Computers & Geosciences 94 (2016) 56–67

Contents lists available at ScienceDirect

Computers & Geosciences

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

Research paper

Realizing parameterless automatic classification of remote sensing imagery using ontology engineering and cyberinfrastructure techniques

Ziheng Sun [a,*], Hui Fang [a], Liping Di [a,*], Peng Yue [b]

[a] Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA 20030, USA
[b] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, Hubei 430079, China

ARTICLE INFO

Article history:
Received 30 September 2015
Received in revised form
9 April 2016
Accepted 6 June 2016
Available online 11 June 2016

Keywords:
Parameterless automatic classification
Ontology database
Cyberinfrastructure
Remote sensing image classification
Geoprocessing Web service
Geoprocessing workflow

ABSTRACT

It was an untouchable dream for remote sensing experts to realize total automatic image classification without inputting any parameter values. Experts usually spend hours and hours on tuning the input parameters of classification algorithms in order to obtain the best results. With the rapid development of knowledge engineering and cyberinfrastructure, a lot of data processing and knowledge reasoning capabilities become online accessible, shareable and interoperable. Based on these recent improvements, this paper presents an idea of parameterless automatic classification which only requires an image and automatically outputs a labeled vector. No parameters and operations are needed from endpoint consumers. An approach is proposed to realize the idea. It adopts an ontology database to store the experiences of tuning values for classifiers. A sample database is used to record training samples of image segments. Geoprocessing Web services are used as functionality blocks to finish basic classification steps. Workflow technology is involved to turn the overall image classification into a total automatic process. A Web-based prototypical system named PACS (Parameterless Automatic Classification System) is implemented. A number of images are fed into the system for evaluation purposes. The results show that the approach could automatically classify remote sensing images and have a fairly good average accuracy. It is indicated that the classified results will be more accurate if the two databases have higher quality. Once the experiences and samples in the databases are accumulated as many as an expert has, the approach should be able to get the results with similar quality to that a human expert can get. Since the approach is total automatic and parameterless, it can not only relieve remote sensing workers from the heavy and time-consuming parameter tuning work, but also significantly shorten the waiting time for consumers and facilitate them to engage in image classification activities. Currently, the approach is used only on high resolution optical three-band remote sensing imagery. The feasibility using the approach on other kinds of remote sensing images or involving additional bands in classification will be studied in future.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

After decades of studies on remote sensing image classification, researchers have recognized that it is not an easy task to automatically and accurately classify the image pixels (Sabins, 1978; Lillesand et al., 2004; Richards and Richards, 1999). Most of the current classification activities need a lot of human intervention to find the best combination of input parameter values for the adopted classification algorithms (Schowengerdt, 1983; Blaschke, 2010). For example, the artificial neural network (ANN) based classification methods (Benediktsson et al., 1990; Heermann and Khazenie, 1992; Bischof et al., 1992) usually require users to specify the number of hidden layers, the number of neurons in each layer, the function on each neuron, the initial weights of each border and training samples before every classification. Any change on these parameters could lead to a completely different outcome. To obtain a satisfying result, experts have to try a lot of possibilities to figure out a relatively better combination. The finding process is very time consuming and full of uncertainties. In the mean time, a combination of parameter values of an algorithm might be only suitable for a particular image. Once the target image changes, the combination need be reset too. Another thing that makes it worse is that even if the same algorithm is applied

* Corresponding authors.
E-mail addresses: szhwhu@gmail.com (Z. Sun), ldi@gmu.edu (L. Di).

http://dx.doi.org/10.1016/j.cageo.2016.06.004
0098-3004/© 2016 Elsevier Ltd. All rights reserved.

upon the same image, different experts may get different results. Because people, who have various preferences, have different definition about "good result".

Due to the heavy workload and uncertainties in pursuing for a good combination of parameters, the work of classifying a remote sensing image seems very difficult to be finished in a both automatic and accurate way. Most current researches still depend on traditional classification strategy with a lot of manual parameter tuning work that is greatly impacted by user preferences and costs too much labor and time. It remains a very big challenge to automate the whole classification process, remove the impact of human preference driven activities on the final results and reduce the overall workload and time cost.

This paper presents an idea of parameterless automatic classification trying to meet the challenge. Parameterless automatic classification requires only one image from consumers as input and outputs a labeled vector automatically. No parameter adjusting is needed. The whole process is a black box to endpoint consumers. An approach is proposed to realize the idea. It needs an ontology database to store the experiences on parameter tuning for various classification algorithms in many circumstances. The knowledge in the database should be repeatedly tested and acknowledged (Mizoguchi and Ikeda, 1998). The knowledge can help avoid remote sensing experts from repeating the time-consuming parameter tuning work on a new arriving image. A sample database is used to record training samples of image segments. Web services in Cyberinfrastructure (Hey and Trefethen, 2005; Sun et al., 2014) are involved to supply the functionality to complete all the basic steps within each classification. Workflow technology is adopted to chain Web services together into a workflow instance that can be executed automatically. The values for the input parameters of each Web service are automatically selected from the knowledge database. The selection is driven by reasoning rules and target image properties such as location, resolution, size, band count, camera, date, format, etc.

To validate the proposed approach, we implemented a Web-based prototypical system named PACS (short for **P**arameterless **A**utomatic **C**lassification **S**ystem). The knowledge database and sample database are currently built for high resolution (HR) and very high resolution (VHR) optical remote sensing images. So for now the system can only process three-band optical HR/VHR remote sensing images. Basically, the spatial resolution of HR optical images is between 1 and 10 meters (e.g., Geoeye-1 Blue/Green/Red, WorldView-2 Blue/Green/Yellow/Red, QuickBird-2 Blue/Green/Red, SPOT-5 PAN, IKONOS Blue/Green/Red, CBERS-2B PAN) while the spatial resolution of VHR optical images is less than 1 meter (e.g., Geoeye-1 PAN, WorldView PAN, Airbone, IKONOS PAN) (Pesaresi et al., 2013).

A few images were tested on PACS. The results prove the feasibility of the proposed approach. It is indicated that the classified results will be more accurate if the two databases have higher quality. Once the experiences and samples in the databases accumulate into the same level as a human expert owns, the approach may get the results with similar quality to those produced manually.

The approach can bring a number of benefits to remote sensing image classification community. First, endpoint consumer only need submit an image and leave the technical things behind. The approach requires no parameter configuration and is totally automatic. Secondly, for remote sensing experts, their work becomes easier as they do not have to make duplicated efforts on tuning input parameter values if the relevant experiences already exist in the database. On the other side, they can also contribute their experiences to augment the knowledge database. When other experts want to classify a new image with similar features, the knowledge will provide valuable references and help quickly obtain reliable classification results. Another benefit is that the approach makes it very possible to process the massively increasing amount of remote sensing images in time. With the quick development of Earth observation sensors, the acquired remote sensing images are overwhelming everyday. Traditional manual image classification method is slow and can hardly discover the valuable information in most of the images in time, especially during an ongoing disaster such as earthquake, flood, typhoon, landslide, debris flow, etc. The proposed approach is automatic and parameterless so that the extra waiting time costs and possible error operations caused by human experts are eliminated. Thus, the approach provides a very good chance to match image processing capability with the massive increasing volume of images.

The remainder of the paper is organized as follows. Section 2 introduces the related work. In Section 3 an approach with three major modules is proposed. Section 4 implements PACS. Section 5 introduces some experiments conducted on the system and evaluates the approach based on the results. Section 6 discusses the pros and cons of the approach. Section 7 concludes the paper and gives the future work.

2. Related work

Parameterless or automatic parameter tuning classification/clustering methods have been studied for decades in other domains such as mathematics (Mangasarian, 2005; Mohd et al., 2012; Jones et al., 1998; Bergstra and Bengio, 2012), statistics clustering (Sheikh et al., 2005), medical science (Yang et al., 2013), biology (Cawley and Talbot, 2006), artificial intelligence (Farrahi Moghaddam and Cheriet, 2012; Berglund and Sitte, 2006; Wang et al., 2013), face recognition (Dornaika and Assoum, 2013) and text detection (Wu et al., 2002). For example, Foss and Zaïane (2002) presented a parameterless clustering algorithm that clusters over a range of resolutions and finds a potential optimum clustering without requiring any parameter inputs. Chung et al (Chung et al., 2005) proposes an efficient and high performance parameterless quadrilateral-based algorithm for image segmentation. The experiment results show that the algorithm performs better than three other segmentation algorithms on over or under segmenting situation. Lezoray et al. (2007) introduced a parameterless discrete regularization framework and applied it in color image filtering. Pătrăucean et al. (2012) present a parameterless line segment and elliptical arc detector. Levachkine et al. (2002) proposed an approach to automatically interpret the semantics of image segments by low level image processing techniques. But these approaches are difficult to be directly transplanted to remote sensing domain due to the complexity, variety, uncertainty of the contained objects in remote sensing images.

In remote sensing domain, some researches have been conducted for automatic classification of geospatial objects like buildings, roads, lakes, rivers, forests, planes and cars from remote sensing imagery. Many attempts have been made but few substantial progresses were achieved. Benz et al. (2004) tried to use object oriented analysis combining with fuzzy methods to semi-automatically vectorize remote sensing imagery and import the results into GIS (geographical information system). Inglada (2007) proposed an automatic recognition approach to discover man-made objects in high resolution optical remote sensing images by using SVM classification based on geometric image features. Sun et al. (2012) use spatial sparse coding bag-of-words (BOW) model to detect targets with complex shape in high resolution remote sensing images. Ren and Chang (2003) developed a spectral-based algorithm for automatic spectral target recognition (ASTR) in hyperspectral imagery with no requirements on a priori knowledge.

Download English Version:

https://daneshyari.com/en/article/506995

Download Persian Version:

https://daneshyari.com/article/506995

Daneshyari.com