DR. JOHN E. ESTES* LESLIE W. SENGER $University$ of *California Suntn Barbara, Cdf. 93106*

An Electronic Multi-Image Processor

A quantitative evaluation of the University of Kansas system shows that it offers a possible manner for processing the imagery from an Earth Resources Satellite.

INTRODUCTION

THE TREMENDOUS AMOUNTS OF multiband coming from the ERTS program and from supporting aerial photographic programs require sophisticated processing techniques so that the data can be combined, correlated and enhanced effectively. A possible answer to these

A multi-image processor is a system either electronic, optical, or a combination of the two, which brings into common register images of an object or an area taken in various spectral bands. Basic cueing for the system is obtained from either tone variations, such as shape and texture, or tone values. First-order, second-order, and third-order systems obtain

ABSTRACT: *The ERTS program will produce large quantities of multiband data requiring sophisticated processing techniques. A possible answer lies with the University of Kansas electronic multi-image processor. A comparative one-way analysis-of-variance test sequence was performed to relate interpretability of imagery obtained with the Kansas system to that of comparable images produced by other multi-spectral combiners and by filtered panchromatic and color films.* The experimental hypothesis was that different types of imagery exhibit different *degrees of interpretability for particular earth features. Interpreters of various levels of skill were tested in terms of correct responses, omission error committed and commission errors committed for features in diverse environments. Results of the experiments indicate that the Kansas system is competitive with other enhancement and photographic systems as an aid to interpretation* of *earth features. Research to advance the state of the art in terms of operational capability is recommended.*

urgent needs lies with the University of Kansas electronic multi-image processor, IDECS. The study reported on here concerned a preliminary evaluation of this system's capability to accomplish these tasks.

* The research presented here is an outgrowth of the thematic experiments conducted under U.S. Army Engineering Laboratories Contract No. DAAK 2-67-C-0435 with the Center for Research in Engineering Science, University of Kansas, "Multi-Image Correlation Systems Study for MGI", and is reported on in "Multi-Image Cor-relation Systems Study for MGI Quantitative Evaluation of Electronic Multi-Image Processor
(U) Supplementary Final Report", U.S. Army Engineering Topographic Laboratories, Contract accomplished by the authors while they were employed by the Forestry Remote Sensing Laboratory, University of California, Berkeley.

cueing from tone variations, tone values, and a combination of tonal variations and tone values respectively.

IDECS is an acronym for "image discrimination, enhancement, combination and sampling." The system has been developed at the University of Kansas as a versatile research tool to be used for interpretation of multiband imagery. Among its first-order interpretation modes are differentiation and monochromatic additive combination. Second-order cueing modes include additive color combination, single and multi-image level selection (spectral selection) and isodensitrace production. The IDECS system is essentially a second order system, its most powerful interpretation tools (color combination and spectral selection) being cued for tone values. However, the system

may be used to simulate higher dimensioned (i.e., larger number) input vector systems and a general third order system by inclusion of one or several intermediate steps. Information from several images may be recorded as a monochromatic additive enhancement and/or textural information may be coded as tone values. These may then be combined to form a third-order processor output. Since processing occurs rapidly (1/30 second per input set), automatic interpretation modes are investigated simultaneously with interpreter-aiding modes-the speed limitation being imagehandling time rather than processing time. Input images may be of any size up to 3 by 4 inches, and output display recording is usually accomplished by taking color photography of the color T.V. screen on which the enhanced imagery appears.

In the thematic experiment performed under a contract that preceded the one here reported upon, an evaluation of IDECS was accomplished by researchers with a knowledge of IDECS operation. However, no attempt was made in the earlier test to compare the interpretability of features on IDECS enhanced imagery with that of conventional black-andwhite photography, color photography, 01 with multiband photography enhanced with equipment other than IDECS. The making of such comparisons was considered essential if we are ever to develop the most efficient operational procedures, from an interpretation standpoint, for the electronic enhancement of images.

EXPERIMENTAL HYPOTHESIS

The research reported on here concerned the use of a comparative test sequence to relate the interpretability of particular images obtained utilizing IDECS to that of comparable images produced by other combiners of multispectral images and by images originally produced on both filtered panchromatic and color films. For each technique or device used, a single enhanced/combined image was selected for each phenomenon (category) at each test site, and comparisons were made of theit interpretability in terms of errors of omission¹

¹ "Omission" is defined as the complement of "correct", i.e., $\%$ Omission= (100- $\%$ C). "Commission", on the other hand, is defined as the actual number of errors committed (e.g., calling a field "alfalfa" where in reality it is a barley field, ex- pressed as a percent. Thus, and commission, correct identification, and time for interpretation. The interpretation tests were designed to (1) obtain statistically valid measures of interpreter performance for studying the different kinds of imagery; (2) measure this performance in terms of both time required and accuracy achieved: and (3) measure performance as a function of the background of training and experience of the photo interpreter.

The theoretical basis for the experiment was the hypothesis that different types of imagery exhibit varying degrees of interpretability for particular earth features. Signatures of earth resource phenomena vary according to that portion of the electromagnetic spectrum which is being imaged. A particular feature may be relatively easy to interpret using one type of imaging system (e.g., sugar beet fields on radar), whereas interpretation of the same phenomena from imagery of another system may be quite difficult (e.g., sugar beets from barley, alfalfa, or wheat on ektachrome infrared film). Such differences should permit the types of imagery to be ranked acccrding to interpretability, feature-by-feature. The experiment was designed to hold all other factors as nearly constant as possible so that if the interpretability of any given feature was being tested the only variable would be the imagery itself. This variability manifested itself quantitatively in terms of both the time and the accuracy of interpretation of each feature as a function of type of imagery. Accuracy ratings took into consideration the correct responses, errors of omission, and errors of commission.

TESTING PROCEDURES

The experiments performed involved the use of nine different kinds of imagery from both photographic and enhancement systems (panchromatic film with a Wratten 25A filter, panchromatic film with a Wratten 58 filter, black-and-white infrared film with a Wratten 89B filter, color aerial film, infrared ektachrome film with a Wratten 15 filter, a Forestry Remote Sensing Laboratory optical combiner image, a University of Kansas electronic combiner IDECS image, a Philco-Ford electronic combiner image and combined interpretation of a color, infrared ektachrome and University of Kansas electronic

Total number of commission errors $\%$ Commission = \times 100

Total possible responses - total possible correct responses

PLATE 1. IDECs enhancement of Apollo imagery
for the Imperial Valley test site.

PLATE 2. Apollo infrared Ektachrome of the Dallas-Fort Worth test site. Training and prediction areas are inked directly on the print's surface. Lettered squares indicate training areas with initials indicating categories,

* Courtesy Academic Senate, University of California, Santa Barbara.

combiner **IDECS** image). Each type of imagery was tested for the interpretation of four discrete categories, and a fifth category which combined all four categories for any of the given areas. This produced a 5×9 array in which there were 45 individual cells, each cell representing a unique combination of a particular category with a particular type of imagery. This required the use of a minimum of 45 interpreters in order to fill each cell with an individual interpreter, minimizing the problem of area familiarization. In addition, it was considered to be more valid statistically to have several replications per cell, at least three replications being a minimum requirement.

The 45 interpreters required for these tests were selected in such a way that they could be placed into three groups of 15 each according to their level of competence-high, medium, low-based on background data sheets filled out by the interpreters. A *high* (highly skilled interpreter) was ranked as such if he had taken a remote sensing course and also had had several years of work experience in the field, a *medium* was one who had taken a course in air photo interpretation or remote sensing but had had little work experience in the field, and a *low* was one who had neither taken a course nor had obtained any work experience in the field. Each person was then randomly assigned to each of three cells such that each of the 45 cells contained a high, medium and low interpreter. This was accomplished in a manner such that no person would look at one of the four areas more than three times. This arrangement kept the problem of area familiarization to a realistic minimum. It reduced the problem of interaction between the categories to be interpreted in a given area, and it also permitted three replications per cell for more valid statistical results. By establishing each cell as equivalent to any other cell in terms of level of competence, this test allowed meaningful comparisons to be made between cells (image types) by category.

At the outset, two methods of presenting images to the interpreter were tested, one employing prints and a second requiring projection of the images on to a viewing screen. One area was used as a control and an equal number of skilled and unskilled interpreters were asked to interpret individual categories in the same manner as detailed in the basic test procedure seen below. Statistically, no significant difference was noted between the interpretations accomplished using prints and those using projection techniques. Consequently, the researchers were able to choose what they considered to be the optimum method for presenting to the interpreter the test material for a given site. The actual tests were administered individually with a particular interpreter looking at one category on a discrete image type at one time.

Where *prints* were employed, the interpreter was asked to identify particular categories in a prediction area based on information given in a training area on the same image. Information for the training samples was dexived from available *ground truth* maps of the test areas. The interpreters were required to record their identifications on a map where the prediction samples and training samples were delimited. An example of an interpretation of Bare Soil from **IDECS** enhanced Apollo imagery of the Imperial Valley can be seen in Plate 1. In this example, Miss Wall was given both the image and a map of the area on which only those fields of bare soil in the training area had been filled in. She then took six minutes to fill in what she considered to be the fields of bare soil on the rest of the image. A comparison with the "ground truth" map for the area (see Figure 1) shows that she correctly identified 16 fields, committed errors in 7 instances and omitted 6 fields which were in actuality bare soil.

FIG. 1. Ground truth category test format pre- sented to interpreters for the **IDECS** enhancement of Apollo imagery and ground truth category test. Miss Wall was given this sheet with only those areas labeled bare soil (BS) in the training area filled in. She then lableed those fields on the rest of the sheet outside the training area that she considered to be bare soil with a BS based on Plate 1. Her time was recorded at the upper right and her test results are shown at the lower left.

The infrared ektachrome image of the Dallas-Fort Worth test site (Plate **2)** represents another method of presenting the data to the interpreter. 8×10 -inch prints were used in these instances. Here the training and prediction areas were inked directly onto the print's surface. Each training area was labeled with an initial indicative of its category (T) for trees, W for wheat, etc.) whereas prediction areas were numbered sequentially from 1 through 21. The interpreter was given an answer sheet on which was specified the name of the category which he was to interpret. Entries on the sheet also indicated the area and image type and provided space for name, date and time. The remainder of the sheet contained numbers followed by blank spaces corresponding to the numbers of the prediction areas labeled. By this means the results obtained could readily be analyzed in terms of correct answers, omission errors, and time required to complete the test.

Where *projection* was employed, as in imagery of the Imperial Valley test area, answer sheets such as those shown in Figure 1 were used. In other instances, however, such as the Meadow Valley test area, an overlay containing the training and prediction areas was attached to a viewing screen and the image was then congruenced to the *ground truth* overlay. Here the interpreters were given numbered answer sheets on which to place letters that would signify how they had interpreted specific areas, with the numbers on the sheet corresponding to those in the prediction area. These interpretation tests were carried out using projection techniques with the interpreters viewing images projected onto a screen in a darkened room. The resultant images measured approximately 18×13 inches and were viewed from a distance of approximately 12 feet, giving a scale on the Apollo imagery of the Imperial Valley of about 1 : 15,000; however, interpreters were encouraged to view the screen at what they considered the best distance. This ranged from a few feet up to 20 feet. Numbers of interpreters varied, with a maximum of 6 individuals taking a particular test at a given time.

STATISTICAL DESIGN

The time and accuracy information was used as raw data for a computer program specifically selected for this experiment (the statistical methodology of this program is described in detail presently). The computer printout consisted of a one-way analysis of variance table which supplied the observed

F-ratio for each category. These F-ratios formed the basis for the image rankings by categories on the computer printout. A Duncan multiple-range test was then applied to the ranked data for each category in order to determine whether differences existed in the data at the 5 percent level of significance. Further analyses of the raw data and image rankings were made by manual calculation.

The purpose of these various calculations was to achieve the following objectives:

- To determine the percent time and accuracy of interpreters at each level of competence.
- To rank the types of imagery by individual categories for each area according to correct responses, errors of commission and time.
- To rank the types of imagery by individual categories for each area according to a composite score of time and accuracy through the linear combination of the percent scores for omission errors, commission errors and time.

In this experiment there were k (8 or 9) types of imagery that were of interest to be ranked for each category within each area. The statistical model for this experiment can be considered a one-way classification with *n(3)* replications. Therefore, a completely randomized design was most suitable for our analysis.

The linear model for this design is:

$$
X_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad i = 1, 2, \cdots k; j = 1, 2, \cdots n
$$

where μ is the general mean, α_i is the main effect for the *i*th image, and ϵ_{ii} is the experimental error for the jth replication on the ith image.

To test the significance of difference of the main effects for the images, the various mean squares and F-ratio need to be computed under the null hypothesis:

$$
H_0 = \alpha_i = 0 \quad \text{for all } i = 1, 2, \cdots k.
$$

Based on the above model, the sums of squares can be obtained as follows: Total sum of square:

$$
(TSS) = \sum_{i=1}^{k} \sum_{j=1}^{n} X_{ij}^{2} - C
$$

$$
C = \frac{1}{k \cdot n} \left(\sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij} \right)^{2}
$$

Between-image groups sum of square:

$$
(BSS) = \frac{1}{n} \sum_{i=1}^{k} \left(\sum_{j=1}^{n} x_{ij} \right) 2 - C
$$

Within-image groups sum of square or error sum of square:

$$
(ESS) = TSS - BSS = \sum_{j=1}^{3} (s_{ij} - \bar{x}_i)^2
$$

TABLE 1.

The analysis of variance for this one-way classification is shown in Table 1.

The computer F -ratio is significant if it exceeds the theoretical F-value that is obtained from an F -table. If F is greater than $Fk - I_{\alpha}$, $(n-1)k$, the null hypothesis that there is no difference between the images can be rejected with a $(I - \alpha)$ level of confidence.

If the null hypothesis is rejected, an additional multiple comparison should be made to rank those images. Because the sizes of the treatment groups are large, the most appropriate method is a multiple range test.

The procedure followed in a multiple range test is:

- 1. Determine $S_{\tilde{x}} = \sqrt{\text{(Mean Square Error/n)}}$ with $DFE = (n-1)k$
- 2. From the table of Significant Studentized Range Points (SSR) obtain Significant Range Distribution

$$
SSR\left[\begin{array}{c}DFE\\ \not p\\ p_s\end{array}\right]
$$

where $DFE = (n-1)k$, \dot{p} (the sizes referring to the number of means involved in a comparison) = 2 to k , ρ_s (level of significance).

3. Compute Least Significant Range (LSR) :

$$
LSR\binom{p}{\rho_s} = SSR\left[\begin{array}{c} DFE\\ p\\ p_s \end{array}\right] \times S\bar{x}.
$$

4. Rank the means $(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_k)$ into ordered means

$$
x(1) \geq x(2) \geq \cdots \geq x(k).
$$

5. Test the differences in the following order: $x(1) - x(k)$

$$
x(1) - x(k)
$$

\n
$$
x(2) - x(k) \cdots x(k-2) - x(k)
$$

\n
$$
x(k-1) - x(k)
$$

\n
$$
x(1) - x(k-1)
$$

\n
$$
x(2) - x(k-1) \cdots x(k-2) - x(k-1)
$$

$$
x(2) - x(k - 1) \cdots x(k - 2) - x(k - 1)
$$

$$
x(1) - x(2)
$$

$$
x(2) - x(3).
$$

If any of these differences exceeds the appropriate

$$
LSR\left(\frac{\dot{p}}{\rho_s}\right),\,
$$

we declare that this difference is significant at an α level of significance.

6. The results of the test can be summarized as several homogeneous subsets. The ranking is then based on those homogeneous subsets. In order to illustrate the above theoretical approach, an example is given in Table **2.**

CONCLUSIONS

The results of the statistical analyses and an examination of the raw data pointed to several conclusions: (1) interpreters are able to nerceive color differences more quickly and accurately than gray level differences; (2) for low resolution imagery, such as orbital photography, discrete color differences are more valuable for interpretation than colors that grade subtly into one another; **(3)** fot high-altitude photography, distinct color variation and **IDECS** quality resolution are competitive with high-resolution systems that exhibit inter- and intra-color gradations; and (4) for low-altitude photography, color differences are more important than gray level differences, and resolution assumes greater importance because interpreters tend to focus their attention on geometry more often than on color or tonal patterns.

The conclusions stated above point to several more specific conclusions than can be made concerning the interpretability of **IDECS** imagery. (1) For high altitude and satellite photography, **IDECS** is highly competitive with other systems tested, and better than many. This high ranking can be attributed to the fact that **IDECS** organizes data into discrete color amalgamations and possesses compatible resolution capabilities for imagery of high altitude and satellite scales. (2) For low altitude photography, the generalizing advantage of **IDECS** is handicapped by the incompatibility of its resolution when

Table 2. Example for Imperial Valley HF (i.e., "High Flight" Photography)-

* In this test, *5%* level of significance is used.

where

 $SSR\left(\begin{array}{c} 16 \\ p \\ .05 \end{array}\right)$

for *p-2* to *8* are obtained from the significant points of the Studentized Range Distributions

$$
LSR\begin{pmatrix}p\\0.5\end{pmatrix} = SSR \begin{pmatrix}16\\p\\0.5\end{pmatrix} \times S_{\bar{x}}.
$$

 (3)

Compare *58* with all the rest:

$$
x(1) - x(8) = 45.333 > 20.767 = LSR \binom{8}{.05} \text{ significant}
$$

$$
x(1) - x(7) = 20.667 > 20.644 = LSR \binom{7}{.05} \text{ significant}
$$

(TABLE *2.-Continued on following page)*

The results of the test can be summarized in one way merely by listing the types of imagery in a dedecreasing order of interpretability, as follows:
58 25A FRSL CIR+1 $CIR+IDECS$ CIR IDECS 89B PF

Any two means not underscored by the same line are significantly different at the *5* percent level. Any two means underscored by the same line are not significantly different from one another. Based on those homogeneous subsets, the overlap part of the two lines can be grouped as one rank. Therefore, the following rankings for percent correct can be made:

compared with the resolution of the input imagery. Although IDECS may be able to detect the important differences that are observable at this scale, it does not have the necessary spatial resolution to display them for interpretation as adequately as the higher resolution input imagery. **(3)** Highly skilled interpreters tend to interpret IDECS imagery faster and more accurately than medium and low-skilled interpreters. (4) This quantitative evaluation experiment indicates that IDECS is competitive with other enhancement and photographic systems as an aid to interpretation of earth features on the basis of time and accuracy. Limitations are evident when IDECS is compared with low altitude, high resolution photography, but this problem can be solved. The preformance of IDECS, in addition, makes a strong case for multiband imagery of the type which will be acquired by the Earth Resources Technology Satellite (ERTS) scheduled to orbit in early 1972, since the ease with which black-andwhite imagery can be exposed and processed

makes repeatable results more obtainable. This repeatability can be expressed in terms of consistency of image tone signature for earth resource phenomena, not only from one photographic mission to the next, but within the imagery from a single mission. This factor is important if we are to automate datahandling techniques with systems such as IDECS.

RECOMMENDATIONS

Based on our research and the conclusions stated in the previous sections, the researchers at the Forestry Remote Sensing Laboratory, University of California, Berkeley, recommended the following as logical extensions of the work performed under this contract. (1) Higher resolution input/output equipment should be acquired in order to upgrade the interpretability of the IDECS output display. (2) An evaluation should be made of the ability to identify features on successive frames of imagery using the same machine settings as proved optimum for identifying

those same features on a *calibration* frame. In this test a training region would be designated on the first frame of the strip similar to the interpreter training regions utilized in the interpreter tests reported on in this project. In this training area the IDECS operator would maximize his machine settings for the identification of individual categories. The IDECS operator would not know the ground truth for the entire area, only the ground truth for that area designated as the training area. The optimum settings for each individual category would then be stored utilizing the disk interface. When all category settings had been made and stored, the total composite image would then be placed on the viewing screen presenting the operator's optimum enhancement of all categories under investigation. The output display would then be photographed for comparison with the ground truth map of the area. Keeping the machine settings for each category constant, the next frame in the flight strip would be placed into IDEcs and the output display recorded and so on with the next frame and the next. When a suitable number of frames had been processed and recorded, an analysis of the results would give an indication of the reliability of automated IDECS interpretation. By comparison of the photographs of the output display with ground truth data, an idea of the fall-off in the accuracy with which *category interpretation* can automatically be made utilizing IDECS could be achieved. Tests could also be performed comparing the ability of image interpreters of varying Ievels of competence to correctly identify the same categories as IDECS was programmed to interpret from several frames of conventional multispectral imagery. The accuracy and time required to preform these interpretations could be compared with the accuracy and time required to program, interpret and process IDECS imagery of the same area. This test would be close to an actual operational test of IDECS ability.

A second type of test which could be incorporated into this procedure or accomplished in conjunction with these tests was suggested by Dr. David Simonett of the Center of Research, University of Kansas. This test would examine the variation in operator machine settings required to obtain the more accurate predictions for the categories utilized in the tests recommended above.

REFERENCES

- Anderson, T. W. 1958. An Introduction to Mul*tivariate Statistical Analysis.* John Wiley & Sons,
- Inc. New York. Cochran, W. G. 1963. *Sampling Techniques.* John Wiley & Sons, Inc. New York.
- Cochran, W. *G.* and G. M. Cox. 1957. *Experimen*tal Designs. John Wiley & Sons, Inc. New York.
- Colwell, R. N. et al. 1966. *A Multispectral Photo*graphic Experiment Based on a Statistical Anal-
ysis of Spectrometric Data. School of Forestry, *ysis of Spectrometric Data.* School of Forestry, University of California, Berkeley, NASA Grant NGR/05-003-080. Final Report. Septem-
ber 30.
- Dalke, G. W. and J. E. Estes. 1968. *Mufti-image Correlation Systems Study for MGI.* Final Re-port. University of Kansas, University of California, Berkeley, U. S. Army Engineering Topo- graphic Laboratories, Contract No. DAAK-02- 67-C-0435. December.
- Dixon, W. J. (ed.) 1968. *BMD, Biomedical Computer Programs.* University of California Press. Los Angeles, California.
- Estes, John E. and Leslie W. Senger, *A Quantita-
tive Evaluation of Electronic Multi-Image Pro*cessors, U.S. Army Engineering Topographic Laboratories, Contract No. DAAK-02-6r-C-0435, Center for Research, Incorp., The Uni-versity of Kansas, Lawrence, Kansas and University of California, Berkeley, California. Final Report, September, 1969. Hogg, R. V. and A. T. Craig, 1965. *Introduction to Mathematical Statistics.* The Macmillan Corn-
-
- pany. New York.
Lent, J. D. 1969. "Additive Color Enhancement-Techniques and Equipment." In *Analysis of Earth Resources on Apollo 9 Photography by R.N.* Ecolwell et al. University of California, Forestry Remote Sensing Laboratory Berkeley. NASA

Contract No. NAS 9-9348. Progress Report, Contract No. NAS 9-9348. Progress Report, July. Ross, D. 1969. "Image Tone Enhancement."
- Paper presented at the 1969 meetings of the American Society of Photogrammetry, Washing-American Society of Photogrammetry, Washington, D.C.

Wouldn't you like to become a member of the American Society of Photogrammetry and help perpetuate its convention programs and this journal? It costs only \$15 per year! Inquire from ASP, 105 N. Virginia Avenue, Falls Church, Virginia 22046.