AD-A059 829 SOUTHEASTERN MASSACHUSETTS UNIV NORTH DARTMOUTH DEPT --ETC F/G 6/4
FEATURE EXTRACTION AND SIGNAL PROCESSING APPROACH TO REALTIME P--ETC(U)
AUG 78 C H CHEN AFOSR-76-2951
UNCLASSIFIED EE-78-5 AFOSR-TR-78-1279 NL

| OF |
AD
A059829

END
DATE
FILMED
12-78
DDC

TECHNICAL REPORT SERIES IN INFORMATION SCIENCES
(Dr. C. H. Chen, Principal Investigator)

SOUTHEASTERN MASSACHUSETTS UNIVERSITY · UNITATE · 1895 · FORTIOR

SOUTHEASTERN MASSACHUSETTS UNIVERSITY
ELECTRICAL ENGINEERING DEPARTMENT

NORTH DARTMOUTH, MASS. 02747 U.S.A.

(2)

Grant AFOSR-76-2951
FINAL REPORT,
TR EE-78-5

FEATURE EXTRACTION AND SIGNAL PROCESSING

APPROACH TO REALTIME PATTERN RECOGNITION.

by

C. H. Chen
(Principal Investigator of the Grant)
Department of Electrical Engineering
Southeastern Massachusetts University
North Dartmouth, Massachusetts 02747

Submitted to

Lt. Colonel George W. McKemie
Directorate of Mathematical
and Information Sciences
Air Force Office of Scientific Research (AFSC)
Bolling Air Force Base, D.C. 20332

2304 A2

AFOSR

TR-78-1279

15p.

August 1978

407 932

LB

FINAL REPORT FOR GRANT AFOSR 76-2951

Summary

The integration of feature extraction for pattern recognition and the digital signal processing into one study as performed in this project has resulted in advances in both areas, and the discovery of many new ideas which are beneficial to both areas. There are common problems, such as the finite sample size effect, in both pattern recognition and signal processing. For example, digital signal processing techniques are much needed in extracting effective features while statistical pattern recognition can be useful in image processing. More specifically, this research has carefully examined the fundamental problem of the finite sample size and its effect on feature selection and classification rules. Most effective features for seismic pattern recognition have been developed through the signal modelling study. In the image recognition work, new results include the rotationally invariant digital Laplacian operation and a new adaptive Kalman filtering technique for efficient realtime image processing. Detailed computer results have been developed and documented to support the theoretical study. Finally for image classification, the specific problem of contextual information is examined and a decision tree procedure is developed which can process both the statistical and structural features for effective classification.

FINAL REPORT FOR GRANT AFOSR 76-2951

I. Statistical Feature Extraction

The work on the distance measures has been completed. Major effort has been made on the study of finite sample size in statistical pattern recognition. Appendix I has a more complete detail on this study. Under small sample size, many existing theoretical results based on infinite sample assumption are not valid. Many analytical problems under small sample size do not have a solution yet. For example, the performance of the nearest neighbor decision rule is not available at finite sample size except for specific cases. These are difficult analytical problems. The experimental results are very much dependent on the data used, but they do give us an idea of the performance. There are certain decision rules which perform better at small sample size than the others. And sometimes the degradation in performance due to finite sample size is not significant. So the problem is important and further work is necessary.

II. Seismic Pattern Recognition

Two best sets of features for automatic seismic classification are the short-time spectral features, and the parameters of the autoregressive moving average model. The ARMA provides a fairly good spectral matching to many seismic record. The use of the AR model alone however is not adequate. It is noted that learning samples should be chosen properly as there are large within-class variations of the seismic records due to a number of reasons such as different geographical locations for various events. Although a new set of seismic data tape was provided by Seismic Data Analysis Center, the time limitation of the project would not make it possible to pursue such study.

Signal modelling appears to play an increasingly important role. This is a subject which is important to pattern recognition and signal processing.

III. Image Pattern Recognition

The modified gradient method, approximation to rotationally-invariant digital Laplacian and an adaptive Kalman filtering method are the three techniques which are theoretically sound and experimentally proven by using both the aerial reconnaissance and FLIR imagery which we have available. For the adaptive filtering method some new results are shown in Appendix II. The filter has the capability to monitor the object boundary and make proper adjustment in filter parameters. In case the transition matrix is unknown, it can be estimated by using an on-line estimation method which simultaneously estimate the parameters and states. This filtering method thus requires little or no prior knowledge to begin with and the processing is very fast and suitable for realtime needs.

The rich contextual information in images makes it necessary to extract both statistical features and the structural features. The best way to utilize both kinds of features in classification is the binary classification trees which process the features sequentially according to their ranking. The decision will be based on majority vote. For pre-designed trees the required decision time may be a fraction of a single stage classifier. Optimal tree design technique is available. We feel that the sequential decision tree is very promising for use in complex recognition systems.

IV. Bibliography - List of Publications Prepared Under the Grant

1. C. H. Chen, Editor, "Pattern Recognition and Artificial Intelligence", Academic Press, Inc. 1976.

2. C. H. Chen, "On Statistical and Structural Feature Extraction", in the book, "Pattern Recognition and Artificial Intelligence", ed. C. H. Chen; Academic Press, Inc. 1976.

3. C. H. Chen, "Seismic Pattern Recognition", in the book, "Computer-Aided Seismic Analysis and Discrimination, ed., C. H. Chen, Elsevier Scientific Publishing Company, 1978.

4. C. H. Chen, "On Digital Signal Modelling and Classification With the Teleseismic Data", Record of IEEE ASSP Society Conference, 1978.

5. C. H. Chen, "Note On A Modified Gradient Method for Image Analysis", Pattern Recognition Journal, Vol. 10, No. 4, 1978.

6. C. H. Chen and C. K. Lau, "Feature Formulation On Reconnaissance Imagery Analysis", Proceedings of the Seventh Annual Symposium on Automatic Imagery Pattern Recognition, May 1977.

7. C. H. Chen, "A Review of Statistical Pattern Recognition" submitted as a technical report of the grant; also to appear in the Proceedings of the 1978 NATO ASI on Pattern Recognition and Signal Processing edited by C. H. Chen.

8. C. H. Chen, "Introduction to Seismic Signal Processing", submitted as technical report of the grant; also presented at the 1978 NATO ASI on Pattern Recognition and Signal Processing.

9. C. H. Chen, "A New Look at the Statistical Pattern Recognition", submitted as a technical report of the grant; also presented at the IEEE Computer Society Workshop on Pattern Recognition and Artificial Intelligence, April 1978.

10. C. H. Chen, "Some New Results on Image Processing and Recognition", submitted as a technical report of the grant; also presented at the IRIA Seminar on Pattern Recognition and Image Processing, Le Chesnay, France, June 22, 1978.

11. C. H. Chen, "Performance Bounds of a Class of Sample-Based Classification Procedures", Proceedings of IEEE Computer Society Conference on Pattern Recognition and Image Processing, June 1977.

12. C. H. Chen, "A New Class of Derivative Operators for Image Recognition" submitted as a technical report of the grant, January 1978.

13. C. H. Chen, "Evaluation of Classification Rules for Imagery Recognition", Proceedings of the Eighth Annual Symposium on Automatic Imagery Pattern Recognition, April 1978.

14. C. H. Chen, Interim Scientific Report of the Grant, submitted February 1977.

15. C. K. Lau, "Study of a Class of Imagery Recognition Problems", Master's Thesis, S.M.U., Electrical Engineering Department, work supported by the grant, May 1977.

16. Che Chang Chen, "Seismic and Imagery Pattern Recognition" (tentative title), Master's Thesis, S.M.U., Electrical Engineering Department, work supported in part by the grant; expected completion date: September 1978.

17. T. Kwan, "Spectral Estimation, Modelling, and Classification Study for the Seismic Data" (tentative title), Master's Thesis, S.M.U., Electrical Engineering Department, work supported in part by the grant, expected completion date: December 1978.

Appendix I for Final Report

FINITE SAMPLE CONSIDERATIONS IN STATISTICAL PATTERN RECOGNITION

C. H. Chen

Electrical Engineering Department
Southeastern Massachusetts University
North Dartmouth, Massachusetts 02747

Abstract

Most research on statistical pattern recognition has been based on the assumption of large/infinite sample size and thus the asymptotic performance is of primary interest. In practical recognition problems the sample size is often limited and the actual performance may be quite different from that theoretically predicted. The design and evaluation of recognition systems must take into account the finite sample constraint. There has been little considerations of the finite sample effects in statistical pattern recognition because exact solutions are generally unavailable. The close relation between the dimensionality and sample size further complicates the problem. This paper is concerned mainly with the limitation of learning sample size. The finite sample effects are considered in three major problem areas: distance and information measures, classification rules and contextual analysis.

1. Introduction

A fundamental assumption frequently made in statistical pattern recognition is that the number of samples available for learning (training) or classification is large or infinite. Thus the asymptotic performance is of primary interest, based on which the recognition system is designed. In practice the sample size is limited because the samples are costly or may not be easily available. The actual performance may become quite different from that theoretically predicted. There has been little theoretical considerations of the finite sample effects in statistical pattern recognition because exact solutions are in general unavailable. The close relation between the dimensionality and sample size further complicates the problem. In this paper, the effects of finite sample size will be considered especially in the problem areas of distance and information measures, classification rules, and contextual analysis when the learning sample size is limited. This study will enhance the understanding of the fundamental behaviors of statistical recognition systems so that better systems can be designed.

2. Finite Sample Distance Measure

Distance measures are useful for feature selection and extraction and for error bounds of Bayes error probability. They have been extensively examined in recent years under the assumption of large sample size (see e.g. [1]). To determine the distance measures from a limited number of samples, a maximum likelihood estimation procedure may be used [2]. The discussion here will be limited to independent Gaussian samples for divergence and Bhattacharyya distance but can be extended to other cases.

Consider first the case of two univariate Gaussian densities with means m_1 and m_2 and the same variance σ^2 which is known. Let "^" denote the quantity evaluated by using the sample estimates. Then the difference between the estimated and known divergence is

$$\hat{J} - J = \frac{1}{\sigma^2} [(\hat{m}_1 - \hat{m}_2)^2 - (m_1 - m_2)^2] \tag{1}$$

which has the expected value

$$E(\hat{J} - J) = \frac{1}{N_1} + \frac{1}{N_2} > 0 \tag{2}$$

where N_1 and N_2 denote the numbers of samples for classes 1 and 2 respectively. The positive bias given by Eq. (2) indicates that the divergence evaluated by using a finite number of samples can lead to an over optiministic estimate of the error probability. The variance of the estimate is

$$E(\hat{J} - J)^2 = 3(\frac{1}{N_1} + \frac{1}{N_2})^2 + 4J(\frac{1}{N_1} + \frac{1}{N_2}) \tag{3}$$

which approaches zero as the sample sizes approach infinite. Thus $\hat{J}$ is a consistent estimate of J.

Next consider the univariate Gaussian densities with zero means and variances σ_1^2 and σ_2^2. The divergence based on the sample estimated parameters is

$$\hat{J} = \frac{\hat{\sigma}_1^2}{2\hat{\sigma}_2^2} + \frac{\hat{\sigma}_2^2}{2\hat{\sigma}_1^2} - 1 \tag{4}$$

The ratio $\omega = \hat{\sigma}_1^2/\hat{\sigma}_2^2$ has the F-distribution with (N_1, N_2) degrees of freedom. The expected error due to the finite sample size is

$$E(\hat{J} - J) = \frac{\sigma_1^2}{\sigma_2^2} \frac{1}{N_2 - 2} + \frac{\sigma_2^2}{\sigma_1^2} \frac{1}{N_1 - 2} \geq 0 \quad (5)$$

where the positive bias can be significant for small sample sizes. It can be shown that Eq. (4) is also a consistent estimate.

The Bhattacharyya distance based on the sample estimated parameters is

$$\hat{B} = \frac{1}{2} \log\left(\frac{\hat{\sigma}_1}{\hat{\sigma}_2} + \frac{\hat{\sigma}_2}{\hat{\sigma}_1}\right) = \frac{1}{4} \log \frac{(1+\omega)^2}{4\omega} \quad (6)$$

By using the Taylor series expansion of $\hat{B}$ with respect to the true value B, and retaining terms up to the second order in the expression, we obtain:

$$E(\hat{B} - B) = \frac{\sigma_2^2 - \sigma_1^2}{4(\sigma_2^2 + \sigma_1^2)}\left(1 - \frac{2}{N_2 - 2}\right) + \frac{\sigma_2^4 + 2\sigma_1^2\sigma_2^2 - \sigma_1^4}{8(\sigma_2^2 + \sigma_1^2)^2} \left[1 - \frac{4}{N_2 - 2} + \frac{N_2^2(N_1 + 2)}{N_1(N_2 - 2)(N_4 - 4)}\right] \quad (7)$$

which is negative for $\frac{\sigma_1^2}{\sigma_2^2} \geq 1 + \sqrt{2}$ and positive otherwise. As the sample sizes approach infinity, the bias is not zero because of the series truncation. However the sample size effect is evident from Eq. (7).

Next consider the multivariate Gaussian densities for p-dimensional measurements. Let $\bar{x}_1$ and $\bar{x}_2$ be the sample mean vectors corresponding to the true mean vectors μ_1 and μ_2 of classes 1 and 2 respectively. Also let S be the sample estimate of the common covariance matrix $\sum$ given by

$$S = \frac{1}{N_1 + N_2 - 2} \left\{ \sum_{i=1}^{N_1} (x_i - \bar{x}_1)(x_i - \bar{x}_1)' + \sum_{i=N_1+1}^{N_1+N_2} (x_i - \bar{x}_2)(x_i - \bar{x}_2)' \right\} \quad (8)$$

where x is the vector measurement from either class 1 or class 2.

For infinite sample size the exact value of the divergence is known and given by

$$J = (\mu_1 - \mu_2)' \textstyle\sum^{-1}(\mu_1 - \mu_2) \quad (9)$$

which is the same as the Mahalanobis distance. The divergence using sample estimated parameters is

$$\hat{J} = (\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 - \bar{x}_2) \quad (10)$$

where $E[S] = \sum$. The covariance matrix $\bar{x}_1 - \bar{x}_2$ is

$$E[(\bar{x}_1 - \mu_1 - \bar{x}_2 + \mu_2)(\bar{x}_1 - \mu_1 - \bar{x}_2 + \mu_2)'] = \textstyle\sum \left(\frac{1}{N_1} + \frac{1}{N_2}\right)$$

Let $k = \frac{1}{N_1} + \frac{1}{N_2}$. The random variable $\hat{J}/k$ has a Hotelling's T^2 non-null distribution (see e.g [3]) with $N_1 + N_2$ samples and $f = N_1 + N_2 - 2$ degrees of freedom given by

$$H_p'\left(\frac{\hat{J}}{k}\middle| f, \frac{J}{k}\right) d\frac{\hat{J}}{k} = e^{-J/2k} \sum_{r=o}^{\infty} \frac{(J/2k)^r}{r!} \frac{1}{B\left(\frac{p}{2} + r, \frac{f-p+1}{2}\right)} \frac{(\hat{J}/kf)^{(p/2)+r-1}}{\left(1 + \frac{\hat{J}}{kf}\right)^{(1/2)(f+1)+r}} d\left(\frac{\hat{J}}{kf}\right), \; \hat{J} > 0. \quad (11)$$

By using the formula

$$\int_0^\infty \frac{x^{\mu-1}}{(1+x)^v} dx = B(\mu, v - \mu)$$

the expectation of $\hat{J}$ can be written

$$E[\hat{J}] = \frac{kfp}{f-p-1} + \frac{fJ}{f-p-1} > J \quad (12)$$

which approaches J when the sample sizes become infinity. Also we can show that

$$E(\hat{J} - J)^2 = \frac{f^2}{(f-p-1)(f-p-3)} [J^2 + 2Jk(p+2) + p(p+2)k^2] + J^2 - \frac{2kfpJ + 2fJ^2}{f-p-1} \quad (13)$$

which approaches zero as the sample sizes approach infinity. Thus in the multivariate Gaussian case, $\hat{J}$ is also a consistent estimate of J. For equal covariance but unequal mean vectors, the Bhattacharyya distance experiences the same effect as divergence as B = J/8.

For two multivariate Gaussin densities with zero mean vectors and covariance matrices $\sum_1$ and $\sum_2$ whose unbiased estimates are V_1 and V_2 respectively, the divergence based on sample estimated parameters is

$$\hat{J} = \frac{1}{2} \operatorname{tr}(V_1 V_2^{-1} + V_2 V_1^{-1}) - p \quad (14)$$

Since the samples from the two classes are independent,

$$E[\hat{J}] = \frac{1}{2} \operatorname{tr}\{E(V_1)E(V_2^{-1}) + E(V_2)E(V_1^{-1})\} - p$$

Both V_i and V_i^{-1} follow the Wishart distribution with expectations

$$E(V_i) = \textstyle\sum_i, \; E(V_i^{-1}) = \frac{N_i}{N_i - p - 1} \sum_i^{-1}, \; i = 1,2$$

Thus

$$E[\hat{J}] = J + \frac{1}{2}\,\mathrm{tr}\Big(\frac{p+1}{N_2 - p - 1} \Sigma_1\Sigma_2^{-1} + \frac{p+1}{N_1 - p - 1}\Sigma_2\Sigma_1^{-1}\Big) \tag{15}$$

where the bias term coincides with Eq. (5) for $p = 1$, i.e. the univariate case.

The above discussion clearly illustrates the effect of finite sample size on the bias of the sample estimated distance measures. In general the estimated divergence has a positive bias while the behavior of estimated Bhattacharyya distance is less predictable. It is noted also that direct estimation of the distance measures is possible if nonparametric density estimate is employed; but it would be more difficult to study the small sample behavior.

3. Finite Sample Information Measures

For feature selection, more informative features result in low classification errors. However, if the sample size is limited, information measures estimated from samples may not be as effective. Consider the equivocation for m classes defined as

$$H = - E\,[\sum_{i=1}^{m} P(\omega_i/x)\log P(\omega_i/x)] \tag{16}$$

where $P(\omega_i/x) = P_i$ is the a posteriori probability of the ith class and the expectation is taken to the space of x.

The sample-based equivocation using the estimated a posteriori probability $\hat{P}_i$ is

$$\hat{H} = - E\,[\sum_{i=1}^{m} \hat{P}_i \log \hat{P}_i] \tag{17}$$

Let θ_i be the parameter of the ith class, and $\hat{\theta}_i$ its estimate. Assume that the sample size effect is small so that we need consider only the first two terms in the Taylor series expansion of $\hat{P}_i$,

$$\hat{P}_i \simeq P_i + P_i'(\hat{\theta}_i - \theta_i) \tag{18}$$

where P_i' is the partial derivative of $\hat{P}_i$ with respect to $\hat{\theta}_i$ evaluated at $\hat{\theta}_i = \theta_i$. The difference between the estimated and true equivocations can be written as

$$\hat{H} - H \simeq E \sum_{i=1}^{m} [P_i'(\hat{\theta}_i - \theta_i)(1 + \log P_i) + P_i'(\hat{\theta}_i - \theta_i)^2] \tag{19}$$

which is still a function of $\hat{\theta}_i$. It is noted that both P_i and H are **expanded** by using the first order approximations. If $\hat{\theta}_i$ is an unbiased estimate of θ_i, then the expectation of the difference with respect to the estimated parameter depends only on the variance of $\hat{\theta}_i$ which is usually inversely proportional to the sample size. The variance of $\hat{H}$ given by $E(\hat{H} - E(\hat{H}))^2$ where the expectations are with respect to $\hat{\theta}_i$ is approximately proportional to the variance of $\hat{\theta}_i$ or inversely proportional to the sample size. Thus $\hat{H}$ given by Eq. (17) is an asymptotically unbiased and consistent estimate of H. To examine the small sample behavior, specific expressions for $\hat{P}_i$ and P_i are needed in order to evaluate Eq. (19).

4. Finite Sample Discriminant Analysis

Although there is an enormous statistical literature on discrimination in the Gaussian case, the available small sample results are few and inconsistent. The proposed effort will concentrate on some special cases including the class of exponential densities. The linear discriminant function resulting from equal covariance matrices is the most important special case. The common covariance matrix may be determined from training samples of both classes (Eq. 8) which is the assumption made in many statistical literatures. The difference between the error probabilities can be approximated by truncated Taylor series as

$$\int_{\frac{\sqrt{\hat{J}}}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\,dy - \int_{\frac{\sqrt{J}}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\,dy$$
$$= \frac{-(\hat{J} - J)}{4\sqrt{2\pi J}} \exp - \frac{J}{8} + \frac{(\hat{J} - J)^2}{64\sqrt{2\pi J}}\left(1 + \frac{4}{J}\right)\exp - \frac{J}{8} \tag{20}$$

The expected value of the difference can be obtained by using Eqs. (12) and (13). A sample calculation for $N_1 = N_2 = 5$, $p = 2$ and $J = 4$ gives the expected difference $- 0.0027$ which is close to the expected difference of $- 0.0312$ by using an expression due to McLachlan [4] which is computed in [5]. If J increases, the sample sizes must also increase to maintain a good approximation in Eq. (20). Similar analysis of error probability for other discriminant functions using estimated parameters may not be available however. In most cases computer simulation is necessary to determine the relations among performance, sample sizes and dimensionality. A good example is the quadratic discriminant function for unequal covariance matrices [6]. Unfortunately consistent results have not been reported in the literature[5][7] other than the linear discriminant function discussed above. In some cases including the exponential densities, good theoretical approximation of error probability under finite sample size is possible. The individual cases must be examined separately and computer simulation must be used when necessary. There does **not** appear to have a unique solution procedure suitable for all cases.

5. Finite Sample Nearest-Neighbor Decision Rules

The nearest-neighbor decision rule (NNDR) is attractive in the sense that the NN-risk is upper bounded by twice the Bayes risk when the sample size approaches infinity. For a given sample size the 1-NNDR is uniformly better than the k-NNDR.

The small (finite) sample NNDR is important because the data storage and computational requirements can easily be met when the sample size is small. Also in realtime processing, the number of samples that can be processed at a given time must be limited. However the small sample behavior of the NNDR is very much unknown and much study is needed. So far only the following restrictive cases have been considered: Fix and Hodges [8] investigated the small sample performance of 1-NNDR for univariate and bivariate Gaussian distributions. Kanal, et.al. [9] derived the NNDR error probability for binary patterns. Levine, et.al. [10] showed that the performance for small sample sets from uniform distributions is close to its asymptotic value. For multivariate Gaussian densities and allowing the sample size to increase with k, the number of nearest-neighbors, it is shown [5] numerically that the k-NNDR has a performance very close to the Bayes linear discriminant analysis. This indicates that under medium or large sample condition, the NNDR is comparable to the Bayes rule using the estimated parameters.

For 1-NNDR, the conditional error probability given the measurement x and its nearest-neighbor $\mathbf{x}_j$ is [11]

$$r(x, x_j) = P(\omega_1/x)P(\omega_2/x_j) + P(\omega_2/x)P(\omega_1/x_j) \quad (21)$$

Now the usual assumption that $P(\omega_i/x_j)$ approaches $P(\omega_i/x)$ asumptotically does not hold in the small sample case. For a given parametric or non-parametric density, the NN-risk for small sample size can be obtained by taking the expectations of Eq. (21) with respect to x_j and x. Similar expressions can be written for k-NNDR. As the closed form expressions are generally not available for the expectations involved, tight bounds must be established.

It should be noted here that the small sample NNDR behavior examined here is a different problem from the edited or condensed NNDR considered elsewhere (see e.g. [12][13]). However the idea of using a small set of selected learning samples is important. Our experimental results with the teleseismic data [14] have shown that there is always a small subset of good learning samples that dominate the performance. In other words the performance would be insensitive to sample size for good quality learning samples. Thus the small sample NNDR performance need not be worse than the asymptotic performance by properly selecting a small set of learning samples.

6. Contextual Analysis for Image Recognition

A major weakness of statistical pattern recognition is the difficulty to take the contextual relations into account in the recognition process. Character recognition is not considered here as it requires a somewhat difference contextual analysis [15]. An imagery pattern is rich in contextual information part of which is statistical in nature. A formal statistical approach to this problem is the compound decision theory. The finite sample constraint in digital imagery patterns is caused by the limited number of image samples available and the limitation in spatial resolution and quantization levels. In image interpretation and classification study, an image is usually partitioned into a number of subimages. A vector measurement may be taken from each subimage. By assuming dependence on the nearest four neighboring subimages, the compound decision rule is to choose the class which maximizes [16][17],

$$P(x_o/\omega_k)P(\omega_k) \prod_{j=1}^{4} p(x_j/\omega_k) \quad (22)$$

where $\omega_k = 1, 2, \ldots, m$ and x_o is the measurement of the subimage under consideration. Notice that the part of the expression outside of the product sign is identical to that used in a simple maximum likelihood decision rule without considering neighboring subimages at all. Each multiplier in the product term represents the contextual contribution from an adjacent neighboring subimage. The probability densities required for evaluating Eq. (22) must be either assumed or determined from the gray level histogram.

If we assume a Gaussian density for the measurement x, then the finite sample discriminant analysis is useful for subimage classification when the contextual dependence is not considered. If the contextual information is taken into account, then the product terms in Eq. (22) produce additional terms in the discriminant function causing some complexity in error probability computation. However, the effect of estimated parameters based on finite number of image samples can still be determined under Gaussian assumption. If both the finite learning sample and the quantization and spatial resolution constraints are considered, the direct use of histogram would be more suitable. Let the images of interest consist of objects on a background with probability densities p(z) and q(z) respectively where z denotes the gray level. Suppose further that the objects occupy fraction θ of the image area, so that the background occupies fraction $1-\theta$. Then the normalized histogram of the image is the overall gray level probability density $\theta p(z) + (1-\theta)q(z)$. The thresholding technique [18] can then be used for each subimage to decide for each pixel (picture element) whether it belongs to the objects or background. Eq. (22) can be considered as representing an object or background histogram obtained from the object or background pixels of all five subimages. A minimum error decision threshold can be obtained from the two histograms p(z) and q(z). If the subimages under consideration has more pixels below the threshold then the decision is in favor of the objects, otherwise the decision will be the background.

The above procedure makes it easy not only to implement the compound decision rule given by Eq. (22) but also to determine the finite sample effects. The four neighboring subimages obviously increases the effective total number of pixels used for classification. If the object and background histograms are modelled as Gaussian densities then the error probability of the compound Bayes decision rule can be determined from the Gaussian models using estimated parameters. A more general approach is to use the sampling distribution of the histogram for q quantization levels and a total of n pixels for the image given by

$$\frac{\Gamma(n+q)}{\Gamma(r_1+1)\ldots\Gamma(r_q+1)} \prod_{i=1}^{q} p_i^{r_i} \tag{23}$$

where r_i is the number of pixels belonging to the ith quantization level. The Bayes estimate of p_i, the fractional number of pixels for the ith level is

$$\hat{p}_i = \frac{r_i + 1}{n + q}$$

Then it is possible to determine the mean recognition accuracy [19] taking into account the contextual information.

7. Remarks

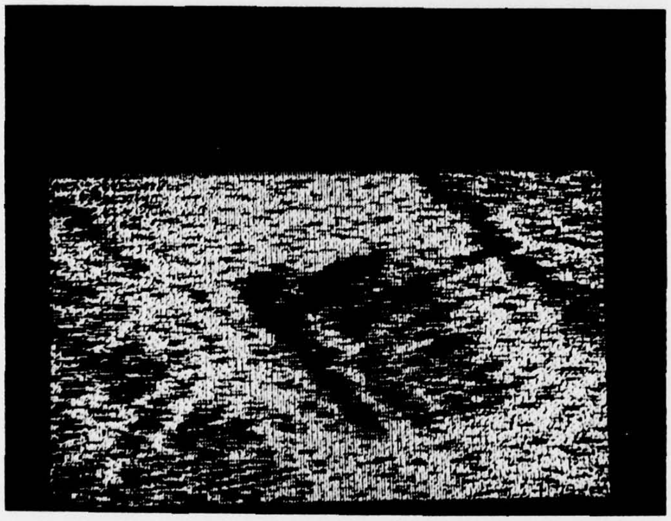
Although it would be desirable to have substantial sample size for all pattern recognition problems considered, limitation in sample size frequently occurs in practice. Except for the uninteresting case of too small sample size, the recognition performance which depends on both sample size and dimensionality need not be poor at small sample size. In designing recognition systems which operate at small learning sample size, classification algorithms which are less sensitive to sample size should be preferred. Unfortunately no single method can be used to examine the finite sample effects in all problems considered. The solutions must be problem dependent. Series expansion and tight error bounds should be used if exact solutions are not available. Distinction among small sample size, medium sample size, and large sample size should also be made in each problem area.

References

1. C.H. Chen, "On information and distance measures, error bounds, and feature selection", Information Sciences Journal, Vol. 10, pp. 159-173, 1976.
2. L.R. Shenton and K.O. Bowman, "Maximum likelihood Estimation and Small Samples", Statistical Monograph No. 38, Charles Griffin & Co., Ltd. 1977.
3. A.M. Kshirsager,"Multivariate Analysis", Marcel Dekker, Inc., New York, 1972.
4. G.J. McLachlan, "Estimates of errors of misclassification on the criterion of asymptotic mean square error", Technometrics, Vol. 16, No. 2, pp. 255-260, May 1974.
5. C.H. Chen, "Performance bounds of a class of sample-based classification procedures", Proc. of Pattern Recognition and Image Processing Conference, pp. 94-74, June 1977.
6. S. Marks and O.J. Dunn, "Discriminant functions when covariance matrices are unequal", Journal of the American Statistical Association, Vol. 69, No. 346, pp. 555-559, June 1974.
7. S. Raudys, "On dimensionality, learning sample size and complexity of classifying algorithms", Proc. of Third International Joint Conference on Pattern Recognition, pp. 166-169, November 1976.
8. E. Fix and J.L. Hodges, "Discriminant analysis, nonparametric discrimination: small sample performance", USAF School of Aviation Medicine, Randolph Field, Texas, Project No. 21-49-004, Report 11, Aug. 1952.
9. L.N. Kanal, et.al., "Adaptive modelling of likelihood classification", Technical Report No. RADC-TR-66-190 AD 636519, June 1966.
10. A. Levine, L. Lustick and B. Saltzberg, "The nearest-neighbor rule for small samples drawn from uniform distributions", IEEE Trans. on Information Theory, Vol. IT-19, No. 5, pp. 697-699, September 1973.
11. K. Fukunaga, "Introduction to Statistical Pattern Recognition", Academic Press, Inc. New York, 1972
12. D.L. Wilson, "Asymptotic properties of nearest neighbor rules using edited data", IEEE Trans. on Systems, Man, and Cybernetics, SMC-2, No. 3, pp. 408-421, July 1972.
13. P.E. Hart, "The condensed nearest neighbor rule", IEEE Trans. on Information Theory, Vol. IT-14, pp. 515-516, May 1968.
14. C.H. Chen, "Seismic pattern recognition", in Special Issue on Computer-Aided Seismic Analysis and Discrimination, edited by C.H. Chen, Geoexploration Journal, Mar. 1978.
15. G.T. Toussaint, "The use of context in pattern recognition", Proc. of Pattern Recognition and Image Processing Conference, pp. 1-10, June 1977.
16. J.R. Welch and K.G. Salter, "A context algorithm for pattern recognition and image interpretation", IEEE Trans. on Systems, Man and Cybernetics, SMC-1, pp. 24-30, January 1971.
17. C.H. Chen, "Statistical Pattern Recognition", Hayden Book Co., Rochelle Park, New Jersey, 1973.
18. A. Rosenfeld and A.C. Kak, "Digital Picture Processing", Academic Press, Inc. New York, 1976.
19. B. Chandrasekaran and A.K. Jain, "Quantization complexity and independent measurements", IEEE Trans. on Computers, Vol. C-23, pp. 102-106, January 1974.

Appendix II for Final Report

original image



filtered result
SNR=0.88

filtered result
SNR=0.55

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 78-1279	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) FEATURE EXTRACTION AND SIGNAL PROCESSING APPROACH TO REALTIME PATTERN RECOGNITION		5. TYPE OF REPORT & PERIOD COVERED Interim FInal
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) C.H. Chen		8. CONTRACT OR GRANT NUMBER(s) AFOSR 76-2951
9. PERFORMING ORGANIZATION NAME AND ADDRESS Southeastern Massachusetts University Department of Electrical Engineering North Dartmouth, Massachusetts 02747		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A2
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332		12. REPORT DATE August 1978
		13. NUMBER OF PAGES 13
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)

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

The integration of feature extraction for pattern recognition and the digital signal processing into one study as performed in this project has resulted in advances in both areas, and the discovery of many new ideas which are beneficial to both areas. There are common problems, such as the finite sample size effect, in both pattern recognition and signal processing. For example, digital signal processing techniques are much needed in extracting effective features while statistical pattern recognition can be useful in

DD FORM 1 JAN 73 1473

UNCLASSIFIED

20. Abstract continued.

image processing. More specifically, this research has carefully examined the fundamental problem of the finite sample size and its effect on feature selection and classification rules. Most effective features for seismic pattern recognition have been developed through the signal modelling study. In the image recognition work, new results include the rotationally invariant digital Laplacian operation and a new adaptive Kalman filtering technique for efficient realtime image processing. Detailed computer results have been developed and documented to support the theoretical study. Finally for image classification, the specific problem of contextual information is examined and a decicion tree procedure is developed which can process both the statistical and structural features for effective classification.