ISPRS Journal of Photogrammetry and Remote Sensing 122 (2016) 179–191

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

A co-training, mutual learning approach towards mapping snow cover from multi-temporal high-spatial resolution satellite imagery

Liujun Zhu [a,b,c,d], Pengfeng Xiao [a,b,c,*], Xuezhi Feng [a,b,c], Xueliang Zhang [a,b,c], Yinyou Huang [a,b,c], Chengxi Li [a,b,c]

[a] Department of Geographic Information Science, Nanjing University, Nanjing, Jiangsu 210023, China
[b] Collaborative Innovation Center of South China Sea Studies, Nanjing, Jiangsu 210023, China
[c] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, Jiangsu 210023, China
[d] Department of Civil Engineering, Monash University, Clayton, Vic. 3800, Australia

ARTICLE INFO

Article history:
Received 27 May 2016
Received in revised form 29 October 2016
Accepted 6 November 2016
Available online 17 November 2016

Keywords:
Co-training
Multitask learning
Multi-temporal
Snow cover
Semi-supervised learning
High spatial resolution image

ABSTRACT

High-spatial and -temporal resolution snow cover maps for mountain areas are needed for hydrological applications and snow hazard monitoring. The Chinese GF-1 satellite is potential to provide such information with a spatial resolution of 8 m and a revisit of 4 days. The main challenge for the extraction of multi-temporal snow cover from high-spatial resolution images is that the observed spectral signature of snow and snow-free areas is non-stationary in both spatial and temporal domains. As a result, successful extraction requires adequate labelled samples for each image, which is difficult to be achieved. To solve this problem, a semi-supervised multi-temporal classification method for snow cover extraction (MSCE) is proposed. This method extends the co-training based algorithms from single image classification to multi-temporal ones. Multi-temporal images in MSCE are treated as different descriptions of the same land surface, and consequently, each pixel has multiple sets of features. Independent classifiers are trained on each feature set using a few labelled samples, and then, they are iteratively re-trained in a mutual learning way using a great number of unlabelled samples. The main principle behind MSCE is that the multi-temporal difference of land surface in spectral space can be the source of mutual learning inspired by the co-training paradigm, providing a new strategy to deal with multi-temporal image classification. The experimental findings of multi-temporal GF-1 images confirm the effectiveness of the proposed method.

1. Introduction

Snow cover extent is an essential input for snow hydrological models and snow hazard monitoring. Consequently, snow cover maps have been extracted using Moderate Resolution Imaging Spectroradiometer (MODIS) (Hall et al., 2002), Landsat Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+) (Crawford et al., 2013), Operational Landsat Imager (OLI) (Zhu et al., 2015) images, and ground based digital camera (Bernard et al., 2013). However, these sensors can unilaterally reach the high spatial or temporal resolution required to capture the seasonal spatial and temporal variations of snow cover. The new generation of satellites in the form of a constellation, such as Europe's Sentinel-2 and China's GF-1/6, can provide relatively high spatial resolution and frequently revisited observations, thereby providing spatial and temporal details of snow cover characteristics.

Current snow cover extraction approaches are mainly based on the special spectral characteristics of snow, i.e. high reflectance at visible wavelengths and low reflectance at shortwave-infrared wavelengths (Warren, 1982). A series of thresholds based on Normalized Difference Snow Index (NDSI) and/or spectral band ratio, are sufficient to separate snow from snow-free (Dozier, 1989; Hall et al., 1995; Riggs et al., 1994). Snow cover in mountain areas in shadow has a large overlap with the snow-free region in the spectral space (Dozier, 1989; Rosenthal and Dozier, 1996). Consequently, topographic correction based on the digital elevation model (DEM) is used to alleviate the influence of mountain shadow (Dozier, 1989; Negi et al., 2009; Rosenthal and Dozier, 1996; Sirguey et al., 2009). Alternatively, some studies apply shadow masks achieved from DEM to exclude the shadow areas (Selkowitz and Forster, 2016). To represent the subgrid snow cover heterogeneities, empirical relationships based on NDSI

* Corresponding author at: Department of Geographic Information Science, Nanjing University, China.
E-mail address: xiaopf@nju.edu.cn (P. Xiao).

http://dx.doi.org/10.1016/j.isprsjprs.2016.11.003
0924-2716/

(Salomonson and Appel, 2004, 2006) and subpixel unmixing models (Painter et al., 2009, 1998; Rosenthal and Dozier, 1996) are proposed to produce fractional snow cover maps. In addition, machining learning techniques, e.g. Artificial Neural Network and Support Vector Machine (SVM), are applied to train more robust models or classifiers for extracting both binary and fractional snow cover (Dobreva and Klein, 2011; Simpson and McIntire, 2001; Zhu et al., 2014).

Despite the great advance, it is still challenging to extract snow cover from high-spatial and -temporal resolution remote sensing images (HSTRRS). Commonly, shortwave-infrared wavelengths are not covered by high-spatial resolution optical sensors. As a result, subgrid snow cover heterogeneities become less critical and NDSI is not available. Some indexes based on visible and near-infrared bands provide promising alternatives for snow cover extraction in plains (Hinkler et al., 2003, 2002) but cannot obtain sound results in the mountain areas because of the severe influence of mountain shadow in HSTRRS. What's worse, high-spatial resolution DEM with satisfactory quality is not commonly accessible, making topographic correction difficult.

In our previous study (Zhu et al., 2014), a SVM based decision tree was proposed to extract snow cover from a single high-spatial resolution image without topographic correction, where snow cover influenced by mountain shadow was treated as an independent class in the classification procedure. Similar to most of the classification methods, the main limitation of this method is its heavy dependency on the quality of the ground-truth samples. Collecting a sufficient number of representative samples is impractical. Moreover, even if a satisfactory classifier was trained for an image with adequate samples, it cannot be directly applied to other acquisitions, because the observed spectral distributions of different images can be different for many reasons, e.g. variations in the observation geometry and mountain shadow. Therefore, a more robust method without heavy dependency on labelled samples and cumbersome topographic correction is needed.

Domain adaptation (also known as transfer learning) is one of the most promising methods to solve this problem. In the domain adaptation paradigm, a strong classifier is trained for a specific image (source domain) with adequate labelled samples and this classifier is then applied to a new acquisition (target domain) with the assistance of unlabelled samples (Liu and Li, 2014). This kind of method has been used in remote sensing classifications (Kurtz et al., 2014; Liu and Li, 2014), especially for the automatic updating of land cover maps (Bahirat et al., 2012; Bruzzone and Marconcini, 2009; Matasci et al., 2015). The main challenge for the application of this method to extract snow cover maps may be that it still needs sufficient labelled samples for a specific image or at least a part of an image.

Semi-supervised learning is another kind of promising approach, which can use a few labelled samples together with unlabelled samples to increase the reliability and accuracy of a classifier. Four paradigms of semi-supervised learning are encountered in literature, i.e. generative models (Shahshahani and Landgrebe, 1994), low density separation algorithms (Joachims, 1999; Vapnik, 1998), graph-based methods (Jordan, 1998), and co-training algorithms (Blum and Mitchell, 1998). All these semi-supervised methods have been successfully applied in remote sensing classifications with a small group of labelled samples (Bruzzone et al., 2006; Camps-Valls et al., 2007; Dalponte et al., 2015; Jackson and Landgrebe, 2001; Tan et al., 2014). However, these methods are merely suitable for the classification of single images and needs to be extended to deal with the multi-temporal classification.

In this study, we proposed a strategy to extend aforementioned co-training (also known as multi-view learning) algorithms from single image classification to multi-temporal ones so that a few labelled samples are sufficient to extract snow cover maps from multi-temporal images simultaneously. In the original concept of co-training, the feature set (e.g. spectral or texture features in terms of remote sensing data) should be split into two subsets, where each subset should be sufficient for training a strong classifier, and these classifiers are conditionally independent of each other for a given class label (Blum and Mitchell, 1998). The process of co-training is rather simple. Two classifiers are trained on two subsets for the same task first. Then these classifiers provide each other with labels for the unlabelled data. The unlabelled samples here serve as a "platform" for information exchange (Zhou and Li, 2010). Further studies showed that the assumption of two conditionally independent feature subsets was not necessary (Wang and Zhou, 2007). The key for the co-training approaches to succeed is that there exists a large difference between the classifiers, while it is not crucial in which the difference is introduced (Dasgupta et al., 2002; Wang and Zhou, 2007). Many variants of co-training have been proposed, e.g. an improved algorithm combining co-training with Expectation-Maximization (Co-EM) (Nigam and Ghani, 2000) and a further integration with SVM (Co-EM-SVM) (Brefeld and Scheffer, 2004).

For the snow cover extraction, the multi-temporal images provide multiple descriptions (multiple feature subsets) of the same snow cover area and these feature sets of snow cover can be different for many reasons, e.g. ageing of snow, contamination caused by dust, change of illumination, and observation geometry. As a result, the classifiers respectively trained on different images have a large difference, and the mutual learning based on the difference can, therefore, be used to improve the reliability of classifiers. Fig. 1 depicts the relationship between the conception of original co-training methods and the multi-temporal extension one for snow cover extraction from HSTRRS. It is expected that a few labelled samples are sufficient to extract snow cover from HSTRRS simultaneously by using the multi-temporal extension of co-training.

However, several issues should be carefully considered in the use of the multi-temporal extension of co-training methods. While different feature subsets split from one feature set naturally have the same labels in the co-training methods, it is not true in multi-temporal cases because of the possible transition between the snow and snow-free areas in different acquisitions. In addition, there are a large number of unlabelled samples in remote sensing. A selection procedure is necessary to choose proper unlabelled samples that can enhance the mutual learning. Furthermore, the co-training methods are inherent two-class methods. Further extension is needed to deal with the multi-temporal multi-class problems. In this study, these issues are addressed to extend the Co-EM-SVM from a single image classification to a multi-temporal one. It is worth noting that other co-training algorithms can also be extended to multi-temporal methods in a similar way.

The rest of this paper is organized in six sections. A brief introduction to Co-EM-SVM, followed by the proposed method, is presented in Section 2. The study area and data are described in Section 3. In Section 4, the experimental design is introduced. The performance of the proposed method is evaluated in Section 5. Sections 6 and 7 are discussions and conclusions, respectively.

2. Methodology

2.1. Co-EM-SVM

Co-EM-SVM (Brefeld and Scheffer, 2004) is an improved variant of co-training (Blum and Mitchell, 1998) and Co-EM (Nigam and Ghani, 2000). In Co-EM-SVM, the available feature set V of a data set is split into disjoint sets V_1 and V_2 (V_i is a collection of some