NASA CR-
147713

(NASA-CR-147713) AN ALGORITHM FOR OPTIMAL SINGLE LINEAR FEATURE EXTRACTION FROM SEVERAL GAUSSIAN PATTERN CLASSES (Rice Univ.) 18 p HC $3.50 CSCL 09B

N76-23883

Unclas

G3/61 28195

ICSA

INSTITUTE FOR COMPUTER SERVICES AND APPLICATIONS

RICE UNIVERSITY

An Algorithm for Optimal Single Linear Feature Extraction from Several Gaussian Pattern Classes

by

S. A. Starks*, R. J. P. de Figueiredo*+,

and D. L. Van Rooy°

*Elec. Eng. Dept., +Math. Sciences Dept.

°I. C. S. A.

ICSA Technical Report No. 275-025-022

EE Technical Report No. 7520

ABSTRACT:

A computational algorithm is presented for the extraction of an optimal single linear feature from several Gaussian pattern classes. The algorithm minimizes the increase in the probability of misclassification in the transformed (feature) space. The general approach used in this procedure was developed in a recent paper by R. J. P. de Figueiredo [1]

Numerical results on the application of this procedure to the remotely sensed data from the Purdue C1 flight line as well as LANDSAT data are presented. It was found that classification using the optimal single linear feature yielded a value for the probability of misclassification on the order of 30% less than that obtained by using the best single untransformed feature. Also, the optimal single linear feature gave performance results comparable to those obtained by using the two features which maximized the average divergence.

Institute for Computer Services & Applications

Rice University

Houston, TX 77001

November, 1975

This work was supported by the Air Force Office of Scientific Research under the Grant 75-2777 and by the National Aeronautics and Space Administration under the Contract NAS 9-12776.

1. Introduction:

Let there be given M pattern classes, where $M \geq 2$, distributed normally in a real n-dimensional Euclidean space E^n. Specifically, let the probability density function in E^n conditioned on pattern class H^j, $j=1, \ldots, M$, be:

$$f_X(x/H^j) = (2\pi)^{-n/2} \left| (\tilde{R}^j) \right|^{-\frac{1}{2}}$$
$$\exp[-\tfrac{1}{2}(x - \bar{x}^j)^T (\tilde{R}^j)^{-1} (x - \bar{x}^j)], \quad (1)$$

where $x = \mathrm{col}\,(x_1, \ldots, x_n)$, the $n \times 1$ vector $\bar{x}^j$ and the $n \times n$ symmetric positive definite matrix $\tilde{R}^j$ are the mean and covariance of H^j in E^n, the superscript T denotes the transpose operation, and $\left| (\tilde{R}^j) \right|$ denotes the determinant of $\tilde{R}^j$. Furthermore, assume that the à priori probabilities P_j for H^j are given for $j=1, \ldots, M$. Endowed with these density functions and à priori probabilities, the space E^n will be called the measurement or data space.

In the linear feature extraction problem as formulated in [1], given an integer m such that $1 \leq m \leq n$, a linear transformation $\hat{A}$ of rank m from E^n to an m-dimensional feature space $E^m(A)$ is sought so that the Bayes risk (and in particular, the probability of misclassification) in $E^m(A)$ is minimized over a class χ of all such transformations $A: E^n \to E^m(A)$, satisfying a suitable constraint.

Let

$$y = Ax, \quad (2)$$

$$\bar{y}^j = A\bar{x}^j \tag{3}$$

$$R^j = A\tilde{R}^j A^T. \tag{4}$$

With these formulas, the probability density functions in $E^m(A)$ conditioned on pattern class H^j, $j=1,\ldots,M$, may be written as:

$$f_Y(y/H^j, A) = (2\pi)^{-m/2} \left|(R^j)\right|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(y - \bar{y}^j)^T (R^j)^{-1} (y - \bar{y}^j)\right] \tag{5}$$

Thus the Bayes risk in $E^m(A)$ can be expressed as:

$$B_R(A) = \sum_{i=1}^{M} \sum_{\substack{j=1\\ \neq i}}^{M} P_j\, C_{ij} \int_{\Omega_i(A)} f_Y(y/H^j, A)\,dy, \tag{6}$$

where $\Omega_i(A)$ is the Bayesian decision region in $E^m(A)$ for H^i and the non-negative numbers C_{ij} are the elements of the Bayes cost matrix. If $C_{ij}=1-\delta_{ij}$, where δ_{ij} is the Kronecker delta, the expression for $B_R(A)$ becomes that of the probability of error in $E^m(A)$.

The constraint that we impose on A is of the form

$$g(A) = \tfrac{1}{2}\,\text{trace}\,(A\,A^T) = \alpha, \tag{7}$$

where α is a positive constant. Incorporating this constraint in the expression for the probability of error, and using Lagrange multiplier theory, one obtains the criterion functional

$$Q_m(A,\lambda) = \sum_{i=1}^{M} \sum_{\substack{j=1 \\ \neq i}}^{M} P_j \int_{\Omega_i(A)} f_Y(y/H^j, A)\, dy \; +$$

$$+ \; \lambda\left(\tfrac{1}{2}\,\text{trace}\,(A\, A^T) - \alpha\right), \qquad (8)$$

where λ is the Lagrange multiplier associated with the constraint on A .

We use an iterative method in determining the extremum of (8). For this purpose, an expression for the gradient of $Q_m(A,\lambda)$ with respect to the elements of A is required. This expression is:

$$\nabla_A \; Q_m(A,\lambda) = \sum_{i=1}^{M} \sum_{\substack{j=1 \\ \neq i}}^{M} P_j \left\{ \left[(R^j)^{-1} \, D^{ij} - E^{ij}\, I \right] \cdot \right.$$

$$\cdot \; (R^j)^{-1} \; A \, \tilde{R}^j \; + \; (R^j)^{-1} \; \cdot$$

$$\left. . \; F^{ij}(\bar{x}^j)^T \right\} + \lambda \; A \; , \qquad (9)$$

where,

$$D^{ij} = \int_{\Omega_i(A)} (y - \bar{y}^j)(y - \bar{y}^j)^T f_Y(y/H^j, A)\, dy, \qquad (10)$$

$$E^{ij} = \int_{\Omega_i(A)} f_Y(y/H^j, A)\, dy \qquad (11)$$

$$F^{ij} = \int_{\Omega_i(A)} (y - \bar{y}^j) \; f_Y(y/H^j, A)\, dy, \qquad (12)$$

and I is the $m \times m$ identity matrix.

In the present paper, the case in which the dimensionality of the feature space is unity, i.e. $m=1$, is considered. As a result, R^j and y^j, $j=1, \ldots, M$, are scalars and $\Omega_i(A)$, $i=1, \ldots, M$, consist of one or more intervals of the real line defined by

$$\Omega_i(A) = \Big\{ y \in E^1(A) : P_i\, f_y(y/H^i, A) > P_j\, f_y(y/H^j, A) \quad \forall j \neq i \Big\}. \tag{13}$$

The boundaries of $\Omega_i(A)$ are chosen from among the roots of (13) where the inequality sign is replaced by an equality sign. In the one-dimensional case, formulas (10), (11), and (12) have closed form solutions thus yielding the following expression for the gradient:

$$\begin{aligned} \nabla_A Q_1(A,\lambda) = \sum_{i=1}^{M} \sum_{\substack{j=1 \\ \neq i}}^{M} \sum_{k=1}^{L_i} & P_j \Big(-(R^j)^{-1} \quad \tilde{R}^j \ A^T \cdot \\ & \cdot (\emptyset_{ik} - \bar{y}^j) - \bar{x}^j \Big) f_Y(\emptyset_{ik} / H^j, A) + \\ & + P_j \Big((R^j)^{-1} \ \tilde{R}^j \ A^T (\omega_{ik} - \bar{y}^j) + \bar{x}^j \Big) \cdot \\ & \cdot \ f_Y(\omega_{ik} / H^j, A) + \lambda A , \end{aligned} \tag{14}$$

where L_i = the number of distinct intervals which compose $\Omega_i(A)$, $\emptyset_{ik}$ and ω_{ik} are respectively the upper and lower endpoints of the k^{th} intervals of $\Omega_i(A)$.

We have used the above expression for the gradient of $Q_1(A,\lambda)$ in an appropriate iterative algorithm to be described in the following section. This algorithm is implemented in the form of a FORTRAN computer program which generates the optimal linear transformation $\hat{A}$.

2. Basic Algorithm

The algorithm that we have just mentioned for computing a $1 \times n$ linear transformation $\hat{A}$ and its corresponding Lagrange multiplier $\hat{\lambda}$, which minimize the criterion functional $Q_1(A, \lambda)$ with $\alpha = \frac{1}{2}$ is presented in Fig. 1 . One of the features of this algorithm is that, at each iterative step, the value of the Lagrange multiplier is updated by a procedure proposed by Tapia in [2].

It should be noted that in block (d) of Fig. 1, a single iteration of the Davidon-Fletcher-Powell unconstrained minimization procedure is performed by leaving the n components of the transformation A free to vary and holding λ fixed. More precisely,

$$A_{k+1} = A_k - K_k \nabla_A Q_1(A_k, \lambda_k), \tag{15}$$

where K_k denotes the stepsize at the k^{th} step of the algorithm. In the block (f) of Fig. 1, the Lagrange multiplier is updated by the formula [2] :

$$\lambda_{k+1} = (< A_{k+1}, A_{k+1} >)^{-1} \cdot \Big(\tfrac{1}{2} (\| A_{k+1} \|^2 - 1) - < A_{k+1}, \nabla_A f(A_{k+1}) > \Big), \tag{16}$$

where $< \cdot , \cdot >$ denotes the inner product in E^n, $\| \cdot \|$ represents the Froboenius norm in E^n, and

$$f(A) = Q_1(A,\lambda) - \tfrac{1}{2}(\|A\|^2 - 1)\lambda . \qquad (17)$$

3. Optimal Single Linear Gaussian Feature Program

The above algorithm was implemented in a double precision FORTRAN procedure consisting of a main program and several subroutines. All software with the exception of the IBM FORTRAN SSP double precision version of the Davidon-Fletcher-Powell algorithm [3] was derived by the authors. The only inputs necessary for the operation of the program are the number of pattern classes, the dimension of the measurement space, initial values for A and λ (if desired), and various control parameters (such as the maximum number of iterations of the basic algorithm to be performed, an estimate on the value of the criterion function at the minimum, etc.)

4. Numerical Results

A procedure was developed for testing the validity of the optimal single linear Gaussian feature algorithm on remotely sensed data. This procedure utilizes the program LARSYS (developed at the Laboratory for Applications of Remote Sensing, Purdue University). The test procedure is outlined in Fig. 2. It should be noted from this figure that the data set is divided into two mutually exclusive subsets; the training subset A, and the classification subset B. These two subsets consist of data from alternate columns of the same data fields. As is reflected in the figure, the subset A is used to generate the statistics used in finding the optimal single linear Gaussian feature. Once the optimal single linear Gaussian feature is found, the subset A is transformed accordingly and statistics for classification in the reduced space are generated. The subset B is used solely for classification purposes. Thus in no case are the same data points used for both the training and classification procedures. Generation of statistics and classification were performed using the STAT and CLASS options of LARSYS. The terms "performance" and "average performance" alluded to in Fig. 2 are defined as follows:

$$\text{Performance} = \frac{\text{\# of correct classifications}}{\text{total \# of classifications}} \,; \tag{18}$$

$$\text{Average performance by class} = \frac{1}{M} \sum_{i=1}^{M} \text{performance for class } i \,. \tag{19}$$

The test procedure has been applied to seven test cases. Five cases employed twelve channel data pertaining to eight pattern classes from the C1 flight line, and the remaining two employed twelve channel data belonging to four pattern classes from the LANDSAT. A typical C1 flight line data set is given in Table I. The twelve dimensional statistics computed by LARSYS for this data set is given in Table II. The resulting optimal single linear transformation $\hat{A}$ is also listed in Table II.

For comparison purposes, in addition to determining the performance of the optimal single linear Gaussian feature, the performances of the best (as computed by utilizing the average interclass Bhattacharyya distance criterion) single feature, the best (as computed utilizing the average divergence criterion) two untransformed features, and all twelve untransformed features were computed. The corresponding results are listed in Table III. It is readily noted that the optimal single linear Gaussian feature gave performance results markedly superior to those obtained by using the best single untransformed feature. Furthermore, the optimal single linear Gaussian feature gave performance results comparable to those obtained by using the best two untransformed features. Thus one could obtain a reduction in storage as well as a reduction in computation effort by using the optimal single linear Gaussian feature instead of the best two untransformed channels.

All calculations were performed using an IBM 370/155 general purpose computer at Rice University's Institute for Computer Services and Applications. The optimal single linear Gaussian feature program operates in 72 K bytes of memory.

In all test cases, the initial guess for the transformation A was made in the following manner. The single feature yielding the highest value for the average inter-class Bhattacharyya distance was found. The component of A corresponding to this feature was then set to a value of one and the remaining components of A were set to zero. Typically for a twelve channel eight class problem, the optimal single linear Gaussian feature program converged within forty iterations and required approximately five minutes of CPU time. The C1 flight line and LANDSAT data as well as the LARSYS program were provided by NASA-JSC. Numerous hypothetical cases were tested and yielded similar results, but these findings are not listed here.

5. Conclusions

The algorithm presented yields encouraging results. A method for finding an optimal n to m transformation, where $1 < m < n$, requires a different algorithmic procedure and the results from this effort will be available in the near future.

Acknowledgements

We are indebted to Dr. R. A. Tapia and Mr. A. D. Sagar for their helpful suggestions on this project.

REFERENCES

[1] de Figueiredo, Rui J. P., "Optimal Linear and Nonlinear Feature Extraction Based on the Minimization of the Increased Risk of Misclassification," Rice University Institute of Computer Services and Applications Technical Report #275-025-014 (June 1974).

[2] Tapia, R. A., "A Stable Approach to Newton's Method for General Mathematical Problems in R^n," Journal of Optimization Theory and Applications, Vol. 14, No. 5 (November, 1974), pp. 453-476.

[3] Kuester, James L. and Mize, Joe H., "Optimization Techniques with FORTRAN," McGraw-Hill, New York, 1973.

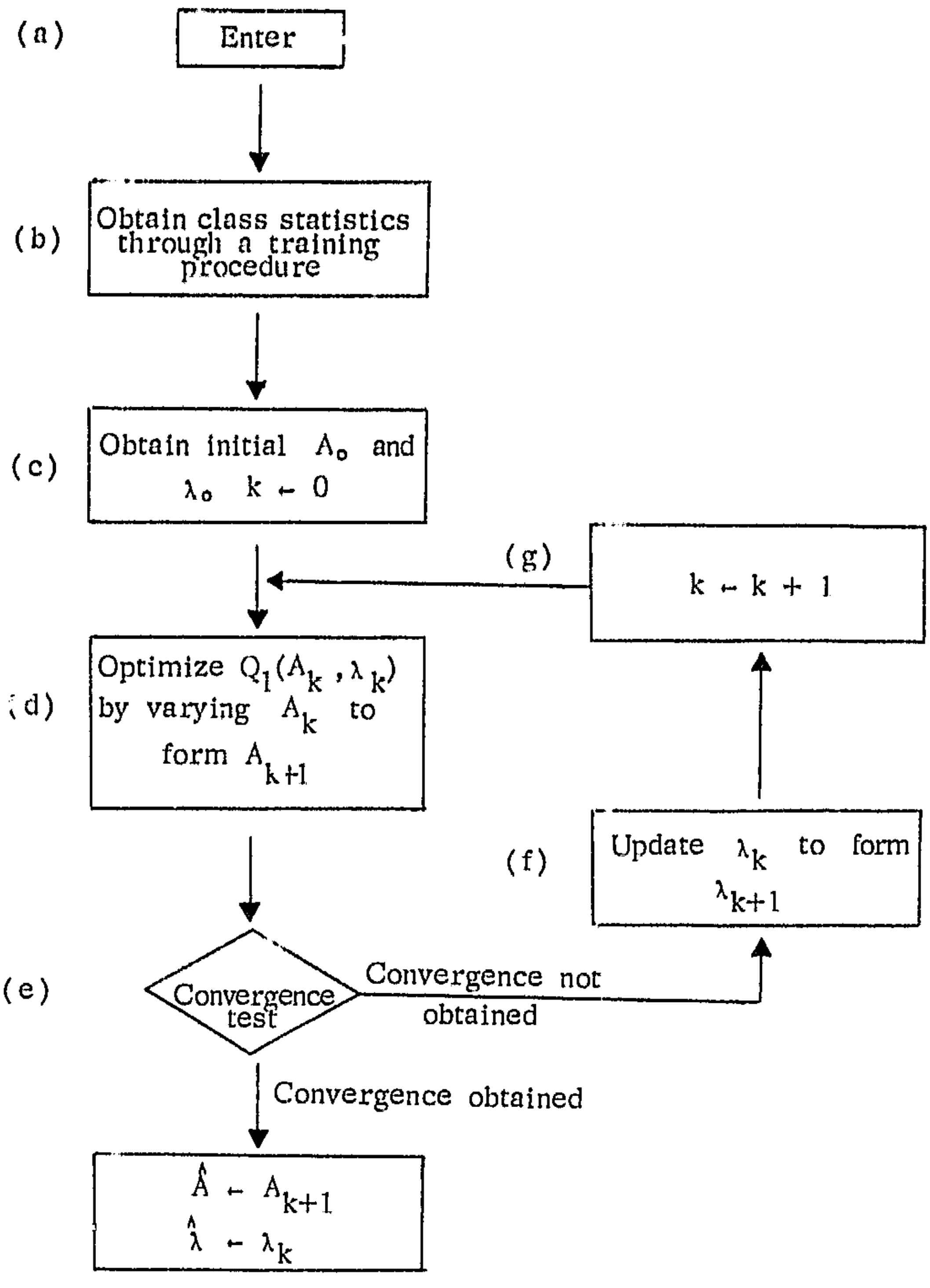

Fig. 1: Basic Optimal Linear Feature Extraction Algorithm

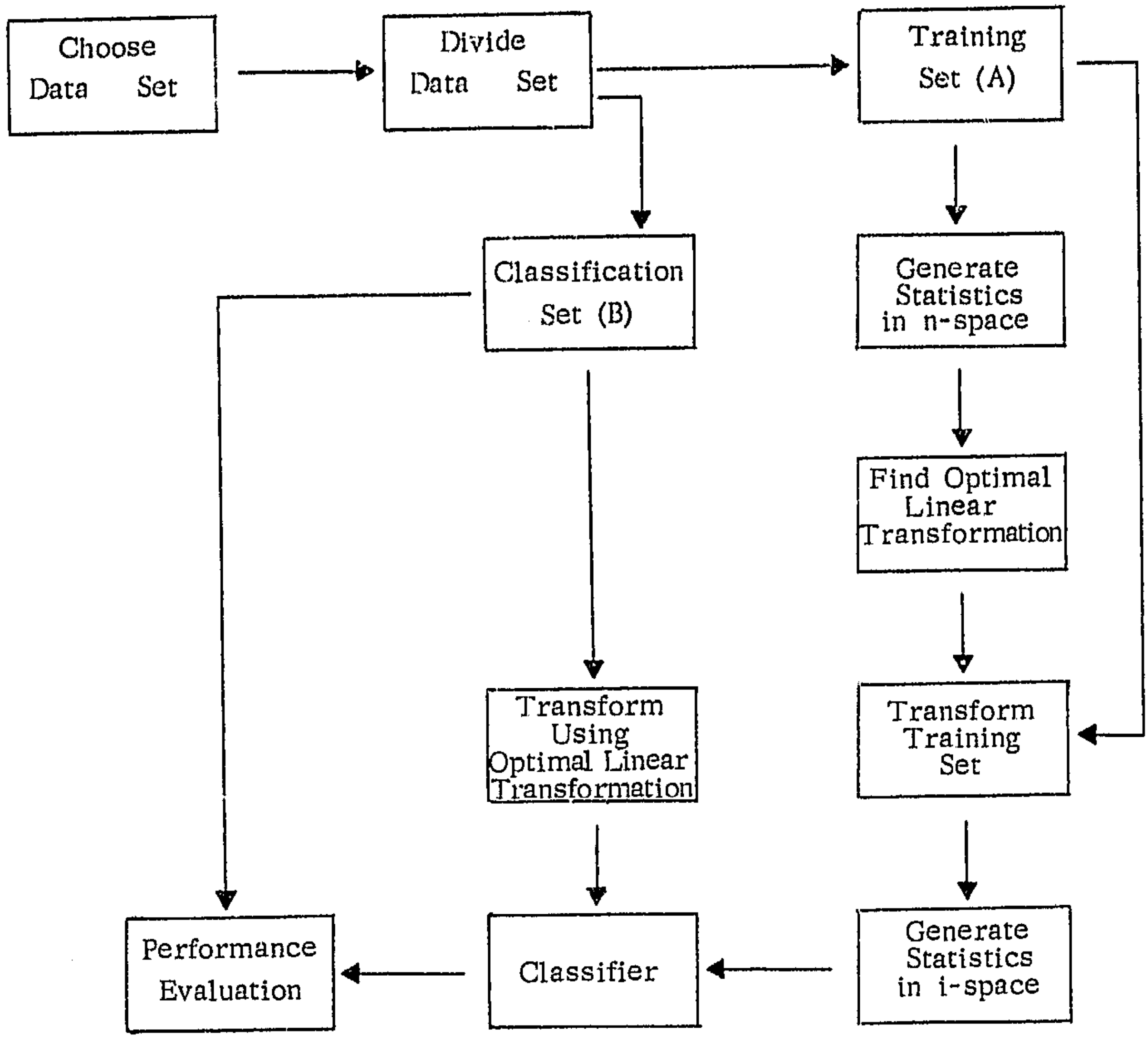

Fig. 2: Test Procedure

Training Set (A) : Flight Line C1 Run No. 6600060

FIELD DESIGNATION	CLASS	LINES (BY 1)	COLUMNS (BY 2)
25- 6-1	Soybeans	65- 81	69- 89
31-13-1	Soybeans	237-253	141-167
36- 7-1	Soybeans	307-327	59- 81
7-23-1	Soybeans	773-777	135-179
36- 1-1	Bare Soil	97-115	49- 85
36- 4-1	Corn I	167-177	33- 77
36- 9-1	Corn I	267-283	45- 61
36- 8-1	Corn I	319-341	21- 31
12- 9-1	Corn I	603-625	13- 33
31-12-1	Wheat I	295-303	135-175
6-14-1	Wheat I	471-495	177-201
7- 2-1	Wheat I	607-665	203-211
12-10-1	Wheat I	656-695	17- 41
6- 2-1	Oats	365-375	145-185
1-11-1	Oats	421-455	63- 83
7- 1-1	Oats	591-599	135-181
6-10-1	Red Clover I	439-447	139-183
1- 2-1	Red Clover I	539-565	175-195
1- 3-1	Red Clover I	599-619	69- 95
6- 8-1	Rye	527-569	127-155
7-24-1	Alfalfa	731-737	129-177
7-24-3	Alfalfa	749-755	131-171
7-22-1	Alfalfa	809-817	155-183

Classification Set (B) : Flight Line C1 Run No. 6600060

FIELD DESIGNATION	CLASS	LINES (BY 1)	COLUMNS (BY 2)
25- 6-2	Soybeans	65- 81	70- 88
31-13-2	Soybeans	237-253	142-166
36- 7-2	Soybeans	307-327	60- 80
7-23-2	Soybeans	773-777	136-178
36- 1-2	Bare Soil	97-115	50- 84
36- 4-2	Corn I	167-177	34- 76
36- 9-2	Corn I	267-283	46- 60
36- 8-2	Corn I	319-341	22- 30
12- 9-2	Corn I	603-625	14- 32
31-12-2	Wheat I	295-303	136-174
6-14-2	Wheat I	471-495	178-200
7- 2-2	Wheat I	607-665	204-210
12-10-2	Wheat I	656-695	18- 40
6- 2-2	Oats	365-375	146-184
1-11-2	Oats	421-455	64- 82
7- 1-2	Oats	591-599	136-180
6-10-2	Red Clover I	439-447	140-182
1- 2-2	Red Clover I	539-565	176-194
1- 3-2	Red Clover I	599-619	70- 94
6- 8-2	Rye	527-569	128-154
7-24-2	Alfalfa	731-737	130-176
7-24-4	Alfalfa	749-755	132-170
7-22-2	Alfalfa	809-817	156-182

Table I : Typical Data Subsets

PATTERN CLASS 1

MEAN VECTOR

85.05	80.06	61.72	62.49	85.97	87.95	64.49	84.12	70.47	82.13	92.61	73.01

COVARIANCE MATRIX

6.68	5.38	4.03	3.81	7.55	6.46	4.48	7.21	5.52	5.75	3.35	0.98
5.38	7.83	3.92	4.21	8.36	7.33	4.89	8.64	6.93	7.07	2.43	0.81
4.03	3.92	4.03	2.83	5.51	4.82	3.55	5.69	4.44	4.82	2.71	0.40
3.81	4.21	2.83	3.77	6.03	5.30	3.71	6.35	5.17	5.28	2.44	0.67
7.55	8.36	5.51	6.03	14.12	11.15	7.36	12.79	10.16	10.91	6.42	2.72
6.46	7.33	4.82	5.30	11.15	11.99	7.14	11.44	8.74	10.37	9.82	4.93
4.48	4.89	3.55	3.71	7.36	7.14	5.66	7.83	6.30	7.30	5.94	2.53
7.21	8.64	5.69	6.35	12.79	11.44	7.83	16.00	12.55	13.07	3.71	1.67
5.52	6.93	4.44	5.17	10.16	8.74	6.30	12.55	12.24	11.17	0.07	0.06
5.75	7.07	4.82	5.28	10.91	10.37	7.30	13.07	11.17	14.26	6.11	3.20
3.35	2.43	2.71	2.44	6.42	9.82	5.94	3.71	0.07	6.11	37.44	14.37
0.98	0.81	0.40	0.87	2.72	4.93	2.53	1.67	0.06	3.20	14.37	13.19

PATTERN CLASS 2

MEAN VECTOR

90.04	85.32	65.41	66.89	92.52	89.47	67.18	95.11	83.47	90.24	71.74	60.36

COVARIANCE MATRIX

3.23	1.23	1.16	0.57	1.17	1.01	0.75	0.71	0.43	0.95	1.01	0.10
1.23	3.32	0.65	0.74	1.17	1.12	0.67	0.66	0.85	0.86	0.74	0.34
1.16	0.65	1.65	0.31	0.72	0.53	0.67	0.60	0.23	0.73	1.00	-0.02
0.57	0.74	0.31	1.36	0.82	0.76	0.48	0.58	0.63	0.71	0.70	0.35
1.17	1.17	0.72	0.82	3.39	1.29	0.63	0.88	0.95	1.25	0.78	0.71
1.01	1.12	0.53	0.76	1.29	2.34	0.80	1.21	1.23	1.40	1.05	0.62
0.75	0.67	0.67	0.48	0.63	0.80	1.50	0.75	0.59	1.05	1.26	0.11
0.71	0.66	0.60	0.58	0.88	1.21	0.75	3.44	1.26	1.64	1.25	0.65
0.43	0.85	0.23	0.63	0.95	1.23	0.59	1.26	2.11	1.63	0.98	0.83
0.95	0.86	0.73	0.71	1.25	1.40	1.05	1.64	1.63	3.61	1.84	0.93
1.01	0.74	1.00	0.70	0.78	1.05	1.26	1.25	0.98	1.84	7.76	0.43
0.10	0.34	-0.02	0.35	0.71	0.62	0.11	0.65	0.83	0.93	0.43	3.04

PATTERN CLASS 3

MEAN VECTOR

84.14	77.84	59.69	60.00	82.31	86.73	62.45	76.64	61.89	72.92	106.62	81.28

COVARIANCE MATRIX

9.57	6.91	4.44	3.99	7.63	4.99	3.00	5.71	4.42	3.53	-2.60	-2.49
6.91	8.91	3.79	4.04	7.16	5.14	3.15	6.38	5.25	4.21	-3.94	-2.50
4.44	3.79	3.43	2.21	4.35	2.81	1.98	3.52	2.76	2.58	-1.43	-1.49
3.99	4.04	2.21	3.18	4.11	3.06	2.08	3.89	3.41	2.80	-3.33	-1.82
7.63	7.16	4.35	4.11	9.85	6.01	3.58	7.12	5.71	5.44	-2.56	-1.85
4.99	5.14	2.81	3.06	6.01	6.05	3.40	5.99	4.66	5.11	0.25	0.56
3.00	3.15	1.98	2.08	3.58	3.40	3.41	4.17	3.80	3.97	-0.70	-0.60
5.71	6.38	3.52	3.89	7.12	5.99	4.17	9.66	7.65	7.28	-6.67	-3.56
4.42	5.25	2.76	3.41	5.71	4.86	3.80	7.65	8.51	6.66	-8.62	-4.15
3.52	4.21	2.58	2.80	5.44	5.11	3.97	7.28	6.66	8.95	-3.00	-1.78
-2.60	-3.94	-1.43	-3.33	-2.55	0.25	-0.70	-6.67	-8.62	-3.00	54.04	23.51
-2.49	-2.50	-1.49	-1.82	-1.85	0.56	-0.60	-3.56	-4.15	-1.78	23.51	21.18

PATTERN CLASS 4

MEAN VECTOR

76.31	73.29	57.78	58.43	79.85	84.55	63.70	85.53	71.71	88.43	107.38	84.97

COVARIANCE MATRIX

9.15	6.61	4.72	4.05	9.95	9.87	6.02	7.23	3.71	4.42	3.12	-1.92
6.61	8.24	4.18	4.46	9.44	9.09	5.59	8.62	6.39	6.34	-2.17	-4.31
4.72	4.18	3.94	2.76	6.33	5.44	3.81	5.93	4.28	4.56	-1.06	-3.18
4.05	4.46	2.76	3.88	6.27	5.83	3.95	6.67	5.68	5.39	-3.54	-4.04
9.95	9.44	6.33	6.27	15.92	13.28	8.17	12.82	9.13	9.01	-3.67	-6.89
9.67	9.09	5.44	5.63	13.28	15.15	8.22	10.44	6.10	6.60	2.63	-2.58
6.02	5.59	3.81	3.95	8.17	8.22	6.27	7.79	5.90	6.17	-0.77	-3.48
7.23	8.62	5.93	6.67	12.82	10.44	7.79	18.89	16.73	16.24	-15.07	-12.71
3.71	6.39	4.28	5.68	9.13	6.10	5.90	16.73	19.43	16.62	-21.22	-14.96
4.42	6.34	4.56	5.39	9.01	6.60	6.17	16.24	16.62	18.09	-15.71	-11.65
3.12	-2.17	-1.06	-3.54	-3.67	2.63	-0.77	-15.07	-21.22	-15.71	53.23	25.77
-1.92	-4.31	-3.18	-4.04	-6.89	-2.58	-3.48	-12.71	-14.96	-11.65	25.77	24.36

Table II : Typical Twelve Dimensional Statistics

PATTERN CLASS 5

MEAN VECTOR

72.70	69.45	54.52	54.30	72.36	80.04	58.20	68.93	55.40	76.90	141.11	107.80

COVARIANCE MATRIX

4.31	2.71	2.09	1.92	4.06	3.99	2.62	3.91	2.68	4.02	4.91	3.28
2.71	4.29	1.81	2.04	4.11	4.32	2.58	4.14	3.12	4.22	2.62	2.23
2.09	1.81	2.31	1.35	2.77	2.65	2.02	2.83	2.03	2.80	2.47	0.99
1.92	2.04	1.35	2.20	2.86	3.00	2.05	3.18	2.48	2.83	-0.27	0.21
4.06	4.11	2.77	2.86	7.68	6.82	4.18	6.64	4.63	6.80	6.26	3.46
3.99	4.32	2.65	3.00	6.82	9.95	5.19	6.53	4.09	9.11	20.91	12.36
2.62	2.58	2.02	2.05	4.18	5.19	4.16	4.21	2.99	5.20	8.13	4.45
3.91	4.14	2.83	3.18	6.64	6.53	4.21	8.54	5.77	6.13	-4.24	-1.88
2.68	3.12	2.03	2.48	4.63	4.09	2.99	5.77	5.80	3.92	-9.13	-4.85
4.02	4.22	2.80	2.83	6.80	9.11	5.20	6.13	3.92	11.65	28.38	16.44
4.91	2.62	2.47	-0.27	6.26	20.91	8.13	-4.24	-9.13	28.38	213.05	112.47
3.28	2.23	0.99	0.21	3.46	12.36	4.45	-1.88	-4.85	16.44	112.47	77.52

PATTERN CLASS 6

MEAN VECTOR

75.69	75.02	60.66	62.67	87.72	[illegible]	68.84	101.93	90.37	103.95	88.69	72.29

COVARIANCE MATRIX

23.56	25.03	18.36	19.86	47.06	50.09	34.30	58.84	43.02	54.03	46.51	25.44
25.03	32.00	21.46	24.35	56.54	61.10	41.57	72.16	53.43	66.42	56.36	31.49
18.36	21.46	17.08	17.68	41.55	44.46	30.72	53.12	39.17	49.21	41.71	22.76
19.86	24.35	17.68	21.02	46.68	51.11	35.23	61.18	45.80	57.17	47.62	25.93
47.06	56.54	41.55	46.68	112.64	119.83	82.04	143.26	104.91	131.46	111.47	61.55
50.09	61.10	44.46	51.11	119.83	134.00	91.52	157.72	115.44	144.94	125.61	67.77
34.30	41.57	30.72	35.23	82.04	91.52	64.27	109.10	80.49	101.59	88.00	46.86
58.84	72.16	53.12	61.18	143.26	157.72	109.10	193.74	142.44	180.90	151.68	83.63
43.02	53.43	39.17	45.80	104.91	115.44	80.49	142.44	111.10	138.98	113.43	63.54
54.03	66.42	49.21	57.17	131.46	144.94	101.59	180.90	138.98	162.26	149.19	85.01
46.51	56.36	41.71	47.62	111.47	125.61	88.00	151.68	113.43	149.19	146.08	78.40
25.44	31.49	22.76	25.93	61.55	67.77	46.86	83.63	63.54	85.01	78.40	52.20

PATTERN CLASS 7

MEAN VECTOR

80.12	79.79	63.68	65.36	93.92	96.71	72.57	102.91	86.13	99.84	96.05	75.86

COVARIANCE MATRIX

4.79	3.53	2.80	2.46	4.28	2.98	2.46	5.04	4.84	4.90	-1.01	-0.46
3.53	6.57	2.84	3.30	5.34	4.36	2.98	6.66	6.74	6.43	-1.25	-0.75
2.80	2.84	3.34	2.17	3.77	2.49	2.35	4.62	4.34	4.49	-0.60	-0.68
2.46	3.30	2.17	3.17	3.92	3.17	2.50	5.24	5.41	5.15	-1.05	-0.84
4.28	5.34	3.77	3.92	8.71	5.07	3.79	8.73	8.56	8.48	-1.85	-1.30
2.98	4.36	2.49	3.17	5.07	5.61	3.34	7.35	7.54	7.18	-0.57	-0.43
2.46	2.98	2.35	2.50	3.79	3.34	3.74	5.51	5.75	5.66	0.17	-0.79
5.04	6.66	4.62	5.24	8.73	7.35	5.51	14.45	13.27	12.80	-2.69	-2.27
4.84	6.74	4.34	5.41	8.56	7.54	5.75	13.27	15.21	13.18	-3.19	-2.56
4.90	6.43	4.49	5.15	8.48	7.18	5.66	12.80	13.18	14.92	-1.45	-1.44
-1.01	-1.25	-0.60	-1.05	-1.85	-0.57	0.17	-2.69	-3.19	-1.45	17.51	3.79
-0.46	-0.75	-0.68	-0.84	-1.30	-0.43	-0.79	-2.27	-2.56	-1.44	3.79	6.18

PATTERN CLASS 8

MEAN VECTOR

77.18	71.98	55.74	55.82	75.69	85.19	60.78	70.25	55.04	78.57	151.63	111.78

COVARIANCE MATRIX

3.51	2.63	2.19	2.03	3.53	1.82	1.79	4.50	3.97	2.08	-13.45	-8.45
2.63	5.07	2.52	2.79	4.62	2.65	2.48	6.38	5.95	3.12	-18.62	-11.52
2.19	2.52	3.02	2.00	3.57	1.88	2.10	4.39	4.07	2.37	-12.85	-8.55
2.03	2.79	2.00	3.04	3.65	2.14	2.10	5.05	4.83	2.47	-15.78	-9.15
3.53	4.62	3.57	3.65	8.08	3.99	3.56	9.02	7.74	4.70	-25.46	-15.88
1.82	2.65	1.88	2.14	3.99	4.29	2.79	5.47	4.85	3.50	-13.33	-7.88
1.79	2.48	2.10	2.10	3.56	2.79	3.34	4.84	4.47	3.05	-13.26	-8.19
4.50	6.38	4.39	5.05	9.02	5.47	4.84	13.32	11.30	6.25	-36.68	-21.31
3.97	5.95	4.07	4.83	7.74	4.85	4.47	11.30	11.91	5.68	-35.37	-20.27
2.08	3.12	2.37	2.47	4.70	3.50	3.05	6.25	5.68	5.66	-14.70	-8.74
-13.45	-18.62	-12.85	-15.78	-25.46	-13.33	-13.26	-36.68	-35.37	-14.70	157.50	81.92
-8.45	-11.52	-8.55	-9.15	-15.88	-7.88	-8.19	-21.31	-20.27	-8.74	81.92	59.72

$$\hat{A} = (-.44040 \quad -.38796 \quad .04587 \quad -.23900 \quad -.08005 \quad -.23782$$
$$.12532 \quad .43812 \quad .51823 \quad -.03791 \quad -.07618 \quad -.21596)$$

Table II (contd.)

Test Case	Total # of Points in Data Set	Performance Results			
		Optimal Single Linear Gaussian Feature	Best Single Feature	Best Two Features	All Twelve Features
I -Cl	10,944	.740	.553	.797	.990
II -Cl	9,941	.759	.648	.892	.992
III -Cl	9,941	.667	.620	.886	.986
IV -Cl	14,997	.803	.683	.857	.993
V -Cl	14,997	.799	.664	.849	.993
VI -LANDSAT	1,051	.850	.834	.864	.918
VII-LANDSAT	1,051	.847	.831	.875	.933
		Average Performance by Class Results			
		Optimal Single Linear Gaussian Feature	Best Single Feature	Best Two Features	All Twelve Features
I -Cl	10,994	.739	.575	.813	.990
II -Cl	9,941	.739	.610	.860	.993
III -Cl	9,941	.706	.615	.871	.975
IV -Cl	14,997	.782	.651	.854	.993
V -Cl	14,997	.784	.653	.848	.993
VI -LANDSAT	1,051	.860	.827	.868	.923
VII-LANDSAT	1,051	.856	.803	.869	.935

Table III: Test Case Results