Automatic Large Scale Topographical Map Updating using Open Street Map (OSM) Data within NoSQL Database Platform

W. Tampubolon[a*], W. Reinhardt[a]

[a]AGIS, Institute for Applied Computer Science, Universität der Bundeswehr, Werner-Heisenberg-Weg 39, 85577, Neubiberg, Munich, Germany,- winhard.tampubolon@unibw.de, wolfgang.reinhardt@unibw.de

KEY WORDS: Topographical, Signature, Crowd-sourcing, Updating

ABSTRACT:

Basically there are two major important aspects of topographical maps, which can be considered as key performance indicators: geometric quality and up-to-datedness' of the map. In this context, adequate map updating is of special interest to achieve effective and efficient geospatial data provision over time. Crowd-sourcing information can potentially also be used for the map updating purpose because of its public participatory manner and its dynamic growth in terms of data volume and schema.

As a matter of fact, each dataset has its own features and structures depending on its geospatial data originator with their own technical specifications. Open Street Map (OSM) uses tags as a reference for feature identification in which users can modify or even create their own definition. In this case, the geospatial feature definition or conversion among different datasets is considered as a crucial stage.

NoSQL database technology has been introduced as a potential alternative solution to existing SQL databases which is supposed to grow more rapidly in the near future. It has the perspective to combine the powerful capability of GIS data processing with an approach of non-relational Data Base Management System (DBMS). This type of data warehouse can potentially accommodate variety of information over the World Wide Web (www) space with different structures into one single geodatabase. MongoDB as one instance of NoSQL database introduces an open source document storage empowered by a replication using a data partitioning approach across multiple machines.

This paper discusses the role of geospatial signatures for topographical object recognition from different vector datasets in an automated way. Its characteristic can be used to identify common objects i.e. points which subsequently play a role in geometrically transforming OSM data into official Large Scale Topographical Map. Hence, the proposed approach for an automation process concentrates on three steps, namely geospatial signature identification, transformation and finally vector data integration. The implementation of this approach is currently ongoing using Python scripting environment in the ArcGIS platform.

1 Introduction

1.1 Background

Topographic map data can be considered as essential and fundamental because of its high importance not only for the planning purpose but also for the evaluation of the ongoing infrastructure development such as road, building, utilities, etc. This basic type of geospatial data consists of natural and manmade objects which can be created with straightforward interpretation in the feature object classification e.g. land cover, geographical names. In other words, the uncertainty

---

* Corresponding author.

for feature identification within the topographical themes is considered not significant. As an example, land cover theme has to classify the general land object type such as forest, settlement, paddy field. It is not necessary to classify each specific type of forests in topographical map production. However, topographical maps include also 3D terrain information i.e. Digital Elevation Model (DEM). According to experience, the provision of DEM requires higher efforts compared to planimetric features such as roads, buildings, administrative boundaries, etc.

In particular for large scale topographical mapping, special emphasis is put on the update of maps of urban regions. This is specifically the case for third world countries, which can allocate only a limited budget to map production and update. Therefore, they concentrate on urban and dense populated areas. On the other hand, public participatory of geospatial communities over the internet can provide tremendous results within the context of crowd-sourcing activities. The main reason for this is the legal and physical openness in geospatial data sources which shares the attractive high resolution imageries as well as vector data sources. Open access geo-information has been widely used as the alternative technical reference for disaster preparedness and emergency response in rapid mapping activities (Tampubolon, 2012). It requires a frequent and dynamic update process to provide recent and actual geospatial data sets, especially during emergency situations. As the demand of geospatial dataset in post disaster i.e. reconstruction phase is increasing, the update mechanism for large scale topographical maps which include detail infrastructure features such as road network must be defined at the first place.

In the era of big geospatial dataset utilization nowadays, there are some prominent data aspects involved namely dynamic structure, high volume, and multi sources. Those aspects require a processing platform which can support real time geospatial data integration among different geospatial data sources over time. NoSQL databases introduce a new approach to overcome the problem of data structure inconsistency as a challenge to the conventional Relational Data Base Management System (RDBMS). The most important factors that trigger NoSQL technologies were the uprising of crowdsourcing and technology driven demands (Harrison, 2013).

MongodDB as an instance of NoSQL database implements schema flexibility by using Java Script Object Notion (JSON) format and a document based approach (Mongodb, 2014). Actually those two approaches will combine the advantages of flexible data structure and data transfer capabilities.

1.2 Research Objectives

The objectives of this research can be divided into three major parts as follows:

- Development of algorithm to recognize topographical object changes using GIS approach within NoSQL database platform;
- Development of an Automatic Large Scale Topographic Map (LSTM) updating mechanism by using ESRI platform i.e. Arctoolbox based on Python scripting environment;
- Technical implementation using Open Street Map (OSM) data to update LSTM road network dataset.

1.3 Technical Workflow

This research is conducted by combining theoretical and empirical approaches in order to achieve proper geospatial data quality in the context of large scale topographical map updating. Initially, the official large scale topographical map dataset has been selected as the benchmark for performing geospatial data assessment (Figure 1).

Using GIS data as a priori information has already been implemented in the map updating process (Bailloeul, 2003). It is assumed at this point that a topographical map is considered to be more accurate and complete in a comparison with other geospatial data sources. Therefore it has been used as a reference in updating purposes for both, the geometrical accuracy and the thematic correctness.

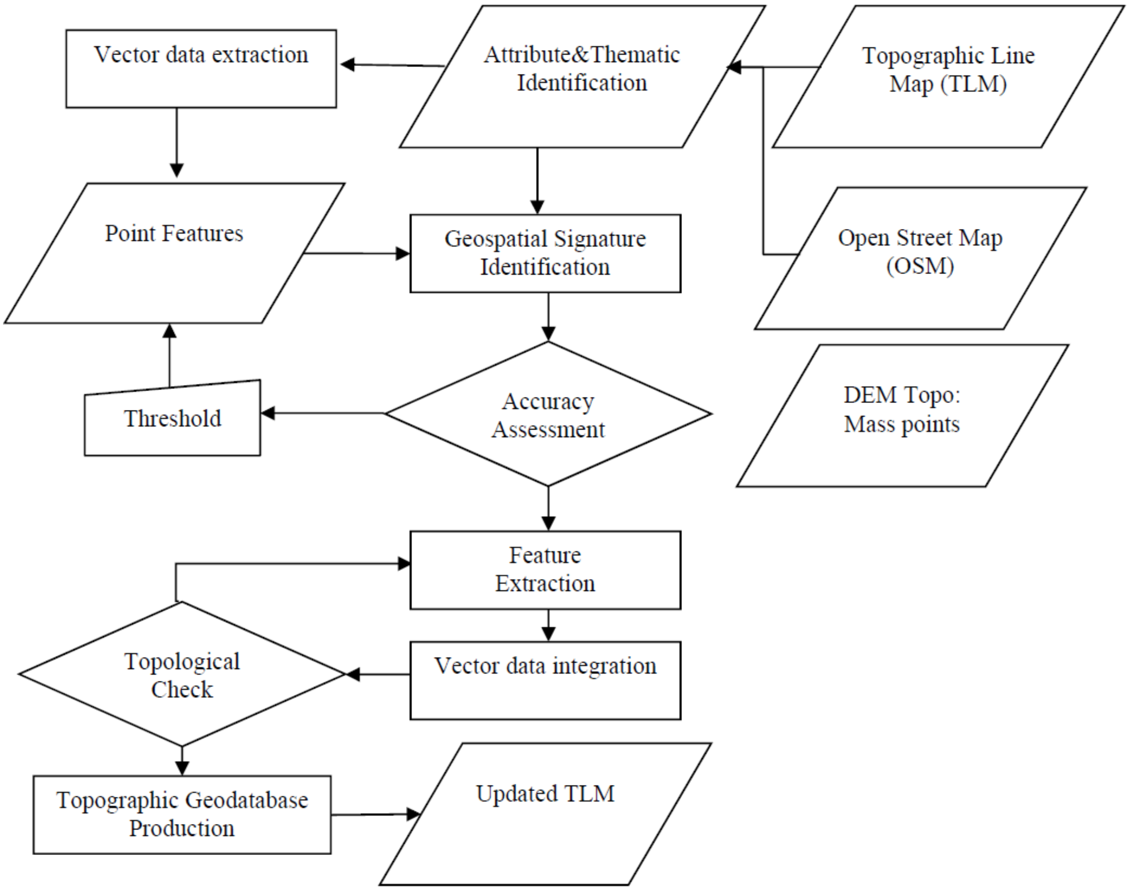

Figure 1: Algorithm workflow.

2 Map Updating Concept

The update mechanism used in this research comprises two major parts as described in the following. The first part has a close connection to the thematic and feature classification of geospatial datasets while the second part deals with the geometrical aspect.

2.1 Thematical update

OSM uses a free tagging system that allows participants to include any unlimited number of attributes for describing their feature interpretations. However, there are some agreements on certain fields, keys and values especially for the mostly used definitions and their combinations. This informal rule plays an important role to establish such a standard for creating geospatial data.

Nevertheless, users can define new features to improve the cartographic style of the map or to contribute on unmapped attributes of the features. In this case, short descriptions of each tag that relate to particular themes or interests can be provided using the feature pages as the reference.

Thematic aspects define the information content within a geospatial data set in which it can describe attribute fields including semantic aspects of the features. NoSQL databases offer advantages for this task e.g. by its schema flexibility and can therefore play a major role in future.

2.2 Geometrical update

Geometrical update has a strong connection with the newly created form of features in the real world. Although the spatial resolution of Very High Resolution Satellite imageries i.e. used by OSM is increasing, it does not always guarantee the geometrical accuracy as expected. In this case, it is necessary to assess geometrical accuracy of OSM data at the first place. Within the proposed approach it has to be considered also detect geometrical changes of topographical features.

Since different technologies provide various types of geospatial data quality, geometrical analysis will focus on self-assessment process to provide partial geometric accuracy for each dataset. Each single assessment will be subsequently combined with other assessments in order to identify geospatial relationships among them.

3 Technical Implementation

In order to detect changes, it is important to define criteria and mechanism with a logical argument as a reference. There are two approaches implemented for this purpose. The first part is how to define geospatial signature which shall be unique for each geospatial object. The second part focuses on quality measurement for the whole map updating mechanism scheme.

3.1 Geospatial signature identification

Change detection in this paper has a main objective for automatic feature distinction based on geospatial aspects. The term geospatial signature focuses on quantifying object identification using GIS approach as a tool. This tool operates regional based calculation by combining different datasets comparison.

$$G_{sign}(dX, dY) = \left(\sum_{n=1}^{m}(dist * \sin \propto), \sum_{n=1}^{m}(dist * \cos \propto)\right) \quad (1)$$

$$Dev_{sign}(dX, dY) = \left(\min_m \sqrt{dX^2 + dY^2}\right) \quad (2)$$

Where m = the number of points within regional based approach

- dist = distance from each point to the analyzed point
- α = azimuthal angle from analyzed point to each point
- dX = deviation in X axis
- dY = deviation in Y axis

Single assessment process for each dataset uses equation (1) to calculate point-based inter connection using angle and distance as the arguments. This calculation is followed by measuring minimum mean value for each point-based moving region as inferred from equation (2). In this case, the minimum mean value exposes common objects while maximum mean value gives different objects. By using this approach, updated features i.e. road segment (in yellow) can be detected as shown in Figure 2, in which maximum mean value exceeds geospatial signature threshold (G.Sig = 0.754).

Figure 2: Geospatial signature identification

3.2 Accuracy Assessment

The tolerances for accuracy assessment used here have been developed based on the National Mapping Accuracy Standard (NMAS), the horizontal tolerance accuracy can be seen in Table 1 (FGDC, 1998).

Map Scale	Tolerance at Publication Scale	Tolerance at Ground Distance
1:1,000	1/30 inch = 0.85 mm	0.85 m
1:5,000	1/30 inch = 0.85 mm	4.25 m
1:10,000	1/30 inch = 0.85 mm	8.5 m
1:25,000	1/50 inch = 0.5 mm	12.5 m
1:50,000	1/50 inch = 0.5 mm	25m

Table 1: NMAS Horizontal Accuracy Tolerance

4 Vector data integration

After the updated objects have been recognized, the planimetric (2 D) compilation can be implemented for updating the official Topographical Map Technical Specification in 1:5,000 map scale produced by BIG. This mechanism has provided ArcToolbox package in which it can directly utilize Mongodb documents storage to perform statistical analysis for geospatial signature identification purpose (Figure 3).

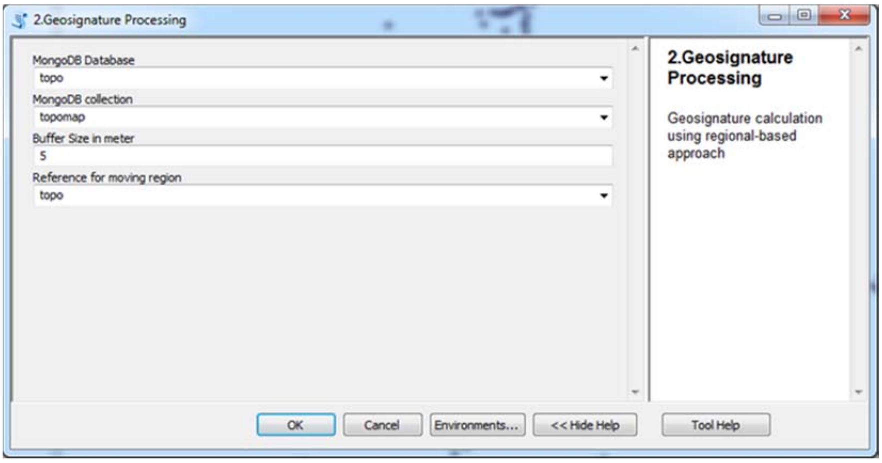

Figure 3: Python toolbox using arcpy.

5 Closing Statement

5.1 Conclusion

Finally this research has introduced the large scale topographical mapping updating mechanism by utilizing NoSQL database both as the GIS processing unit and the geospatial data warehouse. The accuracy of OSM data can be detected automatically in order to fulfill the large scale topographical mapping requirements.

Region-based approach is still considered as the best solution to identify geospatial signature both for geospatial data accuracy measurements and updated object detection. Algorithm created in this paper recognizes identical features and updated features for different purposes. The first is to measure geometric accuracy and subsequently the latter is to detect updated features in the automated way.

5.2 Recommendation

For the future works, in order to improve the challenging task for automatic large scale mapping updating, it is recommended to proceed with the following tasks:

Further algorithm development not only for accuracy assessment but also for OSM data transformation into official large scale topographical maps data.

Extension of geospatial signature identification approach for other purposes e.g. Ground Control Point (GCP) detection between different datasets.

REFERENCES

Boucher, C., 2012. Automatic Detection of Topological Changes for Digital Road Map Updating. In: Instrumentation and Measurement, IEEE Transactions, 61(11), pp. 3094-3102

Bailloeul, T., 2003. Urban digital map updating from satellite high resolution images using GIS data as a priori knowledge. In: 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, pp. 283-287

Federal Geographic Data Committee, 1998. Geospatial Positioning Accuracy Standards Part 3: National Standard for Spatial Data Accuracy (version 2.0), FGDC-STD-001-1998: Federal Geographic Data Committee, pp.3-4 Washington, D.C.

Federal Geographic Data Committee, 1998b. Spatial Data Transfer Standard. Computer Products Office, FGDC-STD-002-1998: Federal Geographic Data Committee, National Technical Information Service, Springfield, VA

Harrison, J., 2013. Using NoSQL & HTML5 Libraries to rapidly generate interactive web visualizations of high-volume spatiotemporal data, Association for Geographic Information (AGI), London, UK. http://www.agi.org.uk/storage/GeoCommunity/AGI13/Speakers/ppts/JackHarrison.pdf (accessed 02.10.2014)

MONGODB, 2014. Top 5 Considerations When Evaluating NoSQL Databases http://www.info.mongodb.com/rs/mongodb/images/ 10gen_Top_5_NoSQL_Considerations.pdf (accessed 02.10.2014)

Tampubolon, W., Kiefl, R., Strunz, G., 2012. Analysis and integration of open access geoinformation in the Spatial Data Infrastructure for emergency response and disaster preparadness, Applied Geoinformatics for Society and Environment, Johor Bahru, Malaysia