classified separately from non-urban.

3. When classifying at Level II or higher order categories, computer-aided techniques appear to have a decided advantage over conventional image interpretation. As evidence of this, the following is the only quantitative comparison of classification accuracy which was conducted in this study at the Level II:

	Classification Product	
Land Use Category	JSC Color Composite (%)	Computer-Aided (%)
Residential	64.7	83.0
Commercial, Industrial, and Transportation	43.8	60.2
Open and Other	-	72.0

The above was conducted only within the urbanized area and was achieved by comparing the classification product with a standard base derived from image interpretation.

4. Investigations at JSC have indicated that computer-aided classification appears to offer greater potential for Level II and higher order category classification than conventional image interpretation. This advantage is assumed to be related to scale and resolution and is probably directly attributable to the capability of the computer-aided classification technique to work near the spatial resolution limits of the sensor system. That is, the human eye-brain combination can readily organize general categories over large areas, as for example Level I at a 1:250,000 scale, but this ability becomes taxed as the number of categories increase and the average area of the units to be classified, decreases. In contrast, the computer-aided classification technique can theoretically detect and potentially identify (classify) each 0.5 Hectare (1.2 acre) picture element as some land surface feature, independent of scale. However, the "mixture" or "unresolved object" picture element problem must be better understood before this becomes practical. Experience at JSC in using computer-aided techniques in classifying surface water from everything else, using Bands 4 and 7, has shown that when three contiguous picture elements have been classified as surface water, there is a high probability that surface water of 4 Hectares (10

acres) or greater in areal extent is present.

Three conclusions may be derived from this:

a. If it can be concluded that computer-aided land use classification works best when used for Level III or IV classification with aggregation back through Level II to Level I, a sampling strategy would have to be developed and used, due to the magnitude of the data processing problem when dealing with the mapping of large areas. The sampling strategy would also have to take into consideration the "mixed" picture element problem and would require a modification to the existing Circular #671 categories.

b. Improvements may be needed in boundary location accuracy due to the degradation caused by "mixed" picture elements which will be found on the perimeter of each classified category.

c. Improvements are needed in the accuracy of area measurement due to the "mixed" picture element problem, especially for classified features less than 100 Hectares (250 acres) in size, since the proportion of the perimeter picture elements increases as the area of the feature decreases.* Classification errors of comission as well as omission must be considered as part of this area measurement problem.

5. When efforts were concentrated on mapping coastal zone features and allowing many of the Level I and II categories to be considered "Other", the following features were classified using the ISOCLS non-supervised clustering procedure:

Clearest Water	Mixed Hardwood and Pine Forest
Less Clear Water	Pine Forest
Most Turbid Water	Wet Bare Soil
Shell Beach and Concrete	Fresh & Brackish Marsh & Wet Rice Fields
Swamp	Other (Urban, Grassland, Dry Cropland)

6. Temporal data sets of Band 7 from two overpasses, registered to each other, were used to automatically determine changes (decrease or increase) in surface water and changes in the tidal limits.

SUMMARY

This study and those of other investigators in the ERTS-1 Program have demonstrated the feasibility of using ERTS-type satellite scanner data for conducting regional and national land use surveys. They have also indicated that monitoring of land use change can be accomplished at

---

* Work by Nalepka and Hyde (Reference 2) in assuming a hypothetical 100 x 100 meter spatial resolution found that the "mixed" picture element problem could lead to a 25% error in the determination of area of a 44 Hectare (110 acre) square field and a 51% error in the area of a 68 Hectare (170 acre) rectangular field.

intervals commensurate with the requirements of States and other agencies. Although the feasibility has been demonstrated, there has not been as yet a systematic application of land use classification, location, and area measurement procedures, using either conventional image interpretation or computer-aided techniques. This systematic application will require evaluating the performance and accuracies necessary to meet the specific needs of land resource management agencies, prior to development of an operational capability.

References

1. "A Land-Use Classification System for Use with Remote-Sensor Data," Anderson, James R. et al, Geological Survey Circular #671, Washington, D.C., 1972.

2. "Estimating Crop Acreage from Space-Simulated Multispectral Scanner Data," Nalepka, R. F. and Hyde, P. D., Environmental Research Institute of Michigan Report ERIM 31650-148-T, Ann Arbor, Mich., August 1973, pp. 7 and 8.

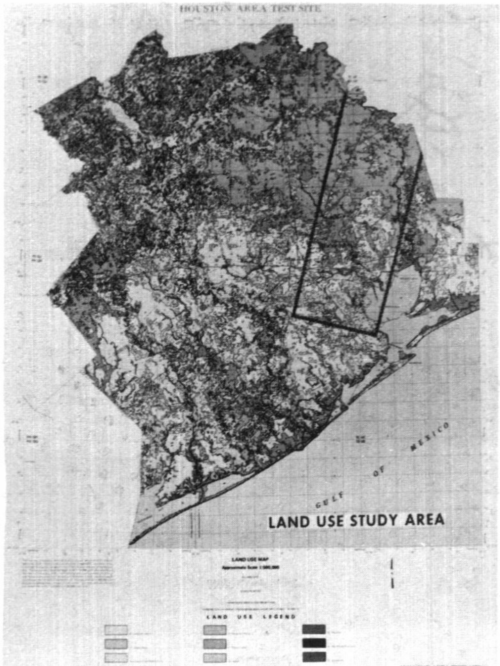

Figure 1 -
The 18 County Houston Area Test Site. Outlined Area Indicates The Prime Study Area Covered By One ERTS-1 Computer Compatible Tape. Houston Is At The Lower Left of This Study Area.

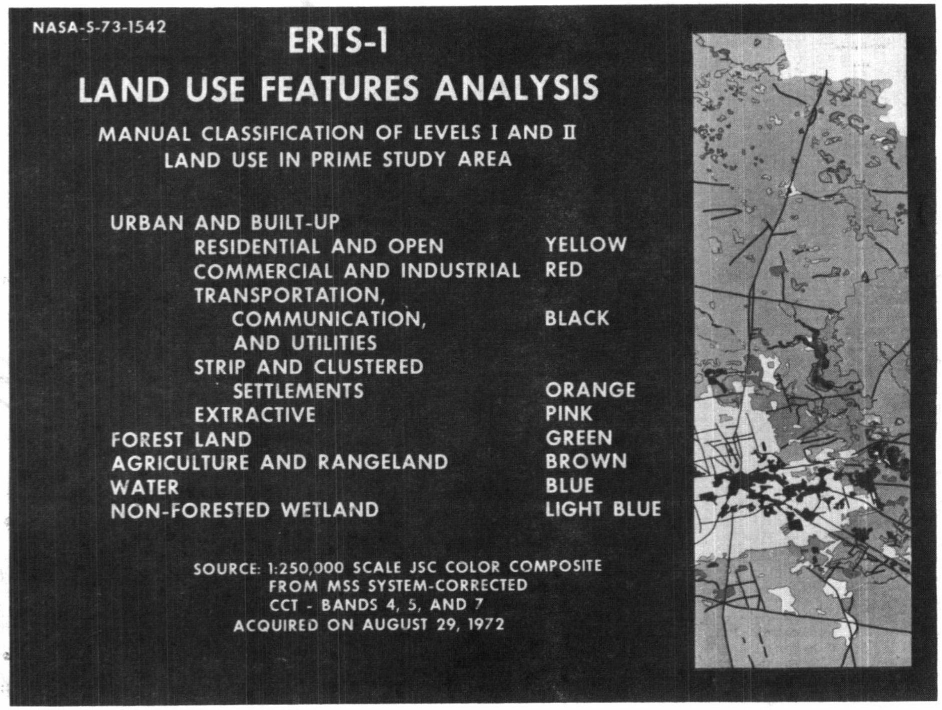

Figure 2 - Land Use Classification of Prime Study Area Using Conventional Image Interpretation of Color-Composite Generated at JSC From Digital Tape.

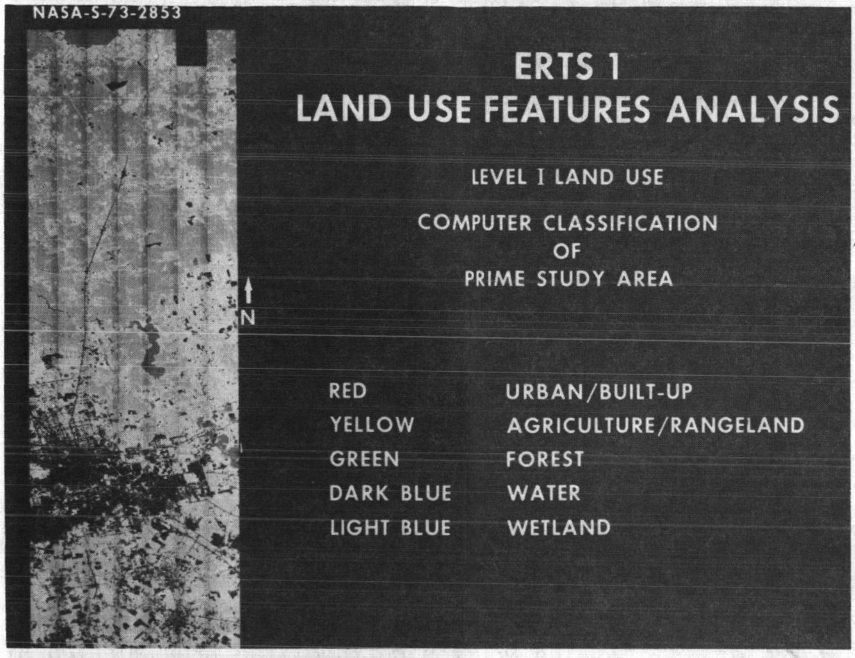

Figure 3 - Level I Computer-Aided Classification of Entire Prime Study Area.

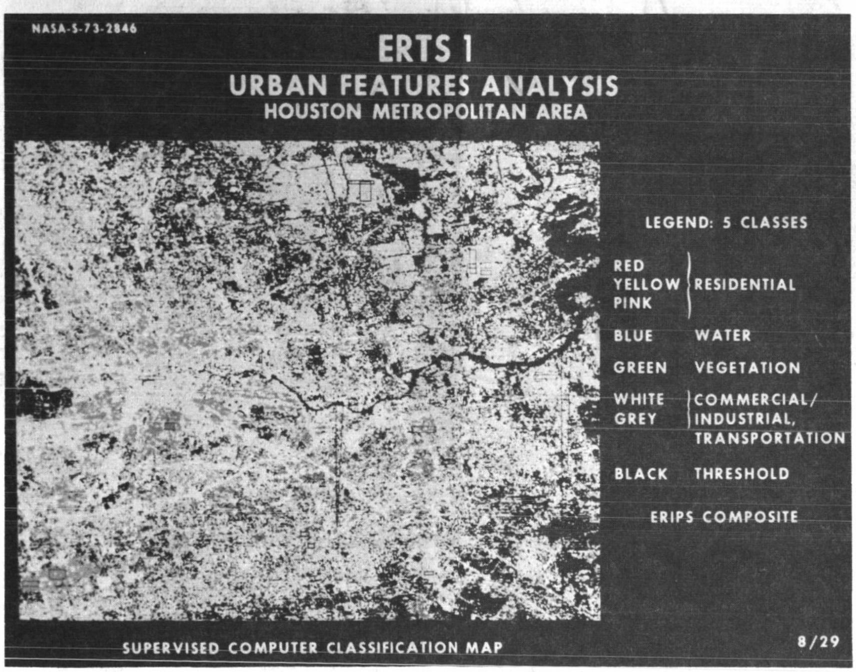

Figure 4 - Supervised Computer-Aided Classification of The Houston Metropolitan Area Portion of The Prime Study Area.

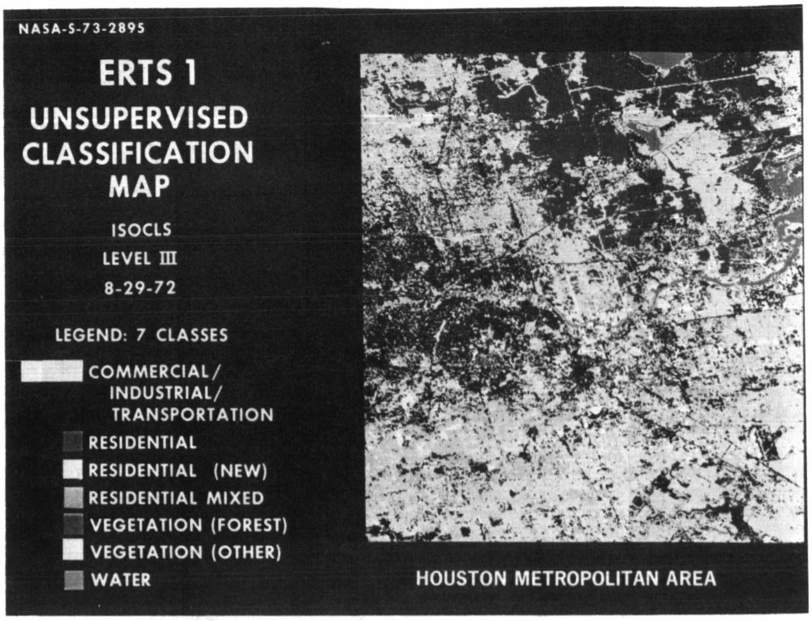

Figure 5 - Nonsupervised Computer-Aided Classification of The Houston Metropolitan Area Portion of The Prime Study Area.

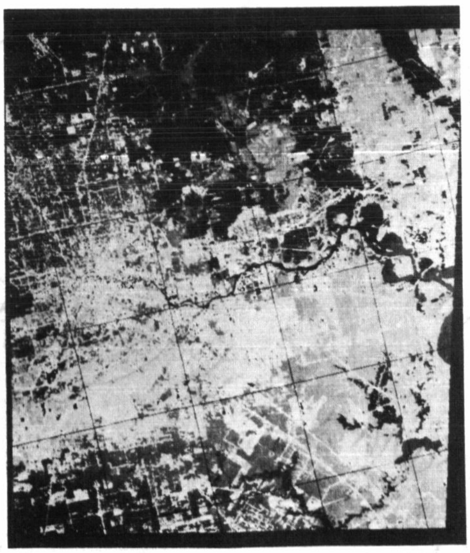

Figure 6 - 1:250,000 Scale Black and White Photograph of The Color Composite Generated At JSC From Part Of A Digital Tape. A Ten Square Kilometer UTM Grid Is Automatically Registered On The Film Recorder Output.

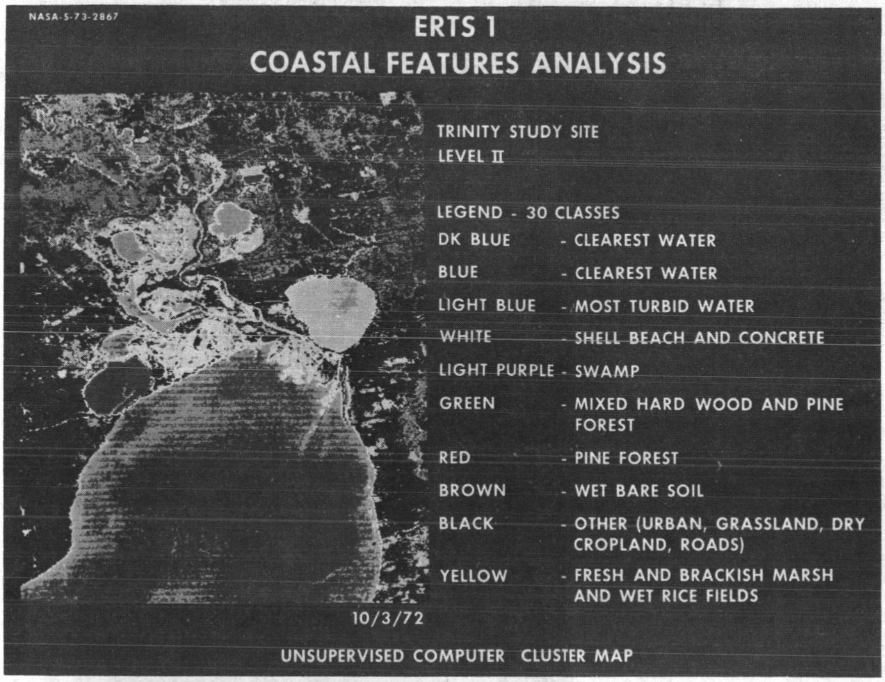

Figure 7.

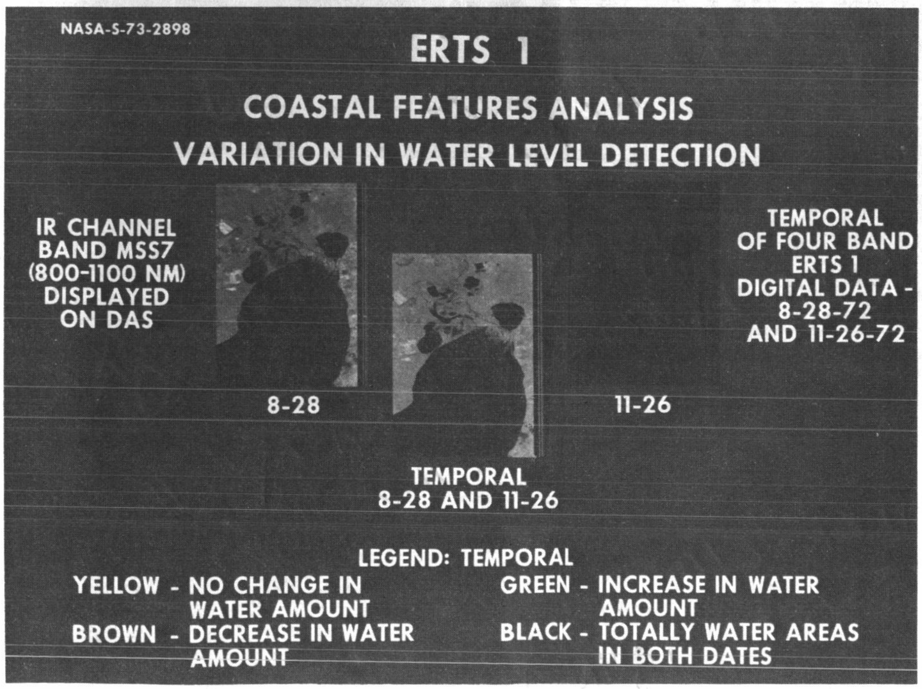

Figure 8.