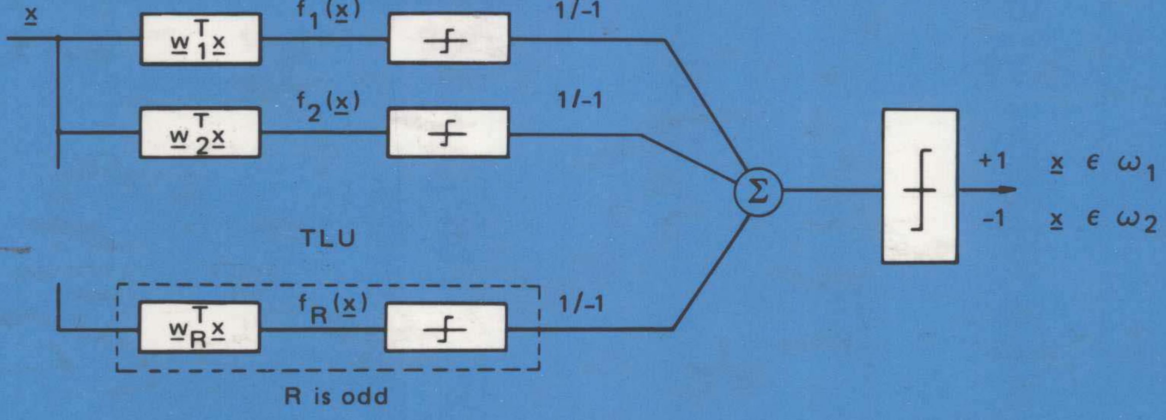
Electrical Engineering and Electronics/23

Pattern Recognition

Applications to Large Data-Set Problems



Sing-Tze Bow

PATTERN RECOGNITION

Applications to Large Data-Set Problems

Sing-Tze Bow
Department of Electrical Engineering
The Pennsylvania State University
University Park, Pennsylvania

MARCEL DEKKER, INC. New York and Basel

Library of Congress Cataloging in Publication Data

Bow, Sing-Tze, [Date]
Pattern recognition.

(Electrical engineering and electronics ; 23)
Bibliography: p.
Includes index.
1. Pattern recognition systems. I. Title.
II. Series.
TK7882.P3B69 1984 001.64'42 84-11420
ISBN 0-8247-7176-1

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

Preface

The purpose of writing this book is twofold: (1) to provide readers with the basic principles involved with the approaches currently employed in pattern recognition; and (2) to demonstrate use of the theories with relevant practical problems so that the theories may be better understood.

The materials collected in this book are grouped into two main parts. Part I emphasizes the principles of pattern recognition, and Part II deals with the preprocessing of data for pattern recognition. There are five chapters in Part I. Chapter 1 presents the fundamental concept of pattern recognition and its system configuration. Selected applications, including weather forecasting, handprinted character recognition, speech recognition, medical analysis, and satellite and aerial-photo interpretation, are discussed briefly. Also in Chapter 1, the two principle approaches used in pattern recognition, the decision theoretic and syntactic approaches, are described and compared. The remaining chapters in Part I focus primarily on the decision theoretic approach. Because of space limitations, the syntactic approach is left for another publication. Chapters 2 and 3 discuss some principles involved in nonparametric decision theoretic classification and the training of the discriminant functions used in these classifications. Chapter 4 introduces the principles of statistical decision theory in classification problems. In recent years, a great many advances have been made in the field of clustering, but because of space limitations and the need to be systematic, the material in Chapter 5 is selected and organized to make readers aware of current trends and the main thrust in attempts to solve pattern recognition problems, so that readers will have no difficulty following the current literature after they read this book.

In Part II emphasis is on the preprocessing of original data for accurate and correct pattern recognition. Appropriate preprocessing of original data has a considerable effect on proper selection of a method for pattern recognition. Chapter 6 discusses dimensionality reduction and feature selection, which are necessary measures in making machine recognition feasible. In that chapter attention is given to the optimum number of features and their ordering, to canonical analysis and its application to large data-set problems, and to the nonparametric feature selection method, which is applicable to pattern recognition problems based on mixed features. Chapters 7 and 8 are devoted primarily to the methodology employed in preprocessing a large data-set problem. More concretely, complex problems such as scenic images are used for illustration. Processing in both the spatial and transform domains is considered in detail.

A set of seven 512 × 512 256-gray-level images are included in Appendix A. These images can be used as large data sets to illustrate many of the pattern recognition and data preprocessing concepts developed in the text. They can be used in their original form and can be altered to generate a variety of input data sets.

This book is the outgrowth of two graduate courses developed for the Department of Electrical Engineering of The Pennsylvania State University: "Principles of Pattern Recognition" and "Digital Image Processing." This material has been rewritten to suit both graduate students and senior undergraduates with high grade-point averages. The book can be used as a one-semester course on pattern recognition by omitting coverage of some of the material. It can also be used as a two-semester course with the addition of some computer projects similar to those suggested herein. The book can also serve as a reference for engineers and scientists involved with pattern recognition, digital image processing, and artificial intelligence.

The author is indebted to Dale M. Grimes, Head of the Department of Electrical Engineering of The Pennsylvania State University, for his encouragement and support during the preparation of manuscript. The author also wishes to thank George J. McMurtry, Associate Dean of the College of Engineering at The Pennsylvania State University, for his valuable discussions and his generous permission to freely use some of his class notes in Chapters 3 and 4 and in the discussion of canonical analysis and its application to large data-set problems.

Sing-Tze Bow

Contents

part I
PRINCIPLES OF PATTERN RECOGNITION

1

Introduction

1.1 PATTERNS AND PATTERN RECOGNITION

The patterns we encounter can fall into two categories: abstract and concrete. Examples of abstract items include ideas and arguments. Recognition of such patterns, termed *conceptual recognition*, belongs to another branch of artificial intellegence and is beyond the scope of this book.

Examples of concrete items include characters, symbols, pictures, biomedical images, three-dimensional physical objects, target signatures, speech waveforms, electrocardiograms, electroencephalograms, and seismic waves. Some of these items are spatial, whereas others are temporal. In the last couple of decades, interest has focused on two types of pattern recognition problems:

1. The mechanism of the pattern recognition system possessed by living organisms. Psychologists, physiologists, biologists, and neurophysiologists have devoted considerable effort toward exploring how living things perceive objects. Most of their results have been reported in the literature of bionics and other relevant disciplines.
2. The development of theories and techniques for computer implementation of a given recognition task. This is a subject that currently challenges both engineers and applied mathematicians. There is no unifying theory available that can be applied to all kinds of pattern recognition problems. Most techniques are problem oriented. Systematic presentation of the theories and techniques forms the basis of this book.

1.2 CONFIGURATION OF THE PATTERN RECOGNITION SYSTEM

1.2.1 Three Phases in Pattern Recognition

In pattern recognition we can divide an entire task into three phases: data acquisition, data preprocessing, and decision classification, as shown in Fig. 1.1. In the data acquisition phase, analog data from the physical world are gathered through a transducer and converted to digital format suitable for computer processing. In this stage, the physical variables are converted into a set of measured data, indicated in the figure by electric signals, $x(r)$, if the physical variables are sound (or light intensity) and the transducer is a microphone (or photocells). The measured data are then used as the input to the second phase (data preprocessing) and grouped into a set of characteristic features (x_N) as output. The third phase is actually a classifier which is in the form of a set of decision functions. With this set of features (x_N) the object may be classified. Figure 1.2 is a schematic diagram of an actual aerial multispectral scanner and data analysis system. The set of data at B, C, and D are in the pattern space, feature space, and classification space, respectively.

1.2.2 Representation of Pattern and Approaches to Their Machine Recognition

In Multidimensional Vector Form

As discussed in Sec. 1.2.1, there will be a set of collected, measured data after data acquisition. If the data to be analyzed are physical objects or images, the data acquisition device can be a television camera, a high-resolution camera, a multispectral scanner, or other device. For other types of problems, such as economic problems, the data acquisition system can be a data tape.

One function of data preprocessing is to convert a visual pattern into an electrical pattern or to convert a set of discrete data into a mathematical pattern so that those data are more suitable for computer analysis. The output will then be a pattern vector, which appears as a point in a pattern space.

To clarify this idea, let us take a simple visual image as the system input. If the image is scanned by a 12-channel multispectral scanner, we obtain, for a single picture point, 12 values, each corresponding to a separate spectral response. If the image is treated as a color image, three fundamental color-component values can be obtained, each corresponding, respectively, to a red, green, or blue spectrum band.

Each spectrum component value can be considered as a variable in n-dimensional space, known as *pattern space*, where each spectrum component is assigned to a dimension. Each pattern then appears as a point in the pattern space. It is a vector composed of n component

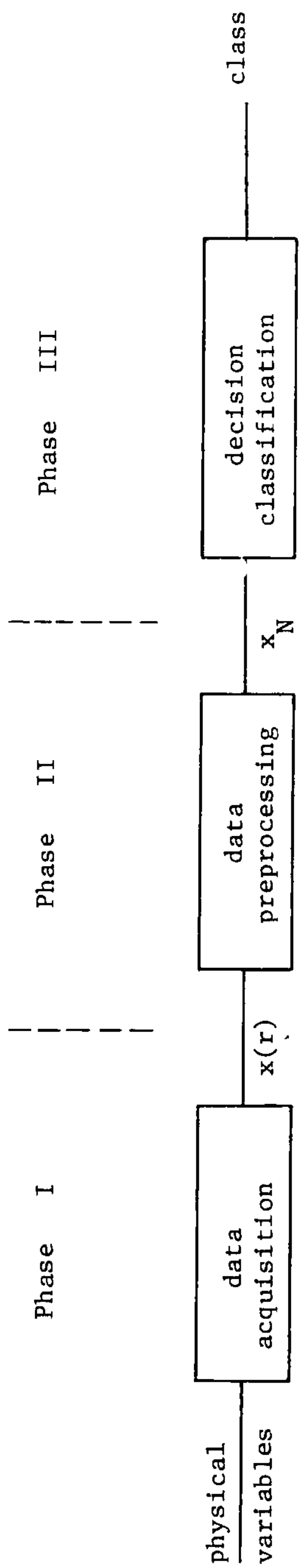


FIG. 1.1 Conceptual representation of a pattern recognition problem.

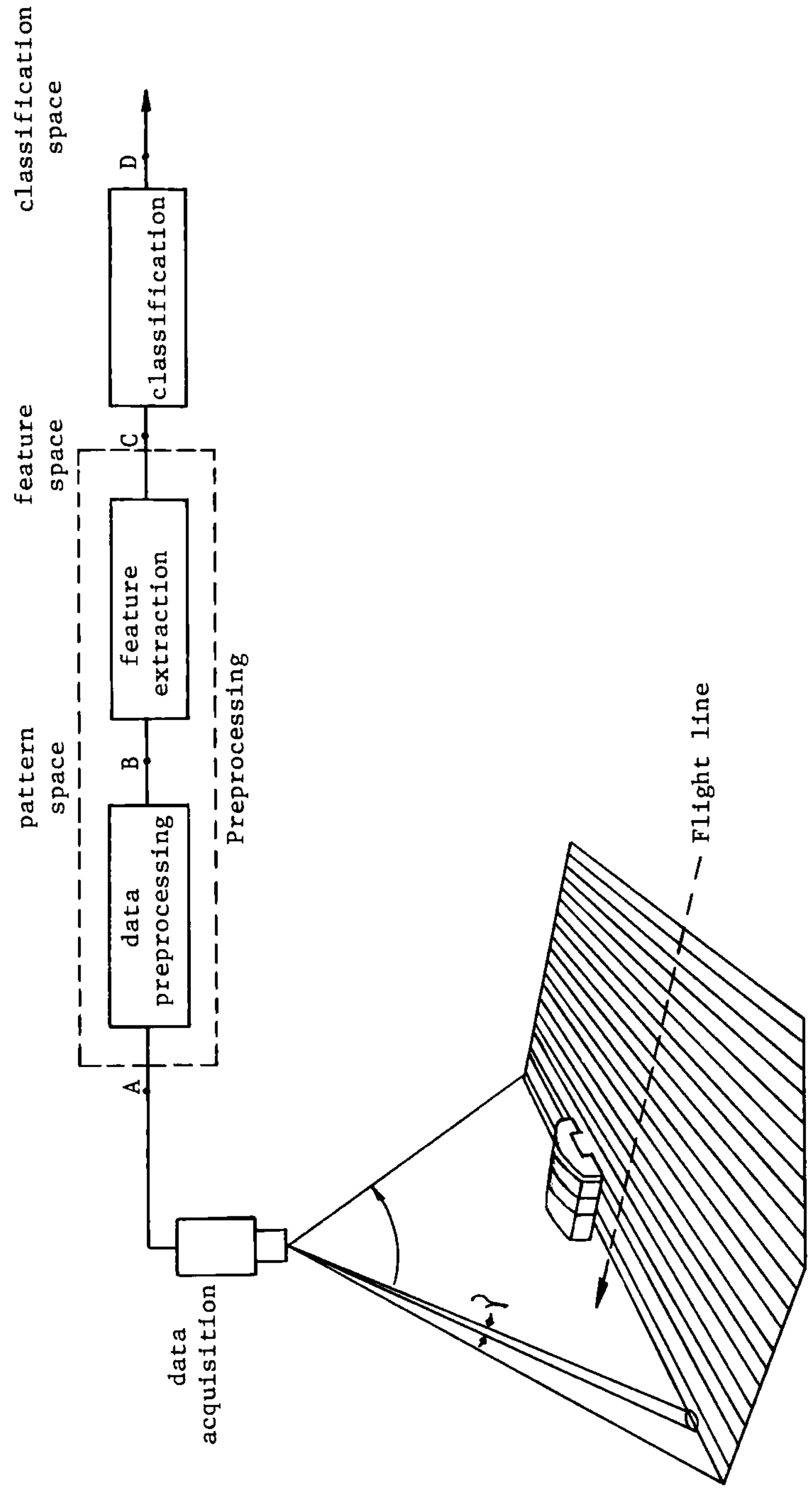


FIG. 1.2 Multispectral scanner and data analysis system.

values in the n-dimensional coordinates. A pattern $\underline{x}$ can then be represented as

$$\underline{x} = \begin{vmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{vmatrix} \tag{1.1}$$

where the subscript n represents the number of dimensions. If $n < 3$, the space can be illustrated graphically. Pattern space $\underline{X}$ may be described by a vector of m pattern vectors such that

$$\underline{X} = \begin{vmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \cdot \\ \cdot \\ \cdot \\ \underline{x}_m^T \end{vmatrix} = \begin{vmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{vmatrix} \tag{1.2}$$

where the superscript T after each vector denotes its transpose, and the $\underline{x}_i^T = (x_{i1}, x_{i2}, \ldots, x_{in})$, $i = 1, \ldots, m$, represent pattern vectors.

The objective of the feature extraction shown in Fig. 1.2 functions as the dimensionality reduction. It converts the original data to a suitable form (feature vectors) for use as input to the decision processor for classification. Obviously, the feature vectors represented by

$$\underline{x}_i^T = (x_{i1}, x_{i2}, \ldots, x_{ir}) \qquad i = 1, \ldots, m \tag{1.3}$$

are in a smaller dimension (i.e., $r < n$).

The decision processor shown in Fig. 1.2 operates on the pattern vector and yields a classification decision. As we discussed before, pattern vectors are placed in the pattern space as "points," and patterns belonging to the same class will cluster together. Each cluster represents a distinct class, and clusters of points represent different classes of patterns. The decision classifier implemented with a set of decision function serves to define the class to which a particular pattern belongs.

The output of the decision processor will be in the classification space. It is M-dimensional if the input patterns are to be classified into M classes. For the simplest two-class problem, M equals 2; for aerial-photo interpretation, M can be 10 or more; and for alphabet recognition M equals 26. But for the case of Chinese character recognition, M can be more than 10,000. In such a case, other representations have to be used as supplements.

Both the preprocessor and the decision processor are usually selected by the user or designer. The coefficients (or weights) used in the decision processor are either calculated on the basis of complete a priori information of statistics of patterns to be classified, or are adjusted during a training phase. During the training phase, a set of patterns from a training set is presented to the decision processor, and the coefficients are adjusted according to whether the classification of each pattern is correct or not. This may then be called an *adaptive* or *training* decision processor. Note that most of the pattern recognition systems are not adaptive on-line; this is so only during the training phase. Also note that the preprocessing and decision algorithms should not be isolated from each other. Frequently, the preprocessing scheme has to be changed to make the decision processing more effective.

As discussed previously, a priori knowledge as to correct classification of some data vectors is needed in the training phase of the decision processor. Such data vectors are referred to as *prototypes* and are denoted as

$$\underline{z}_k^m = \begin{vmatrix} z_{k1}^m \\ \cdot \\ \cdot \\ \cdot \\ z_{ki}^m \\ \cdot \\ \cdot \\ \cdot \\ z_{kn}^m \end{vmatrix} \qquad \begin{aligned} k &= 1, 2, \ldots, M \\ m &= 1, 2, \ldots, N_k \end{aligned}$$

where $k = 1, 2, \ldots, M$ indexes the particular pattern class; $m = 1, 2, \ldots, N_k$ indicates the mth prototype of the class ω_k; and $i = 1, 2, \ldots, n$ indexes its component in the n-dimensional pattern vector. M, N_k, and n denote, respectively, the number of pattern classes; the number of prototypes in the kth class, ω_k; and the number of dimensions of the pattern vectors.

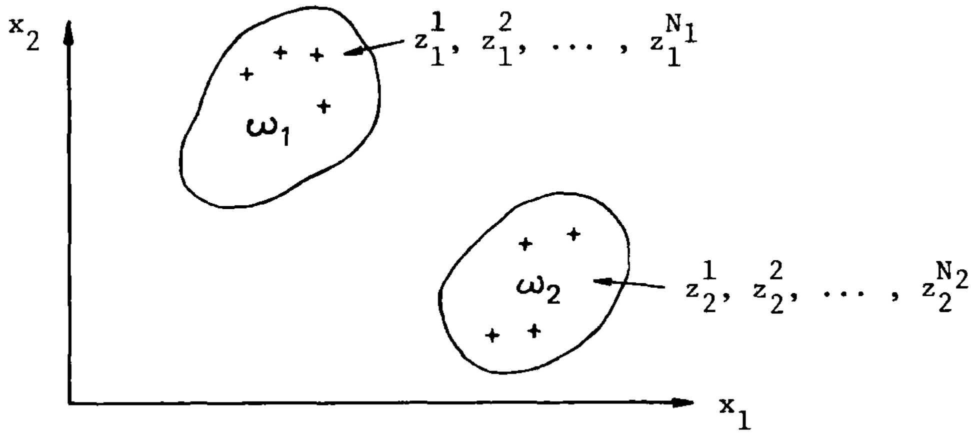


FIG. 1.3 Simple two-dimensional pattern space.

Prototypes from the same class share the same common properties and thus they cluster in a certain region of the pattern space. Figure 1.3 shows a simple two-dimensional pattern space. Prototypes $\underline{z}_1^1, \underline{z}_1^2, \ldots, \underline{z}_1^{N_1}$ cluster in ω_1; prototypes of another class, $\underline{z}_2^1, \underline{z}_2^2, \ldots, \underline{z}_2^{N_2}$, cluster in another region of the pattern space, ω_2. N_1 and N_2 are the number of prototypes in classes ω_1 and ω_2, respectively. The classification problem will simply be to find a separating surface that partitions the known prototypes into correct classes. This separating surface is expected to be able to classify the other unknown patterns if the same criterion is used in the classifier. Since patterns belonging to different classes will cluster into different regions in the pattern space, the distance metric between patterns can be used as a measure of similarity between patterns in the n-dimensional space.

Some conceivable properties between the distance metrics can be enumerated; thus

$$d(\underline{x},\underline{y}) = d(\underline{y},\underline{x})$$

$$d(\underline{x},\underline{y}) \leqslant d(\underline{y},\underline{z}) + d(\underline{z},\underline{x})$$

$$d(\underline{x},\underline{z}) \geqslant 0$$

$$d(\underline{x},\underline{y}) = 0 \quad \text{iff} \quad \underline{y} = \underline{x}$$

where $\underline{x}$, $\underline{y}$, and $\underline{z}$ are pattern vectors and $d(\cdot)$ denotes a distance function. Details regarding pattern classification by this approach are presented in subsequent chapters.

In Linguistically Descriptive Form

What we have just discussed is that each pattern is represented by a feature vector. The recognition of each pattern is usually made by