Int. J. Electron. Commun. (AEÜ) 69 (2015) 198–205

Contents lists available at ScienceDirect

International Journal of Electronics and Communications (AEÜ)

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

An efficient method to solve the classification problem for remote sensing image

Jianqiang Gao *, Lizhong Xu

College of Computer and Information Engineering, Hohai University, Nanjing 210098, PR China

ARTICLE INFO

Article history:
Received 12 March 2014
Accepted 4 September 2014

MSC classification:
68T10
68U10

Keywords:
Remote sensing image
Projection matrix
Support vector machine
Image classification
Kernel method

ABSTRACT

Remote sensing image classification is different to identify the best classification model due to lack of a suitable classification method. Most traditional approaches only focused on using the spectral or spatial information to train classification model. However, these methods may ignore the related information of the image-itself. Remotely data not only a mere collection of independent and identically distributed pixels. Therefore, an efficient classification method is introduced in this paper. The proposed method deals with the information provided by the remote sensing image. Based on the idea of fisher linear discriminant analysis (FLDA), a definition of the same areas and different areas are considered in images, the information of same areas associated with each pixel is modeled as the within-class set, and the information of different areas associated with mean pixel of each same areas is modeled as the between-class set. Therefore, a projection matrix (PM) can be obtained by using within-class and between-class sets with the help of FLDA criterion. Then the PM is jointly used for the classification through a support vector machine (SVM) or K-nearest neighbor (KNN) classifiers formulation. Experiments on two remote sensing images are performed to test and evaluate the effectiveness of the proposed method.

1. Introduction

Image classification is a complex process that may be affected by lots of factors, such as the availability of high-quality images, ancillary data, proper classification procedure, and the analytical ability of scientists. For a particular study, it is often difficult to identify the best classifier due to the lack of a guideline for selection and the availability of suitable classification algorithms to band. So, in order to improve classification accuracy, many researchers have made great efforts [1–10]. In many reports, supervised, semi-supervised and unsupervised are the three popular leaning methods for remote sensing image classification, such as, maximum-likelihood classifiers, neural networks and neurofuzzy models [11–13]. However, for hyper-spectral remote sensing image, there is an important Hughes phenomenon [14]. Therefore, in this case, the above phenomenon is a disadvantage. It needs a long time to deal with the high-dimensional data for the computer.

The purpose of classification is to estimate the different species of each geographic region in remote sensing image. It is usually formulated as a segmentation task where an appearance model is first used to filter the pixel and then threshold setting strategies are utilized to infer the affiliation of the pixel in current frame. Therefore, how to effectively model the appearance of the target region and how to accurately infer the affiliation from all ground-based ancillary data are two key steps for a successful classification system. Although a variety of classification algorithms have been proposed in the last decades, remote sensing image classification still cannot meet the requirements of practical applications. The main difficulty of remote sensing image classification is designing a powerful model which should not only contain the main discrimination information from remote sensing image but also be robust to its variations. Lots of elegant features in the field of pattern recognition can be used to discriminate the category from image and ancillary data. However, the extraction of useful information and modeling are very difficult to achieve because remote sensing image exhibits more complex spectral character, such as high-dimensional data analysis, the selection of band and the fusion of obtained information. Therefore, traditional methods to achieve robust classification are not always feasible.

Recently, lots of methods are proposed, such as multiscale segmentation method [15], which is combining the probabilistic neural network (PNN) with the multiscale autoregressive model. Zhang et al. [16] proposed a hybrid classifier for polarimetric SAR images. Using PNN method, the span image, H/A/α decomposition and the gray level co-occurrence matrix (GLCM)-based texture features of feature set are jointly used for the classification. Many

* Corresponding author. Tel.: +86 18751991606.
E-mail addresses: dr.jq.gao@gmail.com, jianqianggaohh@126.com (J. Gao).

http://dx.doi.org/10.1016/j.aeue.2014.09.001
1434-8411/

researchers believe that kernel-based method is one of the best tools for solving high-dimensional data classification [17–20] or pattern recognition problem [21–23]. In fact, the kernel-based method [24] is applied to project the original data into a feature space in which the data become linearly (or approximate linearly) separable. Fauvel et al. [17] proposed a spatial-spectral kernel-based approach with the spatial and spectral information were jointly used for the classification. Gao et al. [25] proposed a kernel-based block matrix decomposition approach for the classification of remote sensing image.

In addition, due to in hyperspectral remote sensing imagery data analysis, the data modeling process (DMP) plays an important role in almost all supervised or unsupervised classification methods which require the estimation of all kinds of statistic. Therefore, the main goal of this paper is to establish a new and efficient classification method applicable in mathematical theory for remote sensing image. The proposed method is a three-step process, firstly, the information of same areas associated with each pixel is modeled as the within-class set (produce within-class scatter matrix, **Sw**), and the information of different areas associated with mean pixel of each same areas is modeled as the between-class set (produce between-class scatter matrix, **Sb**). And then, using FLDA method, a projection matrix (PM) with very important information can be obtained by solving an optimal problem. Finally, the PM was applied to project the original data into a new feature space, then SVM (or KNN) classifier was used in last stage.

The remainder of this paper is organized as follows. Section 2 briefly reviews the formulations of FLDA, KNN and SVM. In Section 3, the derivation process of the proposed method is described in detail. The effectiveness of the proposed method is demonstrated in Section 4 by experiments on several real remotely sensed images. Finally, Section 5 concludes this paper.

2. Review of FLDA, KNN and SVM

2.1. Fisher linear discriminant analysis (FLDA)

The main idea of FLDA is to perform dimension reduction while preserving as much information as possible. Linear discriminant analysis aims to find the optimal PM such that the class structure of the original high-dimension space is preserved in the low-dimensional space. However, the FLDA cannot be directly applied because the **Sw** has zero eigenvalues. So, many methods are proposed to solve this problem, such as LDA/QR [26], null subspace method [27], range subspace method [28] and median method [29,30].

In this subsection, we first introduce some important notations used in this paper. Let c be the number of classes, N_i be the number of samples from ith class, N be the number of total samples from each class, A_j^i be the jth sample from ith class and m_i be the mean of ith class samples.

$$N = \sum_{i=1}^{c} N_i, \tag{1}$$

$$m_i = \frac{1}{N}\sum_{j=1}^{N_i} A_j^i, (i = 1, \ldots, c). \tag{2}$$

The optimal PM $W = [w_1, w_2, \ldots, w_r]$ can be obtained via maximizing the following criterion [31]. Where r is at most min($c-1$, N).

$$J(W) = \frac{W^T \mathbf{Sb} W}{W^T \mathbf{Sw} W}, \tag{3}$$

where **Sb** and **Sw** are the between-class and within-class scatter matrices, respectively. m_0 is the global mean of all classes samples.

$$\mathbf{Sb} = \sum_{i=1}^{c} (m_i - m_0)^T (m_i - m_0), \tag{4}$$

$$\mathbf{Sw} = \sum_{i=1}^{c}\sum_{j=1}^{N_i} (A_j^i - m_i)^T (A_j^i - m_i), \tag{5}$$

$$m_0 = \frac{1}{c}\sum_{i=1}^{c} m_i. \tag{6}$$

2.2. K-nearest neighbor classifier (KNNc)

KNNc is a nonparametric approach for classification. It does not require the priori knowledge such as priori probabilities and the conditional probabilities. It operates directly toward the samples and is categorized as an instance-based classification method. The main idea of KNNc is as below. Assume that the training samples data X with class identification as $X = \{(x_1, l_1), \ldots, (x_n, l_n)\}$. If there is minimum distance between test sample x and k samples $x_1, \ldots, x_k$ of training samples data, and then, the category of test sample x belongs to l_i ($i = 1, 2, \ldots, k$). Details can be found in [32,33].

2.3. Support vector machine classifier (SVMc)

In this subsection, we briefly review the support vector machine classifier. Given a labeled training set $\{(x_1, y_1), \ldots, (x_N, y_N)\}$, where $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, +1\}$, and $\Phi(\cdot)$ is a nonlinear mapping. Generally speaking, to a high dimension space $\mathbb{H}$, $\Phi : \mathbb{R}^d \to \mathbb{H}$, the SVM algorithm solves

$$\min_{w,\xi_i,b} \left\{ \frac{1}{2} \| w \|^2 + C \sum_i \xi_i \right\}, \tag{7}$$

constrained to

$$y_i(< \Phi(x_i), w > +b) \geq 1 - \xi_i, \quad \forall i = 1, 2, \ldots, N. \tag{8}$$

$$\xi_i \geq 0, \quad \forall i = 1, 2, \ldots, N. \tag{9}$$

where w and b define a linear classifier in the feature space. According to the Cover's theorem [34], the nonlinear mapping function Φ is used in the transformed samples feature space. The parameter C controls the generalization capabilities of the classifier and it must be selected by the user, and ξ_i are positive slack variables enabling to deal with permitted errors.

In mathematics, the primal problem (7) is solved through its Lagrangian dual problem (10) because of the high-dimension of vector w.

$$\max_{\alpha_i} \left\{ \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j < \Phi(x_i), \Phi(x_j) > \right\}, \tag{10}$$

constrained to $0 \leq \alpha_i \leq C$ and $\sum_i \alpha_i y_i = 0$, $i = 1, 2, \ldots, N$. Where auxiliary variables α_i are Lagrange multipliers corresponding to constraints in (8). All Φ mappings are performed in the form of inner products. So, a kernel function K need to be defined as (11).

$$K(x_i, x_j) = < \Phi(x_i), \Phi(x_j) > . \tag{11}$$