A DYNAMIC INTEGRATION METHOD FOR BORDERLAND DATABASE USING OSM DATA

Xiao-guang Zhou*[a], Yu Jiang [a], Kaixuan Zhou [a], Lu Zeng [a]

[a] School of Geosciences and Info-Physics, Central South University, Changsha, Hunan 410083, PR China

Abstract: Spatial data is the fundamental of borderland analysis of the geography, natural resources, demography, politics, economy, and culture. As the spatial region used in borderland researching usually covers several neighboring countries' borderland regions, the data is difficult to achieve by one research institution or government. VGI has been proven to be a very successful means of acquiring timely and detailed global spatial data at very low cost. Therefore VGI will be one reasonable source of borderland spatial data. OpenStreetMap (OSM) has been known as the most successful VGI resource. But OSM data model is far different from the traditional authoritative geographic information. Thus the OSM data needs to be converted to the scientist customized data model. With the real world changing fast, the converted data needs to be updated. Therefore, a dynamic integration method for borderland data is presented in this paper. In this method, a machine study mechanism is used to convert the OSM data model to the user data model; a method used to select the changed objects in the researching area over a given period from OSM whole world daily diff file is presented, the change-only information file with designed form is produced automatically. Based on the rules and algorithms mentioned above, we enabled the automatic (or semiautomatic) integration and updating of the borderland database by programming. The developed system was intensively tested.

Xiao-guang Zhou: zxgcsu@foxmail.com

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-4/W3, 2013
ISPRS/IGU/ICA Joint Workshop on Borderlands Modelling and Understanding for Global Sustainability 2013,
5 – 6 December 2013, Beijing, China

1. INTRODUCTION

Spatial data is the fundamental of borderland analysis of the geography, natural resources, demography, politics, economy, and culture. As the spatial region used in borderland researching usually covers several neighboring countries' borderland regions, the data is difficult to achieve by one research institution or government. In the past few years there has been a very rapid growth of interest in volunteered geographic information (VGI) or crowd sourcing data. VGI has been proven to be a very successful means of acquiring timely and detailed global spatial data at very low cost(Goodchild, 2007; Coleman,et al, 2009). Therefore VGI will be one reasonable source of borderland spatial data. But VGI is produced by amateurs (or called 'neogeographers') voluntarily without strict regulations. The data model is different from the traditional authoritative model. For example, OpenStreetMap (OSM) has been known as the most successful VGI resource. But in OSM all kinds of streets or paths are identified by the tag "highway" in OSM, which is far different from the traditional authoritative geographic information and common sense. Therefore the original OSM data cannot meet the application of Borderland research. In order to solve this problem, we present a dynamic integration method for borderland data in this paper. At first we transfer the crowdsourcing data to be represented using authoritative data model at a certain scale, transfer the features to suitable classes. Because OSM data is collected by free tagging system, there are many unusual feature type tagged by neogeographers according to their understanding. These unusual feature types are difficult to transfer to authoritative classes with the traditional knowledge. In order to transfer the unusual type feature to the suitable classes automatically or semi-automatically, we developed a tool to assign OSM feature to authoritative class interactively and remember this transfer knowledge as a rule automatically, using this method to form a model transformation rule base, then the rule base is used to transfer the other OSM data set to user data model automatically. Using the data model transformation methods we can get a snapshot of the research borderland area at a certain time. But the real world is changing fast. It is need to update the data base of the research borderland area incrementally. VGI still be the low cost and worldwide change-only information resource. However, OSM does not provide methods to download the change file for a given region over certain temporal, but OsmChange provides daily diff data of the whole world. Thus, we developed a method to download the diff files for a given period, and extract the change objects in a given region, and merge the differ files to one change-only information file with designed format. Then the change-only information file is used to update the researching borderland data base automatically.

This paper is organized with 6 sections. Section 1 is introduction. We then discuss the dynamic integration strategy for borderland database in Section 2. The model transformation method is described in section 3. The change-only information extraction and incremental updating methods are discussed in section 4. An experimental test of this method is presented in Section 5. Section 6 provides a summary and concludes the discussion.

2. STRATEGY FOR THE DYNAMIC INTEGRATION USING OSM DATA

XML is the only format one can download from http://www.openstreetmap.org/. But XML is not suitable for borderland analysis. As shape file has been used in GIS widely, we use it as the borderland data format in this study.

There are three basic geometric primitives in OSM data model: nodes, ways, and relations, they are used to describe the spatial component of the features; the semantic information is described by tagging with key-value pairs. Nodes are the only primitives having coordinates information. Nodes with tags are used to represent the point features, and the others are used to describe the location of the ways and relations. Ways are ordered list of nodes, simple ways (not close, not self-intersect) are used to describe linear features, closed ways represent area features. Relations are lists of primitives and other relations. They are used to describe the topology, restriction, and complex regions (with holes). The features in OSM are categorized to three classes, i.e., Primary features，References，Additional properties. Primary features include "aerialway, aeroway, amenity, barrier, boundary, building, craft, emergency, geological, highway, historic, landuse, leisure, man_made, military, natural, office, place" 18 categories; References include "power, public transport, railway, route, shop, sport, tourism, waterway" 8 categories; Additional properties are used to describe the Addresses, Annotation, Name, Properties, Restrictions,etc properties. Therefore Primary and References features are mainly used in borderland applications. In this study, we mainly transfer Primary and References features to traditional authoritative model. As the 18 Primary features and 8 References features can be transferred to shape files according to the feature type and the geometric primitive type automatically. But OSM is a worldwide, inclusive map, there are many unusual things in OpenStreetMap and it is difficult to convert them to the user appropriated level automatically. Therefore, a 26 level model is used to as a middle data model, and a machine studying mechanism is used to convert the unusual things to the user appropriated level. Therefore we can get the base state shape-file map by model transformation.

For borderland application, the base state map from OSM usually is not sufficient, the scientists usually need to integrate other source of data to form a suitable data set. With the change of the real world, the borderland base state data need to be updated. OSM will still be the low cost and worldwide change-only information resource. However, OSM does not provide methods to download the change file for a given region over a certain period. But OsmChange provides the diff data of the whole world daily. Therefore, we can mine change-only information of the researching region from the diff file.

Based on this analysis, a dynamic integration method for borderland database using OSM data is presented in this paper. In this method, the XML format OSM data for a researching region is downloaded, the primary and reference features are converted to 26 primary levels (may still include point, line, and area sub-level, called middle data model in the following) shape-file format according to OSM feature type automatically. A machine-study mechanism is designed to convert the middle data model to the user designed destination data model for borderland application，A method is developed to download the whole world daily diff files for a given period automatically, the change objects in the given region of each diff files are extracted and stored in new DailydiffInP files; the DailydiffInP files are merged to form the change-only information file with designed format. Then change-only information file is used to update the researching borderland data base automatically. The strategy of this study is shown in Figure 2.
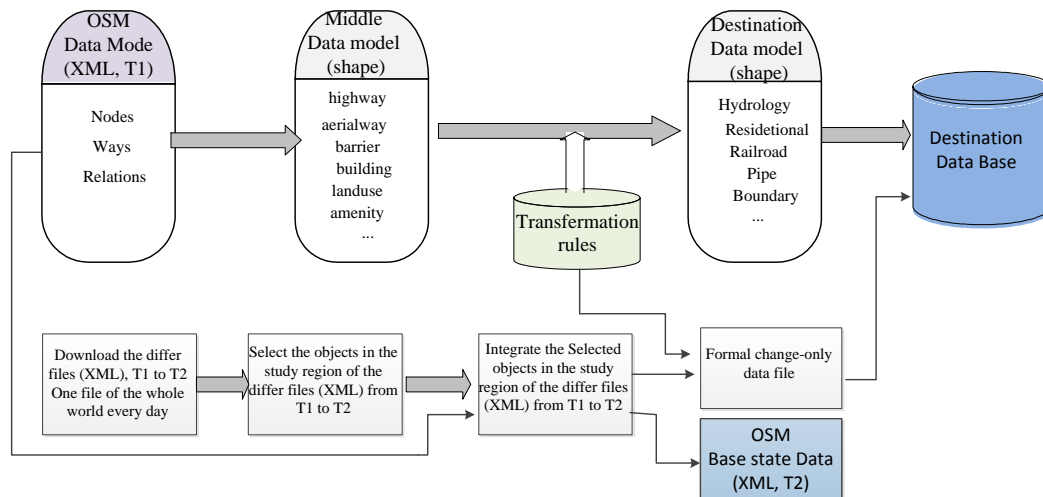
The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-4/W3, 2013
ISPRS/IGU/ICA Joint Workshop on Borderlands Modelling and Understanding for Global Sustainability 2013,
5 – 6 December 2013, Beijing, China

Figure 1. The strategy for dynamic integration of borderland database using OSM data

3. MODEL TRANSFORMATION

As mentioned above, OSM data model is usually different from the borderland research data model. OSM XML data can be converted to 26 primary levels shape-file format (middle model) automatically. Therefore, the key issue of model transformation is how to convert the middle data model to user data model automatically (or semi-automatically), especially how to convert the unusual feature types tagged by neogeographers to appropriate classes of user model.

In order to transfer the middle model to user model automatically (or semi-automatically), a rule-based transformation is presented. The rules are described as table 1 using Chinese 1:50000 national fundamental topographic data model as an example. Rule 1 can be interpreted as:

Rule 1 If OSMtag.k= Landuse && OSMtag.V= reservoir && OSMGeoPrim=way && Beginnode EqualsEndnode =Yes, then TargetLayer= HydrologyArea.

Table 1 Examples rules for converting the middle data model to user destination model

Rule Number	OSMGeoPrim	Beginnode Equals Endnode	Mdl Layer OSM tag.k	OSM Tag.V	TargetLayer
1	Way	Yes	Landuse Area	reservior	Hydrology Area
2	Way	No	Waterway Line	river	Hydrology line
3	Node	_	Natural Point	sea	Hydrology point
4	Way	No	Highway Line	secondary	Railroad line
5	Node	_	Amenity Point	bus_station	Railroad point
6	Way	Yes	Building Area	terrace	Residential Area
7	Way	No	Landuse Line	residential	Residential Line
8	Node	_	Building Point	apartments	Residential point

In this study, we use two methods to form the transformation rules. One method uses the Key, Value, Comment, rendering, and photo described in the OSM Map Feature description to form the rules. The unusual features are not defined in OSM Map Feature document. In order to solve this problem, the other method uses machine study mechanism, i.e., the researcher assigns the unusual features to the suitable classes interactively, the machine remembers the assignations to rule data base, and which can be used in the other data' transformation automatically. Therefore, we developed a tool to assign the unusual features to user classes interactively and remember this transfer knowledge as a rule automatically. Using this model

transformation method, the OSM data can be converted to the user borderland data model.

4. THE CHANGE-ONLY INFORMATION EXTRACTION AND INCREMENTAL UPDATING

As the real world is changing fast, borderland data needs to be updated incrementally. OSM data will still be the low cost and worldwide change-only information resource. However, in many borderland applications, the credibility and completeness of OSM data usually is not sufficient, the scientists usually have to enhance the data quality and integrate other source data to form a new data set. There are two methods to update the user

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-4/W3, 2013
ISPRS/IGU/ICA Joint Workshop on Borderlands Modelling and Understanding for Global Sustainability 2013,
5 – 6 December 2013, Beijing, China

borderland. One method is converting the new OSM data directly using the model transformation method mentioned in Section 3, checking the converted data, correcting the errors, and integrating the other source data again; The other methods is to exact the change-only information from OsmChange and using it to update the integrated user borderland database. In our opinion, the second method is more reasonable. Thus we will discuss the extraction of change-only information from OsmChange and the incremental updating in this section.

4.1 Selecting and integrating the objects in the differ files

OsmChange provides XML format diff file of the whole world daily. Similar to the OSM base state XML data, in OSM diff file the spatial property of the features are described by nodes, ways, and relations, the semantic information is represented by tagging with key-value pairs. OSM diff files have three types of change section, i.e., modify, delete, create, all objects are belong to one change section. These sections begin with "modify", "delete", "create", and end with "/modify", "/delete", "/create". The changed objects are located in the sections, as figure 2 shows.

```
<create>
    <way  id="220084146"  version="1"  timestamp="2013-05-05T04:50:05Z"  uid="427807"
user="samden" changeset="15979419">
        <nd ref="2292058739" />
        <nd ref="1111315307" />
        <tag k="highway" v="service" />
    </way>
</create>
```

Figure 2. OSM diff file format using "create" section as an example

As the diff files include the change information of the whole world, we have to extract the changed objects in the researching region. In OSM XML files, the coordinates are only stored in the Node data; the location of the way objects is described by node; the location of the relation objects is described by nodes and ways. It is assumed that, Polygon is used to describe the researching region; NodeInP, WayInP and RelationInP are containers used to store the data of the nodes, ways and relations in the researching polygon respectivly; NodeIDList and WayIDList are containers used to store the ID of the nodes and ways in the researching polygon respectively; DiffInP is the daily diff file for the objects in the studying region. Therefore the process for selecting the objects in the researching region from OSM daily diff files is shown in figure 3 using "create" section as an example.
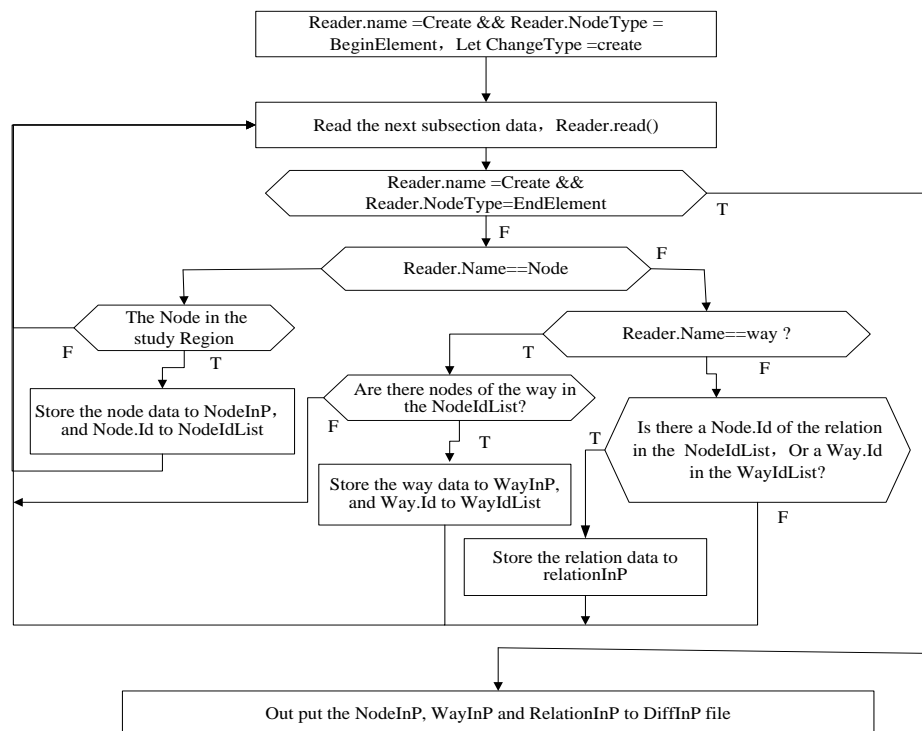

Figure 3 Selecting the objects in the researching region using "create" section as an example

The procedure for selecting the objects in the researching region is as follows:

Step 1 Read a change section begin-flag, store the change type;

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-4/W3, 2013
ISPRS/IGU/ICA Joint Workshop on Borderlands Modelling and Understanding for Global Sustainability 2013,
5 – 6 December 2013, Beijing, China

Step 2 Read one subsection data;

Step 3 Charge the data is a record of the changed object or the end-flag of the section. If it is the end-flag, turn to Step 5; Else

Step 4 Charge the record is a node, way or relation. If it is a node, charge if it is in the study region? If it is "yes", store the node data to NodeInP and Node.ID to NodeIDList; else turn to Step 2; if it is a way, Charge there is any reference nodes in the

way is in the NodeIDList, if it is "yes", store the way data to WayInP and Way.ID to WayIDList, else turn to Step 2; if it is a relation, charge there is any reference node.ID in the relation is in the NodeIDList, or any way.ID is in the WayIDList, If there is "yes", store the relation data to RelationInP;

Step 5 Putting out the NodeInP, WayInP and RelationInP to the daily diff file, i.e., daily DiffInP.

4.2 The Producing of change-only information file and updating of researching region data

As mentioned above, there is no coordinate for the way objects in OSM diff files. But in updating the spatial information is essential. Therefore one key issue for producing the change-only information file is to determine the coordinate for the way objects. In order to update the user database automatically, the changed objects in the diff files also need to assign to the appropriate classes of user model. Thus, another issue is to determine the classes for the changed objects. Because in OSM, there are two types of changed way objects, there are (is)

creation or deletion nodes for first one type of way, this kind of way objects can be found in the diff files; the coordinate of one or more nodes in the second kind of way modified, but there is no creation or deletion nodes, therefore it is not stored in the diff files, it is needed to exact from the OSM base state file.

It is assumed that, NodeDiff, WayDiff, NodeBase, WayBase are containers used to store the data of the nodes, ways from diff file and base state file respectivly. Therefore, the process for Producing of change-only information file is shown in figure 4.
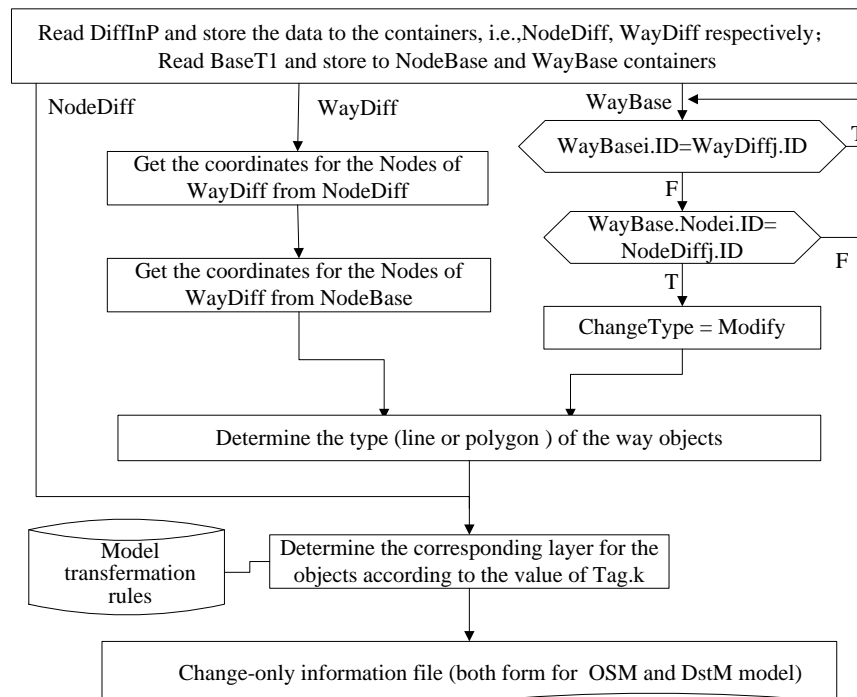

Figure 4 Process for Producing of change-only information file

The procedure for producing of change-only information file is as follows:

Step 1 Read the DiffInP files, store the data to the NodeDiff and WayDiff containers; read the Basestate OSM XML file at T1, store the data to the NodeBase and WayBase containers;

Step 2 For NodeDiff, turn to step step 4; for WayDiff objects, getting the coordinate for the nodes from NodeDiff and NodeBase. For WayBase object, charge if the WayBase.ID is equal to one of the WayDiff.ID, if it is, this way object has been stored in the WayDiff, then charge the next object; else charge if there is Node.ID in this way object is in the NodeDiff, if the answer is "no", this way is unchanged, turn to the next objects; else it is a node modification way;

Step 3 Determine the way object is a line or a polygon;

Step 4 According to the model transformation rules to determine the corresponding layer for the objects;

Step 5 Store the objects to the change-only information file.

OSM XML state file after change will be the source of the node coordinates for the new diff file. It is necessary to update the OSM state file after updating the user data base. In fact, the updating is easy, eliminating the deletion objects, replacing the modified objects, and creating the new objects (ZHOU et al, 2007; 2009).

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-4/W3, 2013
ISPRS/IGU/ICA Joint Workshop on Borderlands Modelling and Understanding for Global Sustainability 2013,
5 – 6 December 2013, Beijing, China

5. EXPERIMENTAL APPLICATION

Based on the rules and algorithms mentioned above, we enabled the automatic (or semiautomatic) integration and updating of the borderland database by programming with Visual C# 2010. The developed system was intensively tested using Chinese national 1:50000 fundamental geographic information data model as the user model and extension the name from Chinese name to include English name and the mother language name; using the OSM data of Vientiane, Laos as experiment data （LonMin=102,LonMax=103.16,LatMin=17.83,LatMax=18.37）, and using OSM data of May 3rd, 2013 as the base state, the differ data from May forth to sixth July). Figure 5 shows the changed roads (red color) of Vientiane extracted from OSM daily diff from May forth to sixth July, 2013.
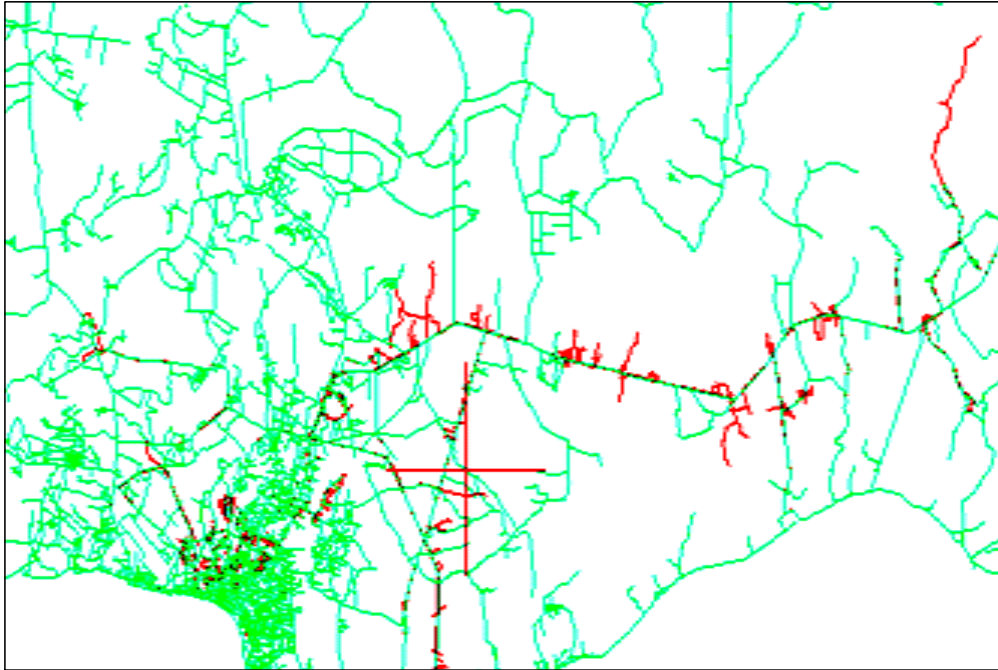

Figure 5 The changed roads (red color) of Vientiane extracted from OSM daily diff from May forth to sixth July, 2013

6. CONCLUSIONS AND DISCUSSION

In this article, we present a dynamic integration method for borderland database using OSM data. In this method, a machine study mechanism is used to convert the OSM unusual data to the user data model; a method used to select the changed objects in the researching area over a given period from OSM whole world daily diff file is presented, the change-only information file with designed format is produced automatically. With the change-only information file, we can update both of the user borderland database and OSM XML file for a researching region automatically.

Although this model was developed to integrate and update the borderland database, the method and algorithms can also be used to integrate and update the other user database. As OSM data is produced by amateurs (or called 'neogeographers') voluntarily, the creditability of the volunteers affect OSM data quality. How to evaluate the creditability of the volunteers will be a key issue in the future.

REFERENCES

[1]   http://wiki.openstreetmap.org/wiki/Map_Features

[2]   Coleman, D., Y. Georgiandou, and J. Labont. 2009. Volunteered geographic information: the nature and motivation of produsers. International Journal of Spatial Data Infrastructures Research 4:332-358.

[3]   Goetz M., Lauer J., Auer M., 2012. An algorithm based methodology for the creation of a regularly updated global online map derived from volunteering geographic information. GEOProcessing, 2012: The fourth international conference on advanced geographic information system, application, and services: 50-58

[4]   Goodchild M. F.，2007, Citizens as Sensors: The World of Volunteered Geography [J].GeoJournal, 69(4):211-221.

[5]   Goodchild M. F. 2009, NeoGeography and the Nature ofGeographic Expertise [J]. Journal of Location BasedServices, 3(2):82-96

[6]   Goodchild M. F., Li L., 2012. Assuring the quality of volunteering geographic information. Spatial Statistics, 2012(1):110-120

[7]   Hall G. B., Chipeniuk R., Feick R.D., Leahy M. G. and Deparday V., 2010. Community-based production of geographic information using open source software and Web 2.0, International Journal of Geographical Information Science, 24 (5): 761 — 781

[8]   Hardya D., Frewa J., and Goodchild M. F., 2012. Volunteered geographic information production as a spatial process. International Journal of Geographical Information Science, 2012.4, pp.1-22

[9]   Hagenauer ,J.,& Helbich ,M., 2012. Mining urban land-use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks, International Journal of Geographical Information Science, 26:6, 963-982

[10]   Kraak M.J., 2011. Is There a Need for Neo-Cartography? Cartography and Geographic

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-4/W3, 2013
ISPRS/IGU/ICA Joint Workshop on Borderlands Modelling and Understanding for Global Sustainability 2013,
5 – 6 December 2013, Beijing, China

Information Sciences, 28 (2): 73-78

[11] Mourad etc. 2010. Automatic change detection of buildings in urban environment from very high spatial resolution images using existing geodatabase and prior knowledge, ISPRS Journal of Photogrammetry and Remote Sensing, 65(1),Pages 143-153

[12] Murakami S., Takemoto T., and Ito Y., 2012, Data updating methods for spatial data infrastructure that maintain infrastructure quality and enable its sustainable operation, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XXXIX-B4, 2012 XXII ISPRS Congress, pp.29-33

[13] OpenStreetMap Wiki. Stats[EB/OL]. http://wiki.openstreetmap.org/wiki/Stats, 2012,10

[14] Paudyal D. R., McDougall K., and Apanab A., 2012, Exploring the application of volunteered geographic information to catchment management: a survey approach, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume I-4, 2012 XXII ISPRS Congress, pp.275-280

[15] Peter Mooney, Padraig Corcoran. 2012,Using OSM for LBS–An Analysis of Changes to Attributes of Spatial Objects, Geoinformation and Cartography,,pages:165-179

[16] ZHOU Xiao-guang, CHEN Jun, 2009，Dynamic operations for spatio-temporal database based on change mapping, Journal of remote sensing, 13（4）：647-652,

[17] Zhou Xiao-guang, Chen Jun, JIANG Jie, Zhu Jianjun, LI Zhilin, Event-based incremental updating of spatio-temporal database，Journal of Central South University of Technology (English Edition), 2004, VOL.11（2），192-198

ACKNOWLEDGEMENTS