AD-A043 495 RICE UNIV HOUSTON TEX DEPT OF ELECTRICAL ENGINEERING F/G 6/4
A NEW APPROACH TO STRUCTURE PRESERVING FEATURE EXTRACTION.(U)
FEB 77 S A STARKS, R J DE FIGUEIREDO AF-AFOSR-2777-75
UNCLASSIFIED EE-TR-7707 AFOSR-TR-77-1142 NL

1 OF 1
AD A043495

END
DATE FILMED
9-77
DDC

(2)

RICE UNIVERSITY

GEORGE R. BROWN SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL ENGINEERING

HOUSTON, TEXAS

A NEW APPROACH TO STRUCTURE PRESERVING FEATURE EXTRACTION

by

Scott A. Starks* and Rui J.P. de Figueiredo*†

*Department of Electrical Engineering
†Department of Mathematical Sciences
Rice University, Houston, Texas 77001

February 15, 1977

TECHNICAL REPORT EE 7707

Note: This paper is to appear in the Proceedings of the 1977 Conference on Information Sciences and Systems, Johns Hopkins University, Baltimore, Maryland.

ABSTRACT:

This paper presents an approach to nonlinear feature extraction based on certain graph theoretical considerations (such as the minimal spanning tree, maximally complete subgraphs, inconsistent edges and diameter edges) and topological considerations (such as interpoint distance measures). After appropriate introductory sections, the feature extraction algorithm is developed in section 4. The algorithm is hierarchical in nature and offers considerable savings in terms of computer computation and storage requirements. An outline of the computer procedure is also included.

This work was supported by the AFOSR Grant 75-2777.

(See 1473)

Preprint of a paper from the Proceedings of 1977 Conference on Information Sciences and Systems held at The Johns Hopkins University, Baltimore, Md., March 30-April 1, 1977.

A NEW APPROACH TO STRUCTURE PRESERVING FEATURE EXTRACTION *

Scott A. Starks
Dept. of Electrical Engineering
Rice University
Houston, Texas 77001

Rui J. P. de Figueiredo
Dept. of Electrical Engineering and
Dept. of Mathematical Sciences
Rice University
Houston, Texas 77001

Abstract

This paper presents an approach to nonlinear feature extraction based on certain graph theoretical considerations (such as the minimal spanning tree, maximally complete subgraphs, inconsistent edges and diameter edges) and topological considerations (such as interpoint distance measures). After appropriate introductory sections, the feature extraction algorithm is developed in section 4. The algorithm is hierarchical in nature and offers considerable savings in terms of computer computation and storage requirements. An outline of the computer procedure is also included.

1. Introduction

In the design of any practical pattern recognition system, the problem of developing an efficient feature extractor is critical. A number of methods for optimal feature extraction exist. These are based on a variety of optimality criteria. A good review of the methods of feature extraction can be found in [1].

Quite often in a real world situation one is presented with the problem that the dimensionality of the measurement space is large, while the number of training samples is small. In this case, obtaining good estimates of the class conditional statistics is virtually impossible. As a result, all feature extraction algorithms which base their optimality criteria upon these statistical quantities are severely hindered. What is needed is a feature extraction method which does not rest solely on statistical considerations but which is based also on structural attributes present in the training data.

In order to accomplish the above, this paper presents a new method for nonlinear feature extraction. We begin this work with the general problem statement. Let there be given a set of data vectors $X = \{X_1, X_2, \ldots, X_N\}$ where each $X_i \in R^n$. We wish to find a corresponding set of vectors $Y = \{Y_1, Y_2, \ldots, Y_N\}$ where each $Y_i \in R^m$ (where m is an integer satisfying $1 \leq m < n$), in such a way that the structure present in X is optimally preserved under the transformation from X to Y, that is, we wish that the mapping of X to Y takes place with the least degradation in structure. Of course, the central question arises as to what constitutes structure in a data set. Thus, the present paper thrusts much of its attention toward this question.

* Supported by the AFOSR Grant 75-2777.

2. Background

In [2], J. W. Sammon developed a point-to-point mapping from a given space to one of lower dimensionality. In his algorithm, Sammon, by an iterative technique, optimized a criterion functional based upon interpoint distances in each space. If d_{ij} represents the Euclidean distance between points X_i and X_j in R^n (this distance will also appear at times as $d(X_i, X_j)$) and d^*_{ij} is the corresponding Euclidean distance between Y_i and Y_j in R^m, then the mapping criteria is defined as:

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

where

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

Generally speaking, this indicates that in optimizing this criterion, one finds the configuration of points Y in R^m whose interpoint distance best match their conterparts in R^n. Further examination of this criterion functional yields that it is a function of (Nxm) variables y_{ij}: i = 1,...N; j= 1,...m.

In certain cases, the product (Nxm) may become so large that optimizing such a criterion functional would be prohibitive in terms of computational complexity. To combat this problem, a heuristic relaxation method was devised by Chang and Lee [3] to perform point-to-point feature extraction. Despite the fact that this algorithm is computationally less demanding than the one developed by Sammon, it still is based upon the same criterion. As a result, it too treats interpoint distance as the only depository of information present in a data set.

Calvert [4] and White [5] have also contributed to the study of point-to-point mappings by applying orthogonal projection theory to the problem and investigating the effect of using L_1 or city-block distances instead of Euclidean distances, respectively.

3. The Search for Structure

When one is presented with the problem of designing a practical pattern recognition system under the constraint of a small number of training samples from a space of high dimensio-

nality, he often resorts to cluster analysis to learn something about the structure of the data. Much work has been done in this area and due to the fact that many clustering techniques might be best described as "ad hoc" or heuristic, there exists a number of clustering procedures. For good reviews of clustering techniques see [6], [7], and [8].

One large heading for clustering procedures is hierarchical clustering. To begin, it is necessary to pose the problem of hierarchical clustering. Let us consider the problem of clustering a set of M objects $Z = [Z_1, \ldots Z_M]$ into C clusters: $W_1, W_2, \ldots W_c$. We might think of clustering as no more than the process of partitioning M objects into C groups $W_1, \ldots W_c$ where $\bigcup_{i=1}^{M} W_i = Z$ and $W_i \cap W_j = \phi$ if $i \neq j$.

The process of hierarchical clustering is begun by partitioning the M objects of Z into M singleton clusters (clusters with only one member). At the next level of the procedure, a new partition is obtained consisting of M-1 clusters. At each successive level, the number of clusters is reduced by one and this process is terminated when the desired number, C, of clusters is obtained. If the procedure is such that whenever any two samples are in the same cluster at a given level, they also are in the same cluster at any higher level, then the clustering procedure is called hierarchical. This type of clustering is widely used in biological toxonomy and other classificatory sciences.

There are two basic types of hierarchical clustering algorithms: agglomerative and divisive. The method described above in which two clusters are merged at each level is called agglomerative. Divisive clustering occurs when at each level one cluster is split into two. Regardless of the type of hierarchical scheme used, the way one defines the distance between two clusters is critical because in agglomerative scheme the two nearest clusters are joined at each stage and in divisive schemes a single cluster is split to yield two clusters which are the farthest apart.

Some of the more widely used distance formulas are:

$$d_{min}(\chi_i, \chi_j) = \min_{X \in \chi_i,\ X' \in \chi_j} d(X, X') \qquad (3\text{-}1)$$

$$d_{max}(\chi_i, \chi_j) = \max_{X \in \chi_i,\ X' \in \chi_j} d(X, X') \qquad (3\text{-}2)$$

$$d_{mean}(\chi_i, \chi_j) = d(m, m') \qquad (3\text{-}3)$$

where
m = mean of χ_i, m' = mean of χ_j, χ_i and χ_j are clusters, d(.,.) represents some distance measure such as Euclidean distance.

Consider the use of d_{min} as the measure of distance between clusters in an agglomerative scheme. We may think of the data points as nodes in some graph. For a review of graph theory see Ore [9]. When d_{min} is used, nearest neighbors determine the nearest subsets or clusters. In the terminology of graph theory, the merger of χ_i and χ_j corresponds to adding an edge between the nearest pair of nodes, one in χ_i and one in χ_j. Since such edges linking clusters always go between distinct clusters, it is obvious that the resultant graph never contains any closed loops or circuits. Such a graph is referred to as a tree. If this edge linking procedure is continued until there is a path from any node to any other node, the resulting graph is said to be a spanning tree. In addition, it can be shown that the sum of the edge lengths of this spanning tree will never be greater than the sum of the edge lengths of any other tree which spans the nodes. Thus this graph is said to be the minimal spanning tree (MST) for the data set and this procedure is called the nearest-neighbor or single-linkage clustering algorithm.

When the d_{max} measure of distance is used in an agglomerative hierarchical clustering scheme, we obtain what is referred to as the furthest-neighbor algorithm. Applying this algorithm to a set of data can also be described in terms of graph theory. At each stage of the hierarchy we produce a graph in which edges connect all the pairs of nodes of a given cluster. In graph theoretic terminology, each cluster is said to form a complete subgraph. Referring to the definition of d_{max}, we deduce that the distance between two clusters is determined by the most distant pair of nodes belonging to their union. This quantity is also referred to as the diameter of the two clusters' union.

As was mentioned previously in connection with the single-linkage algorithm, the concept of the Minimal spanning tree (MST) is an important one and was introduced by Prim in [10]. It is a deceptively simple structure which has amazing properties for cluster analysis. Zahn [11] was the first to demonstrate its powers in dealing with a number of problems which were rendered unsolvable by other methods. He showed that clusters in a two-dimensional space which the eye identified immediately as separate entities could be separated trivially by a clustering algorithm based on the MST. In fact, he went so far as to suggest that the MST was the fundamental mechanism responsible for proximity and Gestalt effects in psychology.

Some other approaches to cluster analysis using tree structures are described and referred to in [12].

The main properties of the MST can be summarized as follows:

1. Any point (node) is connected to at least one of its nearest neighbors.
2. Any subtree is connected to at least one of its nearest neighbors by the shortest available path.
3. The MST minimizes all increasing symmetric functions of interpoint distance.

4. The MST connectivity is invariant under any transformation which preserves the rank order of interpoint distance.
5. The MST structure is easy to compute and resembles a loopless skeleton of the configuration.

Thus we have found two attributes, the MST and the cluster diameter, which in part tell something of the structure present in a data set. In the following action, these two concepts are combined with that of interpoint distance to yield a procedure for nonlinear feature extraction.

4. Feature Extraction Procedure

For convenience, we shall now restate the feature extraction problem:

Let there be given a set of points $X = \{X_1, X_2, \ldots X_N\}$ where each $X_i \epsilon R^n$. We wish to find a corresponding set of points $Y=\{Y_1, \ldots Y_N\}$ where each $Y_i \epsilon R^m$ $(1 \leq m < n)$ in such a manner that the structure contained in the data set X is optimally preserved under the transformation $g : R^n \rightarrow R^m$ which maps X into Y. We take the liberty of expressing $g(X_i)$ as Y_i due to the one-to-one correspondence between the sets X and Y.

For N points, there are N(N-1)/2 independent interpoint distances. After determining these distances, the minimal spanning tree (MST) for the data set can be constructed. Once this is accomplished, a method of feature extraction based on a divisive hierarchical clustering scheme can be implemented in the following fashion.

Let us define the inconsistency measure for an edge of the MST in the following manner. Suppose that there is an edge of the MST connecting nodes represented by the points X_i and X_j. We define the inconsistency measure [11] of this edge as the ratio of length of the edge between X_i and X_j divided by the average length of all MST edges connected to either X_i or X_j (excluding the edge from X_i to X_j). Since for a data set of size N, the MST has N-1 edges, we can store the inconsistency measure for each edge of the MST in an array, E, of dimension N-1.

Since we may think of clustering as merely a process of labeling points according to their cluster or group membership, we can store such membership in an array C of dimension N, where C_i contains the label or membership of point being placed in cluster χ_1. So $C_i = 1$ for all i initially.

To initiate the feature extraction procedure, the array E is searched to find the MST edge with the largest value for inconsistency. We identify the endpoints of this edge and will denote them as X_k^1 and X_ℓ^1. The value for the diameter of the cluster χ_1 is also determined. Let the endpoints of this diameter edge be denoted by X_m^1 and X_n^1 and the diameter be D^1. Once these four points are known, we group them in what we call the active set, A^1. So $A^1 = \{X_k^1, X_\ell^1, X_m^1, X_n^1\}$. This set is termed the active set because it contains the vectors whose images $Y_k^1, Y_\ell^1, Y_m^1, Y_n^1$ are optimally located in the image or feature space at the first stage of the process.

Thus once the membership of the active set is found, the image set can be determined. This is done by finding the set of points

$$(Y_k^1, Y_\ell^1, Y_n^1, Y_m^1) = g(A^1)$$

which minimize the following criterion:

$$Q_1(g(A^1)) = \sum_{i \epsilon I^1} \sum_{\substack{j \epsilon I^1 \\ i<j}} \frac{(d_{ij} - d^*_{ij})^2}{d_{ij}} \qquad (4\text{-}1)$$

under the constraint that

$$d^*_{ij} \leq D^1 \quad \forall \; i, j \; \epsilon \; I^1, \qquad (4\text{-}2)$$

where

$$I^1 = \{ i : X_i \epsilon A^1 \}, \qquad (4\text{-}3)$$

$$d_{ij} = d(X_i, X_j), \qquad (4\text{-}4)$$

$$d^*_{ij} = d(Y_i, Y_j). \qquad (4\text{-}5)$$

The above is the standard Sammon criterion except that we have added the inequality constraint (4-2) and we do not require the distances (4-4) and (4-5) to be Euclidian. Note also that we are applying this criterion to only four points rather than as in [2] to the entire data set.

The optimization of (4.1), (4.2) may be carried out by the iterative method found in the Appendix of this paper.

Once the optimal configuration of $\hat{g}(A^1)$ is found, the location of the points corresponding to Y_k^1, Y_ℓ^1, Y_m^1, and Y_n^1 are fixed and are not allowed to vary through the remaining part of the procedure. As an indication of this fact, the vector B of dimension N is constructed. If the component $B_i=1$, this indicates that the image of X_i has been located. So initially $B_i = 0$ $(i=1,\ldots N)$, but after the first stage $B_i=1$ for $X_i \epsilon A^1$. Before we begin the next stage of the algorithm, we must update the cluster membership on the basis of deleting the most inconsistent edge found earlier and creating two connected subgraphs with the remaining edges of the MST. At this step, some of the components of C will be changed to reflect changes in cluster membership.

In general, we repeat the above procedure until all the points of X have their images mapped into Y. Let us assume that we are at the K^{th} stage of the algorithm. A search of E is conducted to find the K^{th} most inconsistent edge of the MST. The endpoints of this edge, X_k^K and K_ℓ^K, as well as the endpoints of the diameter edge, X_m^K and X_n^K, for the cluster which contains the K^{th} most inconsistent edge are determined. These points are then placed in the active set A^K.

If any points belonging to A^K have already been mapped into images in R^m, it is understood that they will not be moved at this stage of the algorithm. Suppose that the cluster containing the K^{th} most inconsistent edge is designated cluster χ_p and that the diameter of χ_p is D^K. If $\Delta(i,j)$ is the standard Kronecker delta function, the criterion functional for the K^{th} stage is

$$Q_K(g(A^K)) = \sum_{i \in I^K} \sum_{j=i}^{N} \frac{(d_{ij} - d^*_{ij})^2}{d_{ij}} \Delta(p-C_j) \quad (4\text{-}6)$$

under the constraint that

$$d^*_{ij} \leq D^K \; \forall \; i,j \ni X_i \text{ and } X_j \in \chi_p, \quad (4\text{-}7)$$

where

$$I^K = \{ i: X_i \in A^K \}. \quad (4\text{-}8)$$

The above criterion is minimized by the method outlined in the Appendix.

This constraint is added to insure that no two points in the image set of χ_p are farther apart than the diameter of χ_p. Examination of the criteria shows that the configuration $g(A^K) = \{Y^K_k, Y^K_\ell, Y^K_m, Y^K_n\}$ is optimized with respect to intracluster interpoint distances with added emphasis on the diameter edge length and the inconsistent edge length. After the optimization is performed, the cluster membership is updated by setting $C_i = K$ for all values of i representing points belonging to one of the two connected components of cluster χ_p formed by deletion of the K^{th} most inconsistent edge.

This procedure is repeated until all the N points are mapped to R^m. For a flow chart of this entire operation refer to Fig. 1.

5. Determination of the dimensionality of the feature space

The final feature extraction obtained depends upon the value, m, of the dimensionality of the feature space. The problem remains as to how to make a reasonable choice for m.

In[13], Schwartzmann and Vidal present an algorithm for estimating the topological or intrinsic dimensionality of point sets. Since we are concerned in our algorithm with preserving the structure in a data set, it seems only natural that we concern ourselves with determining the topological dimensionality. The algorithm of Schwartzmann and Vidal relies heavily on the MST just as the hierarchical clustering presented earlier. This approach is iterative and is also based on Karhunen–Loéve theory along with the theory of barycentric transformations. As a result, before beginning the feature extraction process, we perform the Schwartzmann-Vidal algorithm to get an estimate of the topological dimensionality to be used as the value for m.

Recall that in the discussion of the MST, it was stated that any transformation of a data set which preserves the rank ordering of the interpoint distances also preserves the MST connectivity. This is intimately connected with the work of Shepherd [14] who was concerned with preserving monotonic relationships. Since the MST has been shown to have important clustering and Gestalt properties [11] it would be desirable to preserve its connectivity under transformation. To do so it would be required that the image set of Y be such that for each inequality $d(X_i, X_j) < d(X_k, X_\ell)$ we would have to have $d(Y_i, Y_j) < d(Y_k, Y_\ell)$. With N(N-1)/2 interpoint distances, this translates to

$$(N(-1)/2) \cdot (N(N-1)/2 - 1)/2$$

pairwise inequality constraints. In most cases, such a large number of constraints would greatly deter optimization.

With this in mind, an NxN array R is constructed so that R^*_{ij} gives the rank order of the distance between Y_i and Y_j. We can then use the value for $\|R-R^*\|$ as a means for determining the "goodness" of the transformation. If the value of $\|R-R^*\|$ is too large, the feature extraction would be repeated only with the value for the dimensionality of the feature space increased. Increasing the value of m would allow for a greater number for the degrees of freedom and would thus enable one more freedom in preserving the rank order of interpoint distance.

Conclusions

An algorithm for nonlinear feature extraction has been presented. It emphasizes structures from graph theory such as the minimal spanning tree, maximally complete subgraphs, inconsistent edges, and diameter edges as important entities to be preserved under transformation. The procedure is hierarchical in nature. By solving a number of small problems instead of one large problem, this procedure greatly reduces the complexity of computer implementation. This approach shows great promise especially when one is not sure of class conditional statistics for it is based on other structures present in the data. Presently, computer programs are being developed to implement this algorithm at Rice University. Numerical results on the application of this algorithm to fluorescence and infrared spectroscopy data for oil spill identification will appear shortly.

APPENDIX

At each stage of the feature extraction algorithm, we are concerned with finding the configuration $\hat{g}(A^K) = \hat{Y}^K$ which optimizes

$$Q_K(g(A^K)) = \sum_{i \in I^K} \sum_{j=i}^{N} \frac{(d_{ij} - d^*_{ij})^2}{d_{ij}} \Delta(p-C_j) \quad (A\text{-}1)$$

where

$$I^K = \{ i: X_i \in A^K \}, \quad (A\text{-}2)$$

under the constraint that $d^*_{ij} \leq D^K$ for all

$i,j : X_i, X_j \in X_p$. (A-3)

This sort of problem is well suited for solution by an algorithm proposed by Fiacco and McCormick [15]. Let us denote $Q_K\ (g(A^K))$ by $F(Y_1, Y_2, Y_3, Y_4)$. Suppose that there are M pairs of points (X_i, X_j) such that both are in X_J. We can express the inequality constraints as

$$G_k\ (Y_i, Y_j) = D^K - (d^*_{ij})_k \geq 0 \quad k=1,\ldots M. \quad (A-4)$$

The procedure developed by Fiacco and McCormick can be applied to this problem as follows:

1. A modified objective function is formulated consisting of the original function to be minimized and penalty functions with the form:

$$P = F - r \sum_{k=1}^{M} \ell n\ G_k, \quad (A-5)$$

where r is a positive constant. As the algorithm proceeds r is evaluated to form a montonically decreasing sequence. $r_1 > r_2 > \ldots > 0$. As r grows small under suitable conditions P approaches F and the problem is solved.

2. Select a starting point (feasible or nonfeasible) and an initial value for r.

3. Determine the minimum of the objective function for the current value of r by an unconstrained gradient technique.

4. Estimate the optimal solution using extrapolation formulas.

5. Select a new value for r and repeat the procedure until convergence criteria are satisfied. A logic diagram for this process is given in Fig. 2.

REFERENCES

[1] Meisel, W.S., Computer-Oriented Approaches to Pattern Recognition. New York: Academic Press, 1972.

[2] Sammon, J.W.Jr., "A Nonlinear Mapping for Data Structure Analysis", IEEE Transactions on Computers, vol.C-18, no.5, pp.401-409, May 1969.

[3] Chang, C.L. and Lee, R.C.T., "A Heuristic Relaxation for Nonlinear Mapping in Cluster Analysis," IEEE Transactions on Systems, Man, and Cybernetics, March 1973, pp. 197-200.

[4] Calvert, T.W., "Nonorthogonal Projections for Feature Extraction in Pattern Recognition", IEEE Transactions on Computers, May 1970, pp. 447-452.

[5] White, I., "Comment on 'A Nonlinear Mapping for Data Structure Analysis'", IEEE Transactions on Computers, Feb. 1972, pp. 220-221.

[6] Hartigan, J.A., Clustering Algorithms, New York: Wiley, 1975.

[7] Duda, R.O. and Hart, P.E., Pattern Classification and Scene Analysis, New York: Wiley - Interscience, 1973.

[8] Johnson, S.C., "Hierarchical Clustering Schemes", Psychometrika, vol. 32, no. 3, pp. 241-54, 1967.

[9] Ore, O., Theory of Graphs, American Mathematical Society Colloquium Publications, vol. 38, 1962.

[10] Prim, R.C., "Shortest Connection Networks and Some Generalizations", Bell System Technical Journal, Nov. 1957, pp. 1389-1401.

[11] Zahn, C.T., "Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters," IEEE Transactions on Computers, vol. C-20, no.1, pp. 68-86, Jan. 1971.

[12] Koontz, W.L., Narendra, R.M., and Fukunaga, K., "A Graph-Theoretic Approach to Non-parametric Cluster Analysis," IEEE Transactions on Computers, vol. C-25, no. 9, pp. 945-949, Sept. 1976.

[13] Schwartzmann, D.H. amd Vidal, J.J.,"An Algorithm for Determining the Topological Dimensionality of Point Clusters," IEEE Transactions on Computers, vol. C-24, no.12, pp. 1175-1183, Dec. 1975.

[14] Shepherd, R.N., "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function," Psychometrika, vol. 27, no. 2, June 1962.

[15] Kuester, J.L. and Mize, J.H., Optimization Techniques with Fortran, New York: McGraw Hill, 1973.

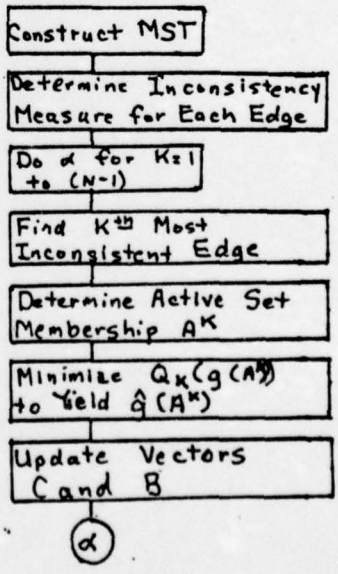

Fig. 1. Flowchart for Feature Extraction Procedure

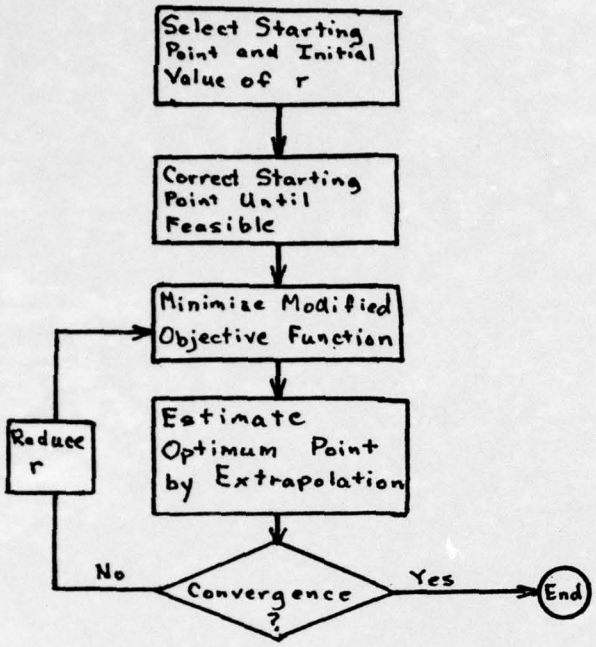

Fig. 2. Flowchart for Optimization Procedure

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR-77-1142	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER: EE-TR-7707	
4. TITLE (and Subtitle): A NEW APPROACH TO STRUCTURE PRESERVING FEATURE EXTRACTION•	5. TYPE OF REPORT & PERIOD COVERED: Interim rept.
	6. PERFORMING ORG. REPORT NUMBER: EE 7707
7. AUTHOR(s): S. A. Starks and R. J. P. de Figueiredo	8. CONTRACT OR GRANT NUMBER(s): AF-AFOSR-2777-75
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Rice University, Department of Electrical Engineering, Houston, TX. 77001	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102 F, 2304/A2
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research (N/M), Bolling AFB, Washington DC 20332	12. REPORT DATE: February 15, 1977
	13. NUMBER OF PAGES: 7
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

For Proceedings of the 1977 Conference on Information Sciences and Systems, held at Johns Hopkins University, Baltimore, Maryland on March 30 - April 1, 1977, pp. 494-500.

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

Feature extraction; pattern recognition; clustering; data structure; non-linear transformation; dimensionality reduction; data compression.

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

This paper presents an approach to nonlinear feature extraction based on certain graph theoretical considerations (such as the minimal spanning tree, maximally complete subgraphs, inconsistent edges and diameter edges) and topological considerations (such as interpoint distance measures). After appropriate introductory sections, the feature extraction algorithm is developed in section 4. The algorithm is hierarchical in nature and offers considerable savings in terms of computer computation and storage requirements. An outline of the computer procedure is also included.

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

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)