ISPRS Journal of Photogrammetry and Remote Sensing 124 (2017) 1–15

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Sparse graph regularization for robust crop mapping using hyperspectral remotely sensed imagery with very few *in situ* data

Zhaohui Xue [a], Peijun Du [b,c,d,*], Jun Li [e], Hongjun Su [a]

[a] School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China
[b] Key Laboratory for Satellite Mapping Technology and Applications of National Administration of Surveying, Mapping and Geoinformation of China, Nanjing University, Nanjing 210023, China
[c] Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing University, Nanjing 210023, China
[d] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing University, Nanjing 210023, China
[e] Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China

ARTICLE INFO

Article history:
Received 20 May 2016
Received in revised form 14 September 2016
Accepted 14 December 2016

Keywords:
Hyperspectral remote sensing
Crop mapping
Sparse graph regularization (SGR)
Heihe watershed
CASI/SASI

ABSTRACT

The generally limited availability of training data relative to the usually high data dimension pose a great challenge to accurate classification of hyperspectral imagery, especially for identifying crops characterized with highly correlated spectra. However, traditional parametric classification models are problematic due to the need of non-singular class-specific covariance matrices. In this research, a novel sparse graph regularization (SGR) method is presented, aiming at robust crop mapping using hyperspectral imagery with very few *in situ* data. The core of SGR lies in propagating labels from known data to unknown, which is triggered by: (1) the fraction matrix generated for the large unknown data by using an effective sparse representation algorithm with respect to the few training data serving as the dictionary; (2) the prediction function estimated for the few training data by formulating a regularization model based on sparse graph. Then, the labels of large unknown data can be obtained by maximizing the posterior probability distribution based on the two ingredients. SGR is more discriminative, data-adaptive, robust to noise, and efficient, which is unique with regard to previously proposed approaches and has high potentials in discriminating crops, especially when facing insufficient training data and high-dimensional spectral space. The study area is located at Zhangye basin in the middle reaches of Heihe watershed, Gansu, China, where eight crop types were mapped with Compact Airborne Spectrographic Imager (CASI) and Shortwave Infrared Airborne Spectrogrpahic Imager (SASI) hyperspectral data. Experimental results demonstrate that the proposed method significantly outperforms other traditional and state-of-the-art methods.

1. Introduction

Crop mapping is vital for agriculture resources monitoring (Becker-Reshef et al., 2010), crop yield evaluation (Doraiswamy et al., 2004), and precision agriculture (Yang et al., 2013). Hyperspectral remote sensing sensors can capture contiguous and dense spectra (i.e., usually hundreds of spectral bands with the wavelength width finer than 10 nm) for the same object on the Earth's surface (Chang, 2003), which provide plenty of useful information for the analysis and interpretation of spectrally similar materials of interest. In this context, hyperspectral data have great potential in accurate identification of crops.

Hyperspectral remotely sensed imagery (HSI) has been extensively applied to crop mapping in the literature. Some traditional parametric classification models, which assume that the input data should obey certain parametric probability distribution, including Gaussian maximum likelihood (Foody et al., 1992), Shannon entropy (Foody, 1995), and discriminant analysis (Galvao et al., 2009; Mariotto et al., 2013; Thenkabail et al., 2013) have been used for pixel-wise classification. However, shortcomings occur when the assumption of parametric probability distribution of the input data is not hold. To this end, non-parametric classification models are more advocated, like distance metric based methods (Rao, 2008; Ran et al., 2015), artificial neural network (Foody, 1996;

* Corresponding author at: Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing University, Nanjing 210023, China.
E-mail address: dupjrs@gmail.com (P. Du).

http://dx.doi.org/10.1016/j.isprsjprs.2016.12.003
0924-2716/

Eddy et al., 2008; Zhang and Wu, 2011), niche hierarchical artificial immune system (Senthilnath et al., 2013), and Support Vector Machine (SVM) with Gaussian kernel function (Foody and Mathur, 2004; Camps-Valls et al., 2004; Mathur and Foody, 2008; Munoz-Mari et al., 2010). Other methods are prone to combining field reflectance measurements (Nidamanuri and Zbell, 2011) and LiDAR data (Koenig et al., 2015), etc.

Although elegant performance has been obtained by using these methods, some possible drawbacks are still existing. On the one hand, the good performance of these methods is usually guaranteed with large number of training data. However, the training data mainly come from imagery interpretation facilitated with *in situ* investigation, which is expensive, time consuming, and difficult, thus making the available training data very limited in practice. On the other hand, most of the current methods were not designed to address the curse of dimensionality issue of HSI. The generally limited availability of training samples relative to the usually high data dimension makes accurate classification of HSI an ill-posed problem (Jackson and Landgrebe, 2002), resulting in the risk of overfitting of the training data and poor generalization capability of the classifier. In this context, it calls for the development of new methods for accurate crop mapping, especially in the absence of sufficient training data and in the scenario of facing high-dimensional spectral space.

Regularization has emerged as a promising paradigm imposed on classifier to prevent it from overfitting in the context of few training data in conjunction with high-dimensional spectral space (Camps-Valls et al., 2014). SVM (Cortes and Vapnik, 1995) naturally equips with regularization through the concept of maximum margin (i.e., an ℓ_2-norm regularizer is adopted in SVM), which is no doubt the most widely used classification method in this community (Camps-Valls and Bruzzone, 2009). However, in order to produce satisfied result, parameter tuning (i.e., cross-validation) is necessary for SVM, which deteriorates the intensive computation issue of the method, especially when dealing with high-dimensional spectral space. Despite Cortes and Vapnik (1995) argued that SVM is resistant against the dimensionality of the input data and classifies well even with few training data, recent study showed that the dimensionality can significantly affect the classification accuracy (Waske et al., 2010). Therefore, limitations in efficiency and accuracy may be experienced when using SVM. Other regularization-based classifiers include regularized linear discriminant analysis (Bandos et al., 2009), Gaussian process based classification (Bazi and Melgani, 2010), relevance vector machine (Mianji and Zhang, 2011), multinomial logistic regression (Li et al., 2011), and sparse methods (Chen et al., 2011; Chen et al., 2013; Zhang et al., 2014; He et al., 2014; Fang et al., 2014; Song et al., 2014; Xue et al., 2015b; Feng et al., 2014; Zhong et al., 2016; Feng et al., 2016), etc. In spite of good performance of these methods, additional information such as non-linearity and homogeneity should be considered in the regularization process, which can greatly improve the generalization performance of the classifier.

Recently, graph regularization opens up new opportunities for HSI classification. In graph regularization based methods, an undirected weighted graph is built by treating samples (both labeled and unlabeled) as vertices, whereas the pairwise similarities between the vertices are regarded as edges. The few vertices carrying class label information are used to predict labels for others via label propagation which assumes label smoothness over the graph (Liu et al., 2012). Graph regularization based methods are non-parametric, discriminative, and capable of modeling the non-linearity (or the manifold data structures) hidden in HSI. Several studies have exploited graph regularization for HSI classification. The pioneering work belongs to Camps-Valls et al. (2007) who employed a graph transduction method, and they further incorporated kernel method and the spatial information to improve the classification performance. Later, Gomez-Chova et al. (2008) adopted a Laplacian support vector machine for multi-spectral imagery classification, where SVM was regularized by graph Laplacian. Similarly, Li et al. (2010) adopted graph Laplacian to regularize the regressors of multinomial logistic regression. Although good classification performance has been observed by these methods, most of them scale poorly with data size due to the high computational complexity. For example, the time costs for these studies carried out by Camps-Valls et al. (2007), Gomez-Chova et al. (2008), and Li et al. (2010) are respectively, $\mathcal{O}(N^3)$ (N is the total number of pixels in the imagery), $\mathcal{O}(N^3)$, and $\mathcal{O}(CB(N'+B))$ (C is the number of classes, B is dimensionality, and N' is the number of training data), which are huge for HSI (i.e., N = 278,139 for the CASI/SASI data used in our study).

More recently, sparse representation (SR) (Wright et al., 2009) has been exploited in graph construction, where the graph structure and the similarity matrix are simultaneously obtained. To this end, each sample is sparsely represented as a linear summation of the rest by minimizing an ℓ_0 (or its variants) problem. A sparse graph can capture both the local structure and the discriminative information of data, which are very useful for clustering, subspace learning, and semisupervised learning (Cheng et al., 2010).

The objective of this research is to develop a novel sparse graph regularization (SGR) method with particular emphasis on addressing robust and efficient crop mapping problems in the context of very few *in situ* data coupled with high-dimensional spectral space. To this end, the sparse unmixing by variable splitting and augmented Lagrangian (SUnSAL) algorithm (Bioucas-Dias and Figueiredo, 2010) was adopted to obtain the fraction matrix generated for the input data with respect to dictionary (few training data). The sparse graph and graph Laplacian matrix are obtained based on the fraction matrix. Next, the predict function estimated for dictionary is obtained by optimizing a typical graph regularization problem embedded with an inductive model inspired from anchor graph regularization (AGR) (Liu et al., 2012). Finally, the inductive model propagates labels from few training data to large unknown data based on the fraction matrix and the prediction function. In addition, graph cuts (Boykov et al., 2001) is used for incorporating spatial information, which improves the homogeneity of the final crop map. In SGR, the *sparsity* inherits from the sparse graph built using SR, i.e., the SUnSAL algorithm. Different from the traditional graph where the vertexes are densely linked to each other (e.g., the commonly used full-link graph built from Gaussian kernel), those vertexes will be sparsely linked when building a sparse graph.

Compared to AGR, the proposed SGR method is more discriminative, data-adaptive, robust to noise, and efficient with time cost of $\mathcal{O}(K^2$ iteration) (K is the number of training data), which is very powerful and well adapted to HSI without complex parameter tuning. Note that, it is unique with regard to previously proposed approaches in this area. It will be shown in this research that SGR provides more accurate crop classification map compared to other traditional classifiers as well as AGR. In addition, our previous studies have validated the good performance of: (1) SUnSAL for HSI decomposition (Xue et al., 2015b); (2) SUnSAL with total variation for HSI classification (Du et al., 2015b); (3) sparse graph for HSI feature extraction (Xue et al., 2015). These studies and conclusions consolidated the theoretical and technical basis of the proposed method. Although this research focuses on crop mapping, the proposed method is quite general and can be extended to other domains.

The remainder of this paper is organized as follows. Section 2 illustrates the study area, hyperspectral data, and the preprocessing procedures. Section 3 presents in detail the mechanism