ISPRS Journal of Photogrammetry and Remote Sensing 97 (2014) 36–45

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

An efficient semi-supervised classification approach for hyperspectral imagery

Kun Tan [a], Erzhu Li [a], Qian Du [b], Peijun Du [c,*]

[a] Jiangsu Key Laboratory of Resources and Environment Information Engineering, China University of Mining and Technology, China
[b] Department of Electrical and Computer Engineering, Mississippi State University, USA
[c] Key Laboratory for Satellite Mapping Technology and Applications of State Administration of Surveying, Mapping and Geoinformation of China, Nanjing University, China

ARTICLE INFO

Article history:
Received 19 June 2014
Received in revised form 29 July 2014
Accepted 12 August 2014
Available online 7 September 2014

Keywords:
Hyperspectral
Semi-supervised learning
Classification
Segmentation
Spectral–spatial feature
SVM

ABSTRACT

In this paper, an efficient semi-supervised support vector machine (SVM) with segmentation-based ensemble (S^2SVMSE) algorithm is proposed for hyperspectral image classification. The algorithm utilizes spatial information extracted by a segmentation algorithm for unlabeled sample selection. The unlabeled samples that are the most similar to the labeled ones are found and the candidate set of unlabeled samples to be chosen is enlarged to the corresponding image segments. To ensure the finally selected unlabeled samples be spatially widely distributed and less correlated, random selection is conducted with the flexibility of the number of unlabeled samples actually participating in semi-supervised learning. Classification is also refined through a spectral–spatial feature ensemble technique. The proposed method with very limited labeled training samples is evaluated via experiments with two real hyperspectral images, where it outperforms the fully supervised SVM and the semi-supervised version without spectral–spatial ensemble.

1. Introduction

Hyperspectral sensors provide abundant spectral information with hundreds of spectrally continuous bands, and the collected hyperspectral imagery can be used to conduct finer classification that could not be achieved by traditional multispectral imagery (Bioucas-Dias et al., 2013; Du et al., 2012; Mountrakis et al., 2011; Tan and Du, 2008, 2011; Tan et al., 2013; van der Meer et al., 2012). Over the past several decades, novel pattern recognition methods have been widely used to remote sensing image processing tasks, such as supervised classification (Li et al., 2013b; Zhong and Zhang, 2012), unsupervised classification (Villa et al., 2013; Yang et al., 2010), feature extraction (Du et al., 2008; Du and Yang, 2008; Jia et al., 2013), target identification (Du and Zhang, 2014; Manolakis et al., 2001), and change detection (Bioucas-Dias and Nascimento, 2008; Camps-Valls et al., 2008). support vector machines (SVM), as one of the research focuses of machine learning, has attracted more attention in remote sensing (Camps-Valls et al., 2006; Fauvel et al., 2008; Melgani and Bruzzone, 2004; Plaza et al., 2009). SVM has advantages such as less rigid requirements for prior knowledge and training samples, fitness to high-dimensional data, and more robustness to noise (Du et al., 2010; Jain et al., 2000). However, it is often difficult for a traditional SVM classifier to offer satisfactory performance in hyperspectral image classification. Specifically, for a high-dimensional hyperspectral image with a limited number of training samples, classification accuracy usually is significantly decreased due to the Hughes phenomenon (i.e., for a limited number of training samples, classification accuracy is decreased with the feature dimension being increased) (Hughes, 1968). Moreover, hyperspectral imagery is usually short of training sets, because sample collection generally involves extensive and time-consuming fieldwork (Dopido et al., 2013; Li et al., 2013a).

A relevant advanced solution to this classification challenge is the introduction of semi-supervised learning (SSL) techniques (Camps-Valls et al., 2007; Chapelle et al., 2008). SSL utilizes a large number of unlabeled data together with the available labeled data, to build classifiers that may be stronger than usual. SSL also requires less human efforts in sample collection, so it is of great importance in practical applications. The research on SLL has experienced a quick development in the past few years, and some popular methods have been proposed, such as generative mixture models (Krithara et al., 2011), self-learning models (Dopido et al.,

* Corresponding author. Tel.: +86 15905159291.
E-mail address: dupjrs@126.com (P. Du).

http://dx.doi.org/10.1016/j.isprsjprs.2014.08.003
0924-2716/

2013), multiview learning models (Culp et al., 2009; Sun and Shawe-Taylor, 2010), transductive support vector machines SVMs (TSVMs) (Bruzzone et al., 2006; Ratle et al., 2010), and graph-based methods (Bai et al., 2013; Camps-Valls et al., 2007).

In particular, semi-supervised SVM (S^3VM) has been successfully applied to remote sensing image classification. Marconcini proposed a novel composite S^3VM for the spectral–spatial classification of hyperspectral images, which can significantly improve classification accuracy compared to both supervised SVMs and progressive S^3VM (Marconcini et al., 2009). Munoz-Mari proposed an algorithm to modify the OC-SVM kernel by modeling the data marginal distribution with a data graph Laplacian (Munoz-Mari et al., 2010). Gu successfully used Laplacian SVM for semi-supervised hyperspectral image classification (Gu and Feng, 2013). A novel approach for semi-supervised learning that adapts available active learning methods to a self-learning framework was presented by Dopido et al. (2013). The aforementioned SSL methods improve classification performance by different strategies. However, as the number of training samples is increased, it may be unbalanced and unbearable for a classifier to correctly exploit all the proper training samples due to computational issues. Furthermore, if a mislabeled sample is added for training, it may degrade classification accuracy. Thus, a semi-supervised algorithm should be designed in a computationally efficient fashion with the objective that classification performance can be truly improved with the use of additional samples.

In this paper, we evaluate the feasibility of adapting spatial information into an SSL paradigm, in which the S^2SVMSE itself selects accurate unlabeled samples for training. It makes use of spatial information in selecting new unlabeled training samples and generating classifier ensemble. It is assumed that samples in a homogeneous spatial segment mostly likely belong to the same class, which helps generate a smooth classification map; meanwhile, samples in the same segment also exhibit within-class variations, which can be useful for classifier training. Up to now, few studies that employ spatial information to select unlabeled samples are reported in the literature. Shi et al. proposed a semi-supervised dimension reduction algorithm, where unlabeled samples were selected based on multilevel segmentation results and t spectrally nearest neighbors similarity measure, and it notes that the number of the neighbors t controls the accuracy of similarity measurement, so they recommended to set a larger t to relax the constraint (Shi et al., 2013). However, it could not perform perfectly with limited or few labeled samples, so α-one-nearest neighbor is proposed to take the action of similarity measurement when labeled samples are very limited. In this work, the mean shift algorithm is used for image segmentation and spatial feature extraction, and a strongly constrained algorithm called α-one-nearest neighbor is applied for integrating spatial information when selecting new training samples. As a result, our proposed approach combines the spectral–spatial information in the SSL strategy.

The paper is organized as follows. In Section 2, the mean shift based segmentation is briefly reviewed, and then the proposed approach is introduced. Section 3 demonstrates the experimental results. Finally, Section 4 concludes with some remarks.

2. Proposed method

Let $X = (x_1, x_2, \cdots, x_l, x_{l+1}, \cdots, x_n) \in R^d$ denote a hyperspectral image with n pixels and d bands. Let $\kappa = \{1, \cdots, k\}$ be a set of k class labels, $y = (y_1, y_2, \cdots, y_l)$ be the labels of l labeled samples $X_L = \{x_i\}_{i=1}^{l}$, and $X_U = \{x_i\}_{i=l+1}^{n}$ be $(n-l)$ unlabeled samples. To perform the proposed semi-supervised approach, a hyperspectral imagery is first segmented using the mean shift method, and then unlabeled samples are selected for training based on the proposed strategy in Section 2.2, finally, classification result is generated based on a spectral–spatial feature ensemble method in Section 2.3. Note that during the processes of unlabeled sample selection and classification result ensemble, spatial information in image segments is considered.

2.1. Mean shift

In this paper, spatial information is used during the selection of unlabeled samples and feature ensemble for classification. Here, the image segmentation method of mean shift (MS) is employed to extract spatial information. MS method is a nonparametric clustering technique without requiring prior knowledge of the number and shape for each cluster. It defines an empirical probability density function, and the modes of densest regions in the space can be estimated by finding the local maxima of the probability density function that is estimated by kernel density estimation method (Comaniciu and Meer, 2002). Once the location of modes is determined, the cluster associated with it is delineated based on the local structure in the feature space. Here, we briefly review the basics of mean shift segmentation (Comaniciu and Meer, 2002).

With the notation defined, given n pixels $x_i, i = 1, \cdots, n$ in the d-dimensional space R^d, the kernel density estimator computed in pixel x is given by,

$$f(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{1}$$

where h is the bandwidth. Here, the kernel $K(x)$ is given by $K(x) = c_{k,d}k(\|x\|^2)$, where the function $k(x)$ is called kernel profile, and the normalization constant $c_{k,d}$ makes $K(x)$ have an integral of one. Defining the function $g(x) = -k'(x)$ with the assumption that the derivative of the kernel profile $k(x)$ exists for all $x \in [0, \infty)$. Using $g(x)$ as the kernel profile, the kernel $G(x)$ is defined as $G(x) = c_{g,d}g(\|x\|^2)$, where $c_{g,d}$ is a normalization constant. The goal of the defined density is to find the modes of densest regions in the space. The modes are located among the zeros of gradient $\nabla f(x) = 0$. Thus, the mean shift vector is defined as:

$$m_{h,G}(x) = \frac{1}{2}h^2 c \frac{\nabla f_{h,K}(x)}{f_{h,G}(x)} \tag{2}$$

$$m_{h,G}(x) = \frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)} - x \tag{3}$$

For a hyperspectral image, it is typically represented as a two-dimensional lattice of d-dimensional vectors. Taking its spatial and spectral information into the mean shift procedure, the spatial and spectral information are represented in the spatial and range domains, respectively, and the multivariate kernel is defined as the product of these two radially symmetric kernels as:

$$K_{h_s,h_r}(x) = \frac{C}{h_s^2 h_r^d} k\left(\left\|\frac{x^s}{h_s}\right\|^2\right) k\left(\left\|\frac{x^r}{h_r}\right\|^2\right) \tag{4}$$

where x^s is the spatial part; x^r is the range part of a feature vector; h_s and h_r are the kernel bandwidths for spatial and range domains, respectively, and c is the normalization constant. MS cannot be used for high-dimensional data (Comaniciu and Meer, 2002; Georgescu et al., 2003), so dimensionality reduction has to be applied first. Here, principal component analysis is used for this purpose.

2.2. Selection of unlabeled samples

In contrast to supervised classification algorithms, semi-supervised methods generally assume that enlarging the training