ISPRS Journal of Photogrammetry and Remote Sensing xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

A web-based system for supporting global land cover data production

Gang Han [a,b,*], Jun Chen [a], Chaoying He [a], Songnian Li [c,d], Hao Wu [a], Anping Liao [a], Shu Peng [a]

[a] National Geomatics Center of China, Beijing 100830, China
[b] College of Resources Science & Technology, Beijing Normal University, Beijing 100875, China
[c] School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China
[d] Department of Civil Engineering, Ryerson University, Toronto, Ontario M5B 2K3, Canada

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Web Service
Global land cover mapping
Data production
Verification
CSCW
Web GIS

ABSTRACT

Global land cover (GLC) data production and verification process is very complicated, time consuming and labor intensive, requiring huge amount of imagery data and ancillary data and involving many people, often from different geographic locations. The efficient integration of various kinds of ancillary data and effective collaborative classification in large area land cover mapping requires advanced supporting tools. This paper presents the design and development of a web-based system for supporting 30-m resolution GLC data production by combining geo-spatial web-service and Computer Support Collaborative Work (CSCW) technology. Based on the analysis of the functional and non-functional requirements from GLC mapping, a three tiers system model is proposed with four major parts, i.e., multisource data resources, data and function services, interactive mapping and production management. The prototyping and implementation of the system have been realised by a combination of Open Source Software (OSS) and commercially available off-the-shelf system.

This web-based system not only facilitates the integration of heterogeneous data and services required by GLC data production, but also provides online access, visualization and analysis of the images, ancillary data and interim 30 m global land-cover maps. The system further supports online collaborative quality check and verification workflows. It has been successfully applied to China's 30-m resolution GLC mapping project, and has improved significantly the efficiency of GLC data production and verification. The concepts developed through this study should also benefit other GLC or regional land-cover data production efforts.

1. Introduction

The need for finer spatial resolution (~30 m) GLC data has been recognized by user groups and remote sensing community for global environmental change studies, biodiversity monitoring, earth system modelling and many other societal benefits areas (Herold et al., 2008; Verburg et al., 2011; Hansen and Loveland, 2012; Gong et al., 2013). Several 30-m resolution GLC projects or initiatives have been launched or announced by national and international agencies in the past few years (Chen et al., 2011; Townshend et al., 2012; Giri et al., 2013). Since the beginning of 2010, China has started to develop 30-m resolution GLC data sets for two baseline years, i.e., 2000 and 2010 (Chen et al., 2011). It is an operational mapping project which aims to produce reliable data products by overcoming the difficulties caused by the spectral confusion among different land cover types and the spectral diversity within the same land cover type (Chen et al., 2014). All available ancillary data and prior knowledge have to be collected and integrated into the mapping process to improve the quality of the land cover characterization. Groups of people located in different geographic places need to perform the data processing, information extraction, verification, quality controlling, post-classification processing, as well as product validation in a collaborative way. Advanced information tools are requested for supporting this massive and operational GLC mapping.

Geographic Information System (GIS) has been considered as an essential tool to integrate ancillary data and prior knowledge, which might vary in data format, accuracy, spatial resolution, and coordinate systems (Ehlers et al., 1989; Lu and Weng, 2007). The major roles of GIS lie in multisource data management (Wang and Howarth, 1994; Heipke et al., 2004), pixel or object-based classification (Dean and Smith, 2003; Bouziani et al., 2010),

* Corresponding author at: No. 1, Baishengcun, Zizhuyuan Road, Haidian District, Beijing 100048, China.
E-mail address: hgbj0001@gmail.com (G. Han).

http://dx.doi.org/10.1016/j.isprsjprs.2014.07.012
0924-2716/

post-classification processing and verification (Huang and Jia, 2102). With the advances of Internet technology, WebGIS has been used recently to integrate distributed imagery and ancillary data from geographically dispersed researchers without considering the data sources (Clark and Aide, 2011), to facilitate on-line data access (Hansen et al., 2011; Li et al., 2011), to aid crowd-sourcing sampling and product verification (Heipke, 2010; Fritz et al., 2012). A recent example is Geo-Wiki, a web-based platform that combines different GLC maps and Google Earth high-resolution images for people to collect more reference and in-situ sample data that are proven very useful for verifying and improving the image classification results (Fritz et al., 2012). On the other hand, computer-supported cooperative work (CSCW) has been introduced to the geo-spatial data production to facilitate real-time access of the project information, control of production process activities, and automation of production procedures through more efficient communication, and collaboration of project participants and coordination of data production processes (Li and Coleman, 2003; Chen et al., 2004). However, a combination of the web service and CSCW technology for supporting GLC mapping has not been discussed and implemented yet.

This paper presents a web-based system for supporting 30-m resolution GLC data production by combining the geospatial web services and CSCW technology. This system is particular useful for better sharing of various datasets, supporting post-classification processing, and facilitating cooperation of data quality checks and verification. It is used to integrate dispersed and heterogeneous data resources, including primary Landsat-like images, preliminary classification results and various ancillary datasets. It also can integrate existing GLC maps and many external web services for such data as Google Earth high-resolution satellite imagery into a single Web 2.0 application. Moreover, this system provides special functionalities for supporting data access and processing to every participant in data production process and production activities, such as identification of land cover types, inconsistency detection and product verification. A "shared environment" is supported by a well-defined collaborative production and verification workflows, which connect participants together, assign them different roles and tasks, and ensure the smooth execution of the underlying processes.

The aim of this paper is to describe the design and development of such a web-based and GIS-enabled system for supporting GLC data production and its utilization in China's 30-m GLC mapping project. The remainder of this paper is organized as follows. Section 2 characterizes GLC data production and reviews related works. Section 3 presents the web service system model describing system components and architecture. Section 4 discusses the implementation of this system, followed by some walk through examples in global 30-m GLC mapping production in Section 5. Finally, some conclusions and future research directions are discussed in Section 6.

2. GLC data production support and related works

A good understanding of the existing GLC data production processes and the supporting requirements is essential for the design and development of the proposed web-based system. The characteristics of 30-m GLC data production and the related research works on supporting systems are examined.

2.1. Supporting requirements from GLC data production

China's 30-m GLC mapping project has collected more than 20,000 scenes of Landsat and other 30-m resolution images for global land cover characterization. However, only use the spectral and texture characteristics of images is not enough for more accurate land-cover classifications at global scale. Previous studies showed that ancillary information can improve the accuracy of image-based land-cover classification (Chen, 1984. Loveland et al., 1999; Muchoney et al., 1999; Mayaux et al., 2003; Lu and Weng, 2007; Huang and Jia, 2102). For instance, the elevation, slope and aspect data can help the identification of forest distribution in mountainous areas with the knowledge of specific vegetation classes and topographic factors, and population density data can be used to distinguish commercial and high-intensity residential areas (Lu and Weng, 2007). It is therefore important to integrate all the available ancillary information from a great variety of sources, and to use them in the identification and verification of land covers during classification processes. A number of ancillary datasets have been collected, such as the existing 6 global land cover datasets, 30-m national land cover data of United States, 30-m CORINE from EU, global DEM, global ecological zones data, etc. Table 1 lists some examples of these data that are from different sources with different specifications. It can be seen that the amount of image data and ancillary data is huge, while most of them have different spatial resolution, format and projection, and some global and regional land cover data sets have different classification classes. Problems arise in dealing with such huge amount of image data and the optimum utilization of the multi-source ancillary data in GLC data production, quality check and data verification. Among them, data distribution, access, format conversion, and the time spent by data producers and contributors in searching for and obtaining data are of major concern. How to enable easy access and use of these data during production process represents a big challenge.

GLC data production is a complex process composed of different kinds of activities, such as sampling connection, image classification, data verification, validation and quality control (Belward et al., 1999; Defries and Townshend, 1999; Estes et al., 1999; Fritz et al., 2003; Lu and Weng, 2007; Tateishi et al., 2011). Each of the processes might need a large group people working on different assigned roles. China's 30-m GLC mapping project partitions the whole world into 5 regional windows, and more than 200 people located in different places participate in the data production. Close collaboration with the project partners is crucial to the success of the project (Chen et al., 2014). This collaborative work requires not only the right data and information to be flowed through the process to the right steps at the right time, but also the coordinated efforts between the people working on individual tasks.

The other important requirement is about the production tools for the operational GLC mapping. Several commercial image processing and GIS software, such as ArcGIS, ERDAS and ENVI, eCognition has been used in the GLC data production to perform required pre- and post-processing of images, image classifications, and handling of ancillary data. Many of them are designed for general purpose without addressing special needs of global land cover characterization in 30 m resolution. On the other hand, a number of algorithms and tools have been developed in house by the GLC mapping project, such as pre-processing tool and water extraction algorithm, etc. But they are scattered in different standalone packages, requiring certain amount of learning effort to use them. It is therefore becoming important to develop some special tools to support collaborative production, especially data dissemination, annotation, inconsistence detection, as well as data verification.

The Internet, as a bridge to connect people and information around the world, makes the globe become smaller and smaller (Butler, 2006). The Internet technology has been used together with GIS to provide maps and image data on the Web, such as OpenStreetMap and Geo-Wiki. With the Internet and Service-oriented computing (SOA), people located at different places can share information easily without worrying about data formats

Please cite this article in press as: Han, G., et al. A web-based system for supporting global land cover data production. ISPRS J. Photogram. Remote Sensing (2014), http://dx.doi.org/10.1016/j.isprsjprs.2014.07.012