Pattern Recognition 41 (2008) 3313 – 3327

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A projection pursuit algorithm for anomaly detection in hyperspectral imagery

José A. Malpica[a,*], Juan G. Rejas[b], María C. Alonso[a]

[a]Dpto. de Matemáticas, Universidad de Alcalá, Campus Universitario, 28871 Alcalá de Henares, Madrid, Spain
[b]Dpto. de Observación de La Tierra, Instituto Nacional de Técnica Aeroespacial (INTA), Teledetección y Atmósfera, Carretera de Ajalvir km 4. 28850, Torrejón de Ardoz, Madrid, Spain

ARTICLE INFO

Article history:
Received 11 November 2007
Received in revised form 9 April 2008
Accepted 15 April 2008

Keywords:
Hyperspectral imagery
Projection pursuit
Simulated annealing
Simplex optimization

ABSTRACT

The main goal of this paper is to propose an innovative technique for anomaly detection in hyperspectral imageries. This technique allows anomalies to be identified whose signatures are spectrally distinct from their surroundings, without any a priori knowledge of the target spectral signature. It is based on an one-dimensional projection pursuit with the Legendre index as the measure of interest. The index optimization is performed with a simulated annealing over a simplex in order to bypass local optima which could be sub-optimal in certain cases. It is argued that the proposed technique could be considered as seeking a projection to depart from the normal distribution, and unfolding the outliers as a consequence. The algorithm is tested with AHS and HYDICE hyperspectral imageries, where the results show the benefits of the approach in detecting a great variety of objects whose spectral signatures have sufficient deviation from the background. The technique proves to be automatic in the sense that there is no need for parameter tuning, giving meaningful results in all cases. Even objects of sub-pixel size, which cannot be made out by the human naked eye in the original image, can be detected as anomalies. Furthermore, a comparison between the proposed approach and the popular RX technique is given. The former outperforms the latter demonstrating its ability to reduce the proportion of false alarms.

1. Introduction

A hyperspectral image is regularly modeled within a Euclidean space where its dimensions are equal to the number of bands in the imagery, and the points in the space represent pixels in the image. This model is called the feature space model [1]. Given the large number of bands in hyperspectral imagery, and its narrowness in the electromagnetic spectrum, most of the data is often redundant and much relevant information are difficult to extract due to the high dimensionality of the space. In order to make hyperspectral imagery more effective by solving practical challenges, methods are necessary that reduce the dimensionality of the data but at the same time keep as much information as possible.

With the introduction of new sensors capable of high spatial and spectral resolution, there has been an expansion of applications for hyperspectral imagery. Specifically, there has been an increasing interest in the use of hyperspectral imagery to detect small objects of interest. This has come to be known as target detection which can be either supervised or unsupervised. In supervised target detection, algorithms lean on prior knowledge, such as the target signature. The detection process for matching signatures is not straightforward due to the complications of converting sensor data acquired in the air into material spectra in the ground. This could be further complicated by the large number of possible objects of interest, as well as uncertainty as to the reflectance or emission spectra of these objects. For example, the surface of an object of interest may consist of several materials, and the spectra may be affected by the background or by weathering processes, or by the image-wide clutter background (see Ref. [2]). In addition, atmospheric transmittance and illumination depend on many factors such as the solar angle, temperature, water vapor concentration, concentration of mixed gases, cloud cover, and so forth.

This paper deals with the unsupervised technique of target detection, also called anomaly detection. Since this technique assumes no prior knowledge about the target or the statistical characteristics of the data, the only available option is to look for objects that are differentiated from the background. Generally, anomaly detection algorithms are used when target models are neither available nor unreliable, when ground truths are unavailable for the calibration of spectral signatures, or when little is known about the size and shape of the objects to be detected. The goal of this research is to find a reduction of background information technique that benefits the detection of atypical points or outliers. For example, man-made objects in the hyperspectral images of rural areas are presented as rare structures in the feature space; these artificial materials usually have very different spectral signatures compared to their natural

* Corresponding author. Tel.: +34 918856761.
E-mail address: josea.malpica@uah.es (J.A. Malpica).

0031-3203/$30.00 © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2008.04.014

surroundings. Furthermore, man-made objects are usually represented by a few pixels compared with the large number of pixels representing the natural background. In the feature space, the pixels representing man-made objects could be considered as outliers of a general sample distribution embodying the totality of pixels in the hyperspectral image.

In the approach set out in this paper, anomaly detection for hyperspectral imagery is the same process as outlier detection in the correspondent feature space for the imagery. In this approach, there is no distinction between anomalies, whether atypical or outlier, although usually the word anomaly refers to the image domain and outlier refers to the statistical domain; they are used synonymously for the purposes of this paper. Outliers are important features that are of special interest to image analysts in their daily routine. Methodologies and algorithms are needed to identify these atypical points, thereby allowing image analysts to decide whether to retain them as interesting information, or whether to classify them as non-interesting information, probably noise, and remove them. Filzmoser [3] suggests that image analysts should be prepared to identify atypical points and then determine how they will be subsequently treated. Points identified as atypical should be carefully evaluated based on the background information to determine their suitability for inclusion in further analyses. Analysts should also determine whether additional or specific information could explain their unusual characteristics. For example, Koltunov and Ustin [4] have presented a sub-pixel thermal anomaly detection for early detection of fire. Since the main objective of our paper is to present an anomaly detector which does not presuppose a spectral signature model, it is submitted that this research provides an effective tool that could aid analysts in performing some of their tasks.

There are an infinite number of projections that could be constructed to convert data in high dimensional space to lower dimensional data. A randomly selected projection of a high dimensional data set will appear similar to a sample from a multivariate normal distribution. However, "normality" is not a pertinent measuring gauge when a researcher is looking for projections where there is some structure in the data, because structured data represents relevant information. Any structure in the lower dimensional data projection can represent the shadow of an actual structure in the full dimensionality. A common method of obtaining projections is the projection pursuit (PP) method which is applied to hyperspectral imagery in this paper. The purpose of PP [5] is to elucidate the structure of the data in a lower dimensional space, such as clusters, shapes or outliers.

A popular anomaly detector for hyperspectral data is commonly referred to as the RX algorithm, named after the initials of its proponents [6]. The RX algorithm is not a PP; rather it is a likelihood ratio detector. Essentially, the algorithm calculates the Mahalanobis distance (MD) from the test pixel to the mean of the background. Schweizer and Moura [7] also proposed an anomaly detection technique based on a maximum likelihood algorithm but used a first-order Gauss–Markov random field model for the background. All these techniques assume a spatially uncorrelated Gaussian distribution for the background. However, such hypotheses are generally not verified in practical scenarios. The PP approach set out in this paper is very different. No assumption of normality is required, since the data is projected in one dimensional space and then analyzed.

It is important to note that a high dimensional space is mostly empty. The curse of dimensionality and the Hughes effect [8] appear as a consequence of this emptiness. Parametric methods of image classification are greatly affected by these two phenomena. Researchers have proposed methods to deal with these issues by using several dimensional reduction techniques (e.g., Refs. [9,10]). Jiménez and Landgrebe [11] applied PP to reduce the dimensionality of the data, and then performed parametric methods in a lower dimensional space. It is submitted that the most important feature of PP is that it is one of the few multivariate methods able to bypass the challenge of dimensionality. Moreover, some of the proposed methods based on inter-point distances to partially avert this problem, such as minimal spanning trees, multidimensional scaling and most clustering techniques, are unable to ignore irrelevant variables characterized with noisy or poor information.

In summary, this paper presents an anomaly detection technique for hyperspectral data based on a PP algorithm. In the next section, the PP used, and the reason for employment of the Legendre coefficients as the index of interest, are presented. The optimization algorithm, a combination of simulated annealing and simplex optimization which is used for optimizing the index in the PP method, is then discussed. In Section 3, the application of the techniques to two types of hyperspectral imagery is demonstrated followed by a discussion of the results. A summary of the main conclusions of this paper is given in Section 4.

2. The PP detector

2.1. Information extraction from high dimensional data

PP is a technique that uses one or more linear combinations of the original features to maximize some measure of "interestingness" called an index. Equivalently, this algebraic definition can be transformed into a geometric intuition which provides that the n-dimensional space of features can be rotated, and that the first few dimensions of the rotated space are then retained. The term for this technique, PP, was coined by Friedman and Tukey [5] who successfully implemented the idea based on work already done by Kruskal [12]. Many of the methods of classical multivariate analysis, such as principal components analysis (PCA) and discriminate analysis appear to be special cases of PP [13]. In the popular technique of PCA, the reduction of data corresponds to linear combinations chosen to maximize the variance of the projected data. Therefore, the measure of "interestingness" is in this case the variance. Goovaerts et al. [14] have proposed a methodology to detect small anomalies in hyperspectral imagery using PCA to reduce the dimensionality of the imagery, at the same time as they employ geostatistical filtering to remove regional background and enhance local signal.

This paper uses the letter X for a random variable with values in $\Re^n$ (which operates in n-dimensional probability densities), for a point cloud, that is, an n-tuple of points $(x_1, x_2, \ldots, x_n)$ in $\Re^n$, or for a pixel in a hyperspectral image with n spectral bands, where $(x_1, x_2, \ldots, x_n)$ represents the values of these spectral bands for the pixel. A linear projection from an n-dimensional space to a k-dimensional space is given by the following equations:

$$Y = G^{\mathrm{T}} X, \quad X \in \Re^n, \quad Y \in \Re^k. \tag{1}$$

where Y is the projected data and G is an $n \times k$ matrix, Y is a vector $k \times 1$, and where k is a small number, usually one or two (for PP). In this work the value of k will be one. This linear combination (or a one-dimensional projection) will be determined so that the probability density built with the values in Y is highly structured. In general terms, it is difficult to specify algorithmically what constitutes structured data. To what extent this structure can be regarded as "interesting" is measured by an index. In this paper, the level of interestingness will be gauged for structures where it is easy to perform outlier detection. Some of the most popular indices that are, or could be, related to hyperspectral analysis, and specifically to anomaly detection, in this kind of imagery are discussed below.