CAPT. JAMES D. TURINETTI
PROF. OLIN W. MINTZER
*Dept. of the Air Force*
*Griffiss AFB, N.Y. 13440*

# Low-Cost Computerized Land-Use Classification

## It was possible to identify correctly land uses from a variety of combinations of multiformat, non-registered multispectral imagery using a densitometer and a computer to perform the numerical evaluations.

### INTRODUCTION

THE BULK OF THE research that has been conducted in the field of automated image interpretation has used registered imagery, some form of image digitizing equipment, and a computer which could store all of these images. This approach is difficult to adapt to the multiformat, non-registered imagery gathered by many independent researchers, and quite often these researchers do not have the sophisticated equipment available to register or digitize their imagery. Thus a definite need exists for an inexpensive method for analyzing multiformat multispectral imagery. The investigation described in this report was designed to:

- Look into a possible low-cost analysis approach to be used with this multiformat situation;
- Look into the value of combining thermal-infrared image information with other remote sensing systems; and
- Look into the utilization of a computer to analyze the data to identify general land use categories.

ABSTRACT: *A low-cost computerized analysis approach is investigated to identify general land-use classifications from a number of multiformat, non-registered, multispectral images. This imagery was combined in the following manner: (1) Day and Night Thermal-Infrared (TIR); (2) Day and Night TIR with Color-Infrared film; (3) Day and Night TIR with those channels of the Multispectral system that covered the same portion of the spectrum that is covered by color film; (4) Day and Night TIR with those channels of the Multispectral system that covered the same portion of the spectrum that is covered on Color-Infrared film; (5) Day and Night TIR with all eight channels of the Multispectral system. The image densities for selected known land-use classes were processed through a Fortran program and used as a model for identifying sample targets of known land-use classes. The evaluation was performed twice: once with three broad land-use categories, and a second time with each broad category subdivided by landform for a total of 12 possible categories. This made it possible to discriminate among soil, vegetation and water based only on image densities. The test area was located in a selected area of Cameron Parish in southwestern Louisiana.*

The investigation was confined to a limited area (known as the study area) of Cameron Parish in Southwestern Louisiana, which was in a coastal deltaic plain that was covered with poorly drained clays, organic soils and a few lenticular bands of sand along the coast. The imagery was from PROJECT SAND.

Table 1 lists the sensors used to gather the

TABLE 1. SENSOR DATA

| *Sensor* | *Image Scale* | *Acquisition Date* |
|---|---|---|
| AAS-18 Thermal Infrared Scanner Nighttime (TIRN) | 1:17,000 | 8 Mar 69 |
| Bendix Thermal Infrared Scanner Day-time (TIRD) | 1:26,000 | 24 Mar 69 |
| Bendix nine channel Emside Multispectral Scanner (MS) | 1:24,000 | 24 Mar 69 |
| Fairchild KC-9 9" × 9" camera Ektachrome (SO-180) Infrared Film (CIR) | 1:20,000 | 19 Mar 69 |

imagery, image scales and acquisition dates.

All three image acquisition dates had similar climatic conditions, and it was assumed that all of the images were taken under similar conditions. The growing season started late in 1969, so that the vegetation was not very well developed on 8 March 1970, but by 24 March there had been a noticeable amount of growth. No corrections were made to compensate for this difference.

A Macbeth TD-102 transmission densitometer with a 1.0-mm aperture was used to take all of the density readings. A Fortran computer program and the Ohio State University's IBM 370 computer were used to form the evaluation models and analyze the data to identify the land uses represented by the densities. This work could also have been performed quite easily on a mini-computer.

Two mathematical models were constructed for this evaluation. The first model had three broad land-use categories: water, soil and vegetation. The densities for all similar land uses were grouped together to form this model, e.g., the densities for sand, asphalt and combination (sand, shell and clay) were all grouped together under the heading of *Soil*. The second evaluation model had the same three main categories as did the first, but this time each major category was subdivided, e.g., the densities for sand, asphalt and combination were all separated and three model subcategories were formed under the land-use category *Soil*.

Each evaluation model was employed to evaluate sensor combinations and then the results were compared internally within each model and between the two models; the conclusions at the end of the paper were made based on this analysis.

## PROCEDURE

The procedure used for this investigation can be divided into four phases: (1) formulation of basic assumptions concerning the imagery, climate and study area; (2) formulation of logic incorporating the assumptions and the image forming characteristics of the remote sensing systems; (3) formulation of a schedule of comparisons to exercise the logic; and (4) implementation of the schedule of comparisons using the densities from the remote sensing systems' imagery.

The following assumptions were made concerning the imagery, climate and study area:

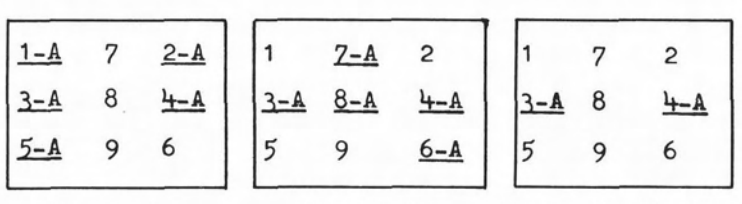


FIG. 1. Sensor Image Identification.

★ There are unique relationships among the densities of land uses formed by remote sensing systems which can be used to identify land uses, differentiating among water, soil and vegetation.

★ The observed imagery was all taken under the same or similar climatic conditions and within a reasonable time frame. (A reasonable time frame was one in which no major seasonal changes had occurred.)

★ Ground truth information was available on the study area and the samples from the imagery were adequate for modeling and classifying land use categories.

The following logic was used to identify land uses from the image densities.

The image densities of known land uses were used to form a model for the specified land uses. Every time the model density was encountered on the imagery, the point was tentatively identified as the given land use. This process was conducted for each land-use category for densities from every piece of imagery. Every time a point was not unanimously identified, that identification was rejected. Figure 1 illustrates this identification process.

Points 1, 2, 3, 4 and 5 are identified as land use *A* on the imagery for Sensor 1. Points 3, 4, 6, 7 and 8 are identified as land use *A* on imagery for Sensor 2. The combination of these two sensors is shown on the Sensor Combination, where only Points 3 and 4 are identified as land use *A*.

Table 2 shows the schedule of sensor combinations used to identify land uses in the various evaluations.

Multispectral Imagery (Color-Infrared) and Multispectral Imagery (Color) used the image densities from those channels of the MS system which scanned the same portion of the spectrum covered on color infrared and color films. Multispectral Imagery, Eight Channel, used the eight channels of imagery from the MS system that were available for this investigation.

The evaluations were performed with two different evaluation models. Evaluation Model I (EMI) had three broad land-use categories: water, soil and vegetation. All samples of a similar nature were grouped into one land-use category to approximate that general land use. A listing of the three land-use categories with the number of areas and the number of samples per category which were used to build EMI are shown in Table 3. Evaluation Model II (EMII) had the same three land-use categories as EMI, but each category was subdivided by landform. The number of possible land-use categories remained at three, but there were now 12 subcategories. Table 4 lists the three land-use

TABLE 2. EVALUATION SCHEDULE

| | | | *MULTISPECTRAL IMAGERY* | | |
|---|---|---|---|---|---|
| | *Thermal Infrared/Day* | *Color Infrared* | *Color Infrared* | *Color* | *Eight Channel* |
| Case I | | X | | | |
| Case II | | | X | | |
| Case III | | | | X | |
| Case IV | | | | | X |
| Case V | X | | | | |
| Case VI | X | X | | | |
| Case VII | X | | X | | |
| Case VIII | X | | | X | |
| Case IX | X | | | | X |

*NOTE: An* X *indicates those sensors used for each case.*

TABLE 3. EVALUATION MODEL I

| *Land Use* | *Number of Sites* | *Total Number of Samples per Sensor* |
|---|---|---|
| Water | 14 | 138 |
| Vegetation | 23 | 289 |
| Soil | 8 | 134 |

TABLE 4. EVALUATION MODEL II

| *Land Use* | *Number of Sites* | *Total Number of Samples per Sensor* |
|---|---|---|
| *Water* | | |
| Beach | 2 | 16 |
| Canal | 3 | 35 |
| Marsh | 8 | 77 |
| Chenier | 1 | 10 |
| *Vegetation* | | |
| Beach | 3 | 25 |
| Marsh | 7 | 85 |
| Chenier | 5 | 48 |
| Terrace - Active | 7 | 116 |
| Terrace - Dormant | 1 | 15 |
| *Soil* | | |
| Sand & Shell | 2 | 44 |
| Asphalt & Shoulder | 2 | 30 |
| Combination | 4 | 60 |

categories and the 12 subcategories with the number of areas and the number of samples per subcategory.

All of the samples used for the land uses in the evaluation models and the targets which were identified with these evaluation models were thoroughly researched and identified before any evaluations were conducted.

Inputs for the evaluation models (known as model sites) were selected according to the following criteria:

- Was the land use of the site known with a reasonable degree of confidence?
- Did the site appear on all pieces of imagery?
- Was the site a good representation of the land use?
- Was the site in a clear, non-degraded portion of the imagery?

Inputs for the targets that were identified with the evaluation models were selected according to the same criteria that were used to select the model sites for the evaluation models, with one additional requirement:

- Was the site small, well defined and easily recognized on all sets of imagery?

The observed densities that were used to develop the evaluation models were taken from model sites which were large enough to allow a minimum of five non-overlapping density readings and still represent only one land-use category.

No effort was made to match up the density from Sensor *A*, sample number two with the density from Sensor *B*, sample number two. A large number of samples were taken for each area on each sensor and then they were all processed to establish the mean value and standard deviation for each sensor for each land use category.

Nineteen sites (known as target points), were selected throughout the study area to be used as targets for identification with the evaluation models. These were prominent, easily recognized sites which could be consistently sampled at the same point on every piece of imagery.

Five separate points and readings were taken for each target point (with maximum overlap), and the readings for each sensor were averaged. This average was then used to represent the density of the target point.

Positive transparencies were used for all density readings. The readings for the black-and-white transparencies were made with the visual filter, Wratten No. 106, on the densitometer, and one reading was all that was necessary to determine the density of the point. The readings for the color transparencies required four readings per time to determine the density of that point. These four readings were made with the red filter (Wratten No. 92), the green filter (Wratten No. 93), the blue filter (Wratten No. 94), and the visual filter (Wratten No. 106), and used in the evaluation model and the target data to represent the color of the land use categories on the CIR.

## THE DYNAMIC MODEL

The term model is used here to describe a set of quantities which approximate or describe something or some object. In particular, the model in this section refers to a set of numbers representing the image densities of land uses on a specific set of remote-sensing imagery. In this context, the term *dynamic* refers to the changing set of conditions that are associated with aerial imagery. The dynamic model was built from image samples which portrayed these various conditions and it was in turn applied to a set of imagery which was gathered under the same conditions.

A number of variables influence every aerial scene and are extremely difficult to model and compensate for. Some of these different conditions include: the atmospheric conditions at the time of the flight, the time of day, the condition of the vegetation and the altitude of the aircraft. These are all variables which directly influenced the density of the images, but were incorporated into the Dynamic Model because the densities for the model and the unknown objects were on the same imagery.

In this investigation, the land uses of a number of sites were known, and it was desired to identify other sites with similar land uses. Image densities of the known sites were used to establish a model which represented the densities of land uses throughout the study area. The target densities were then compared with this model and identified accordingly. The densities for the model sites and the target points were located such that the atmospheric conditions, vegetation, etc., were the same for both the target and model sites. For this investigation the model sites and the target points were intermixed on the same imagery so that not only were the physical parameters of the scene the same, but the image parameters of scale and processing were also the same because of the juxtaposition of the sites. An alternate method of site positioning would have been to select model sites around the perimeter of the image and the target points within the perimeter.

This approach accounted for the specific climatic conditions at the moment of exposure because the atmosphere above the model sites was the same as that over the target points. The altitude was the same, and scene parameters were the same; even the chemical processing of the imagery was the same for both the model site and target point images, as they were both taken from the same piece of imagery. (Note: This is not an unrealistic approach considering the large areas covered on satellite images or the controlled crop studies conducted in the Midwest.

If the inputs to the model are an accurate representation of the larger family of image densities that were used to describe the various land uses, then the model could be used as an accurate land-use discriminator. If the inputs to the model were ambiguous and the model had very general, broad parameters, then the results could also be expected to be ambiguous.

The area covered by the model had to be limited in size so that it did not include any different regional, climatic or atmospheric conditions. If the model included too many variables, the concept of the Dynamic Model approximating a unique set of conditions would have been lost. It was also necessary to have *a priori* knowledge of some of the land uses on the imagery in order to build the model to be extrapolated over the entire area.

## The Fortran Program

The computer program for this investigation was built on a simple scale. A mathematical model was formed from image densities which represented the land-use categories under investigation. The densities of the images of an unknown target were compared with this model, and every time the target and the model were compatible, the target was identified accordingly.

The Fortran program was divided into two main sections. The first section constructed the evaluation model by means of a series of *IF* statements, and identified the land uses of the targets.

The evaluation model was a numerical approximation for the range of densities that might be found on a set of imagery for a set of land uses. It contained the projected high and low values for the densities of the images of known land uses on that same set of imagery. It should be emphasized that this evaluation model related to the specific set of imagery that was used to construct the model, and it was not meant to be applied to any other set of imagery.

The evaluation model was constructed on the assumption that all of the densities for a single category on a single sensor had a normal distribution. (A sampling of the values was plotted, and they did have or tend toward a normal distribution.) A tolerance interval was used statistically to predict the projected high- and low-density values for each image model. (A tolerance interval is a measure of the range of values that are found in a sample of a larger population. The interval states that $X$ percent of the time, $Y$ percent of the population will be within the limits of the tolerance interval.) The limits of the interval were found according to the following formula:

$$Limit = M \pm S \times L$$

where $M$ is the mean of the sample, $S$ is the standard deviation of the sample, $L$ is a multiplication factor determined from a table and based on the number of members in the sample.

A number of tolerance intervals were evaluated, and finally one with parameters of $X = 90$ percent and $Y = 90$ percent was selected. This tolerance interval provided a practical method of identifying the limits for the density values, and was a workable link between the sample size and the limit values. A tolerance interval was not necessarily the only method for computing these limits, but it was one which produced consistent results under a variety of conditions and diversity of imagery. (Note: the flow charts for this program and a discussion of the program can be found in Reference 9.)

The actual land-use identifications were made with an extended Fortran *IF* statement. The upper and lower values of the densities for a specific land-use category for each type of imagery were stored in a matrix and the target densities were evaluated against these limits. If the target density for the first sensor's imagery was within the limits of the densities for that category on that sensor, the *IF* statement was *True* and the program executed the next branch of the *IF* statement. If the sample density for the second sensor's imagery was within the density limits for this sensor, that portion of the *IF* statement was also *True*, and the process continued until all branches of the *IF* statement had been evaluated and found to be *True*, or until the target density and the model density limits were not compatible, and the program exited the *IF* statement. If the sample data were within the limits of the model for all the sensors used in the *IF* statement, the sample was identified as that land-use category, and then the evaluation was repeated with the same target data and a new land-use category until the target data had been evaluated against the models for all of the possible land uses, each being either identified or rejected.

TABLE 5. STATISTICS FOR LAND-USE IDENTIFICATION

EVALUATION MODEL I

| | *Thermal Infrared Day and Night* | *Color Infrared With and Without Thermal Infrared* | | *Multispectral Color Infrared With and Without Thermal Infrared* | | *Multispectral Color With and Without Thermal Infrared* | | *Multispectral Eight Channels With and Without Thermal Infrared* | |
|---|---|---|---|---|---|---|---|---|---|
| | | *With* | *Without* | *With* | *Without* | *With* | *Without* | *With* | *Without* |
| 1. Average Number of Landuses Identified Per Target | 1.74 | 1.42 | 1.63 | 1.26 | 1.63 | 1.32 | 1.53 | 1.32 | 1.31 |
| 2. Percentage of Correct Identifications | 100 | 100 | 94.6 | 89.5 | 84.1 | 89.5 | 79.0 | 89.5 | 72.6 |
| 3. Percentage of Unique Identifications | 31.6 | 63.2 | 31.6 | 68.4 | 57.9 | 63.1 | 31.6 | 63.1 | 52.6 |
| 4. Percentage of Time that a Choice had to be Made | 68.4 | 36.8 | 68.4 | 31.6 | 42.1 | 36.9 | 68.4 | 36.9 | 47.4 |
| 5. Average Number of Possible Landuses to Choose From | 2.32 | 2.21 | 1.92 | 2.20 | 1.95 | 1.99 | 1.76 | 1.98 | 1.67 |
| 6. Percentage of Correct Landuse Identifications in List of Choices | 100 | 100 | 92.7 | 77.1 | 54.6 | 83.2 | 69.4 | 77.1 | 55.7 |

TABLE 6. STATISTICS FOR LAND-USE IDENTIFICATIONS

EVALUATION MODEL II

| | *Thermal Infrared Day and Night* | *Color Infrared With and Without Thermal Infrared* | | *Multispectral Color Infrared With and Without Thermal Infrared* | | *Multispectral Color With and Without Thermal Infrared* | | *Multispectral Eight Channels With and Without Thermal Infrared* | |
|---|---|---|---|---|---|---|---|---|---|
| | | *With* | *Without* | *With* | *Without* | *With* | *Without* | *With* | *Without* |
| 1. Average Number of Landuses Identified per Target | 1.58 | 1.16 | 1.58 | 1.21 | 1.37 | 1.21 | 1.26 | 1.21 | 1.32 |
| 2. Percentage of Correct Identifications | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 3. Percentage of Unique Identifications | 42.1 | 84.2 | 42.1 | 79.0 | 78.7 | 84.2 | 84.2 | 79.0 | 84.4 |
| 4. Percentage of Time a Choice had to be Made | 57.9 | 15.8 | 57.9 | 21.0 | 21.3 | 15.8 | 15.8 | 21.0 | 15.6 |
| 5. Average Number of Possible Landuses to Choose from | 2.00 | 2.00 | 2.00 | 2.20 | 2.47 | 2.31 | 2.68 | 2.20 | 3.00 |
| 6. Percentage of Correct Landuse Identifications in List of Choices | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

TABLE 7. STATISTICAL COMPARISON OF LAND-USE IIDENTIFICATIONS

EVALUATION MODELS I AND II

| | *Thermal* | *Infrared* | *Color Infrared and Thermal Infrared* | | *Multispectrnl Color Infrared and Thermal Infrared* | | *Multispectral Color and Thermal Infrared* | | *Multispectral Eight Channels and Thermal Infrared* | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *Model I* | *Model II* | *Model I* | *Model II* | *Model I* | *Model II* | *Model I* | *Model II* | *Model I* | *Model II* |
| 1. Average Number of Landuses Identified per Target | 1.74 | 1.58 | 1.42 | 1.16 | 1.26 | 1.21 | 1.32 | 1.21 | 1.32 | 1.21 |
| 2. Percentage of Correct Identifications | 100 | 100 | 100 | 100 | 89.5 | 100 | 89.5 | 100 | 89.5 | 100 |
| 3. Percentage of Unique Identifications | 31.6 | 42.1 | 63.2 | 84.2 | 68.4 | 79.0 | 63.1 | 84.2 | 63.6 | 79.0 |
| 4. Percentage of Time a Choice had to be Made | 68.4 | 57.9 | 36.8 | 15.8 | 31.6 | 21.0 | 36.9 | 15.8 | 36.4 | 15.6 |
| 5. Average Number of Possible Landuses to Choose From | 2.32 | 2.00 | 2.21 | 2.00 | 2.20 | 2.20 | 1.99 | 2.31 | 1.98 | 2.20 |
| 6. Percentage of Correct Landuse Identifications in List of Choices | 100 | 100 | 100 | 100 | 77.1 | 100 | 83.2 | 100 | 77.1 | 100 |

## Analysis of Results

The specific identifications of the target points were tabulated, statistically analyzed, and the results are listed in Tables 5 and 6 so that the reader may compare the results of the sensor evaluations with and without day and night thermal-infrared data for both evaluation models. These results demonstrate the improvement in land use identification when TIR data is used.

The following questions were used to make this analysis:

- If an average target was identified, how many possible land-use categories did the computer list for its land-use?
- How often was the correct land use contained in the list of possible land uses referred to in question one?
- How often was the target correctly identified with only one land-use category? (This is called a unique identification.)
- How often did the interpreter have to choose the land use from a list of possible land uses? (This is called an ambiguous identification.)
- When the interpreter had to choose a land use from a list of possible land uses, how many choices did he have to choose from?
- When the interpreter had to choose a land use from a list of possible land uses, how often was the correct answer in that list of choices?

An analysis of Tables 5 and 6 shows that if TIR densities were added to an evaluation model, the land-use identifications from those evaluations were improved over those evaluations made without TIR data. Both models demonstrated this improvement. This indicates that: (1) the figures for the average number of land uses identified per target were lower; (2) the identifications were more accurate for the evaluations with TIR data than for those without TIR data; and (3) the frequency of correct unique identifications was higher if TIR data was used than if it was not used.

Table 7 combines information from Tables 5 and 6 so that the reader may compare the results of the sensor evaluations made with TIR data in combination with the other sensors for both evaluation models.

An analysis of Table 7 shows that: (1) the overall identifications from EMII had fewer choices per target than those from EMI; (2) 100 percent of the time the correct land use was contained in these possible choices; (3) unique identifications occurred more frequently with EMII than EMI; and (4) if a choice had to be made to identify the land use of a target even though the number of choices from EMII was slightly higher than from EMI in two instances, the choices from EMII always contained the correct answer, although errors occurred in the identifications made with Evaluation Model I. These statistics indicated an improvement in the identifications from Evaluation Model II over those from Evaluation Model I.

The difference between the identifications made by the two evaluation models can be attributed to the structure of the models. EMI grouped all similar densities together and used broad general parameters for each land-use category. It was a rough or average approximation of the land uses without any detail or definition, similar to a curve drawn from only a few data points. EMII subdivided each land-use category and provided detailed parameters within the general land-use framework, similar to a curve drawn from a very large number of points. EMII provided a good approximation for the land uses, and as a result the land-use identifications with EMII were better than those from EMI. This was true for all evaluations, both with and without TIR data.

## Conclusions

It was possible to identify correctly land uses from a variety of combinations of multiformat, non-registered multispectral imagery using a densitometer and a computer to perform the numerical evaluations.

The laboratory microdensitometer was a useful tool for gathering the selected density readings. Using the available 1.0–mm aperture ring caused a problem. This aperture was too large and thus sampled too large a ground area. (The reading from the CIR covered a ground distance of 20 meters, and the reading from the MS covered 24 meters on the ground.) It is therefore recommended that a smaller sampling aperture be used for future work of this type.

The combination of TIRD and TIRN with CIR imagery improved the accuracy of land-use identifications over those with only TIRD and TIRN or CIR alone. The combination of TIRD and TIRN with MS imagery also improved the identification accuracies over those of only TIRD and TIRN or MS alone.

The more accurately the evaluation model approximated the actual land uses displayed on the imagery, the better the target identifications.

## Bibliography

1. Buckley, A. A., "Computerized Isodensity Mapping", *Photogrammetric Engineering*, 37:10, A.S.P., Falls Church, Virginia: October 1971.
2. Carnes, R. S. and D. S. Hardin, *Multispectral*

*Imagery Information Content,* The Bendix Corporation, Ann Arbor, Michigan: 1969.

3. Doverspike, G. E., et al., "Microdensitometer Applied to Land Use Classification", *Photogrammetric Engineering,* 31:2, A.S.P., Falls Church, Virginia: March 1965.
4. Nunnally, N. R., "Remote Sensing for Land Use Studies", *Photogrammetric Engineering,* 36:5, A.S.P., Falls Church, Virginia: May 1970.
5. Orr, Donald G., "Remote Sensor Imagery Analysis for Location of Construction Materials in the Mekong Delta", Project SAND (Phase II), Technical Report 52TR, US Army Engineer Topographic Laboratories, January 1970.
6. Orr, D. G. and J. R. Quick, "Construction Materials in Delta Areas", *Photogrammetric Engineering,* A.S.P., Falls Church, Virginia: April 1971.
7. Tarkington, R. G. and A. L. Sorem, "Color and False-Color Films for Aerial Photography", *Photogrammetric Engineering,* 24:1 A.S.P., Falls Church, Virginia: January 1963.
8. Turinetti, J. D., *Computer Analysis of Multispectral Image Densities to Classify Landuses,* RADC-TR-72-237 Technical Report, Sep 1972, Rome Air Development Center, Air Force Systems Command, Griffiss AFB, New York: September 1972.
9. Turinetti, J. D. and O. W. Mintzer, "Computer Analysis of Imagery", *Photogrammetric Engineering,* 39:5, A.S.P., Falls Church, Virginia: May 1973.

---