8a. NAME OF FUNDING / SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL *(If applicable)*	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR 75-2777

8c. ADDRESS *(City, State, and ZIP Code)*	10. SOURCE OF FUNDING NUMBERS			
BLDG 410 BAFB DC 20332-6448	PROGRAM ELEMENT NO. 61102F	PROJECT NO. 2304	TASK NO A2	WORK UNIT ACCESSION NO.

11. TITLE *(Include Security Classification)*
OPTIMIZATION TECHNIQUES FOR FEATURE EXTRACTION IN AUTOMATIC PATTERN RECOGNITION

12. PERSONAL AUTHOR(S)
Rui J. P. de Figueiredo

13a. TYPE OF REPORT Final	13b. TIME COVERED FROM ________ TO ______	14. DATE OF REPORT *(Year, Month, Day)* October 31, 1979	15. PAGE COUNT 18

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

DTIC
ELECTE
NOV 28 1989
S B D

89 11 27 114

AFOSR-TR- 82-1322

OPTIMIZATION TECHNIQUES FOR FEATURE EXTRACTION IN AUTOMATIC PATTERN RECOGNITION

(Final Technical Report for the AFOSR Grant
75-2777)

by

Rui J. P. de Figueiredo

Rice University Technical Report EE-7910
Houston, Texas October 31, 1979

CONTENTS

Page

Abstract

Contents

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification

By
Distribution/
Availability Codes

Dist	Avail and/or Special
A-1	

1. INTRODUCTION

The AFOSR Grant 75-2777 was in effect from November 1, 1974 to May 15, 1979. It had as its main objective the development of theory and techniques for optimally extracting features from data for the purpose of classifying these data with the least error.

The above goal was achieved through the development of the techniques and ancillary results described in the following section. These developments are suppported by the documents listed in Section 3. Various other activities that took place under the aegis of the grant are described in Section 4.

2. RESEARCH RESULTS

Both statistical and nonstatistical approaches to the problems of feature extraction, dimensionality reduction, and image processing were investigated. They are briefly summarized in the following subsections.

2.1. Statistical Feature Extraction

Let x denote a data vector, to be classified, belonging to a n-dimensional real measure space R^n. By a "feature" of x we mean a (linear or nonlinear) measurable functional $\varphi_i: R^n \to R^1$. If we pick or construct m features $\varphi_i: R^n \to R^1$, $i = 1, \ldots, m$, the resultant map $\varphi \overset{\Delta}{=} \text{col}(\varphi_1, \ldots, \varphi_m): R^n \to R^m$ is called a "feature extraction map" and the resultant space R^m a "feature space." Since there is a measure (or measures) given on R^n, there is a corresponding measure (or measures) on the feature space R^m induced by φ. To signal this fact we will denote the measure space R^m by $R^m(\varphi)$.

One of the fundamental problems that we investigated during the period of the grant was:

Problem 1. Given positive integers n and m such that $n > m$; a set of measures on R^n defined by the probability density functions (likelihood functions) $f_X(x/H_j)$ conditioned on pattern classes H_j, $j = 1, \ldots, M$, which are known or estimated from a set of samples; a set of prior probabilities P_j for the pattern classes H_j, $j = 1, \ldots, M$; and a class Φ of admissible feature extraction maps $\varphi: R^n \to R^m(\varphi)$; find a $\hat{\varphi}$ which minimizes the Bayes risk (probability of misclassification) $Q(\varphi)$ in the feature space $R^m(\varphi)$ over all $\varphi \in \Phi$.

A nonlinear programming approach [1][4] was used in formulating this problem mathematically, and the ensuing algorithms for the design of $\hat{\varphi}$ were of the Fletcher-Powell type.

The case in which the probability density functions $f_X(x/H_j)$ are normal and the class Φ is the class Φ_L of linear transformations, was completely solved, and tested [2][3] on C1 flight line data and LANDSAT data provided by the NASA personnel from the Johnson Space Center.

The general case in which $f_X(x/H_j)$ are not necessarily normal and Φ is a general class of nonlinear (not necessarily linear) transformations was investigated in detail in [5].

Specifically, the following results were developed in [5]: (I) Based on the training samples of the pattern classes, the construction of expressions for $Q(\varphi)$ and a suitable estimator for $Q(\varphi)$; (II) a study of errors involved in estimating $Q(\varphi)$; (III) characterization of families Φ of transformations φ; (IV) derivation of conditions for an optimal $\hat{\varphi}$; and (V) techniques for finding an optimal $\hat{\varphi}$.

2.2. Adaptive Estimator of Probability Density Functions

The two basic entities entering in the statistical feature extraction analysis, referred to in the preceding section, are the probability density function (likelihood function) and the Bayes risk. New accurate and efficient ways of estimating these entities were developed. In this subsection we briefly describe our new probability density function estimator and in the following subsection, our new Bayes risk estimators.

Specifically, our objective in [7] was to construct an adaptive orthogonal series estimator for a probability density function $f_X(x)$ associated with a real random variable X (From now on, for simplicity in notation, we will drop the subscript in $f_X(x)$).

So, given a sample set $X_1, \ldots, X_n$ of independent identically distributed real-valued random variables, each with the unknown probability density function $f(\cdot)$, the problem considered in [7] was to estimate f from the sample set. The function f was assumed to be in $L_2(a,b)$; f was not assumed to be in any parametric family. We constructed an adaptive "two-pass" solution to the problem: In a pre-processing step (the first pass), a preliminary rough estimate of f was obtained by means of a standard orthogonal-series estimator. In the second pass, the preliminary estimate was used to transform the orthogonal series. The new, transformed orthogonal series was then used to obtain the final estimate. We established consistency of the estimator and derived asymptotic (large sample set) estimates of the bias and variance. It was shown that the adaptive estimator offers reduced bias (better solution) in comparison to the conventional orthogonal series estimator. Computer simulations were presented which demonstrate the small sample set behavior. A case study of a bimodal density confirmed the theoretical conclusions.

2.3. Computationally Efficient Estimators for the Bayes Risk

A computationally efficient estimator for the Bayes risk is one which achieves a desired accuracy with a minimum of computation. In many problems, for example speech recognition, point evaluations of the class conditional densities are computationally costly. Density evaluations are the single most important factor contributing to the computational effort in Bayes risk estimation, thus the amount of computation required by a Bayes risk estimator is defined as the average number of conditional density evaluations it performs. The accuracy of a risk estimator is defined by its variance.

Existing estimators for the Bayes risk, namely the error count estimator and the posterior estimator, require for each sample X_j, $j=1,2...N$, evaluation of the class conditional density $f_m(X_j)$ for each class $m=1,2...M$, a total of $N \cdot M$ density evaluations. For problems such as speech recognition, where the number of classes M is large and density evaluations costly, these estimators are impractical from a computational aspect.

A new class of estimators of the general form $\hat{Q}(T)$ was proposed in [8]. In [8] an estimator $\hat{Q}(T)$ is defined by associating with each class m a subset T_m of the M classes. For two classes, only the error count and posterior estimators belong to this class. For more than two classes, several new estimators for the Bayes risk are included.

Specifically, the following results are presented in [8]. Estimators requiring fewer density evaluations are derived from the class of estimators of the general form $\hat{Q}(T)$ as follows. A scalar parameter α determines the sets $T_m(\alpha)$ of classes that are "α-close" to each class m, hence an estimator $\hat{Q}(\alpha)$ of the general form $\hat{Q}(T)$. As α varies, the sets $T_1(\alpha),...,T_M(\alpha)$ vary and a family $\{Q(\alpha) : 0 \leq \alpha < \alpha_{max}\}$ of risk estimators is achieved. Each estimator in the family is characterized by the average number of density evaluations it requires and its variance.

The optimal estimator from the family $\{\hat{Q}(\alpha) : 0 \le \alpha < \alpha_{max}\}$ is defined as that estimator with maximum computational efficiency, where the computational efficiency of an estimator is the inverse of the product of the average number of density evaluations it requires and its variance. The optimal estimator requires the least amount of computation to achieve a given accuracy, or, symmetrically, achieves the greatest accuracy with a minimum of computation.

In practice, the true optimal estimator cannot be determined since this would in effect require knowledge of the true risk Q. Thus a technique whereby the first n of the total N samples are used to approximate the optimal estimator is proposed. The n samples should contain enough information on the closeness of the classes to determine an almost optimal estimator. The last N-n samples are used in the approximate optimal estimator to obtain an accurate estimate of the risk with a minimum of computation.

2.4. Structure Preserving Dimensionality Reduction

A substantial effort was devoted to the development of new schemes for dimensionality reduction by means of structure preserving nonlinear maps. The results obtained are described in [9], [10], and [11]. Note that this may be considered as a geometric approach to feature extraction.

Problem 2. Given a set of data vectors $X = \{X_1, \ldots, X_N\}$, where each $X_i \in R^n$, find a set of vectors $Y = \{Y_1, \ldots, Y_N\}$, where $Y_i \in R^m$. (where n is an integer satisfying $1 \le m \le n$) in such a way that the "structure" present in X is optimally preserved under the transformation from X to Y, that is, the mapping from X to Y should take place with the least degradation in the data structure.

To solve the above problem, J. W. Sammon* proposed his well-known "nonlinear mapping algorithm" in which the set Y is selected so that the following criterion function is minimized

$$Q(Y) = \frac{1}{d} \sum_{i<j}^{N} \frac{(d_{ij} - d^*_{ij})^2}{d_{ij}}$$

where

$$d = \sum_{i<j} d_{ij}$$

$$d_{ij} = || X_i - X_j ||$$

$$d^*_{ij} = ||Y_i - Y_j||$$

and $|| \cdot ||$ denotes the Euclidian norm in the underlying space.

Sammon's algorithm considers all points in the set Y in the optimization procedure and hence its computational feasibility is limited by the number of points in Y. Also, there are important features that ought to be considered other than the above interpoint distance considered by Sammon.

Our contributions [9] - [11] to the solution of Problem 2 outlined below are oriented toward removing the above limitations of Sammon's procedure and in fact extending it as indicated.

(I) In [9] and [10] we extended Sammon's procedure by constructing criterion functions which embody not only topological considerations (such as interpoint distances) but also certain graph theoretical considerations

---

*J. W. Sammon, "A Nonlinear Mapping for Data Structure Analysis," IEEE Trans. on Computers, Vol. C-18, pp. 401-409, May 1969.

(such as the minimal spanning tree, maximally complete subgraphs, inconsistent edges, and diameter edges). Our algorithm is hierarchical in nature and offers considerable savings in terms of computer computation and storage requirements.

(II) Considering the set X mentioned above as a "training set," we developed techniques for mapping new data points by extrapolation. This was done based on piecewise linear interpolation in [10] and by a K - nearest neighbor rule in [11].

(III) In [11], the dimensionality reducing nonlinear mapping techniques mentioned above were applied with considerable success to the problem of discrimination of seven sleep stages from the eight lead human electroencephalogram (EEG).

2.5. Oil Spill Identification

In work jointly sponsored by the AFOSR Contract 75-2777 and the U.S. Coast Guard (Dept. of Transportation) contract DOT-CG-81-75-1383, the use of a pattern recognition approach for typing and identification of oil spills was investigated. S.A. Starks and M.L. Curtis worked on the project.

The first phase of this study involved generating an algorithm for giving a probability that two oil samples (the suspect sample and the spill sample) are the same. This was done using fluorescence spectra of some 230 oil samples furnished by the Coast Guard Research and Development Center, Groton, Connecticut. These included crude oils, heating oils, and lubricating oils.

The second phase of the above study involved the independence of infrared and fluorescence spectra of oils. In this phase, infrared and fluorescence spectra of 30 oil samples were used.

The rather satisfactory results obtained are described in [12] - [14].

2.6. Digital Image Processing

Three projects in digital image processing, described under 2.6.1 through 2.6.3 below, were successfully completed. The Principal Investigator was also very heavily involved in obtaining funds from the National Science Foundation and from the University to set up a "Signal Analysis and Image Processing Facility" at Rice, where the AFOSR research work described below might be carried out. This objective was achieved and such a facility created as indicated in 2.6.4 below.

2.6.1. Image Restoration Based on Psychovisual Modeling

A member of our group Mr. G. Leigh Anderson collaborated with Dr. A. N. Netravali of Bell Laboratories, Holmdel, N.J. in developing a technique for digital image restoration which exploits the psychovisual model of masking. Additive noise is less visible in a region of the image containing large brightness changes than in a region with little spatial detail; the rapidly varying brightness "masks" the eye's response to noise. Subjective measurements were made to determine the psychovisual visibility of noise as a function of spatial brightness changes. Then filters were designed which smoothed the image under the control of the experimentally determined noise visibility function. The results were very good; most of the visible noise was removed without appreciably blurring the image. The results appear in reference [15].

2.6.2. Image Restoration by Adaptive Kalman Filtering

As a sequel to the project just described, we successfully embarked into another research study, namely the development of an adaptive nonlinear Kalman-type filter for the restoration of two-dimensional images degraded by general image formation system degradations and additive white noise. A vector difference equation model was used to model the degradation process. The object plane distribution function was partitioned into disjoint regions based on the amount of spatial activity in the image, and difference equation models were used to characterize the object plane distribution function.

The results are described in [16]. In this reference, it is shown that each of the above regions can be unqiuely characterized by their second order statistics. The autocorrelation function for each region is then used to determine the coefficients of the difference equation model for each region. Recursive estimation techniques are applied to a composite difference equation model.

If the images are to be restored for human viewing it is desirable to account for the response of the human visual system as part of the receiver characteristics. This is done by weighting the variance σ^2 of the additive noise by a visibility function, where the visibility function is a subjective measure of the visibility of additive noise in an image by the human visual system. As a consequence, the resulting effective variance depends nonlinearly on the state.

Two additional features are added to the new restoration filter to solve problems arising in the implementation phase. A nearest neighbor algorithm is proposed for the selection of a previously processed pixel for providing the previous state vector for the state of pixel (i.j). Secondly, a two-dimensional interpolation scheme is proposed to improve the estimates of the initial states for each region.

2.6.3. Variable Length Chain Coding of Contours

Coding may be considered to be a form of feature extraction and dimensionality reduction.

As an extension of ordinary chain coding[*] we investigated two new chain coding schemes in which the length of the link in the chain was variable. Details on the results obtained are given in [17]. A specific application of these schemes to template matching was made. Experimental results showed that template matching using the schemes developed by us for representation of contours performed much faster than the ones by conventional techniques.

In the above study, a cluster display system [21] was also developed.

2.6.4. Signal Processing and Pattern Analysis Facility

Research work done under the AFOSR grant 75-2777, inspired and motivated the Principal Investigator to help set up at Rice University a Signal Processing and Pattern Analysis Facility. Such a facility was created in early 1979 with funds received from the National Science Foundation and Rice University. It has as its main elements a DEC PDP-11/55 computer with several accessories and a GENISCO Computers, Inc. GCT-3000 Digital Display System. This system is a programmable, raster scan display with a spatial resolution of 512 by 512 pixels and a color resolution of 256 colors (out of 4096 possible colors) or an intensity resolution of 256 gray-levels.

The above system was used in the research reported under 2.6.2.

---

[*] H. Freeman, "On the Encoding of Arbitrary Geometric Configurations" IRE Trans. on Elec. Comp., EC-10, 260-268, June 1961.

3. PUBLICATIONS PERTAINING TO RESEARCH SUPPORTED FULLY OR PARTIALLY BY THE AFOSR GRANT 75-2777

[1] R. J.P. de Figueiredo, "Optimal Linear and Nonlinear Feature Extraction Based on the Increased Risk of Misclassification," Proc. of the 2nd. Joint Intnl. Conference on Pattern Recognition, Copenhagen, Denmark, 1974.

[2] S. A. Starks, R. J.P. de Figueiredo, and D. L. Van Rooy, "An Algorithm for Optimal Single Linear Feature Extraction from Several Gaussian Pattern Classes," Intnl. Journal of Computer and Information Sciences. vol. 6, No. 1, pp. 41-54, March 1977.

[3] R. J.P. de Figueiredo, K. C. Pau, S. A. Starks, and D. L. Van Rooy, "An Algorithm for Extraction of More than one Optimal Linear Feature from Several Gaussian Pattern Classes," Proc. of the Third Intnl. Joint Conf. on Pattern Recognition, Coronado, CA., November, 1976; also, accepted for publication in the IEEE Trans. on Automatic Control.

[4] R. J.P. de Figueiredo, "Design of Optimal Feature Extractors by Mathematical Programming Techniques," in the book Pattern Recognition and Artificial Intelligence, edited by C. H. Chen, Academic Press, 1977.

[5] A.D. Sagar and R. J.P. de Figueiredo, "A Methodology for Optimal Nonlinear Feature Extraction in Statistical Pattern Recognition," Rice University Technical Report 7711, to be submitted for publication.

[6] R. J.P. de Figueiredo, "Feature Extraction Techniques for Classification and Identification of Spectral Signatures," in the Proc. of the 1976 Milwaukee Symposium on Automatic Computation and Control, pp. 303-304, 1976.

[7] G. L. Anderson and R. J.P. de Figueiredo, "An Adaptive Orthogonal Series Estimator for Probability Density Functions," to appear in the March 1980 issue of the Annals of Statistics.

[8] L. D. Wilcox and R. J.P. de Figueiredo, "Computationally Efficient Estimators for the Bayes Risk," Rice University Technical Report 7804, May 1978.

[9] S. A. Starks and R. J.P. de Figueiredo, "A New Approach to Structure Preserving Feature Extraction," Proc. of the 1977 Conference on Information Sciences and Systems, Johns Hopkins Univ., Baltimore, Md., April 1977.

[10] S.A. Starks and R.J.P. de Figueiredo, "An Algorithm for Optimal Nonlinear Structure Preserving Feature Extraction", presented at the 1977 IEEE International Symposium on Information Theory, Cornell University, Ithaca, N.Y., October 1977; submitted to the IEEE Trans. on Computers.

[11] R. C. Pruett and R. J.P. de Figueiredo, "A Nonlinear Mapping Approach to Discrimination of Sleep Stages from the Human Electroencephalogram," Rice University Technical Report EE 7902, May 1979.

[12] S. A. Starks and M. L. Curtis, "Use of Ultraviolet Spectroscopy in Oil Spill Identification," Proc. of the Twenty-Seventh Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, March 1976.

[13] J. S. Mattson, C. S. Mattson, M. J. Spencer, and S. A. Starks, "Multivariate Statistical Approach to Fingerprinting of Oils by Infrared Spectroscopy," Anal. Chem., 49, 297-302, 1977.

[14] M. L. Curtis, "Use of Pattern Recognition Techniques for Typing and Identification of Oil Spills," Technical Report No. CG-D-38-77, U.S. Coast Guard R & D Center, April 1977.

[15] G. L. Anderson and A. N. Netravali, "Image Restoration Based on a Subjective Criterion," *IEEE Trans on Systems, Man, & Cybernetics*, CMS-6(12), 845-853, December 1976.

[16] Sarah A. Rajala and R. J.P. de Figueiredo, "Adaptive Nonlinear Image Restoration by a Modified Kalman Filtering Approach," Rice Univ. Technical Report EE-7904, June 1979. Submitted to *IEEE Trans. On ASSP*.

[17] K. C. Pau and R. J.P. de Figueiredo, "Variable Length Chain Coding with Applications," Rice University Technical Report EE-7903, May 1979, to be submitted for publication.

[18] A. N. Netravali and R. J.P. de Figueiredo, "On a Class of Minimum Energy Controls Related to Spline Functions," *IEEE Trans. on Automatic Control*, AC-21, October, 1976.

[19] R.J.P. de Figueiredo, "An Interpolative and Smoothing Theory for Hardware Implementation of Feature Extraction", Rice University Technical Report EE-7710, June 1977.

[20] R. J. P. de Figueiredo, "Approaches to the Formulation and Solution of Nonlinear Inverse Problems", Rice University Technical Report EE-7810, November, 1978.

[21] K. C. Pau, "An Interactive Cluster Display System", Rice University Technical Report EE 7709, May, 1977.

4. OTHER ACTIVITIES

Professor de Figueiredo attended and presented invited and contributed papers at various meetings during the tenure of the AFOSR Grant 75-2777. Of these, special mention may be made of the following: He organized and co-chaired the main panel discussion session at the Joint Workshop on Pattern Recognition and Artificial Intelligence held in Hyannis, Massachusetts on June 1-3, 1976. He was a member of the IEEE delegation to the 1977 USSR Popov Society meeting in Moscow.

Also, over the entire period of the AFOSR Grant 75-2777, Professor de Figueiredo had many fruitful technical discussions with members of the Pattern Recognition Group of the Rome Air Development Center (RADC), especially with Mr. H. E. Webb and J. Faust. He participated in projects sponsored by RADC, monitored by Mr. H. Webb. Two of the reports that resulted from this activity are listed as references [19] and [20] in Section 3. Reference [18] was presented at a special RADC one-day meeting, held in New York City in November of 1978, on "Signal Restoration and Ill Posed Problems."

The grant research was enhanced by the visit made to our group by Professor Marian Mazur, member of the Polish Academy of Sciences and a noted cyberneticist, in the months of January and February of 1977. Professor Mazur had extensive discussions with the Principal Investigator on a theoretical framework for the description of the structure and functioning of natural human intelligence, and on possible application of such a framework to the solution of pattern recognition by machines.

Finally, with the full or partial support from the AFOSR grant, six Ph.D. dissertations (of G. L. Anderson, L.D.Wilcox, S.A.Starks, A.D.Sagar, S.A.Rajala, and R.C.Pruett) and two Master of Sciences theses (of K.C.Pau and S.A.Rajala) were completed, and four Ph.D. dissertations now in progress (of S.Fogel, C.L.Hu, T.J.Brzustowicz, and T.C.Chen) were begun.

5. CONCLUSION AND ACKNOWLEDGEMENTS

Over near five year period of the AFOSR grant 75-2777, methodologies as well as techniques for optimal feature extraction, based on statistical and geometrical frameworks, were developed. New image processing techniques, which account for the response of the human visual system as part of the receiver chracteristics, were also developed. Finally, some basic theoretical as well as applied results of general interest were obtained as a consequence of this effort.

We are indebted to the late Lt. Col. Thomas J. Wachowski and to Lt. Col. George W. Mc Kemie for the encouragement and direction provided to our program. The many helpful discussions had with Mr. Haywood E. Webb, Jr. of RADC are also gratefully acknowledged.