Lockheed Electronics Company, Inc.
A SUBSIDIARY OF LOCKHEED CORPORATION
1830 NASA Road 1, Houston, Texas 77058
Tel. 713-333-5411

JSC-15042

80-10012
NASA CR-
160398

Ref: 642-7558
Job Order 73-705-06
Contract NAS 9-15800

TECHNICAL MEMORANDUM

LIST SPECTRAL KEYS STUDY

By

T. B. Dennis

and

M. D. Pore

(E80-10012) LARGE AREA CROP INVENTORY EXPERIMENT (LACIE). LIST SPECTRAL KEYS STUDY (Lockheed Electronics Co.) 19 p HC A02/MF A01 CSCL 08F

N80-13591

Unclas
G3/43 00012

Approved By: T. C. Minter

T. C. Minter, Supervisor
Techniques Development Section

August 1979

LEC-13456

CONTENTS

PRECE PAGE BLANK NOT FILMED

TABLES

ACRONYMS

AI	analyst-interpreter
DSG	discriminant-labeled percentage of small grains
GSG	ground-truth percentage of small grains
LACIE	Large Area Crop Inventory Experiment
Landsat	Land satellite
LIST	Label Identification from Statistical Tabulation
PCL	probability of correct labeling
Pixel	picture element
SPSS	Statistical Package for Social Sciences

1. INTRODUCTION

Label Identification from Statistical Tabulation (LIST) is an analyst's picture element (pixel) labeling procedure for making at-harvest small-grain proportion estimates in the Large Area Crop Inventory Experiment (LACIE). In this labeling procedure, the analyst-interpreter (AI) is required to answer questions about the segment and pixels which relate to simple properties that discriminate small grains from nonsmall grains. The responses, along with pertinent agricultural and meteorological variables, are statistically weighted to develop a discriminant function which is trained on blind site ground-truth labels.

The preliminary development of LIST was analyzed and reported by Pore (ref. 1); from this, a semiautomated operational LIST was developed by Abotteen and Pore (ref. 2) and tested on blind sites in both Kansas and North Dakota (ref. 3). Improvements in the green number and brightness keys used in LIST were made by Dennis and Pore (ref. 4), and the resulting discriminant was tested on LACIE Phase III Kansas data. While testing this discriminant on LACIE Phase III data, an effort was made to use this data to produce alternate keys which would outperform greenness and brightness. Keys were developed for channel ratios and normalized channel rankings which are now in study for use in spectral yield modeling. These new spectral keys are described in section 2. In section 3, the results obtained using these keys alone and using them in conjunction with the greenness and brightness keys are presented, and the interpretations of the results are discussed. Conclusions and recommendations are summarized in section 4.

The earlier LIST studies (ref. 1 through ref. 4) supported development of a semi-automated labeling technique that is now being tested operationally. This study, however, is a research and development study into the possible benefit of implementing an alternate set of spectral keys in the LIST discrimination process.

2. DEVELOPMENT

Cate (ref. 5) and Hay (ref. 6) proposed that the ratio of the land satellite (Landsat) Channel 4 to Channel 2 be used as a possible alternative to greenness as a spectral aid. This ratio and other combinations of spectral data were considered as possible replacements or improvements to the greenness and brightness keys used in LIST. The specific ratios considered were:

a. Channel 4 to Channel 2

b. Channel 3 plus Channel 4 to Channel 1 plus Channel 2

c. One-half of the sum of Channel 1 and Channel 2

d. Channel 1 to Channel 2

e. Channel 4 to Channel 3

In addition to these ratios, another proposal by Cate (ref. 5) was considered. This proposal is a key based on the relative ranking of the four Landsat Channels. The new keys were developed from the LACIE Phase III Kansas blind site data, and all of the blind sites which had four usable acquisitions were used. Listed in table 2-1 are the segments and acquisitions used. The individual channels were first normalized using the method outlined by Cate (ref. 5), and then ratios were obtained. These normalized ratios were standardized as a function of the julian date and Robertson biostage. This standardization, not to be confused with the late normalization procedure above, is the same procedure that was applied to the green number and brightness transformations described in ref. 4; however, in this case, the LACIE Phase III data were used to produce means and standard deviations of the ratios for each satellite sweep.

In table 2-2, the means and standard deviations for both small grains and nonsmall grains obtained for each ratio are listed. Several of these ratios were immediately abandoned for two reasons. First, for each variable, LIST uses four standardized distances, their absolute values, and two trajectory comparisons which give a total of ten new diminsions; for five ratios, the number of dimensions would be too large. Second, several of the ratios showed little or no separation between the means and standard deviations of grains

TABLE 2-1.– DATA USED IN THE DEVELOPMENT

Segment	Acquisitions
1851	6289, 6361, 7067, 7193
1340	6287, 6305, 7101, 7191
1175	6287, 6305, 7101, 7191
1881	6288, 6324, 7066, 7156
1885	6287, 7101, 7155, 7191
1279	6290, 7050, 7158, 7194
1285	6272, 7086, 7158, 7194
1033	6288, 6325, 7084, 7156
1890	6288, 6360, 7066, 7156
1183	7063, 7082, 7099, 7154
1021	6255, 6363, 7159, 7194
1295	6288, 6306, 7066, 7156
1864	6326, 7050, 7122, 7194
1170	6287, 6305, 7101, 7191
1346	6287, 6305, 6323, 7101
1032	7068, 7086, 7158, 7194
1853	6253, 6361, 7067, 7193
1861	6326, 7104, 7158, 7194
1180	6285, 7081, 7099, 7153
1158	6287, 6323, 7101, 7155
1166	7046, 7082, 7154, 7190

TABLE 2-2.— MEANS AND STANDARD DEVIATIONS OF NORMALIZED RATIOS

Acquisition range	Number of samples	Channel 4 to Channel 2				Channel 3 + Channel 4 to Channel 1 + Channel 2				One-half of the sum of Channel 1 and Channel 2				Channel 1 to Channel 2				Channel 4 to Channel 3			
(Julian date)		Grains		Other		Grains		Other		Grains		Other		Grains		Other		Grains		Other	
		μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
6242 - 6260	95	0.79	0.11	1.25	0.60	0.82	0.09	1.17	0.42	20.5	3.8	19.3	3.8	0.98	0.08	1.05	0.13	0.94	0.07	1.03	1.10
6261 - 6278	18	.84	.06	1.22	.85	.85	.06	1.12	.52	18.7	2.1	19.0	3.8	.95	.06	1.04	.16	1.03	.07	1.04	.13
6279 - 6296	556	.87	.16	1.12	.40	.88	.13	1.08	.27	20.8	3.5	19.7	3.8	.99	.10	1.02	.13	.98	.08	1.01	.08
6297 - 6314	271	.93	.19	1.03	.23	.88	.16	1.03	.19	18.0	3.6	20.8	4.2	1.08	.14	1.00	.12	1.04	.15	1.00	.10
6315 - 6332	199	1.09	.21	.97	.14	1.05	.17	.97	.12	19.7	2.8	20.2	3.0	1.05	.11	.99	.10	1.02	.09	1.01	.10
6351 - 7002	177	1.05	.18	.99	.14	1.01	.14	.99	.13	18.9	2.5	20.5	4.1	1.05	.11	1.01	.12	1.03	.10	1.00	.12
7039 - 7056	96	1.06	.10	.99	.09	1.04	.09	.98	.08	19.5	2.2	19.7	3.4	1.02	.08	1.01	.10	1.02	.07	1.01	.09
7057 - 7074	289	1.06	.12	.98	.09	1.05	.10	.98	.08	18.9	2.2	20.3	2.7	1.02	.08	1.00	.07	1.01	.07	1.00	.06
7075 - 7092	143	1.16	.31	.99	.24	1.11	.22	.98	.19	18.4	2.5	19.9	3.4	1.04	.09	1.00	.08	1.04	.07	1.00	.07
7093 - 7110	399	1.40	.54	.92	.37	1.28	.37	.93	.27	18.1	3.5	21.2	3.9	1.11	.16	.98	.12	1.03	.08	.98	.08
7111 - 7128	55	1.43	.42	.89	.41	1.31	.31	.90	.27	17.1	2.5	21.4	2.8	1.11	.10	.98	.10	1.06	.08	.96	.07
7147 - 7164	530	1.07	.35	1.07	.51	1.04	.25	1.03	.36	19.0	3.5	21.0	5.2	1.01	.12	1.03	.17	1.01	.08	.99	.07
7183 - 7200	652	.91	.20	1.23	.77	.94	.16	1.12	.48	22.2	3.1	18.7	4.1	.95	.09	1.09	.18	.98	.05	1.01	.10

and nongrains. Therefore, it was decided to limit the ratios used in the testing to the following:

a. Channel 4 to Channel 2

b. Channel 3 plus Channel 4 to Channel 1 plus Channel 2

In addition to testing the two ratios, the ranking of the four normalized channels was studied. For example, if the reflectance in Channel 4 was greater than the reflectance in Channel 3, which was greater than that in Channel 1, which was greater than that in Channel 2, then there was a better chance that the sample represented healthy vegetation than if some other ranking of channels was observed. In order to develop the channel ratios into a quantitative measure, a rather complicated transformation of the rankings was generated using empirical data to indicate the likelihood that a pixel with a given ranking is a small-grains pixel on a particular pass. A description of this transformation of the rankings follows.

Frequency tables were constructed to show which rankings were indicative of grains at various times of the growing season and which rankings were more indicative of nongrains. The rankings that were observed to favor grains and nongrains at various times and the percentages of grains and nongrains observed in each class are listed in table 2-3. For each acquisition, a pixel was assigned a value of 1, -1, or 0, depending on whether the ranking was favorable for grains, favorable for nongrains, or neutral. These four values obtained from table 2-3 and their sum together with the variables obtained from the above ratios were studied as potential discriminant variables.

The discriminant variables were then analyzed using a stepwise discrimination process to evaluate the discriminability of the ensemble of variables. The objective was to derive the subset of variables that yielded optimum discriminability.

TABLE 2-3.– OBSERVED CHANNEL RANKINGS

Acquisition range (Julian date)	Rankings favoring grains	Percentage of grain and other samples in favorable category		Rankings not favoring grains	Percentage of grain and other samples in unfavorable category	
		Grains	Other		Grains	Other
6242 - 6260	G, H	64.2	23.2	Q, R, W, X	5.3	46.3
6261 - 6278	G, H	66.7	26.4	Q, R, W, X	0	44.4
6279 - 6296	A, B, G, H	75.8	26.9	D, K, L, P, Q, R, T, U, V, W, X	10.9	58.1
6297 - 6314	A, B, C, E, H	59.3	23.5	P	1.1	5.7
6315 - 6332	W, Q	28.1	4.7	A, G, H, J, K, L, P	18.2	47.4
6351 - 7002	W, Q	19.8	8.1	A, G, H, I, J, L, O, P, R, U, X	32.5	55.6
7039 - 7056	W, Q, X	33.3	7.0	A, B, C, G, H, I, J, P	18.7	46.2
7057 - 7074	W, Q, X	31.8	11.0	A, B, G, H, I, J, U	18.7	51.0
7075 - 7092	W, Q, X	44.8	16.0	A, G, H, I	15.4	44.0
7093 - 7110	W, Q, X	66.4	15.9	A, G, H, I	16.9	64.9
7111 - 7128	W, Q, X	81.8	7.6	A, B, G	3.6	80.5
7147 - 7164	J, L, W, Q	38.9	34.4	A, B	9.1	7.3
71[illegible] - 7200	J, K, L, P, R, V	24.3	4.3	A, B, C, F, M, W	17.8	54.3

Symbol	Ranking of channels
A	$1 \geq 2 \geq 3 \geq 4$
B	$1 \geq 2 \geq 4 > 3$
C	$1 \geq 3 > 2 \geq 4$
D	$1 \geq 3 \geq 4 > 2$
E	$1 \geq 4 > 2 \geq 3$
F	$1 \geq 4 > 3 > 2$
G	$2 > 1 \geq 3 \geq 4$
H	$2 > 1 \geq 4 > 3$
I	$2 \geq 3 > 1 \geq 4$
J	$2 \geq 3 \geq 4 > 1$
K	$2 \geq 4 > 1 \geq 3$
L	$2 \geq 4 > 3 > 1$

Symbol	Ranking of channels
M	$3 > 1 \geq 2 \geq 4$
N	$3 > 1 \geq 4 > 2$
O	$3 > 2 > 1 \geq 4$
P	$3 > 2 \geq 4 > 1$
Q	$3 \geq 4 > 1 \geq 2$
R	$3 \geq 4 > 2 > 1$
S	$4 > 1 \geq 2 \geq 3$
T	$4 > 1 \geq 3 > 2$
U	$4 > 2 > 1 \geq 3$
V	$4 > 2 \geq 3 > 1$
W	$4 > 3 > 1 \geq 2$
X	$4 > 3 > 2 > 1$

3. RESULTS AND INTERPRETATIONS

3.1 RESULTS

Stepwise and direct linear discriminant analyses were performed on the LACIE Phase III Kansas blind site data. Both analyses were performed using the Statistical Package for Social Sciences (SPSS) described in reference 7. The stepwise analyses were used primarily to indicate which variables were more useful in the discrimination process and which variables could possibly be removed for a later analysis. The direct analyses were used to compare training accuracies for various sets of discriminant variables. Table 3-1 illustrates the format used for the tables 3-2 through 3-6, which show the results of these analyses. This format is the same as that used for tables in references 3 and 4.

The best accuracy obtained, using only subsets of the new keys and the five AI canopy keys, is presented in table 3-2. The overall accuracy here was significantly lower than accuracies observed in previous studies (ref. 4), and attempts to improve accuracy were made by including the greenness and brightness keys. The accuracy obtained using ratio variables, channel rankings, AI canopy keys, and greenness and brightness keys is shown in table 3-3. The total number of variables involved, however, makes the use of all four sets of keys impractical.

The stepwise portion of this analysis indicated that using both ratios was providing redundant information and that it might be possible to drop the 4 to 2 ratio. Table 3-4 illustrates that identical results were obtained when this ratio was dropped from the analysis. However, the total number of variables was still too great for practical use.

In table 3-5, the accuracy obtained when the (3 + 4)/(1 + 2) ratio was dropped is shown. Only the channel rankings, AI canopy keys, and greenness and brightness keys remained in the analysis. In table 3-6, the results obtained using only greenness and brightness and AI canopy keys are given. These are the variables that were tested on the LACIE Phase II Kansas data and are

TABLE 3-1.— CONTINGENCY TABLE KEY

Ground truth	Direct discriminant		Conditional percent
	Small grains	Other	
Small grains	a	b + e	g
Other	c	d + f	h
Conditional percent	i	j	

PCL =
GSG =
DSG =

Total =

Symbol definitions:

a, b, c, and d = raw pixel counts for the four test segments.

e and f = raw pixel counts for the designated "other pixels."

Total = k = a + b + c + d + e + f.

g, h, i, and j = marginal probabilities (P_r), expressed as percentages, of correct labeling:

$$g = \frac{a}{a + b + e} \times 100 = (1 - P_r \text{ (omission)}) \times 100$$

$$h = \frac{d + f}{c + d + f} \times 100 = (1 - P_r \text{ (commission)}) \times 100$$

$$i = \frac{a}{a + c} \times 100$$

$$j = \frac{d + f}{b + d + e + f} \times 100$$

$$GSG = \frac{a + b + e}{k} \times 100$$ = the ground-truth percentage of small grains

$$DSG = \frac{a + c}{k} \times 100$$ = the discriminant-labeled percentage of small grains

$$PCL = \frac{a + d + f}{k} \times 100$$ = the probability (expressed as a percentage) of correct labeling

TABLE 3-2.— RESULTS OBTAINED USING AI CANOPY KEYS TOGETHER WITH KEYS DEVELOPED FROM NEW RATIOS AND CHANNEL RANKINGS

Ground truth	Direct discriminant		Conditional percent
	Small grains	Other	
Small grains	714	150 + 80	75.2
Other	169	1054 + 1232	93.1
Conditional percent	80.9	90.6	

PCL = 88.1 percent   Total = 3405
GSG = 27.9 percent
DSG = 25.9 percent

TABLE 3-3.— RESULTS OBTAINED USING AI CANOPY KEYS, GREENNESS AND BRIGHTNESS KEYS, AND NEW RATIO AND CHANNEL RANKING KEYS

Ground truth	Direct discriminant		Conditional percent
	Small grains	Other	
Small grains	737	133 + 80	77.6
Other	155	1068 + 1232	93.1
Conditional percent	82.6	91.5	

PCL = 89.2 percent   Total = 3405
GSG = 27.9 percent
DSG = 26.2 percent

TABLE 3-4.— RESULTS OBTAINED USING AI CANOPY KEYS, GREENNESS AND BRIGHTNESS KEYS, AND NEW RATIO AND CHANNEL RANKING KEYS WITH THE 4:2 RATIO REMOVED

Ground truth	Direct discriminant		Conditional percent
	Small grains	Other	
Small grains	732	138 + 80	77.1
Other	150	1073 + 1232	93.9
Conditional percent	83.0	91.4	

PCL = 89.2 percent
GSG = 27.9 percent
DSG = 25.9 percent

Total = 3405

TABLE 3-5.— RESULTS OBTAINED USING AI CANOPY KEYS, GREENNESS AND BRIGHTNESS KEYS, AND CHANNEL RANKING KEYS

Ground truth	Direct discriminant		Conditional percent
	Small grains	Other	
Small grains	730	140 + 80	76.8
Other	155	1068 + 1232	93.7
Conditional percent	82.5	90.6	

PCL = 89.0 percent
GSG = 27.9 percent
DSG = 26.0 percent

Total = 3405

TABLE 3-6.– RESULTS OBTAINED USING ONLY AI CANOPY KEYS AND GREENNESS AND BRIGHTNESS KEYS

Ground truth	Direct discriminant		Conditional percent
	Small grains	Other	
Small grains	680	190 + 80	71.6
Other	186	1037 + 1232	92.4
Conditional percent	78.5	90.6	

PCL = 86.6 percent
GSG = 27.9 percent
DSG = 25.4 percent

Total = 3405

currently in use in LIST. It is interesting to note the difference in accuracies on LACIE Phase II data (ref. 4) and LACIE Phase III data. In fact all accuracies for the LACIE Phase III data are significantly lower than those observed for LACIE Phase II data. An explanation is given in section 3.2.

3.2 INTERPRETATIONS

The difference in accuracies of the greenness and brightness keys for the LACIE Phase II and Phase III data can be partially explained by the fact that these keys were built using LACIE Phase II data and were, therefore, more efficient at explaining Phase II data. This must also explain the difference between the results for greenness and brightness keys and the ratio keys. Because the ratio keys and spectral ranking keys were built using LACIE Phase III data, they should improve the accuracy of greenness and brightness keys on that data. In view of this, the difference between the greenness and brightness accuracy and the new variables accuracy is not considered significant.

4. CONCLUSIONS AND RECOMMENDATIONS

This study indicates that the particular channel ratios studied (section 2: a, b, c, d, and e) are not superior to greenness and brightness for discriminating small grains in Kansas. Furthermore, this study demonstrates that the normalization and ranking system described in section 2, table 2-3 does contribute to the discrimination of small grains from other. However, this was discovered in a developmental mode of analysis and has not been tested on an independent test set.

It is recommended that greenness and brightness be continued in LIST discrimination (in lieu of channel ratios), and further, that the ranking system be tested for possible use in small-grains discrimination.

5. REFERENCES

1. Pore, M. D.: Label Identification from Statistical Tabulation (LIST). LEC-10825, Rev. A, Nov. 1977.

2. Abotteen, R. A.; and Pore, M. D.: Development of Semiautomated Operational Label Identification from Statistical Tabulation (LIST). LEC-11457, Feb. 1978.

3. Pore, M. D.; and Abotteen, R. A.: Label Identification from Statistical Tabulation (LIST) Test and Evaluation. LEC-12381, Nov. 1978.

4. Dennis, T. B; and Pore, M. D.: Improvement of Green Number and Brightness Keys Used in Label Identification From Statistical Tabulation (LIST). LEC-12714, Oct. 1978.

5. Cate, R. B.; Phinney, D. E.; and Kinsler, M. C.: Standardized Classification of Landsat Digital Data Using the Retinex Theory of Color Vision. LEC-13499, May 1979 (Unpublished).

6. Hay, C.: "Manual Identification of Crop Types." Proceedings of the LACIE Symp., National Aeronautics and Space Administration at the Lyndon B. Johnson Space Center, Oct. 1978.

7. Nie, N. H.; Hull, C. H.; Jenkins, J. G.; Steinbrenner, K.; and Bent, D. H.: Statistical Package for the Social Sciences (SPSS). McGraw Hill Book Co., 1975.