1970 Arr.
Num. Ana.

MULTI-IMAGE CLUSTERING

Robert M. Harlick
Engineering Topographic Laboratories
Fort Belvoir, Virginia

ABSTRACT. One technique used for image data analysis is clustering. Clustering procedures group the data points or resolution cells into blocks having members which are highly interrelated or similar. The distinct blocks are hypothesized as having been produced by different environmental entities or processes. Hence, examination of the blocks yields interpretation concerning the nature and diversity of the environment from which the data were gathered. Since the blocks reflect the "natural" structure of the environment as seen by the instruments which collected the data, clustering techniques can help the researcher formulate the concepts needed to deal with the instrument-environment system.

A measurement space and spatial clustering procedure is utilized on a multi-image data set obtained from the northern part of Yellowstone Park with the Michigan 12 channel scanner system. The results indicate that about a half-dozen distinct environmental categories can be recognized without any a priori ground truth information.

I. MULTI-IMAGE CLUSTERING. Clustering is a way to automatically extract information from data, in our case multi-image data (1-10). The traditional way to extract information from imagery is to present the imagery to skilled photo-interpreters for analysis. But this is a slow process intended only for limited amounts of imagery. To analyze the large amounts of data which today's remote sensors gather by airplane and satellite, automatic processing techniques must be used.

The automatic processing techniques fall into two classes: supervised and unsupervised. Clustering is an unsupervised technique. With supervised techniques, one gathers a training set of data for which one knows the correct identification of each distinct entity in the data. One then estimates the necessary conditional probability distribution and determines a decision rule from them. The decision rule can then be employed to identify any other data set gathered under similar conditions. With unsupervised techniques there is no training data set or decision rule. One attempts to determine the structures in a data set. Distinct structures are then interpreted as corresponding to distinct objects or environmental processes.

The advantage of the supervised techniques is that the scientist is able to decide what types of environmental categories among which he wishes to distinguish. The decision rule then determines to which such environmental category an arbitrary data entity belongs. The disadvantage of the supervised techniques is that they are sensitive to mis-calibrations. Any slight difference between the sensor calibrations or state of environment for the training data and the new data will cause error. For instance, if one were analyzing multi-images to determine vegetation or crop type on the basis of multi-spectral image grey tone, then one would find that the grey tone associated with a crop when it is sunny is not the same grey tone associated with it when it is cloudy. Perhaps it had rained and the ground was wet or perhaps it was hot and the ground was dry perhaps the irrigation and fertilizer were different; these differences too would make different grey tones. Perhaps one corn field was planted a few weeks before another or the hybrid of corn was different. Perhaps the film emulsions came from different emulsion batches or the developers were not all of equal strength. It is obvious that one could compile a long list of the intervening variables, to measure them (if he could), which affect the environment-image system calibration.

The advantage of the unsupervised techniques is that they are not sensitive to calibration problems. Two small-area patches of corn growing in the same field are going to be detected as being similar because they have similar grey tone associated with them. The disadvantage of the unsupervised techniques is that they are not able to identify what are the distinct environmental structures they determine.

II. POTENTIAL APPLICATION OF THE MULTI-IMAGE CLUSTERING TECHNIQUE. The use of the potential information contained in images is well documented in the remote sensing literature, a selected sample of which is given in reference 10-24. One can potentially determine vegetation species distribution, location of mineral deposits, worldwide distribution of geomorphic features, sea surface temperatures, location of fish schools and icebergs, spread of pollutants, cloud cover distribution and movement, developing storm systems, water movement, soil and vegetation moisture content, transportation networks, and extent of urbanization to name a few.

What is not clear is what sensor or set of sensors is best for what jobs. By employing multi-image clustering to imagery obtained from a set of sensors, one can determine the distinct environmental structures as seen through the sensor's eyes. By then comparing these structures to what is actually in the environment, as determined from field work, one can tell what categories among which the set of sensors will distinguish best. This procedure is more efficient than one based on supervised techniques since it only has to be performed once while the supervised technique will have to be evaluated for each a priori set of categories chosen.

Once the multi-image clustering technique is developed so that it can be cheaply implemented in real time, another use for it will become feasible: it can be used to pre-process the data performing a feature extraction function. The features obtained will not be sensitive to the calibration problem. Identification decisions are then made as usual on the basis of the extracted features. The result will be a multi-image processing system which can work well despite the calibration difficulties.

III. DATA ANALYSIS. Twelve images taken by the Michigan scanner system were the multi-image data set. These images were taken of a 2 mile by 6 mile area in the northern part of Yellowstone Park at approximate coordinates $100^{\circ}30'$ by $44^{\circ}57'$ on September 19, 1967. Figure 1 shows an old panchromatic photograph of the area taken in 1954. Each image of the multi-image set was, in effect, a picture taken with a different narrow-band filter where the filters passed light in narrow bandwidths from the near infrared band part of the spectrum through the ultraviolet portion of the spectrum. Table 1 tabulates these bands. The images were digitized to 256 levels on a grid of 220 x 1260 for a total of about 270,000 resolution cells for each image. Each resolution cell contains the returns from 12 spectral bands coming from a 20 ft. x 20 ft. small-area ground patch. Successive resolution cells contain returns from small-area ground patches separated by a gap of 20 ft.

In order to reduce computer time, the original twelve images were processed to yield four smaller images, but with most of the statistical and spatial structure preserved. The first part of the pre-processing consisted of a principal components analysis. A principal components analysis may be considered in the following fashion. In any image, some grey tones occur more frequently than others. We may consider the relative frequence of the grey tones on the image as defining a one-dimensional probability distribution. This probability distribution has a mean and variance. Similarly in the twelve image multi-image set each resolution cell has a 12-tuple of grey tones. Some of these 12-tuples occur more frequently than others, and we may consider the relative frequency of the 12-tuple grey tones as defining a twelve-dimensional probability distribution. This probability distribution too has a mean and variance. The variance is called the total variance. A principal components

Figure 1. 1954 Photographic Image Taken of Area

Image	Wavelength band
1	800-1000 milli-microns
2	720-800
3	660-720
4	620-660
5	580-620
6	550-580
7	520-550
8	500-520
9	480-500
10	460-480
11	440-460
12	400-440

Table 1 Tabulates the wavelength bands for each of the images

analysis, as might be expected, determines principal components. The first principal component image is obtained by taking that linear combination or weighted average of the original twelve images such that the variance of the probability distribution of grey tones on the first principal component image is higher than it could be for any other linear combination. The second principal component is obtained by taking that linear combination, orthogonal to the first, such that the variance of the probability distribution of grey tones on the second principal component image is higher than it could be for any other linear combination orthogonal to the first. In general, the kth principal component image is obtained by taking that linear combination, orthogonal to the earlier 1st, 2nd, ...,(k-1)th linear combinations of the original twelve images such that the variance of the probability distribution of grey tones on the kth principal component image is higher than it could be for any other linear combination orthogonal to the earlier 1st, 2nd,...,(k-1)th ones. Because the sum of the variances of all the principal components equals the total variance of the original twelve-dimensional probability distribution, the ratio of the variance of the kth principal component to the total variance is called the variance accounted for by the kth principal component. The variance accounted for by the kth principal component is an indicator of how much statistical structure from the original twelve images is preserved by the kth principal component image.

The principal components provided the following results.It was determined that the first component accounted for 97.4% of the variance, the second component 1.6% of the variance, and the third component .9% of the variance. The respective weights used in the linear combination are listed in Table 2.

Table 2 has an interesting interpretation. The first linear combination has weights which are all positive and which are about the same magnitude. The first principal component image is then very close to what a panchromatic image of the area would be. This should not be surprising since most photo interpreters will prefer a panchromatic image over any narrow-band image because they see more structure in it. The second linear combination weights the infra-red part of the spectrum negatively, the middle of the spectrum hardly at all and the ultra-violet part of the spectrum positively. This weighting trend from the infra-red to the ultra-violet is almost a linear one. It is indicative of the fact that the spectral

Wavelength Band	1st	2nd	3rd
	Principal Component		
	% Variance Accounted for		
	97.4%	1.6%	.9%
800-1000	.15702	-.33153	.27651
720-800	.22758	-.33529	.26796
660-720	.28342	-.31332	.13188
620-660	.23184	-.23019	.11530
580-620	.20199	-.16154	.010509
550-580	.17486	-.094258	.025696
520-550	.32559	-.070442	.25499
500-520	.24756	.031677	.046381
480-500	.42727	.09246	-.12730
460-480	.52815	.22196	-.60118
440-460	.25916	.30966	.043766
400-440	.14878	.65714	.61121

Table 2 Table of weights used to obtain the linear combinations for the 1st, 2nd, and 3rd principal components.

while that for ryolite (a volcanic rock known to be prevalent in the area photographed) is almost flat in that very same region. Hence, weighting the 460-480 milli-micron part of the spectrum negatively and the 400-440 milli-micron end of the spectrum positively will enhance the difference between vegetation and ryolite.

Since the first three principal components accounted for about 99% of the variance, there is no need to cluster twelve images. Almost all the information is contained in 3 linear combinations (principal components) of the original twelve. However, because the clustering procedure treats each image equally, the second and third principal component images, which together only account for a few percent of the total variance, would be unduly emphasized. Therefore, a new set of images, called the principal components dispersed images, was prepared by taking linear combinations of the first three principal components. The linear combinations are mutually orthogonal and result in distributing the variance almost equally among the dispersed images. Table 3 shows the weights used in these linear combinations.

The next step in the preprocessing consisted of reducing the size of the three dispersed principal component images. Each was 220 resolution cells horizontally by 1220 resolution cells vertically. They were reduced in size to 73 resolution cells horizontally by 406 resolution cells vertically by taking every third row of resolution cells and every third resolution cell on each such row taken. To compensate for the roughening effect by such a reduction, a fourth image was prepared by reducing the size of the first principal component image by averaging the grey tones of nonoverlapping 3 x 3 blocks of resolution cells.

The final step in the preprocessing consisted of quantizing the grey tones of the four images to 13 quantized grey tone classes. This was done in two parts: the images were first quantized to 64 grey tone classes by a folded-tail linear quantizing procedure (see Figure 2). The folded tails quantizing is essentially a linear quantizing procedure modified to ignore extreme wild points on the tails of the distribution. In other words, instead of determining the highest grey tone and the smallest grey tone and then equally divide the resulting interval up into 64 pieces as the linear quantizing does, the folded tails linear quantizing determines a "high" grey tone less than the highest and a "small" grey tone greater than the smallest and equally divides that resulting interval up into 64 quantized classes. Of course, grey tones higher than the determined "high" and smaller than the determined "small" get put in the highest and smallest quantized class respectively.

	1st	2nd	3rd
	Dispersed Components		
1st Principal Component	$1/\sqrt{3}$	$1/\sqrt{3}$	$1/\sqrt{3}$
2nd Principal Component	$1/\sqrt{6}$	$-2/\sqrt{6}$	$1/\sqrt{6}$
3rd Principal Component	$1/\sqrt{2}$	0	$-1/\sqrt{2}$

Table 3 Table of weights for the mutually orthogonal linear com combinations used to disperse the variance of the first three principal components.

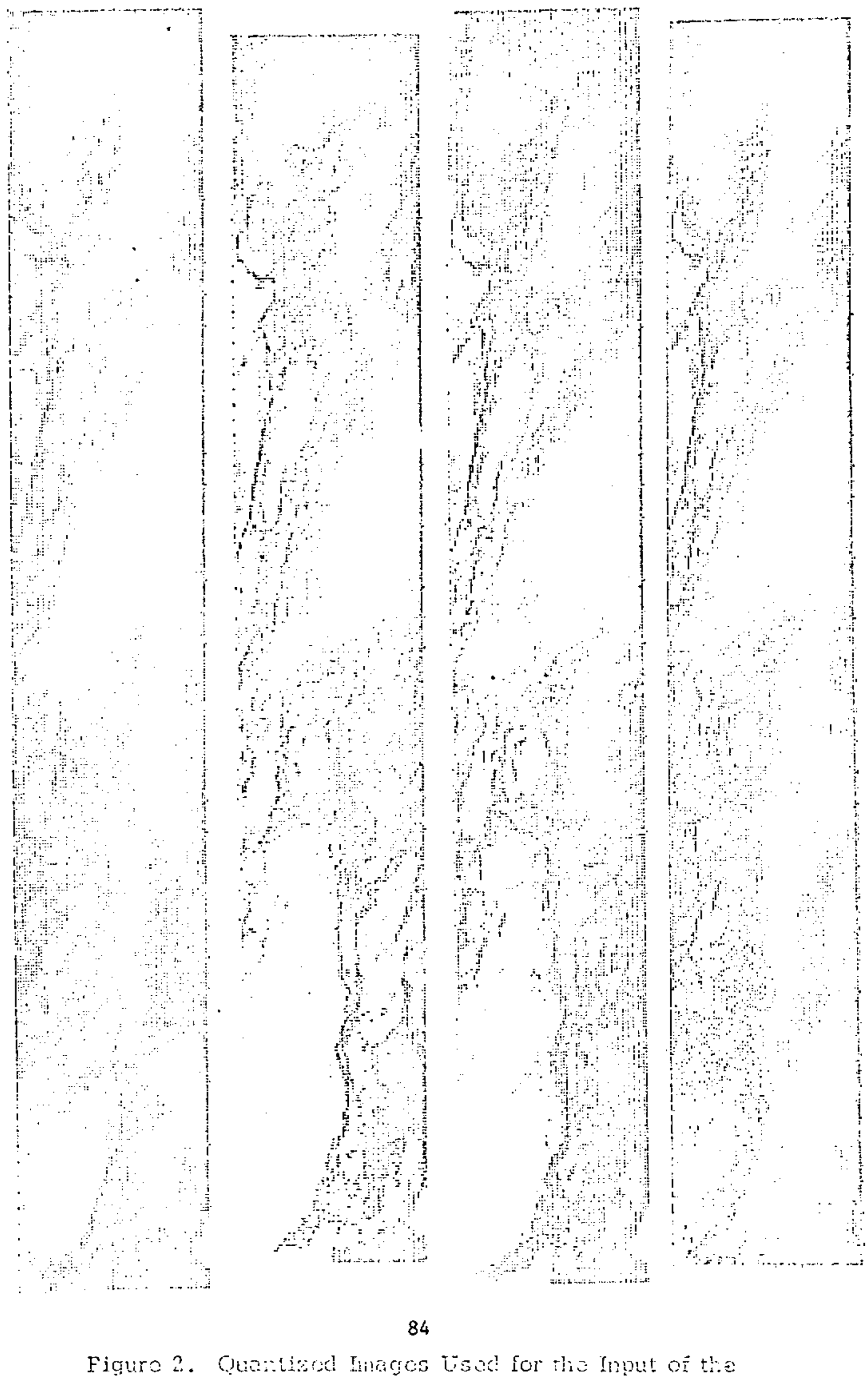

Figure 2. Quantized Images Used for the Input of the Spatial Clustering Procedure

The spatial quantizing procedure further reduces the number of quantized classes from 64 to 13 in a way which capitalizes on the spatial dependence of the 64 quantized grey tone classes. A 64 x 64 Markov transition probability matrix is set up where the (i,j)th element is the probability that a resolution cell having a grey tone in the ith grey tone class will be next to a resolution cell having a grey tone in the jth grey tone class. The spatial quantizing uses the information in this matrix to form quantized classes whose grey tones have a high probability of occurring next to each other. Hence, spatial continuity of grey tones tends to be preserved. These quantizing procedures are fully described in reference 18.

After quantizing the four images to 13 grey tone classes, the images were clustered.

IV. DESCRIPTIVE OUTLINE OF SPATIAL CLUSTERING. The clustering procedure starts by finding sets of resolution cells called center sets. Center sets are good places to build a cluster around. Resolution cells whose grey tone N-tuples are similar enough are sequentially added to the existing center set. When there are no more resolution cells similar enough, the existing cluster is complete, and a new cluster is begun from a new center set.

We assume that each object on the ground produces grey tones which are similar and homogeneous. There can be any spatial distribution of objects; one object may only occur once and another hundreds of times or each object may occur approximately the same number of times. It would be intuitively reasonable to form center sets from those spatial locations which have fairly homogeneous measurement space coordinates and which are representative measurements of a class of objects. However, since the location and extent of objects are unknown to the clustering procedure, it must try to induce this information from the data structure. Since we assume that the set of measurements recorded from any object form a homogeneous set of highly similar grey tones, and the location of these measurements in the image sequence is a small, more or less, spatially connected region, perhaps by breaking up the image sequence into a set of spatially connected subsequences and examining the measurements in each subsequence we can obtain the necessary information. Thus we make each spatially connected subsequence: (1) large enough to include within it a substantial proportion of the measurements recorded from at least one object; and (2) small enough so that the substantial proportion of the measurements recorded from the object make up a large proportion of the measurements in the subsequence. If we can form subsequences in this way, then the empirically observed probability distribution of the grey tone N-tuples in the sequence will be dominated by the substantial proportion of grey tone N-tuples in the sequence recorded from some particular type of object. Thus, if a particular object occurs only once then there will be one subsequence dominated by it. By picking out the kind of measurements which typify that subsequence (i.e. those which have high probability or high self-association in the subsequence), then the set of all the spatial locations containing these measurements is a good center set.

A grey tone N-tuple has high self-association if there is high probability that a resolution cell having that grey tone or a grey tone similar to it will be contiguous to a resolution with the given grey tone N-tuple. Two grey tone N-tuples are said to have high cross-association if there is high probability that a resolution cell having one of the grey tone N-tuples, or a grey tone similar to one of them, will be contiguous to a resolution cell having the other grey tone N-tuple or a grey to similar to the other.

Since the clustering procedure we have proposed starts with center set one, build on it until no more similar measurements can be found, and then starts building on center set two, etc., we must specify how the order is determined for center sets. We should naturally start with the most important center set and here importance can be correlated with self-association. That center set is most important which has the highest self-association of all center sets in the subsequence form which it originates.

Next we must consider exactly how each center set grows. If there exists a resolution cell outside but next to some resolution cell in the center set and if the cross-association of the grey tone N-tuples of these resolution cells are sufficiently high and the self-association of the grey tone in the original center set is not too different from the self-association of the grey tone in the outside resolution cell, then the outside resolution cell is added to the center set. When no such outside resolution cell exists, the growing is terminated and a new center set is defined. A precise description of the clustering procedure can be found in reference (8).

V. RESULTS. A map of the distinct environmental objects as determined from the clustering procedure is displayed in Figure 3. The computer took .3 hour processor time and 1.8 hour wall time to produce these results. Some of the regions on the map can immediately be seen as representing some of the dominant types of categories in the area. A detailed comparison of the map with the ground truth for the region is required to make a good evaluation of these results.

HOMOGENEOUS REGIONS PRODUCED BY THE SPATIAL CLUSTERING METHOD

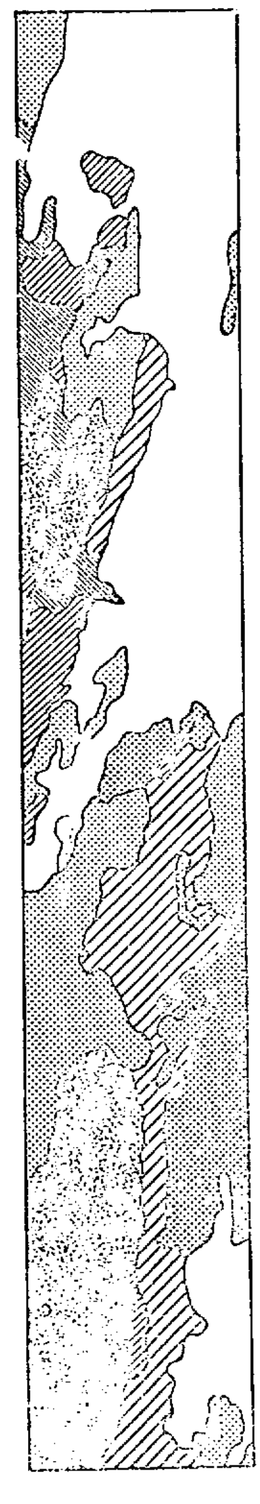

Figure 3

This is the first time that such a clustering algorithm has been tested on a multi-image and there are some problems with the clustering procedure itself which have to be solved. One problem concerns the specification of the clustering parameters. This first time they had been specified on a trial and error basis and it resulted in much wasted computer time. It should be possible to determine them from appropriate probability distributions which the multi-image generates. A second problem concerns boundary delineation. It appears that some boundaries extend too much while others not enough. The problem may be related to either optimal determination of clustering parameters or to the way the clusters themselves grow. A third problem is concerned with the large amount of computer time needed to implement the algorithm. A careful study of the growth of the clusters can probably lead to a new algorithm for which the computer time is cut by an order of magnitude.

REFERENCES

1. Bonner, R. E., "On Some Clustering Techniques," IBM Journal of Research and Development, Jan. 1964.

2. Mattson, R. L., and J. E. Dammann, "A Technique for Determining and Coding Subclasses in Pattern Recognition Problems," IBM Journal of Research and Development, July 1965.

3. Patrick, E. A., and J. C. Hancock, "Nonsupervised Sequential Classification and Recognition of Patterns," IEEE Transaction on Information Theory, Vol, IT-12, No. 3, 1966.

4. Cooper, D. B., and P. W. Cooper, "Nonsupervised Adaptive Signal Detection and Pattern Recognition," Information and Control, Vol. 7, 1964.

5. Ward, J. H., "Hierarchial Grouping to Optimize an Objective Function," American Statistical Association Journal, March 1963.

6. Lance, G. N., and W. T. Williams, "A General Theory of Classificatory Sorting Strategies II. Clustering Systems," Computer Journal, Vol. 10, No. 3, Nov. 1967.

7. Haralick, R. M. and G. L. Kelly, "Pattern Recognition with Measurement Space and Spatial Clustering for Multiple Images," Proceedings of the IEEE, Vol. 57, No. 4, April 1967.

8. Haralick, R. M. Multi-Image Pattern Recognition: Ideas and Results, Ph.D. Dissertation, University of Kansas, 1969.

9. Fralick, S. C., "Learning to Recognize Patterns without a Teacher," IEEE Transactions on Information Theory, Vol. IT-13, No. 1. pp. 57-64, January 1967; Stanford Technical Report 6103-10, SEL-65-011, March 1965.

10. Spragins, J. D., "Learning without a Teacher," IEEE Transactions on Information Theory, IT-12, No. 2, April 1966.

11. Colwell, R. N., et al,. The Usefulness of Thermal Infrared and Related Imagery in the Evaluation of Agricultural Resources, Vol. 1, "Cultivated Crops," NASA Contract No. 12-12-100-8316(20), Berkeley, University of California, 1966, 137pp.

12. Gerlach, A. C., "Geographic Applications of Remote Sensor Data from Aircraft and Spacecraft," Proceedings of the International Geographical Union Regional Conference, Mexico City, Mexico, 1966.

13. LARS, Remote Multispectral Sensing in Agriculture, Annual Report Vol. 3, LARS Publication No. 844, Lafayette, Purdue University Agricultural Experiment Station, September 1968, 175 pp.

14. Menzer, F. E., and C. R. Cooper, "Aerial Photographic Methods of Potato Disease Detection," University of Maine, Agricultural Experiment Station Bulletin 646, 1967, 14pp.

15. Meyer, M. P., and L. Calpouzos, "Detection of Crop Diseases," Photogrammetric Engineering, XXXIV, 6, 1968, 554-557.

16. Simonett, D. S., "Land Evaluation with Remote Sensors in the Infrared and Radar Regions," in Land Evaluation,edited by G. A. Stewart, Canberra, Australia: CSIRO and MacMillan, 1968, 349-366.

17. Thorley, G. A., "Some Uses of Color Aerial Photography in Forestry," in ASP, Manual of Color Aerial Photography, 1968, 392-393.

18. Weiss, M., et al., "Evaluation of an Infrared Sensor as a Clear Air Turbulence Detector," Sixth Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, October 1969.

19. Carneggie, D. M., "Analysis of Remote Sensing Data for Range Resource Management," Forestry Remote Sensing Laboratory, Berkeley, University of California, 1968.

20. Garrison, W. L., et al., "Five Papers on Remote Sensing and Urban Information Systems," Technical Report No. 1, Urban and Transportation Information Systems, Department of Geography, Northwestern University, Evanston, Illinois, 1966.

21. Lewis, T., "Evaluation of Multiple Polarized Radar Imagery for the Detection of Selected Cultural Features," U.S. Department of Interior, Geological Survey, Technical Letter NASA-130,NASA Contract No. W-12589, 1968.

22. Marble, D. F. and F. E. Hortin, "Remote Sensing and the Study and Planning of Urban Areas, NASA, Earth Resources Aircraft-Program Status Review Vol. I, Manned Spacecraft Center, Houston, Texas, 1968.

23. Morain, S. A., and D. S. Simonett, "K-Band Radar in Vegetation Mapping," Photogrammetric Engineering, July 1967, 730-740.

24. Schwarz, D. E., D. S. Simonett, G. F. Jenks, and J. R. Ratzlaff, "The Construction of Thematic Land Use Maps with Spacecraft Photography," Department of Geography, Center for Research in Engineering Science, University of Kansas, U.S.G.S. Contract USGS 14-08-0001-10848, 1968.