ISPRS Journal of Photogrammetry and Remote Sensing 108 (2015) 191–204

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Decision fusion and non-parametric classifiers for land use mapping using multi-temporal RapidEye data

Fabian Löw [a,*], Christopher Conrad [a], Ulrich Michel [b]

[a] Department of Remote Sensing, Würzburg University, Am Hubland, Germany
[b] Department of Geography, University of Education Heidelberg, Germany

ARTICLE INFO

Article history:
Received 24 December 2014
Received in revised form 27 May 2015
Accepted 3 July 2015
Available online 25 August 2015

Keywords:
Agricultural land use
Supervised classification
Decision fusion
High-resolution
Multi-temporal
RapidEye

ABSTRACT

This study addressed the classification of multi-temporal satellite data from RapidEye by considering different classifier algorithms and decision fusion. Four non-parametric classifier algorithms, decision tree (DT), random forest (RF), support vector machine (SVM), and multilayer perceptron (MLP), were applied to map crop types in various irrigated landscapes in Central Asia. A novel decision fusion strategy to combine the outputs of the classifiers was proposed. This approach is based on randomly selecting subsets of the input dataset and aggregating the probabilistic outputs of the base classifiers with another meta-classifier. During the decision fusion, the reliability of each base classifier algorithm was considered to exclude less reliable inputs at the class-basis. The spatial and temporal transferability of the classifiers was evaluated using data sets from four different agricultural landscapes with different spatial extents and from different years. A detailed accuracy assessment showed that none of the stand-alone classifiers was the single best performing. Despite the very good performance of the base classifiers, there was still up to 50% disagreement in the maps produced by the two single best classifiers, RF and SVM. The proposed fusion strategy, however, increased overall accuracies up to 6%. In addition, it was less sensitive to reduced training set sizes and produced more realistic land use maps with less speckle. The proposed fusion approach was better transferable to data sets from other years, i.e. resulted in higher accuracies for the investigated classes. The fusion approach is computationally efficient and appears well suited for mapping diverse crop categories based on sensors with a similar high repetition rate and spatial resolution like RapidEye, for instance the upcoming Sentinel-2 mission.

1. Introduction

Agricultural management increasingly uses thematic maps based on classification of remotely sensed data sets for monitoring systems (Justice and Becker-Reshef, 2007). Maps of agricultural land use provide important information to support decision makers and agricultural policies, for instance to verify claims by farmers who apply for public subsidies, or assisting in the practice of precision agriculture (c.f. Alganci et al., 2013). Challenges in operational use of satellite remote sensing for crop mapping include consistently creating accurate maps across different landscapes and years with a high spatial and thematic detail. Such maps should provide accurate information on spatial cropping pattern or crop acreage over larger geographical areas (c.f. de Wit and Clevers, 2004), because they are often used as input to spatially explicit assessments like crop water requirement calculation (Conrad et al., 2013) or crop yield modelling (Lobell et al., 2003).

For this reason, two issues become vitally important, namely the sources of satellite images, which determines the spatial and temporal resolution, and the choice of appropriate classifier algorithms. Crop mapping based on mono-temporal data sets is often deemed inaccurate, for instance when rotations with two or more crops are present (Conrad et al., 2014; Löw et al., 2013) and because agricultural landscapes are characterized by high intra-class variability of multiple and complex land use types that are difficult to separate spectrally due to low inter-class separability (Atzberger, 2013). Hence, multi-temporal data sets of different kind are usually being used for this purpose (Wardlow and Egbert, 2008; Waske and Braun, 2009). A distinction between different land uses in such multi-temporal data sets is based on the crop-specific temporal signatures, e.g. characteristics that correspond to key phenological stages (e.g. green-up, peak, senescence). Up to a certain level, classification accuracies can be increased by

* Corresponding author.
E-mail addresses: fabian.loew@uni-wuerzburg.de, fabian.loew@gmx.de (F. Löw).

http://dx.doi.org/10.1016/j.isprsjprs.2015.07.001
0924-2716/

increasing the quantity of images (Conrad et al., 2011; Murakami et al., 2001; Wardlow et al., 2007).

However applications with sensors like MODIS (250 m), which have such a high revisit frequency, can be prohibitive in landscapes where the field size is smaller than the pixel size and thus misclassification increases (Hsieh et al., 2001; Löw and Duveiller, 2014; Ozdogan and Woodcock, 2006). With recent missions like RapidEye with 6.5 m multi-spectral resolution (Tyc et al., 2005), SPOT 6/7 (with 6 m multi-spectral resolution and 1.5 m panchromatic resolution), Landsat-8 with 30 m resolution (15 m panchromatic) and PROBA-V with up to 100 m at nadir for the visible near infrared (VNIR) resolution (Francois et al., 2014; Sterckx et al., 2014), or up-coming missions like Sentinel-2 with 20 m resolution (Drusch et al., 2012) and VENμS with 10 m multi-spectral -resolution (CNES, 2015), regular and regional assessments of agricultural production are theoretically possible as these sensors provide a high spatial resolution plus a high revisit frequency that is necessary to identify and distinguish agricultural crops from other land uses. For instance, much can be expected from the global coverage at 20 m every 5 days using the Sentinel-2 platforms, provided with a relatively wide swath of 290 km.

Although still often used (e.g. El-Magd and Tanton, 2003; Turker and Ozdarici, 2011), conventional parametric methods like the maximum likelihood classifier (MLC) are often not optimal for classifying such multi-temporal (e.g. Löw et al., 2012) or multisource data sets (Waske and Benediktsson, 2007) because they make assumptions like parametric distribution of the input data. Non-parametric algorithms like support vector machines (SVMs) (Cortes and Vapnik, 1995) or random forests (RFs) (Breiman, 2001) have been successfully used for the classification of diverse remote sensing data sets like multi-temporal optical data (Conrad et al., 2014; Duro et al., 2012; Löw et al., 2013), SAR data (Waske and Braun, 2009), or a combination of both (Waske and van der Linden, 2008). The achieved overall accuracies are often high and significantly improved, compared to conventional statistical classification algorithms like MLC (Pal and Mather, 2005; Pal, 2005; Waske et al., 2010a). Still, non-parametric classifiers can show considerable disagreement in classifying land use in maps, which means that they produce complementary results (Fauvel et al., 2006; Liu et al., 2004) and hence are accurate in different locations in a map (e.g. Clinton et al., 2015).

By making use of such complementary behaviour, ensemble based systems (Policar, 2006) have been developed in remote sensing, termed "consensus classification" (Benediktsson and Sveinsson, 2003), "multiple classifier systems" (Benediktsson et al., 2007; Du et al., 2012) or "decision fusion" (Benediktsson and Kanellopoulos, 1999; Fauvel et al., 2006). One strategy is to create variants of the same classifier algorithm, for instance through boosting, i.e. the sequential reweighting of training samples (Freund and Schapire, 1996), or bagging, i.e. the generation of training sample subsets (Breiman, 1996). However, boosting can be computationally demanding because the data is processed in series and it can perform less efficient than bagging, which can minimize the sensitivity of the classification algorithm to noise in feature data and labelling errors in training data. The training of variants of a classifier based on several, randomly generated feature subsets (Bryll et al., 2003; Ho, 1998) can provide very accurate results (Waske et al., 2010b). Another strategy is combining different types of data sets, for instance optical and SAR, or to fuse the outputs from different classifier algorithms (Doan and Foody, 2007; Waske and Benediktsson, 2007). Different strategies were developed to combine the results, for instance algebraic combination rules, majority voting strategies, or the use of meta-classifiers (Kittler et al., 1998; Policar, 2006; Waske and Benediktsson, 2007; Wolpert, 1992).

Despite this history of research (c.f. Du et al., 2012), a significant increase in classification accuracy is not warranted (Foody et al., 2007; Giacco et al., 2010). For instance, many fusion approaches do not explicitly take input account the reliability of the single classifier algorithms during the fusion process, which can impact classification accuracy (Fauvel et al., 2006; Jeon and Landgrebe, 1999). Only few studies have been reported to evaluate high-resolution, multi-temporal data sets in this context, although there is indication that pixel-based classification accuracies can be more accurate with spatial resolutions higher than 10 m (Alganci et al., 2013; Turker and Ozdarici, 2011). Further, classification uncertainty can increase when pixel sizes exceed the size of agricultural fields (Löw and Duveiller, 2014) and even in object-based classification the presence of sub-field plots with different crop types can increase the classification uncertainty (Schorcht et al., 2012). In this situation, data sets that combine a high temporal with a high spatial resolution, like RapidEye, are probably the most valuable data source. But even with non-parametric algorithms like SVM there can be speckle in maps, which could be reduced by decision fusion (e.g. Waske and van der Linden, 2008; Waske et al., 2010a,b). Yet, investigations about the performance of different classifier algorithms including decision fusion in classifying data sets with both, a high spatial and temporal resolution like RapidEye or time series of Quickbird are rare (e.g. Du et al., 2013), in particular for agricultural crop mapping.

This study therefore examines the performance of different classifier algorithms and fusion techniques, applied to multi-temporal data sets from RapidEye for mapping different crop categories in irrigated agricultural landscapes in Central Asia. A selection of different, non-parametric classifier algorithms was evaluated: DT, RF, SVM, and multilayer perceptron neural networks (MLPs), which have only attained little attention in the context of agricultural land use classification yet. A comparison of these approaches is worthwhile, regarding the numerous applications that are based on these algorithms as well as the general differences of the underlying classifier concepts. In addition, a novel method to fuse the probabilistic outputs of these algorithms is proposed. The innovative aspect of this fusion is that it adapts the idea of random feature selection with an approach that takes into account the reliability of each single classifier algorithm during the combination process. It is evaluated which classifier performs the most accurate and whether or not decision fusion can enhance the accuracy and quality of land use maps. Moreover, the impact of training sample size on the classifier performance and the amount of speckle in the maps is investigated. Finally, the transferability of the approaches is evaluated in four different agricultural landscapes with different spatial extents and in different years.

2. Study area

This study is based on different agro-ecological landscapes in Central Asia (Table 1 and Fig. 1). They are located alongside the Amu-Darya and Syr-Darya rivers and are characterized by vast agricultural systems, which were extensively developed under the aegis of the former Soviet Union during the second half of the 20th century (Saiko and Zonn, 2000). The methods were tested in four study sites with a size of 30 km $\times$ 30 km, hereafter referred to as "small" test sites. Further, the spatial transferability of methods was tested in two larger study sites.

The first small site is located in the Khorezm region (KHO) in the northwestern part of Uzbekistan (Fig. 1). The agricultural landscape appears fragmented due to a comparatively high diversity of crops (e.g. cotton, rice, sorghum, and maize, winter wheat, and fruit trees). Multiple cropping is sometimes practiced (growing