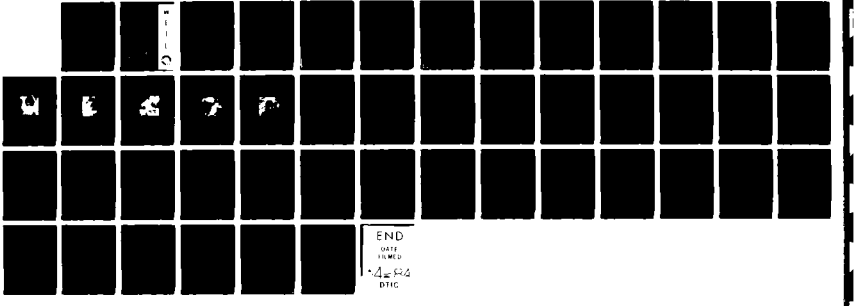
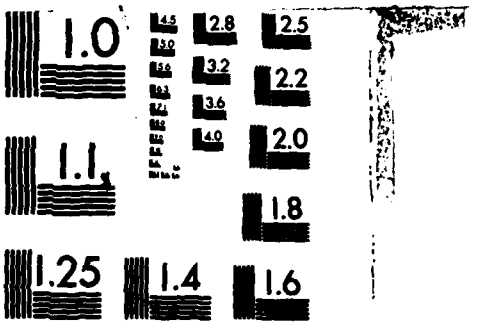
AD-A138 366 FEATURE ANALYSIS AND REDUCTION OF LAWS TEXTURE MEASURE 1/1
(U) ARMY ENGINEER TOPOGRAPHIC LABS FORT BELVOIR VA
R S RAND ET AL. OCT 83 ETL-0343

UNCLASSIFIED F/G 20/6 NL





ETL-0343

AD A138366

Feature analysis and reduction of laws texture measure

Robert S. Rand

James A. Shine

OCTOBER 1983

U.S. ARMY CORPS OF ENGINEERS
ENGINEER TOPOGRAPHIC LABORATORIES
FORT BELVOIR, VIRGINIA 22060

E T L

Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

The citation in this report of trade names of commercially available products does not constitute official endorsement or approval of the use of such products.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ETL-0343	2. GOVT ACCESSION NO. AD-A138366	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) FEATURE ANALYSIS AND REDUCTION OF LAWS TEXTURE MEASURE		5. TYPE OF REPORT & PERIOD COVERED Research Note 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Robert S. Rand James A. Shine		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Army Engineer Topographic Laboratories Fort Belvoir, VA 22060		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 4A762707A855, B, 00026
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Engineer Topographic Laboratories Fort Belvoir, VA 22060		12. REPORT DATE October 1983 13. NUMBER OF PAGES 43
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Bayes Classifier	Laws Texture Measure
Digital Imagery	Pattern Recognition
Divergence	Principal Components

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Laws Texture Measure was tested in an autoevaluation experiment and compared to some of the texture measures previously studied at ETL. In addition, four methods of component reduction were evaluated. Laws Texture was found to be the most effective measure for identifying a Buildings/Roads class; however, it was less effective on other classes. A technique that transforms the original components into principal components and then arranges the transformed components in an order that maximizes the divergence was selected as the most effective component reduction method.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

PREFACE

This study was conducted under DA Project 4A762707A855, Task B, Work Unit 00026, "Topographic Mapping Techniques."

The study was done during 1982 under the supervision of Mr. Dale E. Howell, Chief, Information Sciences Division, and Mr. Lawrence A. Gambino, Director, Computer Sciences Laboratory.

Special thanks are extended to Mr. Edward Bosch who assisted in the divergence analysis of the feature component data.

COL Edward K. Wintz, CE, was Commander and Director and Mr. Robert P. Macchia was Technical Director of the Engineer Topographic Laboratories during the report preparation.

CONTENTS

ILLUSTRATIONS

TABLES

FEATURE ANALYSIS AND REDUCTION OF LAWS TEXTURE MEASURE

INTRODUCTION

The Engineer Topographic Laboratories (ETL), Computer Sciences Laboratory (CSL), is investigating the problem of finding an inexpensive set of window operations that can be used to generate multicomponent data from digital or digitized imagery. The quality of this data must be such that it can be used effectively by a classifier to segment a scene on an image into feature categories. The effort is part of a comprehensive "feature extraction study" at CSL, which is taking a digital approach to solving the Defense Mapping Agency's production requirements for Mapping, Charting and Geodetic (MC&G) data. However, because of the preliminary nature of this part of the study, only simple feature categories such as forests, fields, buildings, and roads are considered. Also, the source images have so far been limited to digitized panchromatic and infrared aerial photographs.

Previous experiments supporting the effort have been published.[1,2,3] In these experiments, image descriptors were defined through window operations such as the Max-Min texture measure, edge texture measures, and a few simple "Ad-Hoc" measures (average gray shade, standard deviation of gray shades, and range of gray shades). These window operations were used to generate data from the various source images. The generated data was then processed in a few algorithms that estimated its information content and segmenting properties; the divergence measure and Bayes classifier were the primary tools in this analysis. In addition to test regions, a few complete scenes were classified using some of these descriptors.

---

[1] M. A. Crombie, R. S. Rand, and N. J. Friend, An Analysis of the Max-Min Texture Measure, U.S. Army Engineer Topographic Laboratories, Fort Belvoir VA, ETL-0280, January 1982, AD-A116 768.

[2] M. A. Crombie, R. S. Rand, and N. J. Friend, Scene Classification Results Using the Max-Min Texture Measure, U.S. Army Engineer Topographic Laboratories, Fort Belvoir, VA, ETL-0300, July 1982, AD-A123 496.

[3] M. A. Crombie, N. J. Friend, and R. S. Rand, Feature Component Reduction Through Divergence Analysis, U.S. Army Engineer Topographic Laboratories, Fort Belvoir, VA, ETL-0305, October 1982, AD-A123 474.

This research note discusses an experiment that was done to study another window-type operation, the Laws texture measure.[4] CSL's motivation for studying this texture measure is that it is particularly suited to the computer; the primary operations are convolution and moving-window functions, both of which are simple and fast. However, even though a large number of components can be generated quickly, the cost of processing these components also increases. Therefore, the feasibility of using this texture measure depends to some extent on the success in reducing the number of these components through some selection process, which might or might not require a coordinate transformation into some alternative representation. Because of this need to reduce components, considerable attention is given to methods of component reduction.

DESCRIPTION OF EXPERIMENT

Overview. The experiment is divided into two parts. In the first part, raw texture data is generated by a procedure developed by Laws and is then used directly by a Bayes classifier on test areas of five scenes. The results of this part are intended to measure the effectiveness of the data in image segmentation and to determine whether Laws texture is a good competitor with some of the other texture measures mentioned above, specifically the Max-Min texture and the two-component Ad-Hoc measure. In the second part, component reduction techniques are applied to the raw data and the resulting data is used by the classifier. Here, the results are intended to measure the effectiveness of alternative "reduced" representations of the data in image segmentation. As mentioned above, because of the large amount of data and the large number of computations involved, the feasibility of using Laws texture depends on the success in finding an effective "reduced" representation.

[4] Kenneth Ivan Laws, Textured Image Segmentation, Image Processing Institute, University of Southern California, Los Angeles, CA 90007, USCIPI Report 940, January 1980.

In support of the second part of the experiment, two techniques of component reduction are considered--principal components and divergence analysis. Principal components are components of the feature vectors in a transformed coordinate representation. This representation has two notable properties: first, the energy flux (variance) of the data is contained within a few components, and second, the components are uncorrelated over the sample space from which they were derived.

The divergence is a measure of information that is suitable for selecting the components with the most discriminating power. In this experiment, the divergence is used to define an "order of importance" for components in both the original representation (coordinate frame of the raw data) and in the principal component representation. Also, an order of importance is defined for the principal components using their energy compression property. By virtue of the technique from which they are generated, the first principal component contains the highest percentage of the variance, the second principal component contains the second highest percentage of the variance, etc.

Three methods are considered for component reduction: Method I arranges the raw components in an order that maximizes the divergence, Method 2 uses the principal components in an order that maximizes the energy flux, and Method 3 arranges the principal components in an order that maximizes the divergence. In addition, another method--Method 4--is briefly considered; this method uses principal components generated by a different transformation technique and orders the components according to their energy flux (similar to Method 2).

After defining sets of reduced components from these methods, a Bayes classifier is used to classify the areas of the scenes. Also, in the interest of reducing processing cost, additional classification runs are made to test the possibility of exploiting the uncorrelated nature of principal components. Since uncorrelated components have the property of diagonal covariance matrices, the use of this diagonal property in a classifier will significantly reduce the number of computations needed during implementation.

A diagram that illustrates the procedure described above is shown in figure 1.

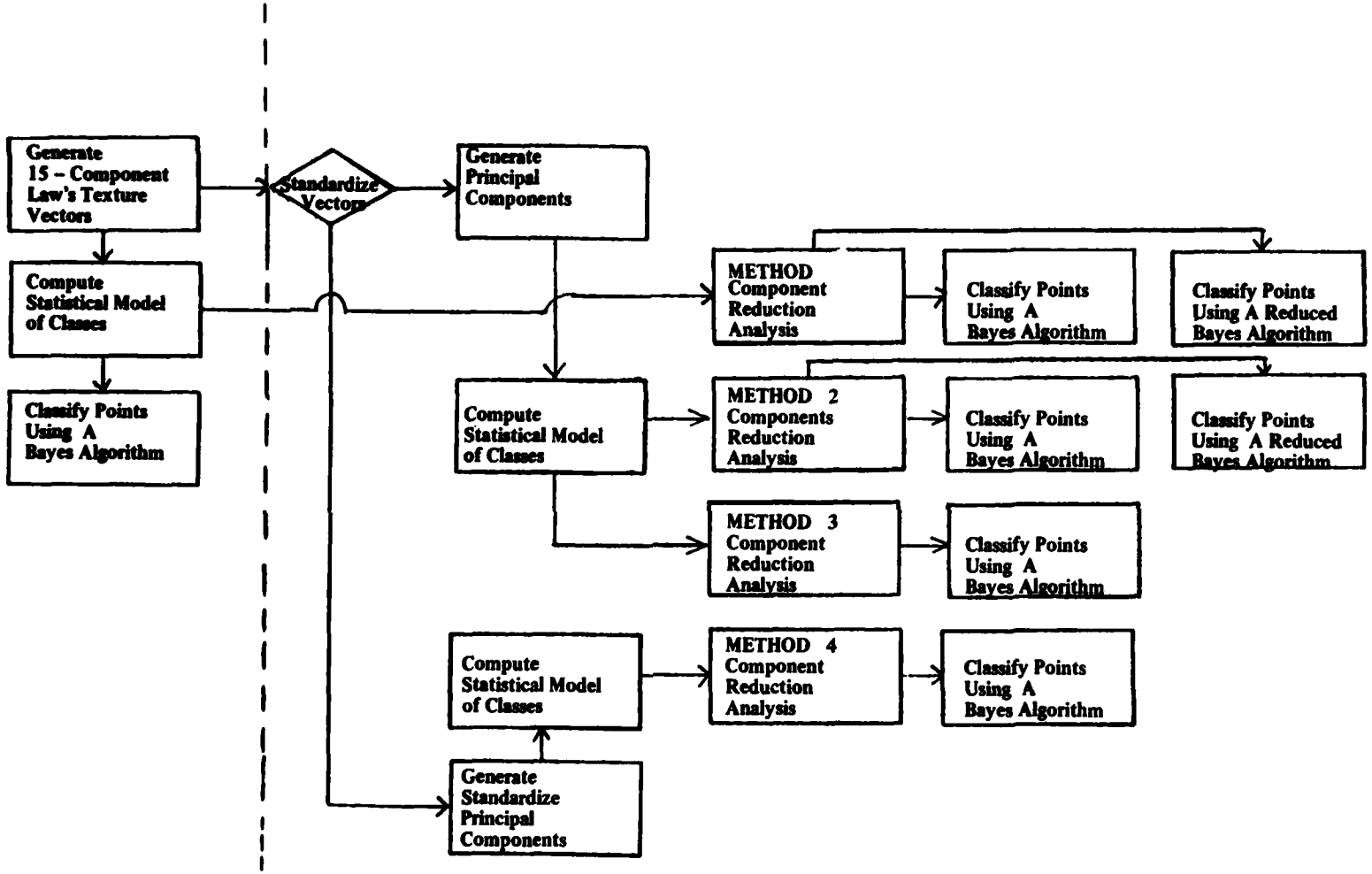
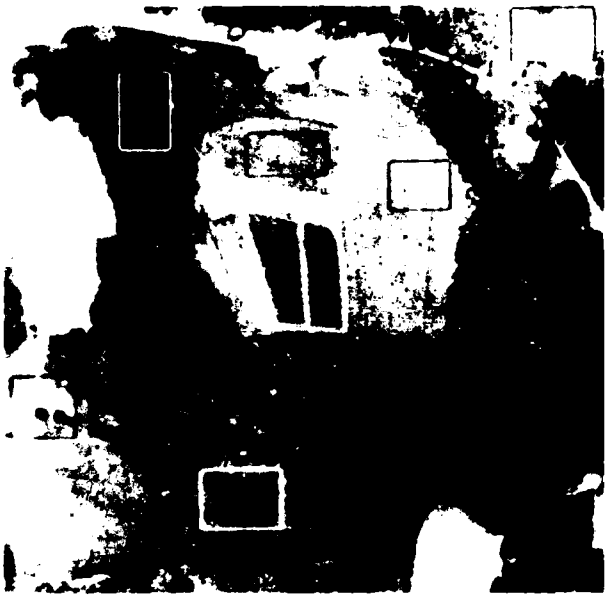


FIGURE 1. Diagram of experimental procedure.

Test Images. Five digital subscenes, each containing 1024 by 1024 picture elements (pixels), are used in the experiment as source data. These source images are panchromatic exposures digitized to 1-meter resolution and stored on disk with 8-bit accuracy. Reference to the scenes are made as A, B, C, E, and H. Training regions for the scenes are selected to represent a building-road class and various forest, field, and scrub classes. These regions were used in earlier reports, and although it was later decided that some were not very good samples, they are used again here for consistency. Because the experiment is preliminary in nature, only the training regions used as samples are processed by the classifier. At a later time--and if Laws texture is found to be feasible--complete images will be processed. Figures 2 through 6 show the five scenes. The rectangular training regions are outlined on each scene, and the classes that each will represent are listed below it.

Generation of Laws Texture Data. By following a procedure suggested by Kenneth Laws,[5] 15 component vectors are generated as data for the experiments. A three-step procedure is used. The first step is to convolve the desired image points with 16 different masks, resulting in a convolved image plane for each mask. The set of masks are defined by the cross-product computations described in appendix A. In the second step, each point in the 16 convolved images is transformed to a measure of texture energy by a moving window operation that computes the standard deviation of the 15 by 15 points surrounding it. In the third step, the texture energy planes are ratioed to the first plane, resulting in 15 invariant texture energy planes. The data is then converted to vector format. Once the data is in vector format, it can be used to develop statistical training models and processed in a classification algorithm (part I) or used to generate principal component data, which would then be modeled and classified in an alternative representation (part II).

[5] Kenneth Ivan Laws, Textured Image Segmentation, Image Processing Institute, University of Southern California, Los Angeles, CA 90007, USCIPI Report 940, January 1980.

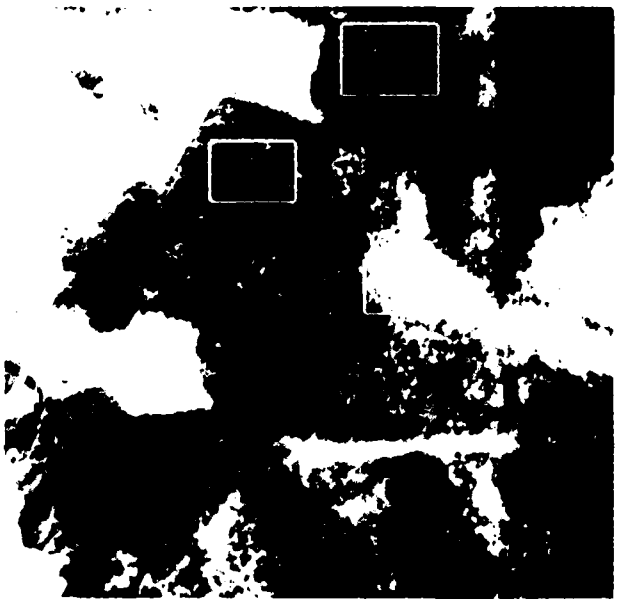


Class Label (Training Region)	Characteristic Feature	Number of Points
1	Building and Roads	81
2	Gray Field	72
3	Rough Field	104
4	Heavy Forest	84
5	Light Field	63
6	Light Forest	108

Figure 2. Scene A, Panchromatic, Exp. 54.

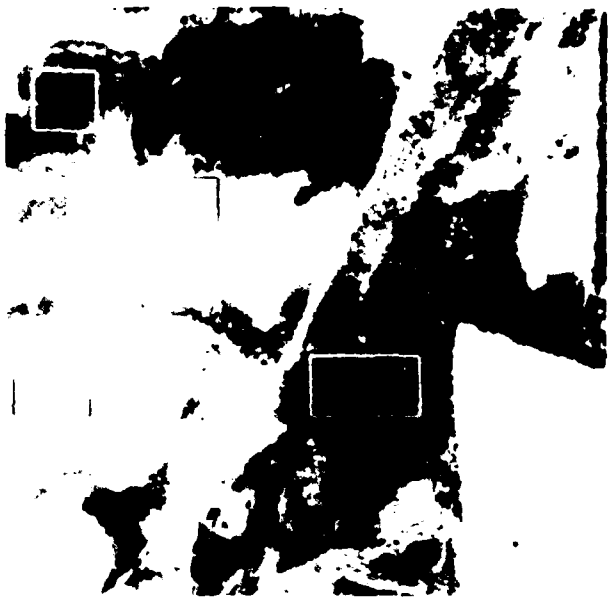
Class Label (Training Region)	Characteristic Feature	Number of Points
1	Heavy Forest	121
2	Scrub	99
3	Field Building and Road	81
4	Dark Field	77
5	Light Field	117
6	Light Forest	81

Figure 3. Scene B, Panchromatic, Exp. 54.



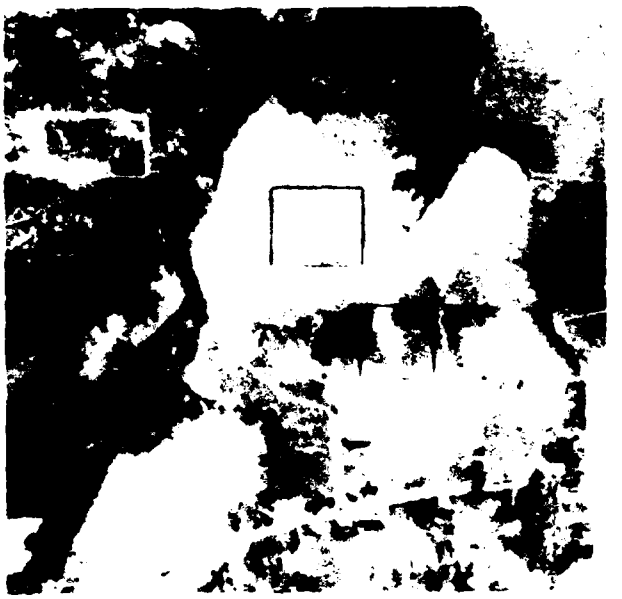
Class Label (Training Region)	Characteristic Feature	Number of Points
1	Heavy Forest (Light)	117
2	Heavy Forest (Dark)	165
3	Light Forest	104
4	Light Field	84
5	Gray Field	104

Figure 4. Scene C, Panchromatic, Exp. 54.



Class Label (Training Region)	Characteristic Feature	Number of Points
1	Dark Field	81
2	Light Field	99
3	Heavy Forest	153
4	Scrub	81
5	Building and Road	30
6	Light Forest	81

Figure 5. Scene E, Panchromatic, Exp. 54.



Class Label (Training Region)	Characteristic Feature	Number of Points
1	Dark Field	121
2	Light Field	168
3	Scrub	153
4	Building and Road	90

Figure 6. Scene H, Panchromatic, Exp. 54.

Statistical Models. Each task in this experiment--whether it is the generation of principal components, the divergence analysis, or the classification--requires the use of a statistical model. The generation of principal components uses the covariance estimate S or the correlation estimate R of the entire data set that is to be transformed. The divergence analyses and the classification runs require a statistical model for each of the classes (feature categories) that is defined.

The models that are used for defining classes make the assumption that the corresponding populations are Gaussian, an assumption that implies that second-order statistics are sufficient to specify completely the class distributions. Each class is modeled with a multivariate normal distribution using the parameters $\bar{u}_i$ as the estimate for the mean vector (with k-components) and S_i as the estimate for the covariance matrix (with k x k symmetric elements). The dimension of the distribution is k, where k is the number of texture components. The estimates for the parameters are unbiased estimates for the parent populations, and because of the Gaussian assumption, they reduce simply to sample means and sample covariances. Therefore, $\bar{u}_i$ and S_i are computed as

$$\bar{u}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \bar{X}_{ij} \tag{1}$$

$$S_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\bar{X}_{ij} - \bar{u}_i)(\bar{X}_{ij} - \bar{u}_i)^T \tag{2}$$

where X_{ij} is the j^{th} sample vector from class i, and n_i is the number of sample vectors from class i. The samples used to create these models are extracted from the training regions discussed in the section on "Test Imagery" and are shown in figures 2 through 6.

As will be discussed below, the principal components are generated using either the covariance estimate S or the correlation estimate R. Since the data set consists of the union of all the class samples, S can easily be computed using the individual class covariances S_i:

$$S = \sum_{i=1}^{N} w_i S_i$$

where N is the number of sample areas and w_i is a weight function. The correlation matrix is extracted using the combined covariance matrix; elements of R are computed as

$$R_{ij} = \frac{S_{ij}}{\sigma_i \sigma_j} \tag{3}$$

where σ_i and σ_j are the standard deviation for the i^{th} and j^{th} components, respectively.

Divergence Analysis. In two of the four methods of data reduction considered in this experiment, the texture components are arranged in an order of importance using a distance-type measure called divergence.[6] The divergence, $J(\sigma_i\ \sigma_j)$, is a measure of information that indicates the difficulty in discriminating between two classes, i and j. The value of $J(c_i\ c_j)$ is a scalar and measures the amount of discriminating information in the data. A large value suggests that it should be easy to discriminate between two classes; whereas, a small value suggests that the discrimination will be difficult. The expression for the divergence $J(c_i\ c_j)$ is

$$J(C_i, C_j) = \int_{-\infty}^{\infty} [P(x/C_i) - P(x/C_j)] \log \left\{ \frac{P(x/C_i)}{P(X/C_j)} \right\} dx \qquad (4)$$

To compute this measure, one must know the class-conditional probability distributions $p(x/C_i)$ and $p(x/C_j)$. These distributions, in general, are not usually known and are usually estimated according to some assumption. In the model of this experiment, the assumption is that the classes have multivariate normal distributions. Assuming such distributions, one can compute the divergence as

$$J'(C_i,C_j) = \tfrac{1}{2}\,\mathrm{Tr}(K_i - K_j)(K_j^{-1} - K_i^{-1}) + \tfrac{1}{2}\,\mathrm{Tr}(K_i^{-1} + K_j^{-1})(\bar{u}_i - \bar{u}_j)(\bar{u}_i - \bar{u}_j)^T \qquad (5)$$

The assumption of normality is made in this experiment recognizing that $J'(C_i\ C_j)$ will at best give only an approximation to the real solution $J(C_i, C_j)$. Although the distribution of a Gaussian population can be completely specified by the mean and covariance parameters, other distributions require either different parameters or higher order statistics. Therefore, fitting this model to any non-Gaussian population results in only an estimate.

[6] See books such as Harry Andrews Introduction to Mathematical Techniques in Pattern Recognition, John Wiley & Sons, Inc., 1972.

Classification Algorithm. A Bayes classifier that makes the assumption of normality is used in the classification exercises.[7] The resulting decision function used to classify vector x is

$$d = \max \{d_i;\ i = 1,N\} \tag{6}$$

where

$$d_i = \ln p(c_i) - \tfrac{1}{2} \ln S_i - \tfrac{1}{2} (\bar{x} - \bar{\mu}_i)^T S_i^{-1} (\bar{x} - \bar{\mu}_i) \tag{7}$$

and N is the number of classes. The vector x belongs to the class i for which $d_i = d$.

The parameter $p(c_i)$ is the a priori probability that a vector belongs to class i, and the parameters $\bar{\mu}_i$ and S_i contain the class statistical elements discussed in the last two sections. However, the function d is shortened in the classification exercises by setting the a priori probabilities $(P(c_i);\ i = 1,N)$ equal. This has the effect of dropping the term $\ln p(c_i)$ from the equations.

In the last few classification exercises, a simplified Bayes classifier is tried on uncorrelated principal component data. The savings in computation by using this algorithm can be understood by looking at the term $(\bar{x} - \bar{\mu}_i)^T S_i^{-1} (\bar{x} - \bar{\mu}_i)$ in equation 7. This term must be computed N times for each point that is labeled, and the number of computations required by this term, if all the elements of the matrix S_i^{-1} are included, is proportional to K(K+1)/2 (the number of elements in a symmetric matrix of order K), where K is the number of components. The simplified classifier, however, uses only the diagonal elements and is proportional to K. Letting K_s be the number of components used in the simplified classifier and K be the number in the original classifier, an efficiency saving factor can be defined as

$$e(K_s,K) = \left(1 - \frac{2\ K_s}{K(K+1)}\right) x\ 100\% \tag{8}$$

---

[7] See books such as Duda and Hart, Pattern Classification and Scene Analyses, John Wiley & Sons, Inc., 1973.

The effectiveness of the simplified classifier, however, is good only when the off-diagonal elements are small compared to those along the diagonal, i.e. when the components have little correlation. Therefore, the purpose of the last few classification exercises is to study an algorithm's efficiency versus its effectiveness on principal component data.

Generation of Principal Component Data. The output from the Laws procedure is used as input to generate principal component data. The principal component data is used in the second part of the experiment to study methods of component reduction.

The generation of these components is accomplished by an orthogonal transformation $\bar{x}_e = A^T \bar{x}$, and it has the effect of rotating the original components "X" into a coordinate system, which is aligned in a direction that contains most of the data's energy (variance) in a smaller number of dimensions. It also has the effect that the transformed components are uncorrelated. These effects should allow equivalent classification results with a far less number of computations than that required from the higher dimensioned correlated data.

The matrix A consists of the column vectors $(a_1, a_2, \ldots a_n)$, which are eigenvectors of another matrix T. Component reduction Methods 2 and 3 use a matrix T=S to construct the matrix of column vectors A, where S is a matrix containing the covariance estimates for the data set population (the data set is taken to represent a scene as a whole and not individual classes). Component reduction Method 4 uses a matrix T=R to construct A, where R is a matrix containing the correlation estimates for the data. Whereas defining T as T=S will generate a set of principal components without considering the variation of scale in the original data, letting T=R will consider this variation and produce a set of standardized principal components. The principal components resulting from these methods are in general not the same, and the testing of the standardized components in Method 4 is made to check whether their use would improve classification accuracy. For a continued discussion of principal components, see appendix B.

PART I OF EXPERIMENT

Numerical Results. Classification exercises are performed on scenes A, B, C, E, and H. For each scene, the test areas are used to train the classifier; the training is then followed by a classification on these same test areas. Table 1 shows the results as a confusion matrix for each scene. These matrices show the percentage of class labels assigned to each training area. The format of these results is consistent with that of previous ETL reports so that the Laws texture measure can be compared to the other textures considered by CSL--namely the Max-Min texture measure and the two-component Ad-Hoc measure.[8,9] Table 2 gives a comparison between Laws texture and these other two measures.

Classification Results. The confusion matrices shown in table 1 offer a good way to rate the decision errors made during the classification runs. Two types of error associated with making a decision, type I and type II, are shown by these matrices. For each of the class labels, a type I error is made where the classifier rejects the label when it is a correct choice, and a type II error is made where the classifier accepts the label when it is wrong. Summing the non-diagonal elements across a row "I" gives the type I error for C_i; whereas summing the non-diagonal elements down across "I" gives the type II error for C_i. Note that although the rows of the matrices are normalized to 100 percent, the columns are not. This is because the elements for the column sum should really be weighted according to the number of points in the corresponding training areas. However, the qualitative nature of the discussion does not really require this; it should be enough to merely scan the columns for type II errors.

[8] M. A. Crombie, N. J. Friend, and R. S. Rand, Feature Component Reduction through Divergence Analysis, U.S. Army Engineer Topographic Laboratories, Fort Belvoir, VA., ETL-0305, October 1982, AD-A123 474.

[9] M. A. Crombie, R. S. Rand, and N. J. Friend, An Analysis of the Max-Min Texture Measure, U.S. Army Engineer Topographic Laboratories, Fort Belvoir, VA., ETL-0280, January 1982, AD-A116 768.

The percentages of class assignments in each training area are shown in table 1.

Table 1. Autoclassification results

	Test Area	Class 1	2	3	4	5	6
	1	84.0	4.9	7.4	0.0	0.0	3.7
	2	1.4	70.8	4.2	8.3	12.5	2.8
Scene A	3	11.5	1.0	58.7	4.8	7.7	16.3
	4	4.8	4.8	6.0	73.8	2.4	8.3
	5	1.6	9.5	4.8	3.2	79.4	1.6
	6	8.3	1.9	11.1	1.9	2.8	74.1
	1	74.4	12.4	0.8	0.8	0.8	10.7
	2	13.1	74.7	3.0	0.0	2.0	7.1
Scene B	3	4.9	7.4	72.8	2.5	3.7	8.6
	4	0.0	1.3	2.6	79.2	16.9	0.0
	5	0.0	4.3	12.0	10.3	71.8	1.7
	6	7.4	14.8	1.2	0.0	1.2	75.3
	1	79.5	11.1	8.5	0.9	0.0	
	2	14.5	72.7	11.5	0.0	0.6	
Scene C	3	4.8	6.7	84.6	2.9	1.0	
	4	0.0	0.0	4.8	77.4	17.9	
	5	1.0	1.0	3.8	11.5	82.7	
	1	92.6	4.9	2.5	0.0	0.0	0.0
	2	9.1	86.9	3.0	1.0	0.0	0.0
Scene E	3	2.6	0.7	77.8	10.5	1.3	7.2
	4	0.0	0.0	4.9	86.4	3.7	4.9
	5	0.0	0.0	0.0	0.0	100.0	0.0
	6	0.0	0.0	6.2	6.2	1.2	86.4
	1	66.9	14.0	14.9	4.1		
Scene H	2	13.1	78.0	8.9	0.0		
	3	3.9	5.9	79.1	11.1		
	4	0.0	1.1	4.4	94.4		

A comparison of the Laws texture measure, the Max-Min texture measure, and the two-component Ad-Hoc measure is shown in table 2.

Table 2. Comparison of laws texture with alternate texture measures

Scene	Texture Measure	Class (percentage of correct hits)					
		1	2	3	4	5	6
A	Laws	84	71	59	74	79	74
	Max-Min	68	89	61	95	82	60
	2-Component	63	81	90	99	78	90
B	Laws	74	75	73	79	72	75
	Max-Min	97	59	79	95	75	64
	2-Component	78	34	42	96	86	58
C	Laws	79	73	85	77	83	
	Max-Min	96	78	77	88	89	
	2-Component	-	-	-	-	-	
E	Laws	93	87	78	86	100	86
	Max-Min	96	89	80	80	90	96
	2-Component	-	-	-	-	-	-
H	Laws	67	78	79	94		
	Max-Min	69	86	81	90		
	2-Component	-	-	-	-		

The error rate for each of the class labels differs considerably over the five scenes. The type I errors were fairly high and ranged from 14 to 28 percent for forest-type classes, 7 to 33 percent for field-type classes, 14 to 25 percent for scrub-type classes, and 0 to 27 percent for buildings/roads-type classes. Forest classes were confused mostly with other forests and with scrubs. Field classes were confused mostly with other fields, buildings/roads (on Scene B), and with scrubs (Scene H). Scrub types were confused mostly with other forests and with buildings/roads (Scene H). Depending on the scene, the buildings/roads class was confused with either forests, fields, or scrubs. The type II errors were also fairly high and showed similar confusion.

One exception to the hign error rate occurs for the buildings/roads class in Scene E and Scene H. Scene E shows a particularly good response from the classifier. The classifier labels all 30 points in the buildings/roads training area correctly (0 percent type I error) and maintains a very low type II error rate. Scene H also shows a good response with 5 percent incorrect labeling, although the type II error, labeling 11 percent of the scrub area as buildings/roads, is somewhat high. Observing that Scene E has a very tightly controlled training group compared to the other scenes (particularly Scene B) and noticing the classifier's excellent response in this area, as well as its good response on Scene H, an important conclusion can be made--Laws texture is an excellent descriptor to identify a buildings/roads class.

A comparison of three texture measures listed in table 2 shows that Laws texture is more effective on the buildings/roads class than the others. However, except for this, Laws is not as effective. Generally, the type I errors are lower using the other measures. A comparison of the confusion matrix results between the Max-Min texture and the Laws texture (see earlier report for Max-Min results[10]) also shows type II errors are lower when using the Max-Min texture.

[10] M. A. Crombie, R. S. Rand, and N. J. Friend, An Analysis of the Max-Min Texture Measure, U.S. Army Engineer Topographic Laboratories, Fort Belvoir, VA, ETL-0280, January 1982, AD-A116 768.

PART II OF EXPERIMENT

Numerical Results. As mentioned earlier, four methods of component reduction are attempted. Scenes A, B, and E are used in this part of the study. In beginning the study, covariance results for both the raw component data and the principal component data, as well as histograms for the first and second principal components, are reviewed. This information is shown for Scene B in tables 3 and 4 and figure 7. The covariance results are reviewed to check the relationship between texture components. The histograms are reviewed to check the validity of the Gaussian assumptions made in subsequent models.

Following this, component reduction is attempted using four different methods, which resulted--for most of the scenes--in four different sets of eight components. The order of the components in these sets is given in tables 5 and 6. This order for Method 1 is shown in table 6a as a function of the original raw components where, for example, on Scene B the divergence has placed the 14th raw component in the 1st position, the 5th raw component in the 2nd position, etc. The order of components in Method 2 is the same order in which they were produced, according to variance, and table 5 shows the percentage of accumulated variance. For example, in Scene B, 86 percent of the variance occurred in the 1st principal component, and 90 percent of the variance occurred in the 1st and 2nd principal components. The new order of the components for Method 3 is shown in table 6b as a function of the principal components generated by Method 2. The order corresponding to Method 4 is the same as that for Method 2, and since the energy compression results were almost identical to those in table 5, the results are not shown.

After a set of components is defined and ordered according to one of the four methods of reduction, the divergence is used to measure the discriminatory power between classes as a function of the number of added components. Table 7 shows the results of the divergence exercise on Scene E in detail. Similar exercises were done on Scene A and Scene B. This table and the corresponding tables for Scene A and Scene B (tables for A and B are not shown) are simplified in table 8.

Table 3. Covariance results for untransformed components, Scene B

K	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Average
K_{11}	.0002	.0002	.0010	.0007	.0011	.0001	.0006
K_{22}	.0003	.0003	.0006	.0004	.0007	.0002	.0004
K_{33}	.0001	.0001	.0003	.0002	.0004	.0001	.0002
K_{44}	.0000	.0001	.0001	.0001	.0002	.0000	.0001
K_{12}	.0002	.0003	.0007	.0004	.0008	.0001	.0004
K_{13}	.0001	.0002	.0005	.0003	.0006	.0001	.0003
K_{23}	.0001	.0002	.0004	.0003	.0005	.0001	.0003
K_{14}	.0001	.0001	.0003	.0002	.0004	.0001	.0002
K_{24}	.0001	.0001	.0003	.0002	.0004	.0000	.0002
K_{34}	.0001	.0000	.0002	.0001	.0003	.0000	.0001

Table 4. Covariance results for principal components, scene B

K	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Average
K_{11}	.01407	.01813	.05030	.03378	.05761	.01067	.03077
K_{22}	.00082	.00147	.00149	.00206	.00163	.00117	.00144
K_{33}	.00092	.00189	.00127	.00149	.00143	.00079	.00130
K_{44}	.00063	.00086	.00104	.00216	.00190	.00058	.00120
K_{12}	-.00004	.00005	-.00396	.00073	.00295	.00027	.00000
K_{13}	.00125	.00354	-.00202	-.00177	-.00208	.00108	.00000
K_{23}	-.00007	-.00025	-.00010	.00002	.00018	.00022	.00000
K_{14}	-.00008	-.00061	-.00028	.00062	.00097	-.00062	.00000
K_{24}	-.00024	-.00060	-.00038	.00081	.00069	-.00028	.00000
K_{34}	-.00009	-.00026	.00039	.00003	.00012	-.00019	.00000

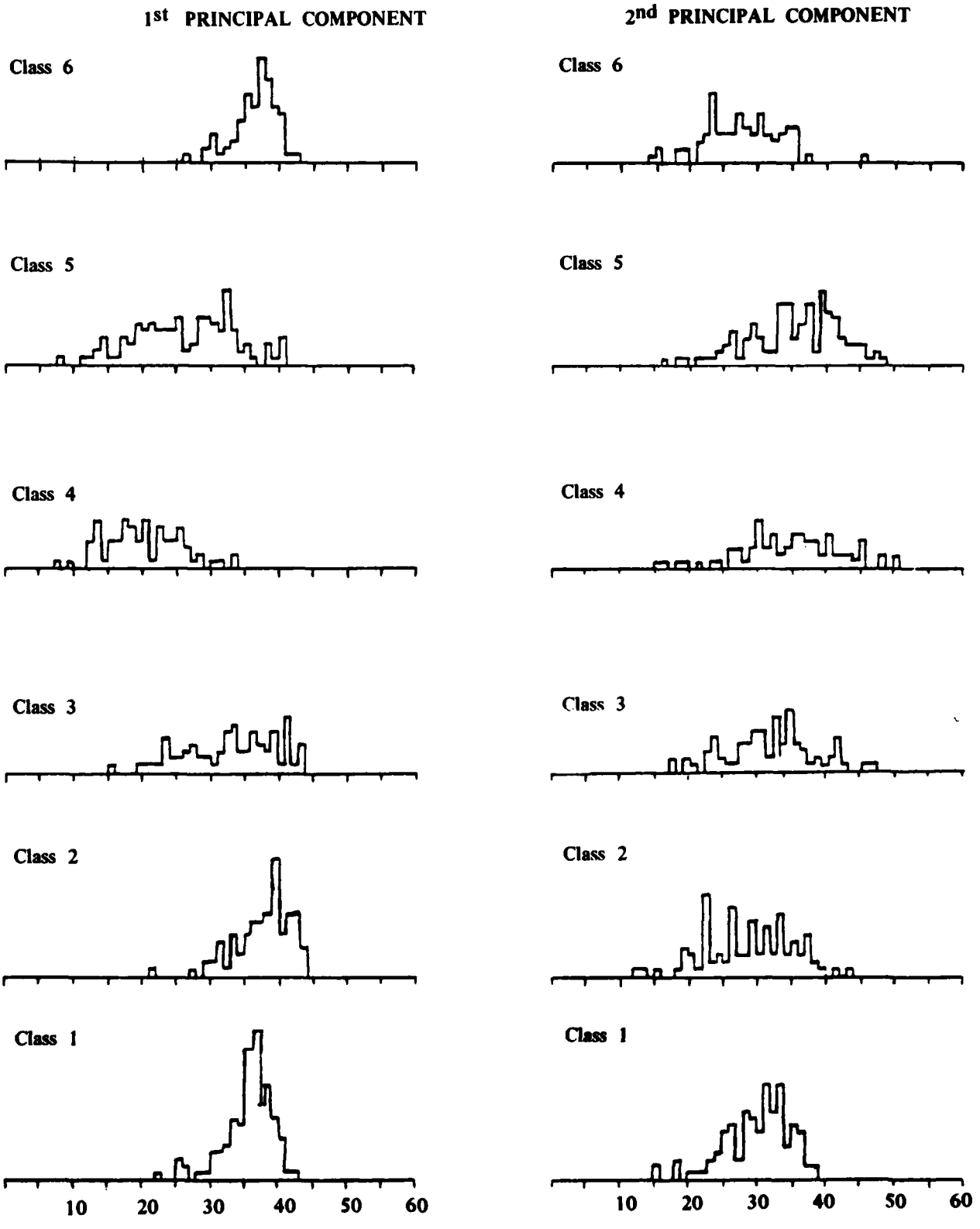


FIGURE 7. Histograms for the training areas in Scene B.

Table 5. Energy compression for principal components

Method 2: The percentage of variance that is accumulated as principal components are added is tabulated below.

	Principal Components							
Scene	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
A	86	91	94	96	98	99	99	99
B	86	90	94	97	99	99	99	100
E	81	87	92	96	98	98	99	99

Table 6. Component selection using divergence

Eight components are arranged in order of importance using the divergence measure as a criteria.

a. Method 1: Untransformed components are arranged in order according to divergence.

	Order							
Scene	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
A	12	2	15	4	3	1	7	11
B	14	5	6	10	9	3	15	7
E	3	4	7	15	1	2	10	9

b. Method 3: Transformed components are arranged in order according to divergence.

	Order							
Scene	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
A	1	11	2	4	5	7	12	9
B	1	3	7	8	5	4	6	2
E	1	4	6	5	14	15	2	3

Table 7. Divergence results during component reduction, scene E

The divergence measure for each class pair as a function of the number of components tabulated is below.

a. Untransformed components were selected using divergence as the criteria.

Number of Components	Class Pairs														
	1-2	1-3	1-4	1-5	1-6	2-3	2-4	2-5	2-6	3-4	3-5	3-6	4-5	4-6	5-6
1	.06	8.5	25	29	30	9.6	28	33	33	2.7	6.3	1.7	1.2	.73	4.3
2	.82	12	39	31	39	11	37	33	38	3.4	8.5	2.4	1.8	2.4	8.0
3	1.2	16	46	77	48	12	41	79	43	4.0	18	3.0	4.7	2.5	12
4	1.5	16	46	87	51	14	43	100	50	4.2	25	4.1	10	3.2	15
5	1.7	18	55	110	53	16	52	130	52	4.7	27	4.4	12	3.7	17
6	2.0	19	57	170	57	17	54	200	59	5.6	43	6.4	15	4.7	24
7	2.3	20	63	190	60	18	65	230	61	7.3	48	6.9	16	6.3	29
8	3.3	22	68	250	65	20	71	280	64	9.0	53	7.1	19	7.3	33
15	9.1	35	110	610	87	29	110	620	87	17	150	13	60	16	120

b. Transformed components were selected using variance as the criteria.

Number of Components	Class Pairs														
	1-2	1-3	1-4	1-5	1-6	2-3	2-4	2-5	2-6	3-4	3-5	3-6	4-5	4-6	5-6
1	.16	5.1	14	26	15	7.7	20	33	21	2.3	8.2	1.0	1.9	.92	6.2
2	.71	9.0	21	30	27	8.7	22	38	30	2.7	9.9	2.4	3.6	1.3	10
3	.86	13	30	45	31	13	31	52	34	3.0	11	2.5	4.3	1.6	11
4	1.2	18	50	58	44	16	51	67	45	4.3	13	3.1	5.2	3.5	12
5	1.6	18	53	71	46	17	53	92	48	4.6	21	3.6	10	3.7	17
6	1.7	19	57	110	50	19	57	160	51	6.0	35	4.1	13	4.6	24
7	2.1	21	60	130	53	19	59	170	53	6.6	38	4.5	16	5.0	30
8	2.7	23	64	140	58	20	62	180	59	7.2	42	6.8	17	6.4	31
15	9.1	35	110	610	87	29	110	620	87	17	150	13	60	16	120

c. Transformed components were selected using divergence as the criteria.

Number of Components	Class Pairs 1-2	1-3	1-4	1-5	1-6	2-3	2-4	2-5	2-6	3-4	3-5	3-6	4-5	4-6	5-6
1	.16	5.1	14	26	15	7.7	20	33	21	2.3	8.2	1.0	1.9	.92	6.2
2	.36	11	25	34	31	12	30	41	34	3.0	9.2	1.8	2.1	2.7	7.4
3	.42	12	32	80	33	13	40	110	38	4.0	23	2.2	4.4	2.9	14
4	.77	13	34	95	35	14	42	140	41	4.2	32	2.6	9.8	3.1	19
5	1.2	13	38	120	36	15	48	180	43	5.0	44	2.8	12	3.6	25
6	2.4	14	41	190	37	15	50	250	48	5.3	66	3.7	18	4.0	44
7	3.4	18	48	200	47	17	53	260	53	6.0	71	4.9	22	4.5	51
8	3.8	21	68	240	55	21	67	300	58	7.4	77	5.6	25	5.8	56
15	9.1	35	110	610	87	29	110	620	87	17	150	13	60	16	120

Table 8. Summary of divergence results

The mean "$\bar{D}$" and standard deviation of divergence values over the C(6,2) = 10 class combinations "S_D" are listed below for three scenes (A,B,E) and four different methods (1,2,3,4).

a. Divergence results for four components:

Scene		Method Used To Generate Components 1	2	3	4
A	$\bar{D}$	7	6	7	-
	S_D	5	4	5	-
B	$\bar{D}$	17	15	15	16
	S_D	16	14	15	15
E	$\bar{D}$	31	26	32	26
	S_D	31	23	38	24

b. Divergence results for eight components:

Scene		Method Used To Generate Components 1	2	3	4
A	$\bar{D}$	13	12	13	-
	S_D	9	8	9	-
B	$\bar{D}$	25	26	26	26
	S_D	21	20	20	21
E	$\bar{D}$	65	48	67	43
	S_D	85	51	87	40

After constructing the sets of reduced components and estimating their effectiveness using divergence, classification exercises are performed similar to those done in Part I. Each scene consists of six test areas, which are first used to train the classifier as a six-class segmentor and then used as the data to be classified. The scenes are classified using 2, 4, 8, and 15 components for each of the three methods (Method 4 is not included), and the results are shown in table 9.

As an additional part of the analysis, the results were allowed to improve by combining similar class types. The confusion matrices generated by the above classifications (like those in Part I, but are not shown) are used to combine similar forest types and to combine similar field types. This combination process allows the effectiveness of the various component sets to be studied under less strict and possibly more realistic requirements. Table 10 lists the test regions that are combined into new classes, and table 11 shows the classification results after such a combination process is applied. A final classification exercise is made that uses a simplified Bayes classifier. This exercise is motivated by the observation that although principal components are in theory uncorrelated, in practice they are only nearly uncorrelated, as is shown in table 4. However, it should still be possible to eliminate the off-diagonal components in the covariance computations with little effect on classification accuracy. Therefore, the purpose of this last exercise is to look at this effect on accuracy when such computations are eliminated.

Table 9. Classification results during component reduction

The percentage of correct hits in each training area as a function of the number of components is tabulated for three methods of component reduction.

Scene A

Class	METHOD 1				METHOD 2				METHOD 3			
	Number of Components				Number of Components				Number of Components			
	2	4	8	15	2	4	8	15	2	4	8	15
1	70	72	76	84	68	68	78	84	70	69	75	84
2	50	49	62	71	51	57	57	71	32	47	58	71
3	22	29	50	59	23	35	49	59	8	29	36	59
4	14	62	61	74	27	43	64	74	36	45	64	74
5	38	51	62	79	33	37	57	79	43	36	62	79
6	68	53	58	74	61	60	60	74	69	57	57	74

Scene B

Class	METHOD 1				METHOD 2				METHOD 3			
	Number of Components				Number of Components				Number of Components			
	2	4	8	15	2	4	8	15	2	4	8	15
1	33	57	63	74	50	65	65	74	62	52	65	74
2	39	38	50	75	25	39	58	75	17	47	58	75
3	37	43	51	73	5	47	59	73	42	47	59	73
4	78	70	73	79	75	74	62	79	78	77	62	79
5	17	32	55	72	30	40	53	72	20	28	53	72
6	60	59	60	75	58	68	57	75	68	53	57	75

Scene E

Class	METHOD 1 Number of Components				METHOD 2 Number of Components				METHOD 3 Number of Components			
	2	4	8	15	2	4	8	15	2	4	8	15
1	62	72	76	93	43	68	74	93	47	57	86	93
2	70	66	78	87	73	73	76	87	69	59	77	87
3	50	52	62	78	56	57	57	78	46	52	64	78
4	52	60	80	86	44	59	70	86	49	58	74	86
5	83	90	100	100	80	83	100	100	80	93	100	100
6	56	67	81	86	52	67	73	86	64	72	78	86

Component Reduction Techniques. The covariance results in tables 3 and 4 show immediately an advantage to utilizing a principal component representation. Because the off-diagonal covariance elements of the original representation are of the same order of magnitude as the diagonal elements, indicating that the original components are highly correlated, the off-diagonal elements of the principal component representation are of a lesser order of magnitude. The smaller the off-diagonal elements, the smaller is the error introduced by a simpler classifier, which does not take the covariance into account. In the Bayes classifier, for example, a considerable amount of computation can be eliminated by using only the diagonal elements of the covariance matrices "S_i", and making use of the approximation might be worthwhile if the corresponding errors were small. The results of some classification exercises that make this approximation and show its effect on accuracy are discussed later in this section.

Note that although the off-diagonal covariance elements of the original representation are reduced in going to principal components, they are not eliminated, even though the covariance of the test scenes were diagonalized. This is easily explained. Although the data for a particular test scene as a whole is uncorrelated, a subset of this data need not be. This explanation is shown to be true by averaging these non-zero elements over the six class samples, as is done in the last column of table 4.

The histograms for the raw components and the principal components, taken over each of the classes, show the distribution to be, at best, only approximately normal. The effect of this situation is that neither the divergence measure nor the classifier used in this experiment is optimal. Even if the populations corresponding to the class samples are normally distributed (which is not known), the corresponding mean and covariance estimates for these populations are not very good. Therefore, at least some of the loss in the separability of classes and in classification accuracy must be attributed to this less-than-optimal situation. One example of the histograms that is typical of both the original and the principal components is shown for the first and second principal components of Scene B in figure 7.

The selection and ordering of the new components by either Method 1 or Method 3, shown in table 5, are found to be different over the three scenes. This difference shows that the methods of ordering by divergence are scene dependent. Neither Method 1 nor Method 3 can establish one set of components that would be valid over all the scenes. The reason is that the discriminating ability of each component is dependent on the class-pairs that are being measured, and a number of these class-pairs are different over the three scenes.

The selection and ordering of components by Method 2 and Method 4 are the same. These components are selected in the order in which they are generated, an order that maximizes the accumulated energy (variance). The Energy Compression results, listed in table 5, show that at least 81 percent of the data's variance is found in the first component and that at least 96 percent of the variance is found in the first four components. These results are also consistent over the three scenes. The Energy Compression results of Method 4 are almost identical to that of Method 2; therefore, these results are not shown.

The divergence results for some of the exercises are presented in tables 7 and 8. These tables show that as a whole--there is no clear advantage of using one method or another. Also, the divergence values as a function of the number of components converge slowly to the final values of the 15-component vectors, regardless of method. However, focusing attention on Scene E and on the buildings/roads class--the area of strongest performance for Laws--shows a different trend. Although the values do not converge any faster to the final values, they do approach a more acceptable value within a few components. For this class, Method 3 shows a definite advantage over the other methods. Averaging values for four components over class-pairs containing buildings/roads results in divergence values of 47, 31, and 59 for Methods 1, 2, and 3, respectively. Averaging values for eight components over these class-pairs gives the values 127, 82, and 148. Note that in table 8, Method 4 shows no advantage over the other three methods. Because of the increased number of computations in the method and its lack of significant improvement over the others, Method 4 is dropped from consideration at this point in the experiment.

Tables 9 and 11 present the classification results (percentage of correct hits) for the three scenes as a function of the number of components. In table 9 few of the results of the reduced sets converge acceptably to those of the 15-component vectors. If some of the similar classes--such as forest-type classes or field-type classes--are combined according to the definitions in table 10, classification accuracies are improved, and many of the reduced component sets become acceptable. Table 11 shows that most of the eight-component sets converge acceptably and that a good number of the four-component sets are also acceptable. The worst exception to these cases is the poorly defined buildings/roads class in Scene B.

The classification exercises using the simplied Bayes classifier (see section entitled Classification Algorithm) demonstrate that the classifier is both effective and efficient on certain component sets. Comparing the results in tables 12 and 13 with those in tables 9 and 11, there are a number of cases where there is savings in efficiency with little effect on accuracy. For example, comparing the eight-component results of the simplified classifier using the data base created by Method 2 with the four-component results of the unsimplified classifier using the same data, the classifications are approximately equivalent; however, the factor $e(K_8,K)$ with $K_8=8$ and $K=4$ shows that the simpler classifier is 20 percent more efficient using twice as many components.

Table 10. Combination of training regions

A new set of classes, which require less discriminatory power, can be defined by combining training regions. Two such sets were defined for scenes A, B, and E; the new features, the regions used in the new combinations, and the number of points in the resulting regions are listed below.

Scene A

Class Label	New Features	Old Regions	# of Points
1	Building and Road	1	81
2	Forest	4,6	192
3	Field	2,3,5	239

Scene B

Class Label	New Features	Old Regions	# of Points
1	Building and Road	3	81
2	Forest	1,2,6	301
3	Field	4,5	194

Scene E

Class Label	New Features	Old Regions	# of Points
1	Building and Road	5	30
2	Forest	3,6	234
3	Field	1,2	180
4	Scrub	4	81

Table 11. Component reduction results for combined classes

The classification results were combined according to the new set of classes listed in table 10. The percentage of correct hits of each new class for the three methods of component reduction on Scenes A, B, and E are shown below.

Scene A

Class	METHOD 1 Number of Components				METHOD 2 Number of Components				METHOD 3 Number of Components			
	2	4	8	15	2	4	8	15	2	4	8	15
1	70	72	76	84	68	68	78	84	70	69	75	84
2	66	77	71	79	68	71	73	79	75	71	73	79
3	67	69	77	80	68	71	74	80	57	69	70	80

Scene B

Class	METHOD 1 Number of Components				METHOD 2 Number of Components				METHOD 3 Number of Components			
	2	4	8	15	2	4	8	15	2	4	8	15
1	37	43	51	73	5	47	59	73	42	47	59	73
2	95	94	94	97	90	95	96	97	95	93	96	97
3	74	70	81	88	83	84	84	88	77	77	84	88

Scene E

Class	METHOD 1 Number of Components				METHOD 2 Number of Components				METHOD 3 Number of Components			
	2	4	8	15	2	4	8	15	2	4	8	15
1	83	90	100	100	80	83	100	100	80	93	100	100
2	75	76	80	88	74	81	79	88	80	82	83	88
3	93	96	97	97	87	98	98	97	94	92	98	97
4	53	61	80	86	44	59	70	86	49	58	74	86

Table 12. Results for simplified bayes classifier, scene B

The components generated from Method 1 and Method 2 were processed in a Bayes classifier that used only the diagonal elements of the class covariance matrices. The percentage of correct hits in each training area as a function of the number of components is tabulated below for Scene B.

	METHOD 1				METHOD 2			
Class	Number of Components				Number of Components			
	2	4	8	15	2	4	8	15
1	39	42	33	25	52	65	65	62
2	20	22	18	23	26	19	37	45
3	10	9	4	7	7	49	51	48
4	75	74	78	77	77	70	71	73
5	9	21	18	20	30	36	43	43
6	47	69	74	72	54	68	60	65

The classification results shown above were combined according to the new set of classes listed in table 10, and these combined results are shown below.

	METHOD 1				METHOD 2			
Class	Number of Components				Number of Components			
	2	4	8	15	2	4	8	15
1	10	9	4	7	7	49	51	48
2	85	89	99	90	89	91	95	95
3	66	76	76	76	82	82	81	79

Table 13. Results for simplified bayes classifier, Scene E

The components generated from Method 1 and Method 2 were processed in a Bayes classifier that used only the diagonal elements of the class covariance matrices. The percentage of correct hits in each training area as a function of the number of components is tabulated below for Scene E.

	METHOD 1				METHOD 2			
Class	Number of Components				Number of Components			
	2	4	8	15	2	4	8	15
1	10	27	32	32	43	59	72	81
2	50	68	67	64	72	73	71	76
3	46	51	49	52	57	61	61	65
4	11	41	39	51	44	56	58	57
5	63	73	77	80	77	80	90	90
6	59	64	56	54	47	60	69	70

The classification results shown above were combined according to the new set of classes listed in table 10 and these combined results are shown below.

	METHOD 1				METHOD 2			
Class	Number of Components				Number of Components			
	2	4	8	15	2	4	8	15
1	63	73	77	80	77	80	90	90
2	79	82	77	78	71	81	79	78
3	60	79	82	81	86	97	97	96
4	11	41	40	51	44	56	58	57

DISCUSSION

Considering the simplicity and the effectiveness of the previously studied two-component Ad-Hoc image descriptor and the lack of overall improvement in performance of Laws texture measure, the use of the Laws measure by itself as an image descriptor is not feasible. Laws texture measure offers a clear advantage for only one class, the buildings/roads class. However, because of the importance of targeting this class, the Laws texture should not be dismissed. A possible alternative is to combine the two-component Ad-Hoc measure with a reduced set of Laws texture components.

Component reduction Method 3 (arranging a set of principal components in an order that maximizes the divergence) is the most effective technique for defining a reduced set of components, if not for all the class-pairs, at least for the class-pairs containing the buildings/roads class. The ordered component sets constructed by this method are scene dependent, but this is possibly due to some of the different classes that are used on the scenes. If instead of maximizing the divergence over all the class-pairs, the divergence is maximized over only the class-pairs common to all the scenes, the component sets might be made scene independent. However, considering the importance of the buildings/roads class in determining the feasibility of Laws texture, another alternative may be more advantageous. Since this texture measure does not offer any advantage in discriminating class-pairs other than the ones containing buildings/roads, why not restrict the process to maximize the divergence on only those class-pairs. This restriction would improve discrimination between such class-pairs at the expense of the others, but the additional use of other simpler texture measures--such as the two-component Ad-Hoc measure--might easily replace this loss.

The computational handicap of increasing the number of components by adding two or more texture measures might not be worth the expense. However, as shown by using the simplified classifier on uncorrelated data, this handicap might not be as bad as it might at first seem. If the additional components can be computed quickly and if they are close to being uncorrelated, they may be worth processing in this or some other simple classifier.

CONCLUSIONS

1. The Laws texture measure is an excellent descriptor for separating class-pairs that contain the buildings/roads class; however, it is lacking for most other class-pairs.

2. The use of the Laws texture measure by itself as an image descriptor is not feasible when recognizing the simplicity and effectiveness of the two-component Ad-Hoc image descriptor considered in an earlier study.

3. A component reduction method that transforms the original components into principal components and then arranges the new components in an order that maximizes the divergence is the most effective of the four proposed methods.

4. The use of many uncorrelated components in a simplified classifier can be just as effective, but more efficient, than a lesser number of correlated components in a standard classifier.

APPENDIX A. Laws Texture Data

The authors generated sets of texture data by using the Laws technique that was developed in 1979. This method combines both spatial and statistical pixel information; spatial information being a pixel's relation to its immediate neighbors and statistical information being a comparison of a pixel to statistics compiled over a large area of the image. The technique involves three steps:

1. A local convolution for spatial information.
2. A standard deviation computation for statistical information.
3. Normalization.

Step 1. Sixteen 5-by-5 convolution "masks" or windows were moved across the image. These windows resulted from all cross-product combinations of the following four five-component vectors:

```
L: ( 1   4   6  4   1)
E: (-1  -2   0  2   1)
S: (-1   0   2  0  -1)
R: ( 1  -4   6 -4   1)
```

The letters stand for level, edge, spot, and ripple; and their vectors performed the best out of a series of such vectors for Laws, also outstripping three-component and seven-component alternatives. Multiplying one vector by a transpose of another (or the same) vector produces the sixteen 5-by-5 windows. When moved across all possible pixels on an M x N image, an (M-4) by (N-4) convolution image results; 16 windows produce 16 convolutions.

Step 2. A standard deviation image was created from each convolution using a 15 by 15 window of points surrounding each pixel. Each (M-4) by (N-4) convolution image becomes an (M-18) by (N-18) standard deviation image. These images measure the texture energy of each of the 16 convolution windows.

Step 3. The LxL^T window was used for normalization since its standard deviation values will be larger than any of the other 15 planes. Each of the other 15 planes was divided by the "LL" plane, resulting in 15 texture energy planes with values between 0 and 1.

As a final step, the data was converted from separate planes (containing one component for each pixel) into a data set containing multiple components for each pixel. If all 15 planes are used, each pixel will become a 15-component vector; if less are used, the number of components will also be reduced.[11]

[11] Kenneth Ivan Laws, Texture Image Segmentation, Image Processing Institute, University of Southern California, Los Angeles, CA 90007 USCIPI Report 940, January 1980.

APPENDIX B. Principal Component Data

The objective of using principal components is to make an attempt at decreasing the number of components necessary to describe image points without eliminating essential information. Letting each vector describing an image point be composed of N components, this method attempts to select a new coordinate system in which M components--where M is substantially less than N--can be used with little loss in descriptive information. If the new components are then used, for example, in a classification exercise, it should be possible to eliminate the remaining (N-M) components without causing objectionable error in classification.

In order to eliminate (N-M) components and simultaneously optimize the amount of retained information, it is necessary to construct an orthogonal transform using some error criterion, usually the mean-square error criterion. The transformation matrix "A" that accomplishes this consists of columns composed of eigenvectors of another matrix "K_X". The matrix K_X can be either the covariance matrix of the image ($K_X = S$) or the correlation matrix of the image ($K_X = R$). If the matrix $K_X = S$ is used, the transformation A performs a diagonalization of the covariance matrix, such that the covariance matrix of the transformed image, $Ky = A^T K_X A$, is a diagonal matrix where elements are eigenvalues of K_X arranged in descending order. If a component is deleted, then the mean-square error increases by a value proportional to the corresponding eigenvalue. Thus, the set of M components with the largest eigenvalues should be selected and the remaining (N-M) components discarded.

In order to get a better understanding of the physical meaning of the matrices K_X and A, consider the following. Let $M=K_X$ be an N by N matrix operator that transforms the N-dimensional vector X into the N-dimensional vector Y through the equation

$$Y = M\,x \tag{B1}$$

Let the matrix A be an operator that transforms the coordinates of one system ($X_1, X_2, \cdots X_N$) to the coordinates of another system ($X_{e_1}, X_{e_2}, \cdots X_{e_n}$) through the equation

$$\hat{X}_i = A\,\hat{X}_{e_i} \tag{B2}$$

Figure B1 shows a two-dimensional example of such an operator A, when A is an orthogonal matrix ($A^T = A^{-1}$) that rotates the axis X_{e_1}, X_{e_2} by an angle θ from ($\hat{X}_1, \hat{X}_2$) such that the axis $\hat{X}_{e_1}$ is aligned in the direction of maximum variance. Substituting equation B2 into equation B1 for any i = 1,N gives $AY_e = MAX_e$. Solving for Y_e, with $K_X=M$, gives the result

$$Y_e = A^T K_X A X_e$$

$K_e = A^T K_x A$ can be shown to be diagonal if A is a matrix whose columns are the unique eigenvector solution of the matrix K_x.

If the matrix K_x is a matrix containing the covariance estimates of an image, that is K_x=S, then the matrix K_y is a diagonal matrix of eigenvalues whose values are the variances of the new transformed vectors ($X_e = A^T X$). Since the covariance elements of K_y are zero, the components are uncorrelated.

Reference to the method of principal components can be made in a number of books on statistical pattern recognition.[12]

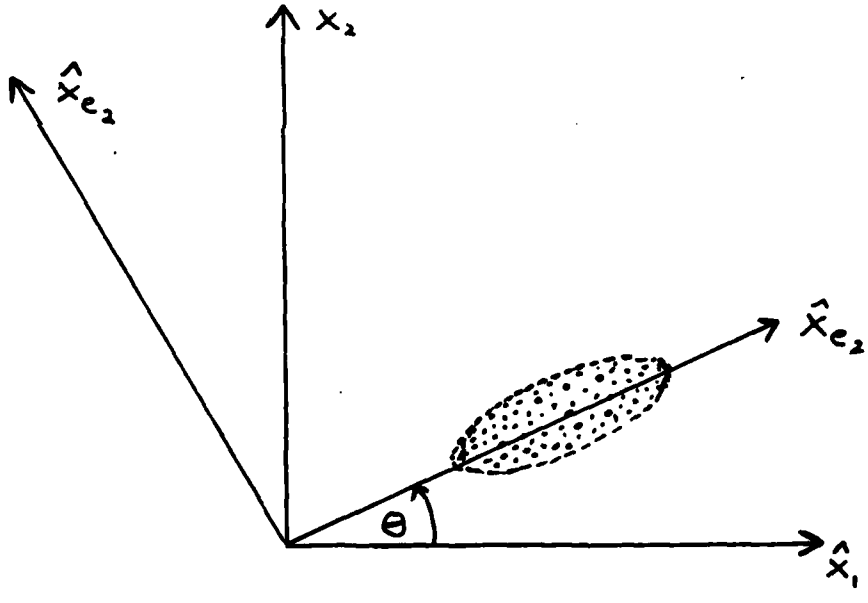
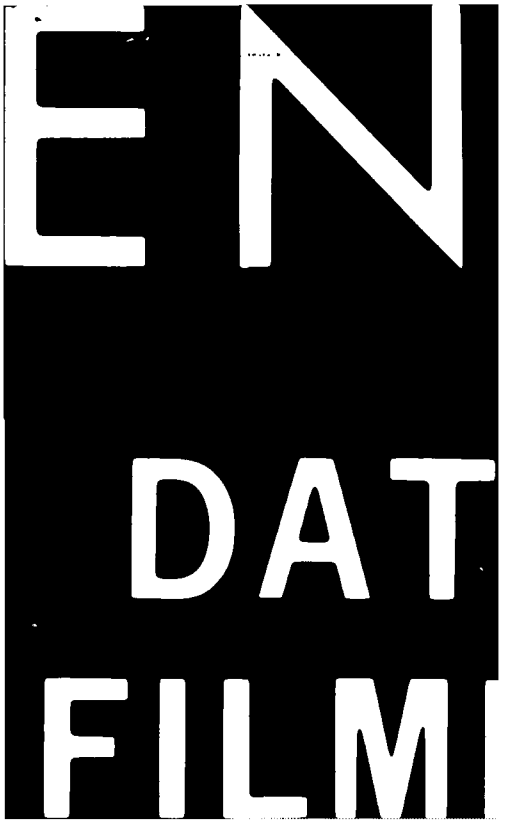


Figure B1. Alignment of Principal Component Data.

[12] One such example: Harry Andrews, Introduction to Mathematical Techniques in Pattern Recognition, John Wiley & Sons, Inc., 1972.


EN
DAT
FILM