Computers and Electrical Engineering 000 (2016) 1–15

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Attribute profile based feature space discriminant analysis for spectral-spatial classification of hyperspectral images ☆

Maryam Imani, Hassan Ghassemian*

Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran

ARTICLE INFO

Article history:
Received 5 June 2016
Revised 21 September 2016
Accepted 22 September 2016
Available online xxx

Keywords:
Spectral-spatial classification
Attribute profile
Discriminant analysis
Hyperspectral

ABSTRACT

An initial feature reduction is necessary to reduce the data dimensionality before applying attribute filters to hyperspectral images. Unsupervised methods such as principal component analysis are not good choices for classification purposes. On the other hand, supervised methods such as linear discriminant analysis have no good efficiency in small sample size situations. In this article, we propose to extract features using feature space discriminant analysis (FSDA), which has been recently proposed in 2015. FSDA only uses the spectral information and ignores the spatial information. In this paper, we overcome this indigenous disadvantage of FSDA and develop FSDA for spectral-spatial classification of hyperspectral images. Our proposed method, called attribute profile based feature space discriminant analysis (APFSDA), extracts spatial features with high class discrimination ability and as little as redundant information. The experimental results on several real hyperspectral images show the superiority of APFSDA compared to some state-of-the-art spectral-spatial classification methods.

1. Introduction

Analysis and processing of remote sensing data is a hot research topic in recent decades [1]. Hyperspectral images acquired by remote sensors contain hundreds of images, corresponding to different wavelength channels for the same scene. The rich spectral information available in hyperspectral images provides possibility to distinguish between different classes of land covers with high details. Supervised classification of hyperspectral images is a difficult endeavor [2]. Generally, the number of training samples used for the learning stage of a classifier is limited compared to the number of spectral bands. So, obstacles, such as the Hughes phenomenon [3], appear as the data dimensionality increases. To address this issue, different methods have been proposed. Feature extraction methods degrade the ill-posed problems with dimensionality reduction [4–7]. Semi-supervised approaches use the ability of unlabeled samples in addition to available labeled samples to improve the classification accuracy [8]. The use of appropriate and efficient classifiers can combat the curse of dimensionality. Recently, sparse representation-based classifiers have been proposed for pattern classification [9]. Moreover, kernel based methods have been widely used due to their insensitivity to the cures of dimensionality [10]. Specifically, support vector machine (SVM) [11], due to its ability to deal with large input space and to produce sparse solutions, is among the state-of-the-art discriminative techniques that can be applied to solve ill-posed classification problems. Multinomial logis-

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. E. Cabal-Yepez.

* Corresponding author.
E-mail address: ghassemi@modares.ac.ir (H. Ghassemian).

http://dx.doi.org/10.1016/j.compeleceng.2016.09.029
0045-7906/© 2016 Elsevier Ltd. All rights reserved.

tic regression (MLR) [12], which has the advantage of learning the class probability distributions themselves, is another powerful approach to deal with ill-posed problem.

To increase classification accuracy, the contextual information should be considered for incorporation into the classifiers. The aim of spectral–spatial classification is to assign each image pixel to one class using a feature vector containing its own value (spectral information) and information extracted from its neighborhood (spatial information). It has been shown that multiple kernels and composite kernels [13] can improve the classification performance by incorporating the spatial information in addition to high rich spectral one. SVM based multiple kernel learning [14] methods generally require convex combinations of kernels, which introduce some practical limitations. In order to overcome these limitations, a generalized composite kernel (GCK) machine has been introduced in [15], which uses the MLR classifier [16] and can linearly combine multiple kernels without any restriction of convexity. So far, different types of features have exploited for spectral-spatial classification. Several methods exploit original spectral features or features linearly derived from them such as linear discriminant analysis (LDA) and principal component analysis (PCA) [17]. On the other hand, several methods exploit features obtained through nonlinear transformations to better model the inherent nonlinearity of original data. Kernel methods, manifold regularizations [18], and morphological analysis [19] for extraction of spatial information are examples of exploiting nonlinear features. It is common to have both the linear and nonlinear class boundaries in the same scene. So, multiple feature learning (MFL) [20] has been developed to integrate multiple features extracted from linear and nonlinear transformations. Similar to GCK, MFL uses the advantages of MLR.

Attribute profiles (APs) [21] are obtained by applying a sequence of attribute filters to a gray-level image to extract spatial features. The extended multiattribute profile (EMAP) generalizes AP to hyperspectral image [22]. In both EMP and EMAP, the multivariate domain of the data is reduced to few dimensions with the PCA transformation. The principal components with the most variance of data are considered for the analysis, while the others are discarded. In the PCA transformation, the components with low variance may contain useful information for class discrimination that by removing of them, the classification accuracy may be degraded. So, unsupervised feature extraction methods such as PCA, which work based on signal representation criteria, may not be very good choices for classification purposes. On the other hand, the use of supervised feature extraction methods such as LDA is not possible in high dimensional data when the number of available training samples is limited. LDA estimates the between-class scatter matrix ($\boldsymbol{S}_\mathrm{b}$) and the within-class scatter matrix ($\boldsymbol{S}_\mathrm{w}$). It maximizes the between-class scatters while minimizes the within-class scatters simultaneously by max $tr(\boldsymbol{S}_\mathrm{w}{}^{-1}\boldsymbol{S}_\mathrm{b})$. The accurate estimation of second order statistics such as scatter matrices needs enough training samples. By using small training set, the within-class scatter matrix becomes singular. In addition to singularity problem in small samples size situation, LDA can extract maximum $c-1$ features where c is the number of classes. We have proposed Feature space discriminant analysis (FSDA) in our previous work [23] for feature extraction of hyperspectral images. FSDA transforms the feature space of data to a new feature space where features are different from each other as much as possible. So, the redundancy between extracted spectral features is minimum. Moreover, the extracted features have high ability in class discrimination. In other words, in the FSDA projected feature space, the within-class scatters are minimum and the between-class scatters are maximum. In addition, FSDA overcomes the singularity problem and has no limitation in the number of extracted features. The original FSDA only considers the spectral information and ignores the valuable spatial information. In this paper, we overcome this disadvantage of FSDA and develop it for spectral-spatial classification of hyperspectral data.

In this work, we propose the attribute profile based FSDA (APFSDA) method. While the conventional EMAP extracts spatial features from the principal components achieved by PCA, we apply the EMAP transformation to the spectral components extracted by FSDA in the proposed method. So, we extract spatial features from the most appropriate spectral components for classification purposes. Applying attribute filters on components extracted by FSDA produces spatial features which have high class discrimination and little redundant information. Therefore, the spectral-spatial features extracted by APFSDA significantly improve the classification accuracy. APFSDA uses the advantages of MLR classifier to deal with ill-posed problems. Note that APFSDA fuses the spectral and spatial features twice: by applying attribute filters on the components extracted by FSDA and by stacking the extracted spatial features on the original spectral bands.

This paper is continued as follows. A brief description of related works is given in Section 2. The proposed method is introduced in Section 3 through three parts. At first, the description of MLR classifier and spectral feature extraction by FSDA are represented with more details. Then, how to extract useful spatial features (with maximum class discrimination and minimum redundancy) by applying EMAP transformation on the spectral components extracted by FSDA is represented. In the next section, the performance of proposed method is compared to some state-of-the-art classification methods. The experimental results are discussed in Section 4, and finally Section 5 concludes this paper.

2. Related works

Some recent spectral-spatial classification methods such as GCK [15] and MFL [20] obtain a feature vector containing spectral and spatial information and feed it to the MLR classifier to provide the classification map. In this paper, we obtain an appropriate spectral-spatial feature vector to classify it with MLR. So, at first we briefly introduce MLR, and then, represent an explanation of GCK and MFL. We compare the efficiency of our proposed method with GCK and MFL in Section 4.