Remote Sensing of Environment 196 (2017) 293–311

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Generating a series of fine spatial and temporal resolution land cover maps by fusing coarse spatial resolution remotely sensed images and fine spatial resolution land cover maps

Xiaodong Li [a], Feng Ling [a,b,*], Giles M. Foody [b], Yong Ge [c], Yihang Zhang [a], Yun Du [a]

[a] Key Laboratory of Monitoring and Estimate for Environment and Disaster of Hubei Province, Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan 430077, China
[b] School of Geography, University of Nottingham, University Park, Nottingham NG7 2RD, UK
[c] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences & Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Article history:
Received 7 August 2016
Received in revised form 25 April 2017
Accepted 12 May 2017
Available online xxxx

Keywords:
Spatial temporal fusion
Super-resolution mapping
Endmember extraction

ABSTRACT

Studies of land cover dynamics would benefit greatly from the generation of land cover maps at both fine spatial and temporal resolutions. Fine spatial resolution images are usually acquired relatively infrequently, whereas coarse spatial resolution images may be acquired with a high repetition rate but may not capture the spatial detail of the land cover mosaic of the region of interest. Traditional image spatial–temporal fusion methods focus on the blending of pixel spectra reflectance values and do not directly provide land cover maps or information on land cover dynamics. In this research, a novel Spatial–Temporal remotely sensed Images and land cover Maps Fusion Model (STIMFM) is proposed to produce land cover maps at both fine spatial and temporal resolutions using a series of coarse spatial resolution images together with a few fine spatial resolution land cover maps that pre- and post-date the series of coarse spatial resolution images. STIMFM integrates both the spatial and temporal dependences of fine spatial resolution pixels and outputs a series of fine spatial–temporal resolution land cover maps instead of reflectance images, which can be used directly for studies of land cover dynamics. Here, three experiments based on simulated and real remotely sensed images were undertaken to evaluate the STIMFM for studies of land cover change. These experiments included comparative assessment of methods based on single-date image such as the super-resolution approaches (e.g., pixel swapping-based super-resolution mapping) and the state-of-the-art spatial–temporal fusion approach that used the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM) and the Flexible Spatiotemporal DAta Fusion model (FSDAF) to predict the fine-resolution images, in which the maximum likelihood classifier and the automated land cover updating approach based on integrated change detection and classification method were then applied to generate the fine-resolution land cover maps. Results show that the methods based on single-date image failed to predict the pixels of changed and unchanged land cover with high accuracy. The land cover maps that were obtained by classification of the reflectance images outputted from ESTARFM and FSDAF contained substantial misclassification, and the classification accuracy was lower for pixels of changed land cover than for pixels of unchanged land cover. In addition, STIMFM predicted fine spatial–temporal resolution land cover maps from a series of Landsat images and a few Google Earth images, to which ESTARFM and FSDAF that require correlation in reflectance bands in coarse and fine images cannot be applied. Notably, STIMFM generated higher accuracy for pixels of both changed and unchanged land cover in comparison with other methods.

1. Introduction

Land cover maps are one of the most fundamental datasets used in many scientific fields and are often produced from remotely sensed images (Bartholomé and Belward, 2005; Friedl et al., 2002). A wide variety of remote sensing systems have been developed, and hence, images are available with different spatial and temporal resolutions, thereby allowing the production of land cover maps at different spatial and temporal scales. With most satellite remote sensing systems, a trade-off typically exists between spatial and temporal resolution. In general, fine spatial resolution remote sensors can acquire images that provide spatially detailed land cover information, but their relatively coarse temporal resolution limits their usage in monitoring rapid land cover changes. By contrast, coarse spatial resolution remotely sensed images can often be acquired at a fine temporal resolution that provides a repetition rate

* Corresponding author at: Key Laboratory of Monitoring and Estimate for Environment and Disaster of Hubei Province, Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan 430077, China.
E-mail address: lingf@whigg.ac.cn (F. Ling).

http://dx.doi.org/10.1016/j.rse.2017.05.011

suitable for the detection of rapid land cover changes but are unable to represent the spatial detail of the land cover mosaic. To realize the full potential of remote sensing as a source of information on land cover change, a method that allows the production of land cover maps with both fine spatial and temporal resolutions is required. Such maps could be obtained by combining all available remotely sensed images of varying spatial and temporal resolution to form a series of fine-resolution land cover maps.

Recently, spatial–temporal image fusion, which aims to produce fine spatial and temporal resolution remotely sensed images from images with different spatial and temporal resolutions, has become a promising means to address the trade-off between spatial and temporal resolution (Gevaert and Garcia-Haro, 2015; Zhu et al., 2016). Spatial–temporal data fusion methods can be categorized into weighted function based methods, unmixing-based methods, and dictionary-pair learning based methods (Zhu et al., 2016). Among the weighted function based methods, the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) proposed by Gao et al. (2006) was developed first and is one of the most popular spatial–temporal image fusion methods. By fusing coarse spatial resolution Moderate Resolution Imaging Spectroradiometer (MODIS) and fine spatial resolution Landsat sensor images, STARFM can predict Landsat-like reflectance images with the spatial resolution of Landsat and the temporal resolution of MODIS. A number of studies have suggested improvements to STARFM, including studies of forest disturbance (Hilker et al., 2009a), and in heterogeneous regions (Zhu et al., 2010), as well as in gap filling to reduce the negative effects of cloud (Gevaert and Garcia-Haro, 2015). STARFM and the improved models based on it have been mainly used to detect reflectance changes caused by processes such as phenology over large areas, and used to generate dense time series of Landsat-like data (Hilker et al., 2009b), enhance land cover classification (Jia et al., 2014), and predict key environmental variations such as evapotranspiration (Anderson et al., 2011) and temperature (Hilker et al., 2009b). Other spatial–temporal image fusion models, such as the unmixing-based algorithm that extracts endmembers on the basis of linear spectral mixture model and assigns the unmixed reflectance to fine spatial resolution pixels (Huang and Zhang, 2014; Zhukov et al., 1999; Zurita-Milla et al., 2009) and the dictionary-pair learning based methods, which capture features from the coarse- and fine-resolution image pairs used for predicting fine-resolution image (Huang and Song, 2012), have also been proposed and applied to Landsat and MODIS images in recent years (Amorós-López et al., 2013; Wu et al., 2012).

Generally, spatial–temporal image fusion models aim to generate a series of continuous reflectance values instead of discrete categorical values. A further image classification step is needed to produce from the reflectance images a corresponding series of land cover maps for the study of land cover class dynamics (Jia et al., 2014). The use of these methods for generating land cover maps and monitoring land cover changes often suffers from two important limitations.

First, most spatial–temporal image fusion algorithms assume that land cover type does not change during the data observation period (Fu et al., 2013; Gao et al., 2006; Zhu et al., 2010). Previous research has shown that STARFM does not deal well with abrupt land cover changes. Song and Huang (2013) showed that STARFM failed to fuse the pixel reflectance accurately in a study of land cover change in an urban area. The Enhanced STARFM (ESTARFM) is often better than STARFM for studies of heterogeneous landscapes (Zhu et al., 2010) but can be worse than STARFM for predicting abrupt changes of land cover type (Emelyanova et al., 2013). The Spatial Temporal Adaptive Algorithm for mapping Reflectance CHange (STAARCH) improves STARFM's performance when land cover type change and disturbance exist, but it is more suitable for spatial–temporal fusion of forest land cover (Hilker et al., 2009a). The Flexible Spatiotemporal DAta Fusion model (FSDAF) can predict Landsat-like reflectance values with both gradual change and land cover type change, but it cannot capture tiny changes in land cover type, such as when only a few fine pixels experienced land cover type change and the change is invisible in the coarse-resolution image (Zhu et al., 2016). Similar to STARFM, the unmixing-based spatial–temporal reflectance fusion methods consider only the change in endmember spectra but not in land cover types (Huang and Zhang, 2014; Zhukov et al., 1999; Zurita-Milla et al., 2009).

Second, most spatial–temporal image fusion methods need one or more observed pairs of coarse- and fine-resolution images for training and require the coarse- and fine-resolution remotely sensed data from different satellite sensors to be mutually comparable and correlated. All the weighted function based methods, including STARFM, ESTARFM, STAARCH, and all the dictionary-pair learning-based methods need one or more observed pairs of coarse- and fine-resolution images, which have comparable reflectance bands, for training (Gao et al., 2006; Gevaert and Garcia-Haro, 2015; Zhu et al., 2010). These methods mainly focus on predicting Landsat-like remotely sensed images with MODIS repetition rates. However, these methods cannot deal with other satellite images, which have uncorrelated reflectance bands, and are thus limited in the use of land cover change analysis. For instance, in regional-scale land cover analysis, the detection of very-high-resolution land cover changes at high temporal resolutions is required. In general, we can obtain a series of Landsat images and a few very-high-resolution images such as panchromatic aerial photograph. The weighted function based and dictionary-pair learning based methods cannot fuse these data because the very-high-resolution images usually have different reflectance bands compared with Landsat images.

The spatial–temporal image fusion methods aim to produce fine spatial–temporal resolution reflectance images rather than land cover maps. The fused fine-resolution images have many applications, such as phenology analysis. If the aim is to generate a sequence of land cover maps from the reflectance images from which land cover change trajectories may be extracted, then a further image classification analysis is still required, which may introduce uncertainty and error in the land cover maps. First, the classification of an image series can be complex and laborious. Training statistics are required to inform classification analysis, and these may vary in quality in time due to issues such as phenology. Moreover, the classification is also problematic, with the potential for different classifiers to generate dissimilar land cover maps from the same training data. Traditional classifiers applied to mono-temporal image may also ignore the temporal information contained in a series of images and thereby produce a classification of sub-optimal accuracy. The spatial–temporal–based image classifier has the advantage in taking both the spatial and temporal links between neighboring pixels (Cai et al., 2014), but is challenging to use for voluminous image series (Liu and Cai, 2012; Liu et al., 2006). Finally, the spatial–temporal image fusion models generate a large volume of fine spatial–temporal resolution reflectance images as the intermediate data to be used for the production of land cover maps. This situation may represent practical challenges in terms of data access and storage.

Given the concerns with the spatial–temporal reflectance fusion model for producing land cover maps, a more appropriate fusion approach could be based on directly downscaling the coarse spatial resolution image series to fine spatial resolution land cover maps rather than reflectance images, with the aid of information derived from a few fine spatial resolution images that may be available. Chen et al. (2015) updated land cover maps from downscaled Normalized Difference Vegetation Index (NDVI) time-series data from MODIS, a current Landsat image, and a Landsat image that pre-dates it. The NDVI time-series data are used as ancillary data to extract changed pixels in the Landsat images, and the labels of changed pixels are determined using the current Landsat image. Thus, this method can update fine-resolution land cover maps with Landsat repetition rates based on available Landsat images, but cannot predict fine-resolution land cover maps with MODIS repetition rates. In addition, a major problem with this approach is that a large proportion of coarse spatial resolution image pixels may be of mixed land cover composition. A possible solution of this problem is to use the fractional land cover class composition images that can be