CROWD-SOURCING APPROACH OF BUILDING GROUND TRUTH DATABASE FOR GLOBAL URBAN AREA MAPPING

Hiroyuki Miyazaki[1], Koki Iwao[2], and Ryosuke Shibasaki[1]
[1] JSPS Fellow, Professor with the Center for Spatial Information Science, The University of Tokyo
5-1-5 Kashiwanoha Kashiwa Chiba, 277-8568, Japan; Tel:+81-4-7136-4307
E-mail: heromiya@csis.u-tokyo.ac.jp, shiba@csis.u-tokyo.ac.jp

[2] Researcher of the GEO Grid Research Group with National Institute of Advanced Industrial Science and Technology,
Tsukuba Central 2, Umezono 1-1-1, Tsukuba, Ibaraki 305-8568, Japan
E-Mail: iwao.koki@aist.go.jp

KEYWORDS: ground truth database, crowd sourcing, Web Mapping Service, Web Feature Service, global urban area mapping

ABSTRACT: We present a practical system for collecting ground truth data needed for global urban area mapping. To collect visually interpreted information effectively, we implemented crowd sourcing, with which the operators collect and send very accurate data generated by visual interpretation through the Internet. The system was constructed with Web Map Service (WMS) and Web Feature Service (WFS), which are standardized scheme of publishing and updating map data over the Internet. Within the system, the conductor and the operators of visual interpretation campaign would save labor and time for transferring data, keeping consistency, and assuring data security. In addition to that, owing to the standardized scheme, the system was flexibly extensible and transferrable to the other purpose. We regard that the system would contribute to improve ground information for global urban area mapping as well as statistical investigation on precision of visual interpretations.

1. INTRODUCTION

Global urban area maps, which represent location, area and shape of urban area of the world, have been used for critical issues on the earth. Many attempts of developing urban area maps in higher resolution and higher accuracy had been performed; however, ground truth data for urban area mapping are still in lack (Miyazaki et al. 2011). Efficient method to collect data is needed for further improvement.

In recent years, several data-intensive science projects adopted crowd sourcing, in which many operators work together on routine data collection over the Internet (Hand 2010). Especially for the field of geographical information, people voluntarily collect environmental and civil information with geographical location derived from Global Positioning System (GPS) and post it to a public database. The information collected by such voluntary activities is called volunteered geographical information (VGI; Goodchild 2008).

The most active VGI is OpenStreetMap (OSM; Haklay & Weber 2008). OSM started with a purpose of providing freely available road map data. They also have interests in quick-response disaster mapping with crowd sourcing. On 11 March in 2011, people in east Japan had a catastrophic disaster caused by Great East Japan Earthquake. As soon as the earthquake occurred, the OSM community started to create the maps for the disaster affected regions. Their efforts are regarded to be important basis for the reconstruction of the regions.

Degree confluence project (http://confluence.org) is an outstanding crowd-sourcing project for building ground truth database. The objective of the DCP is to archive ground information of the latitude and longitude integer degree intersections (the ‘confluences’) by visiting and documenting the state of the surroundings (Iwao et al. 2006). The ground information provided by DCP is statistically usable for ground truth data of land cover classifications (Iwao et al. 2006).

Geo-Wiki (http://www.geo-wiki.org) is another outstanding crowd-sourcing project of building ground truth database. The Geo-Wiki provides web-based interface for mapping validation information of global land cover maps. The user of the Geo-Wiki may determine correct land cover at pixels of disagreement among ground land cover maps by interpreting high-resolution imagery of Google Earth (Fritz et al. 2009).

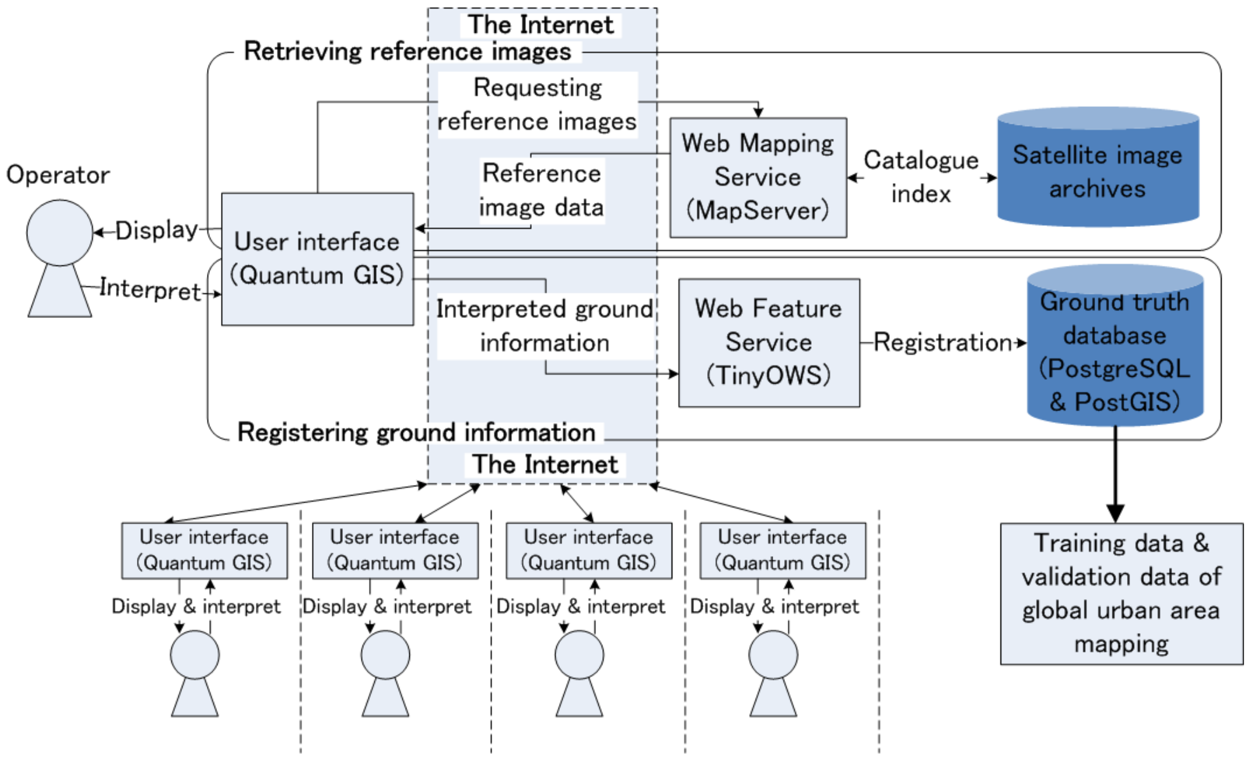

Figure 1. Overview of the system with Web Map Service (WMS) and Web Feature Service (WFS). Texts with brackets indicate the name of the software to implement the service and the interface.

Every project has unique outstanding points; however, they didn't build the system with a common standard, rather they built the system on their 'hand-made' scheme. We regard that a system built with a common standard would be effectively extensible and transferrable for further improvement of the scheme. In addition, such system would be easily extensible not only from server side, but also from user side.

Fortunately, there are internationally standardized schemes for such system useful for crowd sourcing. Those are Web Map Service (WMS; Open Geospatial Consortium 2011b) and Web Feature Service (WFS; Open Geospatial Consortium 2011a). Those enabled us to collect ground information over the Internet without physical communication by mail as well as transferring heavy image files by email or downloading.

In this paper, we present the scheme of crowd-sourcing validation of urban area maps using WMS and WFS and the implementation. It may be a good introduction for the readers to introduce a system with WMS and WFS for the own purpose.

2. METHODOLOGY

2.1. Web Map Service (WMS) of Reference Images

Conventionally, development of ground truth data with many operators living far from the office involved transferring large image files more than hundreds of megabytes by optical media, hard disk drives, FTP, and so on. The conductor has to burn many copies of optical media for the operators or setup FTP server in the office. The operators have to copy the reference images to their computers or wait for finishing download of the reference image data. It would require considerable, but not essential labor and time.

That way also would induce security problems of source data. When transferring reference images to operators, the conductor has to ask the operators to sign covenants not to use the data for other purpose. However, the covenants do not definitely assure security of the reference data. More or less, the risk of technical incidents and human errors is always in backyards.

To save such unessential labor, time, and risks, we focused on using Web Map Service (WMS). WMS is a standard scheme for publishing map data by a server-client model over the Internet. A basic use case is that, when a client requests map images for a certain geographical extent, the server sends reference images only for the requested extent. The most famous WMS service is OnEarth (http://onearth.jpl.nasa.gov/), which provides global mosaic of Landsat, SRTM, and MODIS through the Internet.

The server-client system enabled us to easily control accessing the reference images. For example, if an operator has visual interpretations for a certain extent, the conductor has to consider physical limitation of the media (the operator would get tired when she or he receives tens of optical disks or hard disk drives!) and assure consistency of geographical extent of assigned visual interpretations and the reference images to be interpreted for each operator. By using WMS, owing to the on-demand system, we only had to put reference images into the server.

In addition, because WMS is constructed on a commonly used protocol, called HTTP, it provided us easy access control of operators. The feature was very useful when adding and removing operators for a campaign of visual interpretation.

2.2. Web Feature Service (WFS) of Ground Truth Database

For a campaign of visual interpretation, as well as managing reference images, organizing the posted result from operators would require considerable labor and time of the conductor. The conductor has to manually keep consistency of the transactions of ground information by operator. A typical series of procedures of that are receiving spreadsheet and vector map data by email, associating them, and integrating results from operators. Those would prevent flexible operation of developing ground truth database which required for labor-intensive projects.

For the operation to be flexible, we proposed to introduce WFS to developing ground truth database. WFS is a standard scheme to publish vector map data over the Internet. Addition to that, WFS supports transactions on vector maps between server and clients, like editing the map data over the Internet. WFS have been implemented by several kinds of server software (e.g. GeoServer, TinyOWS).

WFS is basically constructed with geographical information database, thus, the conductor does not need to associate vector data and spreadsheet. Additionally, with a few configurations, WFS automatically records name of user and time of transactions. Such features of the WFS may not be so novel and unique against the way using GIS software.

The powerful feature of WFS is that it accepts requests for transactions of a single geographical database over the Internet. Owing to that, we were able to avoid integrating the records of the visual interpretations separately sent from the operators. Following to that, it also enabled us to flexibly add and remove operators for visual interpretations.

3. IMPLEMENTATION

3.1. Overview of the crowd sourcing system

We implemented the crowd sourcing system with WMS and WFS-T using three open source GIS software: MapServer (http://mapserver.org), PostgreSQL (http://www.postgresql.org), PostGIS (http://postgis.refractions.net), TinyOWS (http://www.tinyows.org), and Quantum GIS (http://www.qgis.org). Figure 1 shows the overview of the implementation.

3.2. Server-side system

MapServer is the most popular server software with WMS. It has three attractive functions so that we selected it for the system. First, it has capability with various image file formats supported by Geographical Data Abstraction Library (GDAL; http://www.gdal.org). Second, it dynamically assembles separated image files into a single layer with least computation for a requested extent. For global urban area mapping, due to sparsely distributed urban area, mosaic operation of the reference images would yield much unnecessary pixels. With the function, we didn't have to perform mosaic operation of the reference images. Third, it dynamically constructs look-up table of RGB composition to enhance contrast within a requested extent. This function is also useful for our project to avoid preparing huge number of reference images enhanced scene by scene.

In this study, we used false color composite of ASTER/VNIR satellite images archived in GEO Grid, AIST, Japan. 11802 scenes of the images for 3372 cities of the world were indexed in the catalogue of MapServer.

For the management of ground truth data, we setup a relational database management system (RDBMS) with spatial data using PostgreSQL and PostGIS. PostgreSQL has a long history of the development by open community and is one of the most leading RDBMS. PostGIS is the spatial data extension of PostgreSQL. It have been

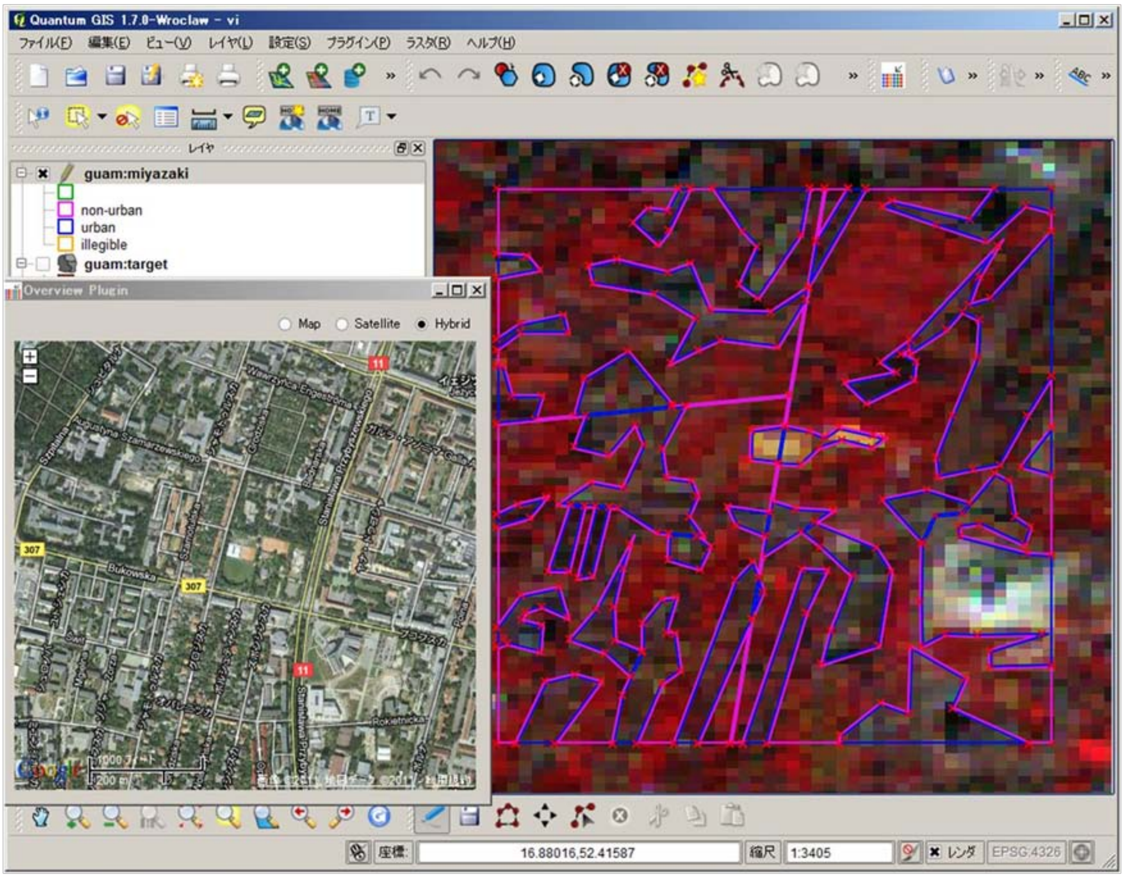

Figure 2. The interface using Quantum GIS for accessing the Web Mapping Service (WMS) and Web Feature Service (WFS).

developed by open community. Owing to the open development, it has flexible compatibility and extensibility with many kinds of software, including WFS. In addition, it has many functions for querying and manipulating spatial data with the geometrical conditions.

For the WFS, we had two options: those were TinyOWS and GeoServer (http://geoserver.org). TinyOWS has only minimum functions, as the name represents; however, it highly performs with customizations for a specific usage. On the other hand, although GeoServer is highly functional and is like an all-in-one package to start up WFS as well as WMS, it would be unstable due to costs on much computer resources required by the rich functions. Therefore, we finally chose TinyOWS for the implementation of WFS.

The server-side system was constructed on Debian GNU/Linux 6.0.2 with Apache 2.2.16. All of the software packages except TinyOWS were provided as compiled binary packages so that stable installation was very easy.

3.3. Client-side system

We described the implementation of server-side system. Here, we describe the client to access to the database through the WMS and WFS. Firstly, we had to decide whether to use existing software or to develop new client software. We regarded that developing system with existing stable software would yield solid progresses; thus, we decided to use existing client software.

There are various kinds of software for accessing over WMS and WFS. Among them, we chose Quantum GIS considering the installation cost and the extensible features. Quantum GIS is developed by the open community and the source code is publicly available. It means that Quantum GIS could be available without any charge. For our project, many persons would be expected to join; thus, the cost of software licenses would be intolerable if we would have introduced non-free software. In addition, owing to the publicly available source code, the software practically works on any operation system of the operator's computer. By the standardization of the interface, labor of preparing instruction could be saved.

It also has a useful feature of customizable plugins written in C++ and Python. Currently, more than 200 plugins are registered in the repository and users can easily install them via internet. For the development of ground truth database, GMap Overview would be useful to acquire detail information from Google Maps. It displays subwindow in which high-resolution satellite images of Google Maps is displayed with synthesis of location to the main window.

4. CONCLUSION

In this paper, we proposed a crowd-sourcing system to collect ground information for global urban area mapping. By introducing WMS and WFS, all of the communication of the data between the conductor and the operators would be achieved only through the Internet with much less traffic than ever. In addition to that, the WMS was expected to improve the data security by the enhanced control on accesses by operators. Also, the WFS was expected to save labor and time for organizing and integrating interpreted information by the consistent transactions of the simultaneous posts to the ground truth database. We constructed the system using open-source GIS software implementing standard schemes of WMS and WFS. By following to the schemes, our system would be effectively extensible and transferrable for further improvement of both server side and client side.

For further study, we will have a practical experiment of building ground truth database for global urban area mapping. Additionally, the system would provide opportunities to boost up the size of the database as well as those to statistical investigation on precision of visual interpretations.

ACKNOWLEDGMENT

This research used ASTER Data beta processed by the AIST GEO Grid from ASTER Data owned by the Ministry of Economy, Trade and Industry of Japan. The study was supported by a grant from Grant-in-Aid for JSPS Fellows (22-2598).

REFERENCES

Fritz, S. et al., 2009. Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover. *Remote Sensing*, 1(3), pp.345-354. Available at: http://www.mdpi.com/2072-4292/1/3/345/ [Accessed July 30, 2011].

Goodchild, M.F., 2008. Commentary: whither VGI? *GeoJournal*, 72(3-4), pp.239-244.

Haklay, M. & Weber, P., 2008. OpenStreetMap: User-Generated Street Maps. *Pervasive Computing, IEEE*, 7(4), pp.12-18.

Hand, E., 2010. Citizen science: People power. *Nature*, 466(7307), pp.685-7.

Iwao, K. et al., 2006. Validating land cover maps with Degree Confluence Project information. *Geophysical Research Letters*, 33, p.L23404.

Miyazaki, H., Iwao, K. & Shibasaki, R., 2011. Development of a New Ground Truth Database for Global Urban Area Mapping from a Gazetteer. *Remote Sensing*, 3(6), pp.1177-1187. Available at: http://www.mdpi.com/2072-4292/3/6/1177/.

Open Geospatial Consortium, 2011a. Web Feature Service. *Open Geospatial Consortium*. Available at: http://www.opengeospatial.org/standards/wfs [Accessed August 31, 2011].

Open Geospatial Consortium, 2011b. Web Map Service. *Open Geospatial Consortium*. Available at: http://www.opengeospatial.org/standards/wms [Accessed August 31, 2011].