Int J Appl Earth Obs Geoinformation 72 (2018) 42–56

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

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

Updating authoritative spatial data from timely sources: A multiple representation approach

Xiang Zhang[a,b,c,*], Weijun Yin[a], Min Yang[a], Tinghua Ai[a], Jantien Stoter[d]

[a] School of Resource and Environmental Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China
[b] Key Laboratory of Geographic Information System, Ministry of Education, Wuhan, China
[c] Key Laboratory of Digital Mapping and Land Information Application Engineering, National Administration of Surveying, Mapping and Geo-information, Wuhan, China
[d] Faculty of the Built Environment and Architecture, Delft University of Technology, The Netherlands

ARTICLE INFO

Keywords:
Volunteered geographic information
OpenStreetMap
Incremental update
Multiple representation databases
Data matching
Cartographic generalization

ABSTRACT

Integrating updates from timely sources such as volunteered geographic information (VGI) into the spatial data maintained at official agencies is becoming a more demanding requirement but presents many challenges. This paper proposes an approach to addressing the technical challenge of propagating updates from timely sources (e.g. OpenStreetMap) to spatial data maintained at separate map scales. The main idea is to establish a multiple representation database (MRDB) for datasets at different scales and time to facilitate incremental update, where linkages between corresponding objects at different datasets are made explicit. First, two ways in which the timely sources can be integrated into official data for incremental update are discussed. To derive the linkages between different datasets, a data matching procedure based on computer vision is presented and fine-tuned to match data in different scale ranges. Furthermore, the generalization history used to produce smaller scale data from the larger ones in official data is inferred based on the linkages, and is then used to guide the update propagation. Finally, a framework for incremental generalization in MRDBs is proposed, where crucial issues like strategies for update propagation, cartographic generalization, and the so-called 'chain reaction' are addressed. The framework is implemented as a fully automated process where operators like simplification, enlargement, compression, displacement and typification are incorporated into the incremental update process. By testing the framework against real world data sets (i.e. OpenStreetMap and official data at 1:10k, 1:50k and 1:100k), we show that the updates are integrated consistently into existing data in terms of spatial relations and cartographic quality. Our work suggests that making use of timely sources by official mapping agencies and companies in a continuous or event-driven data update is technically feasible, with further improvement and extensions discussed.

1. Introduction

Over decades, the acquisition, maintenance and update of geospatial framework data are in the domain of professionals such as national mapping agencies (NMAs) and companies. Although the update cycle of framework data in NMAs has reduced significantly, the update still relies largely on field survey, aerial photo interpretation and interactive generalization (Stoter, 2005; Stoter et al., 2009b) and therefore becomes the bottleneck for data/map production and services.

Recent years have witnessed the proliferation of volunteered geographic information (VGI) (Goodchild, 2007). OpenStreetMap (OSM) is one of the most prominent VGI projects to date and provides access to the open, free and up-to-date digital map data covering the world. In some countries/regions, OSM has received large imports[1] from professional data providers in its early days so that a basic level of data coverage and quality can be guaranteed. Moreover, OSM is constantly updated by large amount of volunteers worldwide and has grown to be a timely data source that covers a rich set of features and semantics (Heipke, 2010; Dorn et al., 2015). However, given its successful applications in the public and scientific domains (Goetz, 2012; Hagenauer and Helbich, 2012), the role of VGI in governmental sectors has not been properly identified until more recently (Haklay et al., 2014).

The major concern lies in data quality. Although comparing VGI with authoritative datasets for quality assessment is informative (Haklay, 2010; Girres and Touya, 2010), some researchers argued for the Linus's law principle from the perspective of big data (Haklay et al.,

* Corresponding author at: School of Resource and Environmental Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China.
E-mail address: xiang.zhang@whu.edu.cn (X. Zhang).

[1] https://wiki.openstreetmap.org/wiki/Import

https://doi.org/10.1016/j.jag.2018.05.022
Received 2 March 2018; Received in revised form 25 May 2018; Accepted 29 May 2018
0303-2434/

2010; Foody et al., 2015), suggesting a different quality assurance approach in VGI communities as compared with NMAs (Elwood et al., 2012). This might explain to some extent NMAs' concerns with VGI. On the other hand, some studies revealed that VGI is accurate enough for use by official agencies in map production (Parker et al., 2012; Olteanu-Raimond et al., 2016). For example, OSM shows at least comparable quality to authoritative datasets in terms of spatial accuracy and coverage in densely populated areas (Haklay, 2010). Besides, since most crowdsourced geographic data (e.g. OSM) have the Open Database license (ODbL), there are many restrictions in using such data.

Given the above issues, there is a recent trend for professional communities to consider the use of crowdsourced geographic information for production (Mooney and Morley, 2014). In a recent survey, Olteanu-Raimond et al. (2016) identified that VGI have been engaged in various degrees in European NMAs. Their survey shows further that NMAs such as Kadaster, the Netherlands and IGN, France have used OSM for change detection, which reduces work load and improves efficiency in traditional methods. This is perhaps in line with the adaptation of the Open Data Policy in many European NMAs (Olteanu-Raimond et al., 2016). Similar progress is reported in North America, where NMAs and companies have used, or plan to use, updates from VGI sources such as OSM instead of surveyors to speed up their production lines (Elwood et al., 2012). Begin (2014) for instance shows that, by providing data to OSM, Canadian NMA could therefore receive updates from OSM.

In this paper, we will discuss the use of timely sources in updating data sets maintained by official agencies from a technical point of view. As discussed previously, framework data are maintained at multiple scales and updated by interactive (manual) generalization. Our general question is therefore: with the timely sources available (e.g. OSM), how can we incorporate the changes and propagate them to data at different scales more consistently and efficiently? This is especially an issue when cartographic generalization has to be used and graphic conflicts need to be handled (Stoter et al., 2009b). Here we propose a multiple representation approach for incremental update of separately maintained datasets. Some concepts and related work is reviewed in the next section.

1.1. Concepts and related work

A multiple representation database (MRDB) is a spatial database that consists of datasets with different levels of abstraction (i.e. map scales), where multiple representations of the same real-world objects are linked by inter-scale connections (Kilpeläinen, 2000). *Incremental update* is a major application of MRDBs.

While incremental update is a more general term from software and database field, *incremental generalization* is used specifically for spatial databases, meaning the propagation of updates across different scales in an MRDB (Kilpeläinen and Sarjakoski, 1995). MRDBs and Incremental generalization are very attractive ideas for both researchers and practitioners (Devogele et al., 1996; Harrie and Hellström, 1999; Hampe et al., 2003; Haunert and Sester, 2005; Müller et al., 2012). However, MRDBs have seldom been implemented for production, especially when the linkage between representations is concerned (Burghardt et al., 2010).

As a key element in MRDBs, the linkage (hereafter referred to as *vertical relation* as in Bobzien et al., 2008) can be used to assess the quality, maintain the consistency between representations, facilitate incremental update, and is useful for multi-scale analysis and visualization (Stoter et al., 2009a; Burghardt et al., 2010). These relations can be complicated, since many-to-many correspondences are possible due to the use of cartographic generalization. For example, objects can be aggregated (resulting in a relation cardinality of n-to-1), deleted (1-to-0), or typified (n-to-m), where n is the number of larger scale objects and m the smaller scale ones.

The vertical relations can be explicitly modeled and recorded in the generalization process, with which the generalization history can also be stored (Burghardt et al., 2010). For instance, Zhou et al. (2009) proposed a model and prototype to log such metadata during the generalization to support incremental update. However, in the absence of such linkages (as in most software systems and NMA datasets), data matching can be used. As for update propagation in MRDBs, Haunert and Sester (2005) identified the issue of 'chain reaction' where an update may reform the linkage structure and hence more objects in the vicinity of the update are influenced. More recently, Müller et al. (2012) proposed to build an MRDB for OSM for potential applications in different domains, but this requires considerable efforts to generalize representations from a single database.

Our work differs from previous research and hence contributes in several aspects. First, we establish the MRDB and hence the vertical relations between data at different scales and time (i.e., OSM and official data at 1:10k, 1:50k, 1:100k) by data matching (Section 2.2). This fits better into current practices in NMAs. The matching is able to handle ambiguous situations caused by cartographic generalization (e.g. displacement and typification). Second, the generalization history of official data is extracted based on the vertical relations and is used to guide the update propagation at individual and group levels (Section 2.3). More importantly, we proposed a framework for propagating updates incrementally to datasets at multiple scales, where cartographic generalization and the so-called 'chain reaction' can be properly handled (Section 2.4).

As a proof-of-concept, we focus on building features in this paper and using OSM as a timely data source. Building footprints in OSM are shown to exhibit higher levels-of-detail than official datasets (Touya and Brando, 2013; Touya and Reimer, 2014; Fan et al., 2014). Also, when compared with our Top10nl (1:10k) we found that OSM buildings have greater details and is more current. Besides, building features require more efforts of generalization and hence pose a more challenging case for incremental generalization. After describing our approach in Section 2, the implementation and results are presented in Section 3 and discussed in Section 4. We draw our conclusions in Section 5.

2. Methodology

2.1. MRDB approach to incremental update of spatial data

2.1.1. An assumption of consistency for multiple representation databases

As a premise, we assume that all representations in an MRDB are logically consistent and synchronous in time. That is, data objects at different map scales represent the identical physical entities of the same time; no change exists between them. This is reasonable because smaller scale data were usually generalized from larger scale data. As a result, differences between data sets in an MRDB are due to map generalization no matter how significant the differences can be (e.g. Fig. 1). This is also a condition to guarantee a logical propagation of updates from larger to smaller scales. If the datasets in an MRDB are not synchronous, it is hard to know whether the differences were caused by the generalization or physical changes, and there is no way in which reliable vertical relations can be established.

2.1.2. Choice of scale in an MRDB for change detection

In an MRDB, changes occur to the data are only reliably detected by comparing the timely data source (e.g. OSM) with data of identical, or similar, map scales. If the scale or LoD of the data deviate considerably from one another, it can be hard (if not impossible) to distinguish physical changes from the discrepancies caused by map generalization.

This can be better illustrated with Fig. 1, where changes can be more reliably identified when the two data (e.g. OSM and Top10nl) are of similar scales (Fig. 1a; areas indicated by arrows are considered physical changes). When comparing OSM and top50nl (Fig. 1b), it becomes obscured if the observed differences qualify as changes or not, as the difference can be a result of map generalization or physical changes,