E7.3 10644
CR-132103

AUTOMATED THEMATIC MAPPING AND CHANGE DETECTION OF ERTS-A IMAGES

Nicholas Gramenopoulos
Optical Systems Division
Itek Corporation
10 Maguire Road
Lexington, Massachusetts 02173

June, 1973

Type I Report
Period Reported: 31 March 1973 - 31 May 1973

Prepared for
GODDARD SPACE FLIGHT CENTER
Greenbelt, Maryland 20771

(E73-10644) AUTOMATED THEMATIC MAPPING AND CHANGE DETECTION OF ERTS-A IMAGES Progress Report, 31 Mar. - 31 May 1973 (Itek Corp.) 7 p HC $3.00 CSCL 08B

N73-25341

Unclas
G3/13 00644

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Automated Thematic Mapping and Change Detection of ERTS-A Images		5. Report Date June 1973
		6. Performing Organization Code
7. Author(s) Nicholas Gramenopoulos PR 504		8. Performing Organization Report No.
9. Performing Organization Name and Address Optical Systems Division Itek Corporation 10 Maguire Road Lexington, Massachusetts 02173		10. Work Unit No.
		11. Contract or Grant No. NAS5-21766
		13. Type of Report and Period Covered Type I Report 31 March-31 May 1973
12. Sponsoring Agency Name and Address Goddard Space Flight Center Greenbelt, Maryland 20771 Technical Monitor: Edmund F. Szajna		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This investigation is concerned with the development of automated interpretation techniques for the recognition and identification of earth resources. The resources will be identified by using both multispectral and spatial signatures. Ground truth data and aircraft under-flight photography will be used to train the recognition algorithms. The data processed will be RBV and MSS images acquired by the ERTS-1 satellite, over six test sites located in the vicinities of: Phoenix, Arizona; Weslaco, Texas; Cascade Mountains, Washington; New Orleans, Louisiana; Salt Lake, Utah; and Salton Sea, California. The processed data will be thematic maps of resources consisting of annotated and outlined images. The seasonal changes of hydrologic and agricultural resources will also be identified. The results are expected to be applicable to a future automatic system of resource inventory and management.

17. Key Words (Selected by Author(s))	18. Distribution Statement
Interpretation Techniques Development, Digital Information Extraction Techniques, Classification and Pattern Recognition	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	5	

*For sale by the Clearinghouse for Federal Scientific and Technical Information, Springfield, Virginia 22151.

DISCUSSION

During this reporting period, very substantial progress was made in this investigation. To give some perspective to this progress, the status of the investigation at the end of March, 1973 will be briefly summarized.

Status of Investigation at the End of March, 1973

ERTS-1 image 1049-17324 which covers an area in southern Arizona, around Phoenix, had been analyzed. Initial attempts to process the digital data of this image through our Pattern and Terrain Classification Software System had failed. Then, detailed analysis of the data showed the presence of noise which was responsible for the failure of the algorithms to recognize reliably the various terrain types. Once the noise was discovered, it was quickly traced to small calibration errors of the MSS detectors (there are six detectors per spectral band). We filtered the noise in the Fourier domain and proceeded to develop new signatures for the various terrain types. Finally, the data was subjected to automatic terrain recognition by the modified PTCS software. The recognition results for most terrain types were very good: 97% for desert, 89% for farms, 80% for mountains, 74% for urban areas, 98% for clouds, 100% for water, 81% for cloud shadows. Only river flood plains which are peculiar geographic features of southern Arizona were recognized poorly (11%). The accuracy of recognition was determined by comparison to existing maps, high altitude (U-2 missions in September and December, 1971) and low-altitude aerial photography (mission 212 of the Earth Resources Aircraft Program, from the Manned Spacecraft Center in Houston, Texas). The accuracy of the cloud recognition was determined by photointerpretation of the ERTS image since the aerial photography was obtained on different dates.

The automated recognition of terrain types was described in a paper given to the ERTS-1 Symposium on March 5-9, 1973 sponsored by NASA/GSFC. The title of this paper is "Terrain Type Recognition Using ERTS-1 MSS Images". The key to the recognition process described above is a heuristically determined algorithm which utilizes the new spatial features developed as well as the brightness information of the MSS-5 band.

Selected Course of the Investigation

At the end of March, the status of the investigation was thoroughly reviewed with the scientific monitor, Mr. William Alford, and the following course was selected:

1. It was decided to combine the multispectral data with the spatial features in the recognition algorithm. The earth resources information present in the ERTS images is conveyed by individual picture elements by their brightness values in the four spectral bands and collectively by the picture element distribution over the image. Both types of information (spectral and spatial) must be exploited.

2. The heuristic algorithm produced excellent results but took a lot of time to develop. Since the purpose of automatic processing is to compete economically with manual photointerpretation, it appeared that the heuristic algorithm was not cost effective when employed with a general purpose computer. It was decided, therefore, to employ the maximum likelihood criterion algoritm instead. This algorithm has been in wide use by many investigators for analyzing multispectral and/or spatial feature data. The results have not always been good. We suspected that this algorithm produces errors when the various resource classes to be recognized do not have Gaussian statistics. It was decided, therefore, that the statistics of the classes be analyzed and that nonlinear transformations of the data be developed which would render this algorithm's results more accurate.

3. It was also decided to introduce "clustering" techniques which are necessary to avoid requiring training data for each ERTS image. The criterion to be used for clustering is also dependent on Gaussian statistics so, as a first step it is necessary to show that the criterion behaves properly.

<u>Status of the Investigation at the End of May, 1973</u>

1. Preprocessing of ERTS-1 Images

   The preprocessing operations have been completed for the following ERTS-1 images:

   1031-17325 from Phoenix, Arizona
   1049-17324 from Phoenix, Arizona
   1103-17332 from Phoenix, Arizona

   1040-18201 from Cascade Mountains
   1077-18260 from Cascade Mountains

   1015-17415 from Salt Lake, Utah

We also tried to preprocess image 1069-17420 but found it was an 800BPI tape mislabeled as a 556BPI. The preprocessing operations involve taking data from the NASA delivered bulk CCT's, combining the data over the area of interest from two CCT's, separating the spectral bands, resampling each band to achieve equal pixel spacing on the ground, filtering the noise and other artifacts such as missing lines and recording individual images in the laser beam recorder to verify that each band has been properly preprocessed and no noise or other problems remain in the data.

2. Six spatial features were combined with three spectral bands (MSS 4, 5 and 7) to produce a 9-dimensional vector for each image cell (32 x 32 picture elements). The maximum likelihood criterion was employed. We found that the covariance matrix for mountains had a singularity and mountains could not be recognized. Two of the spatial features were used for recognizing clouds and water only. At this point we decided to remove these two features and we have been using a seven-dimensional vector ever since. Clouds and water are still recognizable but not by the maximum likelihood criterion.

3. Using the seven-dimensional vectors and image 1049-17324 as training data, the statistics (covariance matrices) for desert, farm, mountains and city were computed. Then, using the statistics, the same input data and the maximum likelihood criterion, the cells were reclassified in one of the four categories. The recognition accuracy was 54% for desert, 92% for farms, 97% for mountains, and 92% for cities. The classification accuracy for farms, mountains and cities improved over the results obtained in March, using the spatial information only and the heuristic algorithm. Desert was poorly recognized (many desert cells were assigned to the mountain category), and we suspected that the statistics for the classes were not Gaussian. Analysis of histograms of the classes showed that no component of any class was even approximately Gaussian. The distributions of each component (spectral band or spatial feature) appeared like the Rayleigh rather than the Gaussian. All components were positive with small means and large standard deviations.

To make the distributions approximately Gaussian various nonlinear transformations were applied on all seven vector components. We found that various logarithms worked fairly well but the resultant distributions are sensitive to amplitude variations in the data. In other words, if the logarithm on a certain base is used for one component (for example, the MSS 7 band) the distribution for each class may or may not become approximate Gaussian depending on solar illumination which changes the range of values obtained in this band. Finally, the nonlinear transformations we have selected are all powers less than 1. For image 1049-17324 the powers for the various components range from 0.8 to 0.025 and have been optimized to produce excellent recognition results in all four classes: desert 89%, farms 97%, mountains, 96%, and city 95%. In arriving at an optimum transformation for each component, Gaussian distributions cannot be achieved for all classes. In other words, as one adjusts the transformation of say the MSS 5 band, the desert distribution may become more Gaussian while the city distribution may become less Gaussian. There is no reason to believe that all distributions can always be made Gaussian in all components by appropriate transformations. To reduce the possibility of such problems and maintain high recognition rates, the number of different classes should be kept small.

Also, the importance of each component in the recognition of each class is not the same. A component may be essential for recognizing one class and relatively unimportant for recognizing other classes. Knowledge of this can be taken advantage of when selecting a transformation for a component. The transformation is adjusted so that an approximately Gaussian distribution is obtained for the class that the component is most important. The effectiveness of each transformation can be judged from the recognition rates achieved. (See above.)

4. There has been considerable work with the "clustering algorithms". In particular, the divergence criterion between classes was examined. This criterion allows a determination of the statistical separability of any two Gaussian classes. If the divergence is greater than 10, then the probability that a vector belonging to one of two classes can be correctly identified, is greater than 80%. The divergence is also dependent on Gaussian statistics. Before the nonlinear

transformations of the vectors (when the classes were highly non-Gaussian) the divergence values varied between 400 and 6×10^8 and seemed to bear no relationship to the recognition rates achieved. After the nonlinear transformations were completed rendering the classes approximately Gaussian, the divergence values took on a more reasonable range (20 - 100) and generally speaking the higher divergence values were associated with pairs of classes between which very few errors were made.

For the next reporting period, we hope to complete the following tasks:

1. Complete all preprocessing operations for three more images

2. Complete clustering algorithm software development

3. Process eight images through the clustering and recognition algorithms.