Computers & Geosciences 47 (2012) 130–142

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

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

Identifying orthoimages in Web Map Services

A.J. Florczyk*, J. Nogueras-Iso, F.J. Zarazaga-Soria, R. Béjar

Computer Science and Systems Engineering Department, Universidad de Zaragoza, María de Luna 1, 50018 Zaragoza, Spain

ARTICLE INFO

Article history:
Received 29 April 2011
Received in revised form
11 October 2011
Accepted 18 October 2011
Available online 25 October 2011

Keywords:
Web Map Service
WMS
Orthophotos
Orthoimage layer detection
Image analysis
Image collection
Web Processing Service
WPS

ABSTRACT

Orthoimages are essential in many Web applications to facilitate the background context that helps to understand other georeferenced information. Catalogues and service registries of Spatial Data Infrastructures do not necessarily register all the services providing access to imagery data on the Web, and it is not easy to automatically identify whether the data offered by a Web service are directly imagery data or not.

This work presents a method for an automatic detection of the orthoimage layers offered by Web Map Services. The method combines two types of heuristics. The first one consists in analysing the text in the capabilities document. The second type is content-based heuristics, which analyse the content offered by the Web Map Service layers. These heuristics gather and analyse the colour features of a sample collection of image fragments that represent the offered content. An experiment has been performed over a set of Web Map Service layers, which have been fetched from a repository of capabilities documents gathered from the Web. This has proven the efficiency of the method (precision of 87% and recall of 60%). This functionality has been offered as a Web Processing Service, and it has been integrated within the *Virtual Spain* project to provide a catalogue of orthoimages and build realistic 3D views.

1. Introduction

Orthoimages are essential in many Web applications to facilitate the background context that helps to understand other georeferenced information. For instance, in the field of disaster management, satellite data play an increasingly important role in supporting decision making (Kwan and Ransberger, 2010; Meisner et al., 2009). Rapid data integration and visualisation are essential to make data accessible and convey them in an easier-to-perceive way (Iosifescu-Enescu et al., 2010). Especially when presenting information to a non-expert audience, visualisation of the data improves the understanding of the situation at hand. Other applications of remote sensing products cover change analysis for monitoring and tracking the type and rate of landscape changes (Julea et al., 2010), urban environment modelling (Krauss et al., 2007), or assessing geospatial information quality (Skirvin et al., 2004).

With the constant improvement of technologies in high-resolution satellite remote sensors, Global Positioning Systems (GPS), databases and geoprocessing sources, there are nowadays increasing amounts of imagery and gridded data. Additionally, thanks to the development and increasing importance of Spatial Data Infrastructures (SDI), the availability and accessibility to these data through standardised and interoperable Web services has increased exponentially in the last years. The Web Map Service (WMS) and Web Coverage Service (WCS) specifications, proposed by the Open Geospatial Consortium (OGC), provide the means for the implementation of services offering visualisation and download of imagery data in well-known formats such as HDF-EOS, GeoTIFF, DTED, NITF or GML. However, it can be observed that there are two main problems that become an obstacle for the access to imagery data on the Web. On the one hand, SDI catalogues (the services provided by SDIs to locate data and services) deployed at national, regional or local levels do not necessarily register all the services providing access to imagery data on the Web. On the other hand, it is not easy to automatically identify whether the data offered by a Web service is directly imagery data or not. With respect to the first problem, the discovery of Web services not directly subscribed in SDI catalogues, some researchers have proposed different strategies based on the crawling of the Web (Li et al., 2010a; López-Pellicer et al., in press). However, the second problem, the automatic categorisation of the content offered by services has received little attention until now.

The purpose of this work is to study this second problem and propose a method for the automatic analysis of WMS services in order to detect if they contain imagery data, i.e. aerial photos or orthorectified satellite images of high resolution. We focus on WMS services because their availability is higher on the Web.

* Principal corresponding author. Tel.: +34 976 762 134; fax: +34 976 761 914.
E-mail addresses: florczyk@unizar.es (A.J. Florczyk), jnog@unizar.es (J. Nogueras-Iso), javy@unizar.es (F.J. Zarazaga-Soria), rbejar@unizar.es (R. Béjar).

0098-3004/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2011.10.017

Currently, the number of WCS service instances is very low in comparison with WMS services (López-Pellicer et al., 2011). Additionally, as WCS services provide full access to data, complex issues about access rights usually restrict their public deployment.

Various heuristic methods for the automatic analysis of WMS services have been investigated in this work. First, a description-based method has been developed. It exploits information offered by the service provider. Then, a content-based method has been investigated. The characteristics of the information that should be detected (i.e. orthoimages) derive from the requirements on radiometric and spatial resolution of sensors that produce it. Therefore, the effort focuses on deterministic methods that exploit image features generated by these kinds of sensors. Additionally, an effective algorithm for the gathering of spatial information from OGC WMS services is an important issue of this content-based method. In the context of this work, an effective retrieval algorithm for content-based analysis means an algorithm for collecting a representative set of fragments of an image (i.e. not empty image fragments), which enables layer analysis.

The developed procedure has been published as a geoprocessing service, which accepts the URL of a WMS capabilities document and returns a list of those of its layers which offer orthoimages and the scale at which the images have been discovered. This geoprocessing service has been applied within the *Virtual Spain* project[1] to develop a catalogue of orthoimages and associated applications. *Virtual Spain* is an R&D project supported by the Spanish Government through the Centre for the Development of Industrial Technology whose objective is to define architectures, protocols and standards for an envisioned 3D Internet focusing specially in 3D visualisation, virtual worlds, user interactions and the introduction of semantic capabilities. In particular, one of the experiments proposed in this project is focused on the crawling of services for access and exploitation of images.[2] The method and the experiment reported in this paper contributes to the filtering of crawled services. Additionally, this project also explores how to link these image services with Virtual Globe applications in order to provide a realistic display of Digital Elevation Models (DEMs) for urban areas, which are produced as the result of another experiment in this project. Virtual Globes provide computer-based representations of the real world that are receiving an increasing interest by experts in the geoscience field (Bailey and Chen, 2011).

The rest of this paper is structured as follows. Section 2 makes an overview of the existing related work. Section 3 describes the method proposed for the analysis of WMS services. Section 4 presents some experiments performed over a crawled collection of WMS services and discusses the efficiency of this method. Section 5 describes the publication of the proposed method as a Web service compliant with the OGC Web Processing Service (WPS) specification. Then, Section 6 presents how it has been used within the scope of the *Virtual Spain* project to develop an orthoimage catalogue. Finally, Section 7 ends with some conclusions and outlook on future work.

2. Related work

The accessibility to geospatial resources influences any geospatial-based research. New approaches, such as on-demand data production (Mansourian et al., 2008), assume discoverability of proper geospatial resources as an input. The discovery of resources within an SDI is based on the digital library paradigm (Béjar et al., 2009). For instance, Li et al. (2010b) present a framework for searching over multiple standardised catalogues, which can provide an integrated search across local and regional SDIs. However, not all Web resources valuable for these kinds of geospatial processes are part of an SDI, e.g. OGC Web services offered by a public administration but not published within any SDI, or resources produced by Web users. In these new scenarios, the automatic discovery of geospatial resources on the Web has recently gained interest within the geospatial community (Li et al., 2010a; López-Pellicer et al., in press). This new approach might provide users with a wider range of geospatial resources. Although standardised geospatial services are thought to be self-descriptive (i.e., their interfaces include an operation to retrieve metadata), recent works indicate there is a lack of good practices. It is common to find inconsistencies between metadata in registries and the metadata offered directly by the service (Wu et al., 2010). As automatic discovery on the Web relies mainly on resource analysis, it might motivate providers to assure more accurate descriptions. For example, in the case of OGC services, it might be supposed that if the discoverability of a resource depends on the quality of its capabilities content, providers would put more effort in the future to generate more valuable information in comparison with the current practices. In the context of this work, WMS instances crawled on the Web are used as input in the experiments to test the performance of our proposed method for filtering orthoimages. As these services have been found on the Web and are deployed by different providers, they represent a realistic scenario that combines heterogeneous metadata of different quality levels.

On the other hand, it must be taken into account that although resources are usually described in conformance with a specific description model, this does not necessarily imply that all potentially interesting information to the user is provided. For example, the WMS capabilities specification does not define a field to state the data presentation scale, i.e. the scale range for which a service renders a map. The capabilities specification just includes an optional field with the range of scales for which it is appropriate to generate a map of a layer. Although this information has different semantics (i.e. it refers rather to the data resolution), it might be assumed that the service renders a map at least at these scales. Additionally, as this field is optional, it may happen that some service providers have included this information in free text fields. In this case, a support tool (e.g. for a keyword-based or semantic-enabled search) is necessary, or an additional effort is required from a potential client (e.g. manual visualisation) to estimate the resource utility. Content-based analysis can be useful to extract additional information from a resource, which is not directly provided within its description. The automatic discovery of geospatial resources based on content analysis instead of relying on keywords in the metadata is gaining interest. For instance, Zhang et al. (2010) present a semantic-based application that is able to perform an intelligent content-based search of Web Feature Services (WFS). Although this approach is not directly applicable for non-textual resources such as imagery, image analysis techniques can be used. Image classification and annotation have been extensively investigated in the field of information retrieval, and machine learning approaches have been commonly applied in this area (Baharudin et al., 2007; Gonzalez-Garcia et al., 2007; Sinha and Jain, 2008). These techniques have been also applied to remote sensing imagery for the classification of multispectral imagery over a semi-urban area (Alonso and Malpica, 2008), or to support the update of existing land use databases (Kressler et al., 2005). The heuristics of our proposed method include both techniques for processing WMS capabilities, and techniques for analysing the content of the layers offered by a WMS service instance.

Finally, it must be noted that any technique for the content-based analysis of a WMS service requires an effective algorithm for collecting a representative set of fragments of an image,

[1] España Virtual project (Virtual Spain project): http://www.españavirtual.org/

[2] http://ev.unizar.es/EV42/