Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 2731–2741
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Statistical pattern recognition in remote sensing

Chi Hau Chen*, Pei-Gee Peter Ho

Electrical and Computer Engineering Department, University of Massachusetts Dartmouth, 285 Old Westport Road, N. Dartmouth, MA 02747, USA

Received 15 February 2008; received in revised form 17 April 2008; accepted 18 April 2008

Abstract

Remote sensing with sensors mounted on satellites or aircrafts is much needed for resource management, environmental monitoring, disaster response, and homeland defense. Remote sensing data considered include those from multispectral, hyperspectral, radar, optical, and infrared sensors. Classification is often one of the major tasks in information processing. For example, we need to identify vegetations, waterways, and man-made structures from remote sensing of earth. The large amount of data available makes remote sensing data uniquely suitable for statistical pattern recognition. This paper will address several issues on statistical pattern recognition that are related to information processing in remote sensing. Though the paper is largely tutorial in nature, some specific issues considered are image models for characterization of contextual information, neural networks for image classification, and the performance measures.

Either to supplement the capability of sensors or to effectively utilize the enormous amount of sensor data, many advances in statistical pattern recognition can be very useful in machine recognition of the data. The potentials and opportunities of using statistical pattern recognition in remote sensing are indeed unlimited.

Keywords: Remote sensing; Statistical pattern classification; Contextual information; Neural networks; Support vector machine; Vector 2-D autoregressive time series; Markov random field

1. Introduction

Basically statistical pattern recognition deals mainly with correctly classifying a pattern into one of several available pattern classes. Basic topics in statistical pattern recognition include: pre-processing, feature extraction and selection, parametric or nonparametric probability density estimation, decision making processes, performance evaluation, post-processing as needed, supervised and unsupervised learning, or training, and cluster analysis.

The large amount of data available makes remote sensing data uniquely suitable for statistical pattern recognition. Though there are many challenges, statistical pattern recognition has found important application in the remote sensing problems. Remote sensing data considered include those from multispectral, hyperspectral, radar, optical, and infrared sensors mounted on satellite or high altitude aircrafts. With improved sensors, a high resolution remote sensing image can easily provide a spatial resolution of 1 foot $\times$ 1 foot, or smaller, per pixel. What is in a pixel, especially its classification information can be important for image interpretation. In this paper we assume each pixel being pure, i.e. it corresponds to a single class. The "mixed" pixel problem would require additional physical information for a good classification. The availability of reliable ground truth is often a problem in remote sensing pattern classification and the problem would be further complicated with mixed pixels. The paper begins with a brief review of the statistical pattern recognition in remote sensing and then addresses the issues of feature extraction, statistical classifiers, contextual image modeling, neural network and support vector machine (SVM) classifiers, and the performance measures.

2. Historical development

The theory of statistical pattern recognition was largely developed during 1960s and 1970s. Some major development includes the formulation of pattern recognition as a Bayes

* Corresponding author. Fax: +1 508 999 8489.
E-mail address: cchen@umassd.edu (C.H. Chen).

0031-3203/$30.00 © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2008.04.013

decision theory problem [1] nearest neighbor decision rules (NNDRs) and density estimation [2], use of Parzen density estimate in nonparametric pattern recognition [3], leave-one-out method of error estimate [4], use of statistical distance measures and error bounds in feature evaluation [5], Fisher linear discriminant and multicategory generalizations [6], supervised parameter estimation and unsupervised learning by decomposition of mixture densities [7], and K-mean algorithm [8], etc. In recent years, there has been much work on combining classifiers (see, e.g. Refs. [9,10]). Also the use of artificial neural networks as well as SVMs for classification has been very popular since the latter part of 1980s. As neural networks and SVMs have been developed in the last 25 years, they are closely linked to statistical classifiers (see, e.g. Refs. [11–13]) though they are not considered in our view as subsets of statistical classifiers.

These developments clearly have significant impact on information processing in remote sensing. The first extensive application of statistical pattern recognition is probably by the Laboratory for Agricultural Remote Sensing (LARS) at Purdue University in 1965. The early paper by Fu [14] demonstrated the feasibility of using statistical pattern recognition in the classification of crops from high altitude aircraft and satellite image data. For a multispectral sensor for land survey typically with eight spectral bands, or component images, each pixel vector $\mathbf{x}$ has eight elements. Fu demonstrated that each component image has a histogram which is approximately Gaussian distributed. The multivariate Gaussian distribution can be written as

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\} \tag{1}$$

where $\Sigma = E\{(\mathbf{x}-\mu)(\mathbf{x}-\mu)^{\mathrm{T}}\}$ is the covariance matrix, $\mu = E(\mathbf{x})$ is the mean of the random vector $\mathbf{x}$, and d is the vector dimension or the number of component images considered. The emphasis on "accurate statistics", i.e. an accurate covariance matrix, in a large amount of work by Landgrebe [15] and his associates has had a major impact on statistical pattern recognition study in remote sensing. The classification work in remote sensing in fact has employed very similar pattern recognition techniques as in character recognition [16] and biometrics [17] and many other areas, though there is less success in remote sensing because of the complexity and diversity of data considered. The use of contextual information in remote sensing image interpretation dates back to the early paper by Welch and Salter [18] and the book by Fu and Yu [19]. They considered the statistical dependence of neighboring pixels which is referred to as contextual information in the paper. However, the compound decision rule they employed is complicated. A more complete contextual information should include the statistical dependence among component images. The covariance matrix stated above carries the second order statistical information among the component images. Texture features based

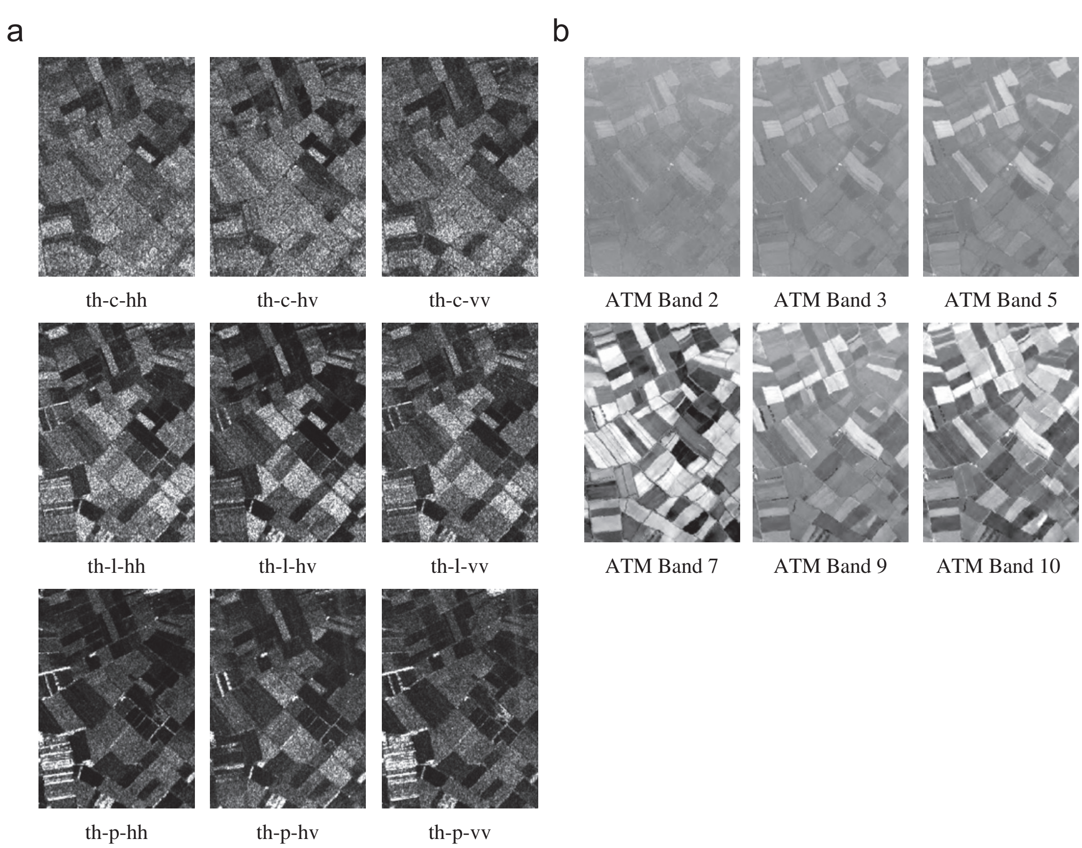

Fig. 1. (a) The nine channel SAR images of UK village. (b) The thermal mapper (ATM) images of UK village.