ISPRS Journal of Photogrammetry and Remote Sensing 138 (2018) 265–280

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Spectral-spatial classification of hyperspectral data with mutual information based segmented stacked autoencoder approach

Subir Paul [a], D. Nagesh Kumar [a,b,c,*]

[a] Department of Civil Engineering, Indian Institute of Science, Bengaluru 560012, India
[b] Centre for Earth Sciences, Indian Institute of Science, Bengaluru 560012, India
[c] Interdisciplinary Centre for Water Research, Indian Institute of Science, Bengaluru 560012, India

ARTICLE INFO

Article history:
Received 3 May 2017
Received in revised form 31 January 2018
Accepted 1 February 2018

Keywords:
Hyperspectral remote sensing
Spectral-spatial classification
Mutual information
Autoencoder
Support vector machine
Random forest

ABSTRACT

Hyperspectral (HS) data comprises of continuous spectral responses of hundreds of narrow spectral bands with very fine spectral resolution or bandwidth, which offer feature identification and classification with high accuracy. In the present study, Mutual Information (MI) based Segmented Stacked Autoencoder (S-SAE) approach for spectral-spatial classification of the HS data is proposed to reduce the complexity and computational time compared to Stacked Autoencoder (SAE) based feature extraction. A non-parametric dependency measure (MI) based spectral segmentation is proposed instead of linear and parametric dependency measure to take care of both linear and nonlinear inter-band dependency for spectral segmentation of the HS bands. Then morphological profiles are created corresponding to segmented spectral features to assimilate the spatial information in the spectral-spatial classification approach. Two non-parametric classifiers, Support Vector Machine (SVM) with Gaussian kernel and Random Forest (RF) are used for classification of the three most popularly used HS datasets. Results of the numerical experiments carried out in this study have shown that SVM with a Gaussian kernel is providing better results for the Pavia University and Botswana datasets whereas RF is performing better for Indian Pines dataset. The experiments performed with the proposed methodology provide encouraging results compared to numerous existing approaches.

1. Introduction

Hyperspectral (HS) remote sensing, also known as imaging spectroscopy, collects information in very fine spectral bands covering a wide range of wavelengths (generally 400–2500 nm) and provides a huge volume of data. HS data form a 3D structure, called hypercube, containing 2D spatial and 1D spectral information. The advantage of HS data is the presence of a large number of continuous spectral bands (approximately 100–250) with very fine bandwidths (5–10 nm) compared to multispectral data, which has only 5–10 spectral bands in some discrete spectral range with 70–400 nm bandwidths. Due to high spectral resolution and continuous spectral bands, HS data is very efficient and resourceful for image classification, identification and preparation of spectral library for different earth surface features (e.g. types of crops, soil, minerals etc.). On the other hand, due to the large number of spectral bands, limited labelled training samples (i.e. curse of dimensionality, also known as "Hughes effect") and presence of redundant information in the data, classification of HS image becomes a very challenging task.

Two primary steps are involved for any type of HS image classification. The first step is dimensionality reduction using either feature extraction (FE) or feature selection approaches to deal with the redundancy present in the HS data, speed up the classification process and reduce the data storage requirements. In the second step, the dimensionally reduced dataset is used in the classifier algorithm for classification of the HS image.

FE algorithms can be categorised into three varieties viz. supervised, unsupervised and semi-supervised (use both labelled and unlabelled samples) (Kianisarkaleh and Ghassemian, 2016). Different unsupervised methods like principal component analysis (PCA) (Ren et al., 2014), independent component analysis (ICA) (Comon, 1994), minimum noise fraction (Green et al., 1988), non-parametric feature extraction (Kianisarkaleh and Ghassemian, 2016), locality-preserving dimension reduction (Li et al., 2012),

* Corresponding author at: Department of Civil Engineering, Indian Institute of Science, Bengaluru 560012, India.

E-mail addresses: subir.paul.iisc@gmail.com (S. Paul), nagesh@iisc.ac.in (D. Nagesh Kumar).

https://doi.org/10.1016/j.isprsjprs.2018.02.001
0924-2716/

wavelet packet analysis and grey model (Yin et al., 2013) are used for FE. On the other hand, in the case of feature selection, important bands are selected based on different band selection criteria such as mutual information (MI) (Guo et al., 2006), and nature inspired algorithms viz. gravitational search algorithm, harmony search, particle swarm optimization, firefly algorithm and bat algorithm (Nakamura et al., 2014). It was reported that FE generally achieves higher classification accuracy, while feature selection can preserve the relevant original information of the spectral bands (Jia et al., 2012).

In FE methods, both linear and nonlinear mapping is applied to original dataset to extract the salient features (Kianisarkaleh and Ghassemian, 2016). In recent years, deep learning methodologies have also emerged as a powerful technique for deep FE and effective dimensional reduction of HS data (Chen et al., 2014; Romero et al., 2016; Zabalza et al., 2016). Although deep learning offers substantial deep feature representation of the high dimensional data which can improve the performance of the classification algorithm, it also increases the complexity of the algorithm and computation time.

PCA is the most widely used linear FE technique, which was also found to be a powerful tool in FE and data reduction (Zabalza et al., 2014). Jia and Richards (1999) had claimed that conventional principal component transformation is dominated by the visible and near-infrared bands. They also proposed a segmented principal component transformation technique to overcome this issue and to reduce the computation time. Autoencoder (AE) is one of the deep architecture based models, which is used to learn deep nonlinear features in an unsupervised manner (Chen et al., 2014). AE is also known as a nonlinear generalization of PCA. Segmented stacked AE (S-SAE) approach reduces the complexity of extracting the local nonlinear features from each spectral segment, where segmentation is achieved based on the inter-band correlation matrix (Zabalza et al., 2016). Correlation based segmentation does not consider the presence of nonlinearity in the data. Parametric dependency measures need to satisfy distributional assumption (e.g. normal distribution). The spectral reflectances of the HS narrow-bands do not always follow the normal distribution. Hence, non-parametric dependency measures are physically more relevant for spectral segmentation compared to the parametric linear dependency measures.

Spectral-spatial classification approaches are widely used in recent years for HS image classification since spatial information is very important for classification. Moreover, consideration of spatial information in the classifier algorithm also improves the classification accuracy (Chen et al., 2015). Use of morphological profiles (MPs) is one of the popularly used methods to take into account the spatial information in HS image classification, where MPs are created applying mathematical morphological operations. Extended morphological profiles (EMPs), created from PCA extracted spectral features (Benediktsson et al., 2005; Fauvel et al., 2008), were used in spectral-spatial classification approach of HS data. Dalla Mura et al. (2011) had used ICA extracted spectral features to create extended morphological attribute profiles (EMAPs) for HS image classification, where they have found ICA performs better than PCA for FE. EMPs can also be created using wavelets instead of PCA or ICA to improve classification performance in support vector machine (SVM) classifier (Quesada-Barriuso et al., 2014). Deep learning methodologies, which are used extensively for deep FE from the HS data, are also applied in spectral-spatial classification approaches based on deep AE (Chen et al., 2014; Ma et al., 2016), deep belief network (DBN) (Chen et al., 2015), attribute profiles (Aptoula et al., 2016) and convolutional neural network (CNN) (Zhao and Du, 2016). Deep learning models, being more robust in handling the nonlinearities of the HS data, can extract more potential features from the data which may improve the classification performance (Ghamisi et al., 2017).

Several parametric and non-parametric classifiers have been developed over the past few years and applied for classification of remote sensing data. Supervised parametric classifier such as maximum likelihood performs well when dealing with unimodal data but in case of multi-modal data, parametric classifiers are not suitable as they assume the data to be normally distributed (Belgiu and Drăguţ, 2016). SVM, a non-parametric classifier, provides a better classification of HS data as compared to traditional parametric and non-parametric classifiers such as maximum likelihood, neural networks, and K-nearest neighbour (Melgani and Bruzzone, 2004; Pal and Mather, 2005). Nonlinear SVM exhibits better classification accuracy compared to linear SVM (Melgani and Bruzzone, 2004). Although the theoretical basis of SVM and the results in some studies suggest that SVM may not be affected by the dimensionality of the dataset, some studies have reported that the classification accuracy can be improved by reducing the dimensionality of the dataset (Pal and Foody, 2010).

A non-parametric classification approach, ensemble classifier is also popularly used in recent studies. In this classification approach, ensembles of classifiers are trained using bagging or boosting approaches on training sample sets. AdaBoost and random forest (RF) are two popularly used tree-based ensemble classification algorithms. Chan and Paelinckx (2008) evaluated RF and AdaBoost classifier for ecotope mapping using airborne HS imagery and found that while both of them attain similar classification accuracy, RF was more robust and faster. Waske et al. (2009) studied the performance of SVM and RF and found that global performances of both the classifiers were quite similar and better than maximum likelihood and spectral angle mapper classifier. They have also reported that parameter selection for RF is computationally less expensive than for SVMs.

Numerous techniques were proposed for FE and classification of HS images in past several years. Most of them are computationally complex and time-consuming. Although, spectral segmentation approach was proposed earlier to reduce the computational time and extract the local features, use of non-parametric dependency measure was not studied for spectral segmentation. Use of locally encoded spectral features to create spectral-spatial features for classification was also not considered in the earlier research works.

Therefore, the main novelties and contributions of the present research work are the extraction of apt spectral-spatial features in less computational time and improvement of classification performance. First, we are proposing spectral segmentation of HS bands based on non-parametric dependency measure MI, which will consider the complete (both linear and nonlinear) inter-band dependence. Local nonlinear spectral features will be extracted from these spectral segments using AE or SAE. Furthermore, we are also promoting the use of encoded features for creating the EMPs to incorporate the spatial information in the spectral-spatial classification approach. Later these features will be used in the radial basis function (RBF) kernel-based SVM (RBF-SVM) and RF classifier for classification. Two most popular non-parametric and nonlinear classifiers are used for HS image classification instead of using the parametric and linear classifier, which can only deal with the data that follows a specific distribution and are linearly separable respectively.

The remainder of this paper is organized as follows. In Section 2, different methods related to the proposed approach are discussed and the proposed approach is introduced. Section 3 provides the details of the used datasets. In Section 4, details of experiments, which are carried out in this study and their results along with important discussions are reported. Finally, Section 5 summarizes the conclusions.