

1-1-1979

# An Approach to Nonlinear Mapping for Pattern Recognition

Nguyen Duong

L. I. Ruh-Ming

Daryl B. Simons

Follow this and additional works at: [http://docs.lib.purdue.edu/lars\\_symp](http://docs.lib.purdue.edu/lars_symp)

---

Duong, Nguyen; Ruh-Ming, L. I.; and Simons, Daryl B., "An Approach to Nonlinear Mapping for Pattern Recognition" (1979). *LARS Symposia*. Paper 289.

[http://docs.lib.purdue.edu/lars\\_symp/289](http://docs.lib.purdue.edu/lars_symp/289)

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

Reprinted from

**Symposium on**

**Machine Processing of**

**Remotely Sensed Data**

**June 27 - 29, 1979**

The Laboratory for Applications of  
Remote Sensing

Purdue University  
West Lafayette  
Indiana 47907 USA

IEEE Catalog No.  
79CH1430-8 MPRSD

Copyright © 1979 IEEE  
The Institute of Electrical and Electronics Engineers, Inc.

Copyright © 2004 IEEE. This material is provided with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the products or services of the Purdue Research Foundation/University. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

# AN APPROACH TO NONLINEAR MAPPING FOR PATTERN RECOGNITION

NGUYEN DUONG, RUH-MING LI, AND  
DARYL B. SIMONS

Colorado State University

## I. INTRODUCTION

Nonlinear mapping has been used in the past for data structure analysis. Many interesting heuristic approaches to map  $n$ -dimensional data to a lower dimensional space such that the local structure of the original data is preserved have been proposed by Kruskal,<sup>1</sup> Sammon,<sup>2</sup> and others. Basically, a criterion for local structure preservation is first defined, then a new data configuration is obtained by an iterative process to minimize the selected criterion. Let the given data sample consist of  $N$   $n$ -dimensional points  $X_1, X_2, \dots, X_N$ , and  $Y_1, Y_2, \dots, Y_N$  are vectors in an  $m$ -dimensional space which correspond to  $X_1, X_2, \dots, X_N$ . Usually  $m$  is smaller than  $n$ . Let  $d_{ij}^*$  denote the distance between  $X_i$  and  $X_j$  and, similarly,  $d_{ij}$  denotes the distance between  $Y_i$  and  $Y_j$ , then the criterion to be minimized is

$$J = \frac{1}{\sum_{i < j} d_{ij}^*} \sum_{i < j} \frac{(d_{ij}^* - d_{ij})}{d_{ij}^*} \quad (1)$$

Sammon used a steepest descent procedure to search for a minimum of  $J$ . His approach requires the computation and storage of the interdistance matrix, which consists of  $N(N-1)/2$  elements. Even for a moderate  $N$ , this matrix is forbiddingly large.

Chang and Lee<sup>3</sup> developed a heuristic "frame method" to accommodate large data sets. From  $M < N$  points chosen from the original sample, a frame in a two-dimensional space is derived through a relaxation method. The remaining  $N-M$  points are then added to the transformed map by adjusting their positions to the points in the frame.

To avoid the inherent distortion of the "frame method," Schachter<sup>4</sup> proposed to partition the feature space into bins (hyper-rectangles) and replace each bin by its center and its weight, which is the number of points falling within it. Thus the number of points to be considered in the mapping

is only  $M < N$ . Again, the relaxation method is used to derive a recurrence relation for non-linear mapping.

All the methods described above cannot handle very large data samples, which are frequently encountered in regional analysis, and cannot be used directly for classification of multispectral Landsat data.

In this paper, a two-stage process to nonlinear mapping for pattern classification is proposed. In the first stage, a data "straightening operation" is used to obtain the new configuration of the training data points in a one-dimensional space, and the Group Method of Data Handling (GMDH) technique, suggested by Ivakhnenko and his co-workers at Kiev in the U.S.S.R.,<sup>5,6</sup> is used to derive the global nonlinear mapping function for the given data sample. In the second stage, pattern classification is done in the one-dimensional transformed space through the use of a minimum-distance decision rule.

## II. DATA STRAIGHTENING OPERATION ON THE TRAINING SET

The following conditions are assumed to be satisfied in the one-dimensional transformed space:

- (1) Each transformed cluster would be a tight cluster; therefore the cluster means can be used to represent cluster location in the transformed space; and
- (2) Local structure, defined as the distances from a cluster to its nearest two neighbors, would be preserved.

If local structure distortion is chosen as an evaluation criterion for selecting the new configuration of the training set in the transformed space, then one would expect to have a data "straightening operation" which starts at one of the two points having the greatest pairwise distance in the original space. The following example will illustrate the above idea.

Let A, B, and C be three points which constitute the training set in the original space (see Figure 1). The data straightening operation is an optimal arrangement of all pairwise distances in the transformed space to minimize the local structure distortion. To start the data straightening operation, a starting node is selected. Then, depending upon the selected starting node and the

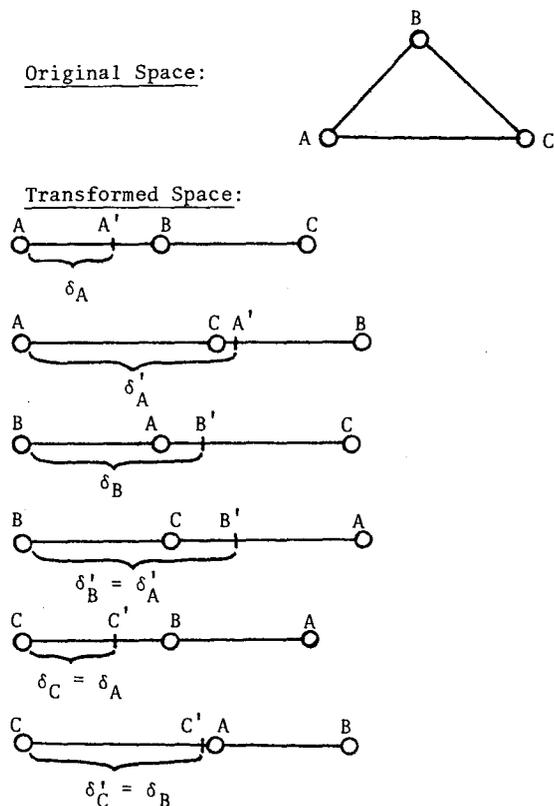


Figure 1. Data Structure in the Original and Transformed Spaces.

direction of displacement, one can have the various paths as indicated by (a), (b), (c), (d), (e), and (f) in Figure 1, where  $\delta_I$  and  $\delta'_I$  represent the amounts of local structure distortion committed by the two paths departing from node I. From these, two observations can be made:

- (1)  $\delta_I$  is always less than  $\delta'_I$ ; i.e., given a starting node I, the shortest path joining I and the other training points will give less local structure distortion compared with other paths departing from the same node.
- (2) The shortest path departing from one of the two points (either A or C) which have the greatest pairwise distance will give minimum local structure distortion.

The following algorithm can then be derived for the data straightening operation on the training set:

- (1) Calculate all pairwise distances in the training set, based on some distance measure;
- (2) Select one of the two points which have the largest pairwise distance as starting node;
- (3) Find the shortest path departing from the selected node and going through all other points in the set; and
- (4) Straighten this path over the transformed (one-dimensional) space.

The problem of finding the shortest path has already been solved by many good algorithms available in the literature. The one proposed by Dantzig<sup>7</sup> is perhaps the most simple but efficient one to be used for this type of problem.

### III. GLOBAL NONLINEAR MAPPING FUNCTION IDENTIFICATION

After straightening the training set in the transformed space, a global nonlinear mapping function has to be derived to map the training data from the original space to its transform in the transformed space. For this purpose, the Group Method of Data Handling (GMDH) technique proposed by Ivakhnenko will be used because of the following advantages:

- (1) It can handle a large number of input variables with strong unknown interaction between them;
- (2) The training sample size may be small which makes the other traditional statistical tests meaningless;
- (3) More than one criterion can be used for model evaluation; and
- (4) It allows the combination of outputs of a number of competing models to improve the mapping.

Following is a brief description of the GMDH concept. Detailed discussions can be found in Mehra's paper.<sup>8</sup>

GMDH uses polynomial models, which are discrete-time analogs of Volterra Series Models introduced into nonlinear analysis by Wiener.<sup>9</sup> The GMDH structure is based on Rosenblatt<sup>10</sup> perceptron concept. A schematic diagram of GMDH is shown in Figure 2. Each layer of GMDH consists of a set of quadratic polynomial functions, called partial descriptions of the final model, of the form

$$y_i = a_0 + a_1 x_j + a_2 x_k + a_3 x_j^2 + a_4 x_k^2 + a_5 x_j x_k \quad (2)$$

with inputs from the previous layer having been passed through a selection layer. The principles of the multilayer GMDH threshold algorithm can be summarized as follows:

- (1) First, include as many variables (i.e. features) as possible that could have (even a remote) influence on y (the desired trans-

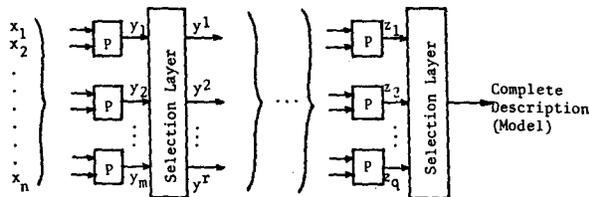


Figure 2. Structure of the GMDH Model.

- formed value) to the first layer.
- (2) Then, the optimal partial descriptions selected to the next layer will be decided by a threshold criterion of goodness of fit.
  - (3) The process is continuously repeated with the imposition of even more rigid thresholds on the norm, so that finally a unique "model" is selected at the output of one of the layers. This "model" corresponds to a minimum of the selection criterion.

In GMDH, three selection criteria have been used frequently by Ivakhnenko and his coworkers, viz.: (a) regularity, (b) unbiasedness, and (c) balance-of-variables. The effects of these criteria on model identification results have been discussed by Mehra.<sup>8</sup> In general, one first uses a "smooth" criterion (i.e., regularity) to select the region of "feasible" models, and then uses a "sharp" criterion (i.e., unbiasedness) to determine one of them exactly.

#### IV. PATTERN CLASSIFICATION IN THE TRANSFORMED SPACE

Let  $F(X)$  be the global nonlinear mapping function derived by the GMDH technique discussed above, then  $F(X)$  will take the necessary values within a specified range for all the elements of each class. Let  $\bar{F}_j(X)$  denote the mean value of  $F(X)$  for all elements of class  $C_j$  in the transformed space, then a decision rule for the classification of a transformed point  $X$  can be constructed as follows:

$$X \in C_j \text{ if } |F(X) - \bar{F}_j(X)| = \min_i |F(X) - \bar{F}_i(X)| \quad (3)$$

for  $i = 1, 2, \dots, N_c$

where  $N_c$  is the total number of classes.

Other classification algorithms can also be used as well. The simplest of these, the histogram thresholding technique, has often shown good classification results due to the enhancement of

the transformed data by the nonlinear mapping process.

In summary, the proposed nonlinear mapping algorithm for pattern classification consists of the following steps:

- (1) Select a training set and calculate class-statistics (i.e., mean and standard deviation);
- (2) Find the shortest path going through all the sample means of various classes and minimize the local structure distortion (as defined previously);
- (3) Generate the transformed training set through data straightening operation, and calculate the "transformed statistics";
- (4) Derive the global nonlinear mapping function based on training data, using the multilayer GMDH thresholding technique;
- (5) Map the complete scene from the  $n$ -dimensional feature-space to the one-dimensional feature-space; and
- (6) Classify the transformed data using decision rule (3) or other selected classification algorithm.

#### V. SAMPLE TEST CASES

The proposed approach to nonlinear mapping for pattern classification has been used to detect crop field boundaries in Landsat/MSS imagery and also for land-use classification in two selected test areas in Colorado. One is an agricultural field near Fort Collins, and the other is a portion of the mountainous area surrounding the Red Feather Lake. Only bands 5, 6, and 7 of the Landsat/MSS data were used in this study. Band 4 was discarded because of its noisy characteristics. Graymaps of the agricultural test area are given in Figures 3a, 3b, 3c, and 3d for bands 4, 5, 6, and 7, respectively.

Based on ground-truth information, a training set was selected for the six crop types in the area, viz.: corn, beets, beans, alfalfa, barley, and oats. The means and standard deviations for these classes are given in Table 1. The new values for class-means after the data straightening operation are given in Table 2. A multilayer sequential GMDH threshold algorithm developed by the authors was then used to identify the global nonlinear mapping function based on the selected training set and the results of the data straightening operation. The basic class-statistics in the one-dimensional transformed space are illustrated in Table 3 and the expression for the mapping function is given by the following equations:

$$\text{First layer } \begin{cases} Y^1 = -10.1929977 \cdot X_3 - 1685.89107 \cdot X_6 + \\ \quad .021669599 \cdot (X_3)^2 + 716.049627 \cdot (X_6)^2 \\ \quad + 2.63359623 \cdot X_3 \cdot X_6 + 1862.17649 \\ Y^2 = -9.46747278 \cdot X_3 - 514.312341 \cdot X_4 + \\ \quad .0242106725 \cdot (X_3)^2 + 189.610666 \cdot (X_4)^2 \\ \quad + 2.66900244 \cdot X_3 \cdot X_4 + 1135.37608 \end{cases}$$



	BAND	MEAN	STDV
CORN	5	33.979	1.163
	6	119.428	2.951
	7	146.326	3.584
BEETS	5	33.029	6.162
	6	143.617	25.512
	7	164.323	29.480
BEANS	5	41.666	5.989
	6	122.666	9.668
	7	135.666	14.841
ALFALFA	5	49.400	14.836
	6	90.800	31.428
	7	88.533	33.521
OAT	5	66.333	6.623
	6	114.333	3.933
	7	116.833	4.446
BARLEY	5	88.111	30.271
	6	109.222	43.303
	7	101.444	39.938

Table 1. Basic Class-statistics in the Original Feature-space.

BEETS	220.72
CORN	250.89
BEANS	264.42
OAT	296.55
BARLEY	323.70
ALFALFA	368.89

Table 2. Results of the Data Straightening Operation.

	MEAN	STDV
BEETS	221.28	4.82
CORN	249.78	1.68
BEANS	269.88	5.96
OAT	307.73	8.24
BARLEY	321.26	5.52
ALFALFA	357.90	14.46

Table 3. Basic Class-statistics in the Transformed Space.

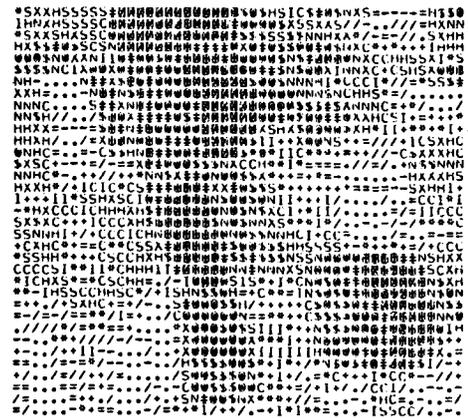


Figure 4. Graymap of the Test Area in the Transformed Space.

Comparing the boundary information in Figure 5d with the combination of boundary information in the three Figures 5a, 5b, and 5c, one can see that most of the information in the original feature-space has been transferred efficiently to the transformed space through the nonlinear mapping operation defined by Equation (4). Using algorithm (3) for pattern classification, the following result was also obtained (see Figure 6). The classification accuracy is a little bit lower compared with a layered classifier using the three original spectral bands 5, 6, and 7. However, higher classification accuracy could be obtained if one would increase the number of layers in the GMDH or refine the selection thresholds.

Similar analysis was also carried out for the Red Feather Lake test area, and the classification result is presented in Figure 8, compared with a graymap from channel 6 (Figure 7).

## VI. CONCLUSIONS

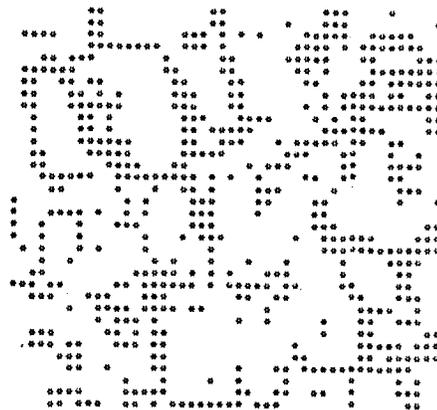
A nonlinear mapping for pattern classification using the GMDH technique has been investigated. This approach offers the following advantages:

- (1) It can automatically identify the unknown relationship between the selected spectral bands and reflect the optimal combination of these bands in the expression of the derived nonlinear mapping function;
- (2) It can be used to compress the information contained in the various spectral bands without major loss; and
- (3) It can be used as a means to combine the boundary information in the various spectral bands.

Future studies on the optimality of the data straightening operation for local structure preservation and on the improvement of the GMDH algorithm



(a)



(b)



(c)



(d)

Figure 5. Boundary Enhancement Results for (a) Band 5, (b) Band 6, (c) Band 7, and (d) Transformed Feature-space.

are desirable. Works are underway at Colorado State University to apply this approach to data structure analysis. For this case, a two-dimensional transformed feature-space is used.

#### VII. REFERENCES

- <sup>1</sup>Kruskal, J. B., "Mathematical Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis," *Psychometrika*, Vol. 29, pp. 1-27, 1964.
- <sup>2</sup>Sammon, J. W., Jr., "A Nonlinear Mapping for Data Structure Analysis," *IEEE Trans. Computer*, Vol. C-18, pp. 401-409, May 1969.
- <sup>3</sup>Chang, C. L. and C. T. Lee, "A Heuristic Relaxa-

tion Method for Nonlinear Mapping in Cluster Analysis," *IEEE Trans. Syst. Man. and Cyber.*, pp. 197-200, March 1973.

- <sup>4</sup>Schachter, B., "A Nonlinear Mapping Algorithm for Large Data Sets," *Computer Graphics and Image Processing*, Vol. 8, pp. 271-276, 1978.
- <sup>5</sup>Ivakhnenko, A. G., "Heuristic Self-Organization in Problems of Engineering Cybernetics," *Automatic*, 1970.
- <sup>6</sup>Ivakhnenko, A. G., "Polynomial Theory of Complex Systems," *IEEE Trans. Syst. Man. and Cyber.*, Vol. 1, No. 4, October 1971.
- <sup>7</sup>Dantzig, G. B., "On the Shortest Route Through a Network," *MAA Studies in Mathematics*, Vol. 11: *Studies in Graph Theory - Part I*; Ed. D. R. Fulkerson.

