

Article

A Progressive Buffering Method for Road Map Update Using OpenStreetMap Data

Changyong Liu ¹, Lian Xiong ¹, Xiangyun Hu ¹, and Jie Shan ^{2,*}

¹ School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China; E-Mails: rscy_liu@whu.edu.cn (C.L.); xionglian@whu.edu.cn (L.X.); huxy@whu.edu.cn (X.H.)

² Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA

* Author to whom correspondence should be addressed; E-Mail: jshan@purdue.edu; Tel.: +1-765-494-2168; Fax: +1-765-494-0395.

Academic Editor: Wolfgang Kainz

Received: 28 May 2015 / Accepted: 20 July 2015 / Published: 27 July 2015

Abstract: Web 2.0 enables a two-way interaction between servers and clients. GPS receivers become available to more citizens and are commonly found in vehicles and smart phones, enabling individuals to record and share their trajectory data on the Internet and edit them online. OpenStreetMap (OSM) makes it possible for citizens to contribute to the acquisition of geographic information. This paper studies the use of OSM data to find newly mapped or built roads that do not exist in a reference road map and create its updated version. For this purpose, we propose a progressive buffering method for determining an optimal buffer radius to detect the new roads in the OSM data. In the next step, the detected new roads are merged into the reference road maps geometrically, topologically, and semantically. Experiments with OSM data and reference road maps over an area of 8494 km² in the city of Wuhan, China and five of its 5 km × 5 km areas are conducted to demonstrate the feasibility and effectiveness of the method. It is shown that the OSM data can add 11.96% or a total of 2008.6 km of new roads to the reference road maps with an average precision of 96.49% and an average recall of 97.63%.

Keywords: OpenStreetMap (OSM); buffer analysis; road map update; road change detection

1. Introduction

The technology of Web 2.0 facilitates the interactions between servers and clients [1]. Specifically, not only can individuals download data from the Web 2.0 Web pages, but they can also access the database and modify its records. As a result of such development, the ways of acquiring geographic data are not any more limited to traditional authorized sources. The data can be from volunteered geographic information (VGI) which is voluntarily provided by a large number of ordinary citizens using network tools [2]. In addition to the convenience in information distribution and access, it is also easier to collect geographic data with non-professional mapping devices, such as the GPS receivers in smart phones or in automobile navigation systems. Technically, all citizens can be regarded as sensors [2] and the sources of geographic information can consequently be as diverse and complicated as the individuals collecting it. In this situation, the data contributors are amateurs and numerous. Goodchild and Glennon [3] suggested that the terminology of crowdsourcing geographic data almost equals to VGI, and the former may be a little more suitable to describe the effectiveness and accuracy attained by common users without expertise using this approach. The enormous potential of this voluntarily generated crowdsourcing data is increasingly gaining great interest. It is believed that VGI should become an indispensable source for geographic data in the coming era of Web 3.0. It will not only expand, complete, and improve existing geographic databases and archives, but will also collect and produce new forms of geographic information that reveal spatio-temporal patterns and help social and economic practices. As a result, research on VGI's potential applications is necessary and significant.

OpenStreetMap (OSM) is one of the most promising projects to provide crowdsourced data. It aims to create a free editable map of the world through the collaboration among individuals [4]. The available OSM data include roads, administrative boundaries, natural polygons, points of interest, and other features. The quality of OSM (road) data has been analyzed by many scholars. Relevant issues about the credibility of crowdsourced data, including OSM data, were discussed by Flanagan and Metzger [5] and Ying *et al.* [6]. They pointed out that geographic data provided by non-experts is also credible. Haklay [7] claimed that OSM data, when compared with the Ordnance Survey (OS) database, had an overlap of more than 80% in London and only 4% of the lines lacked complete attributes. Ather [8] concluded that the positional accuracy of OSM data was very good in comparison to the OS MasterMap ITN layer in England. Girres and Touya [9] assessed the quality of OSM data in France. While demonstrating certain advantages of OSM data, they also pointed out that its intrinsic heterogeneity highly limited their potential applications. Zielstra and Zipf [10] reported their evaluation on the OSM data of Germany, which was assessed by relative positional accuracy and relative completeness. They got the similar results with Girres and Touya [9]. Similarly, Ludwig *et al.* [11] found that 73% of the OSM roads in Germany were within a distance of only 5 m from the reference roads provided by NavTeq. A similar study on quality assessment was also completed in Greece by Kounadi [12], who concluded that OSM data could be used with reliability for several cartographic purposes because it had good thematic and positional accuracy. Helbich *et al.* [13] did similar work by comparing OSM data and Tele Atlas data with field surveying in Germany. They found that the OSM and Tele Atlas data are good enough for small and medium scale mapping applications. They also showed that Tele Atlas data had less distortion than OSM data and the quality

of the OSM data is heterogeneous. A first study in Munich of Germany was presented by Fan *et al.* [14] on the quality of building footprint data in OSM by using four criteria (*i.e.*, completeness, semantic accuracy, position accuracy and shape accuracy) with reference to the German authoritative topographic and cartographic database ATKIS. In terms of general quality evaluation, Mooney *et al.* [15] investigated the quality metrics of OSM data without using reference data source. Recently, Barron *et al.* [16] presented a framework, containing more than 25 different methods, on quality analysis of OSM data based on an OSM-Full-History-Dump. Studies focusing on the usage of OSM data are also being conducted. Online accessibility service based on OSM road data was realized by generating a road network with correct topological relations [17]. Over *et al.* [18] integrated OSM road features into a digital elevation model (DEM) after triangulation and conducted adaptive DEM line modification to generate 3D models. This work showed that OSM has great potential for creating virtual city models but still has room for advancement in absolute accuracy and completeness. Mondzech and Sester [19] generated a route network to implement pedestrian navigation using OSM data and claimed that urban or rural areas may have very different results. Pourabdollah *et al.* [20] successfully conflated the OSM road attributes, including name and reference code, with the Ordnance Survey roads and flagged the new roads in attributes towards an authoritative OSM. Zhang *et al.* [21] conducted one of a few experiments using OSM data in China, whereby they matched OSM data with a commercial database. The reason for the lack of research utilizing OSM data in China may be that the completeness of the OSM roads was poor at that time in China compared to Europe. However, its quantity and quality are ever-increasing with the increasing number of Chinese registered OSM users, making its applications more meaningful and beneficial. More recently, Neis and Zielstra [22] presented a review and their perspectives on the potential of the OSM data.

As new data sources become available, it is necessary to create updated road maps from various sources for specific needs [20,23]. Map conflation is applied to integrate heterogeneous maps overlapped in the same area. To get a timely and cost-effective road map for some applications such as navigation and location based services, we conflate the navigation road map with OSM road data by three main procedures: road matching, change detection and updating. As the biggest developing country, China witnesses a very high rate of road construction compared to developed countries. Finding changes, especially newly built roads and updating the database timely and effectively, is a demanding need [24]. As one of the most important procedures in map conflation, feature (point, polyline, and polygon) matching aims to build the link between the same object in the real world from different maps [25–27]. The buffer growing method is popular in road matching and the determination of buffer radius is a key to the matching outcome [28]. Since the registration quality between potential matching roads may vary from area to area, developing an adaptive buffering method to get a precise buffer radius for road matching is necessary.

The road (base) map in China is a thematic layer from the National Foundation Geographic Information System (NFGIS) at the scale of 1:10,000. Through compilation with reference to newly acquired images and field work this road map is updated routinely to a navigation map. However, there still exist missing roads during compilation because of the limited image resolution or occlusion caused by trees, tall buildings, and cloud *etc.* Also, the update frequency can be low, *e.g.*, once a year in the example of the 1:10,000 road map in Hubei Province. Moreover, the update is often very costly because a large number of crews are needed to conduct field surveying. Our work in this paper

therefore puts forward a simple, yet effective solution to automated change detection and update of road maps by using the OSM data. It takes advantage of OSM data and provides a dynamic road map change detection and update method to generate a complete, new road map. We first take the road maps as the reference roads and compare them with the OSM data after its topology modifications. Then, all the roads in OSM data are found by our proposed progressive buffer analysis, through which the optimal buffer is determined. The detected new roads, either newly mapped or built, are then combined to the reference roads through conflation, yielding an updated reference road map. The reference roads and OSM data of the city of Wuhan in China are used for this study. Based on this extensive study and evaluation, we then conclude our findings and make recommendations for future effort. This work could be of significance in considerably reducing the human and financial investment for road map updates for both government and industry.

2. Methodology

In order to update a reference road map by using the OSM data, the two datasets must be first approximately matched. The most popular method is map conflation technology, which consists of geometric conflation, topological conflation, and semantic conflation [29]. An early application of map conflation was accomplished by using an iterative rubber-sheeting algorithm in a project that merged TIGER (Topologically Integrated Geographic Encoding and Referencing) from the U.S. Census Bureau and the DLG (Digital Line Graph) database from the U.S. Geological Survey in 1983 [25], mostly for a 1:1 matching. Gabay and Doytsher [30] claimed that the topology of lines should be utilized to achieve an m:n matching. Based on the method of buffer growing, Walter [31] used the similarity of the angles, lengths, and shapes of the roads to match the different model data, and the thresholds of the similarity are calculated by statistical investigations. Cobb *et al.* [23] presented an expert system and added non-spatial information to the computation of similarity. A point cluster algorithm assisted by manual interaction was used by Xiong and Sperling [27] to determine matching pairs. Gösseln [32] and Volz [26] used iterative closest point (ICP) to match the points and then connect the corresponding nodes to obtain the matched road line. Four software tools (Conflex, JCS Conflation Suite, MapMerger, and TotalFit) were developed for matching roads from two datasets and to generate a merged one.

We adopt the idea of the buffer analysis method. This is due to the premise that the corresponding OSM data and reference road maps are close to each other geographically since the GPS derived OSM roads or image derived OSM roads have reasonably good position quality and the national NFGIS data were produced under the specification of 1:10,000 topographic maps (5 m in planimetry). Moreover, as we have cleaned up the topologic inconsistency between the two datasets, a simple buffer analysis should suffice for this purpose. For implementation, we first buffer the reference roads and an intersect analysis is followed to find roads that only exist in the OSM data. These roads are regarded as potential new roads because of the fact that the OSM data are updated more frequently than the reference roads. Since the results are affected by the buffer radius, we then compute the optimal buffer radius by examining the results of the new roads under different buffer radii. In the final step, the newly found roads are merged into the reference roads in terms of geometric, topologic, and semantic conflation.

2.1. Buffer and Intersect Analysis

Buffer represents the offset around geographic features. As shown in Figure 1, the reference roads (Figure 1a, black) are buffered to (Figure 1b, blue), which consists of an approximate rectangle at the continuous waypoints and two semi-circles at the discontinuous endpoints. An intersect analysis (Figure 1c) is implemented between the buffered polygon and the OSM roads (Figure 1c, red) to determine possible new roads. If an OSM road is totally contained in the buffer polygon, it is regarded as the unchanged road. However, if it is partly inside or totally outside the polygon, it is considered as a new road (Figure 1d, purple).

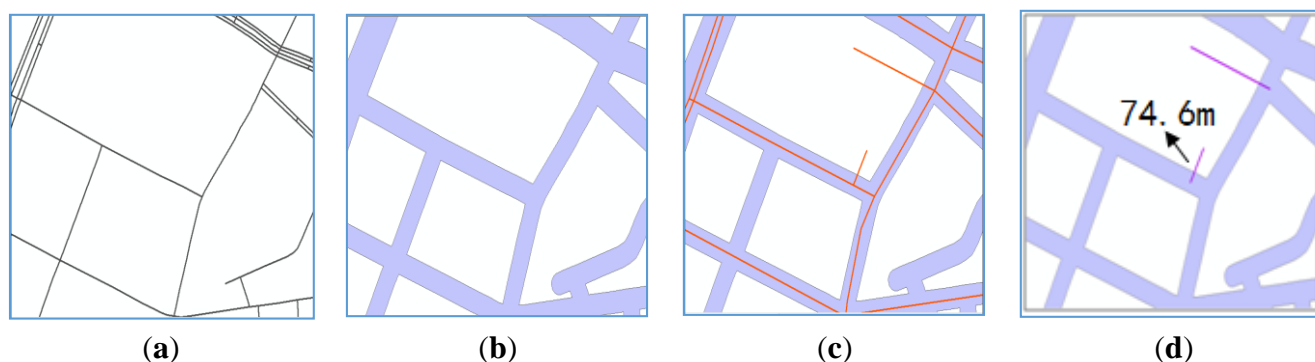


Figure 1. Buffer analysis to find new roads (a) Reference; (b) Buffed reference; (c) OSM roads within the buffer in red; (d) Detected new roads in purple.

However, for complex road intersections, such as roundabouts [33], the buffer analysis may lead to a hole in the center. In Figure 2, the OSM roads that pass through the hole may be mistaken as new roads after intersect analysis. Such holes therefore need to be filled to prevent potential misjudgment. This is accomplished by aggregating the buffer polygons within a certain distance to each other into new polygons.

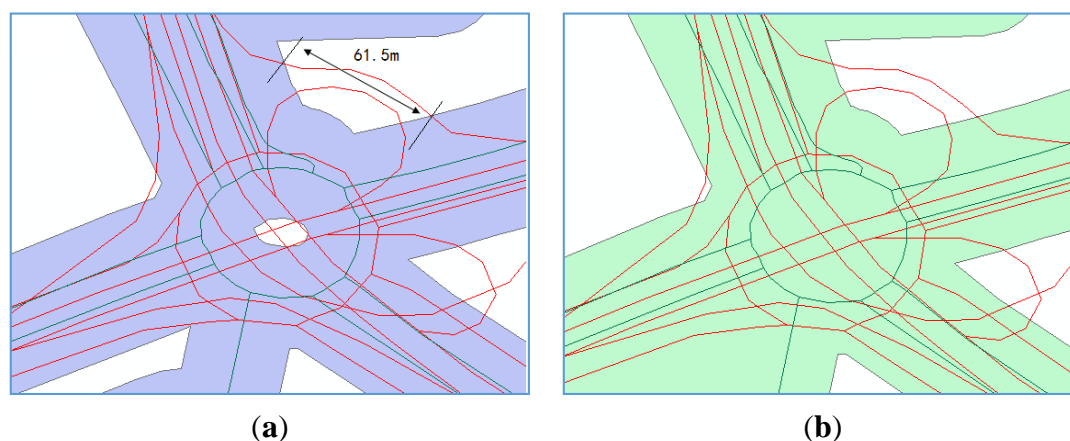


Figure 2. Buffers (a) (in blue) at a roundabout and their aggregation (b) (in green) to fill the hole at the intersect analysis. The OSM roads are in red and reference roads are in green.

2.2. Progressive Buffering

Since the buffer radius plays a key role in determining the new roads, a progressive approach should be implemented, for which we propose a two-step solution. We first determine the range of the buffer radius. Represented as single lines, the OSM roads should lie within the polygons that buffer the reference road centerlines by their widths. The reference dataset in our study uses the centerline of a one direction road to represent the road, i.e., one line for a one-way street. Two lines of similar shape at a certain distance signify a two-way street. When deciding the unchanged roads in the two datasets of the same area, the maximal road width is needed for selecting the range of the buffer radius. The road width varies in different cities. For example, Wuhan has mostly four traffic lanes in the same direction with a width of about 3.75 m each and one bicycle lane with a width of about 3.0 m. Considering the 5.0 m planimetric accuracy of the reference roads, the range of the road width is $(3.75 \times 4 + 3) \pm 5$ m, i.e., from 13 to 23 m. To ensure the correctness of this estimation, we randomly measured the distances of 791 corresponding line pairs in the two datasets. Their histogram is shown in Figure 3. Starting at a proper buffer radius (e.g., 9 m), buffering would include a significant percentage (~58%) of neighboring roads. Further enlarging the buffer radius will essentially include all neighboring roads (e.g., 96% at a 23 meter buffer). Based on the above two consistent estimations, the range of the buffer radius is finally chosen as 11–24 m.

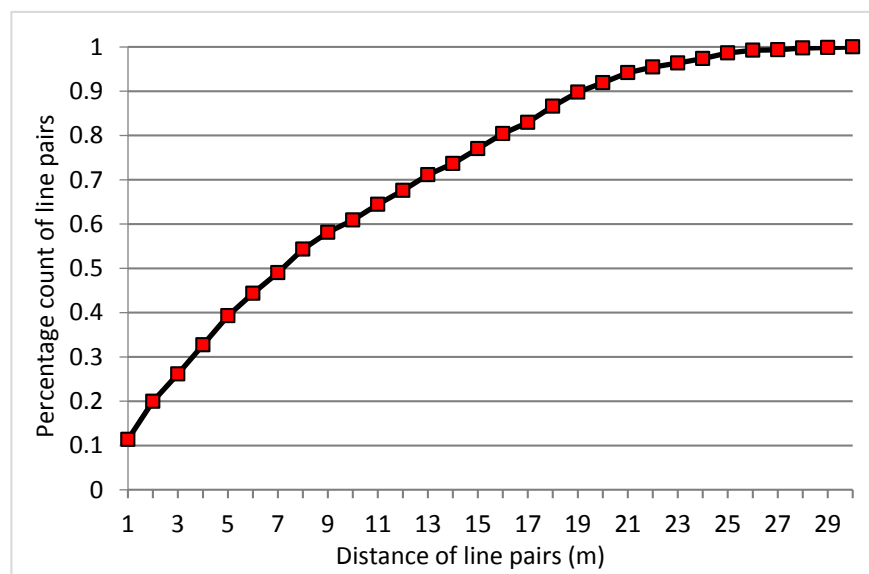


Figure 3. Histogram of the distances between the reference and OSM road pairs.

The second step is to calculate the optimal buffer radius. The optimal buffer radius is within the determined range above. We denote the total count and the total length of the OSM roads in an area respectively as T_n and T_l . Similarly, for the new roads detected under a buffer radius r_i we calculate their total road count $n(r_i)$ and the total road length $l(r_i)$. Using only the count or the length to describe the roads is not proper because the roads recorded in the datasets are segments other than the integral roads in reality. Therefore, the count and length are considered together in Equation.1, where $p(r_i)$ is the overall percentage of the detected new roads

$$p(r_i) = n(r_i) / T_n + l(r_i) / T_l \quad (1)$$

To reach a reliable determination of the optimal buffer radius, we apply a total of N times buffering, *i.e.*, $i = 1, 2, \dots, N$. Figure 4 plots the overall percentage graph for a $5 \text{ km} \times 5 \text{ km}$ area.

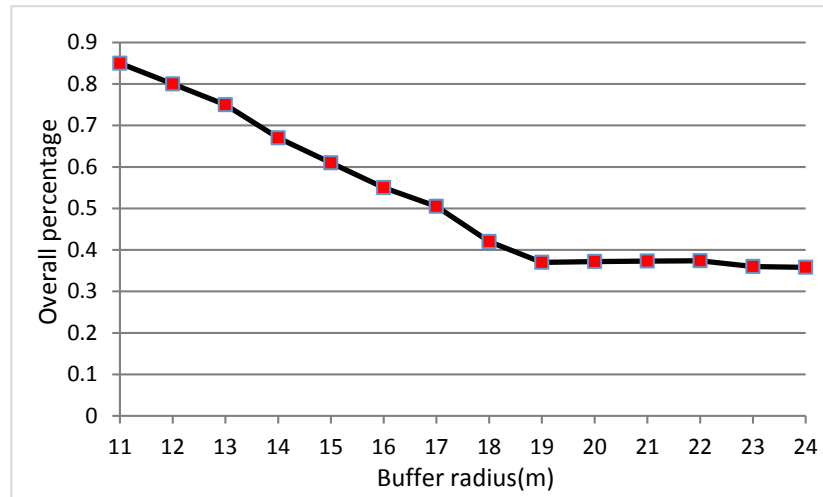


Figure 4. Overall percentage of Equation (1) in a $5 \text{ km} \times 5 \text{ km}$ area.

The optimal buffer radius should reach a stable $p(r)$, *i.e.*, its first derivative

$$dp(r_i) = p(r_{i+1}) - p(r_i) \quad i = 1, 2, \dots, (N-1) \quad (2)$$

should reach a minimum. To reach a reliable solution, we apply a min-max strategy. Starting from r_l , we consecutively divide the $p(r_i)$ into $N - M + 1$ overlapping groups, where M is the size of a group. We then select the max $dp(r_i)$ within each group. These group max(dp)'s are compared to obtain the minimum. The group that has the min(max(dp)) is regarded as our expected optimal interval since its maximal difference value is less than the other groups. It means that the $dp(r_i)$ in this group experiences the least changes, namely, a nearly stable tendency. After the optimal interval is determined, the optimal buffer radius can be acquired simply by averaging the buffer radii in this group. Assume that the first radius in this group is r_k , and the optimal buffer radius r_o is computed as

$$r_o = \frac{1}{M} \sum_{j=k}^{k+M-1} r_j \quad (3)$$

In our study, the size of a group is chosen as $M = 4$. To further verify this idea, the detected new roads are also identified by visual interpretation. The new roads obtained with a buffer radius that is less or more than the values in the optimal interval can lead to mistakes, such as including roads that are not new or excluding roads that are new. Therefore, the optimal buffer radius must be in the optimal buffer interval.

2.3. Add New Roads to the Reference Roads

To form a complete, updated reference road, the newly found OSM roads should be added to the reference road geometrically, topologically, and semantically to the best extent we can. For this purpose, another round of buffer analysis is employed, however, to the new OSM roads this time.

Shown in Figure 5a, if the buffer of an isolated OSM road does not intersect with any reference road, the OSM road can then be directly added into the reference road. Figure 5b shows a reference road completely falling into the OSM road buffer. They are then compared in terms of their significance, e.g., length and direction, so as to decide if we need to keep both or update the reference one with the OSM one. In the scenarios shown in Figure 5c and d, an OSM road barely touches a reference road within a certain threshold; *i.e.*, it either undershoots or overshoots a reference road within its buffer. The new road should then be snapped topologically to connect to the reference road.

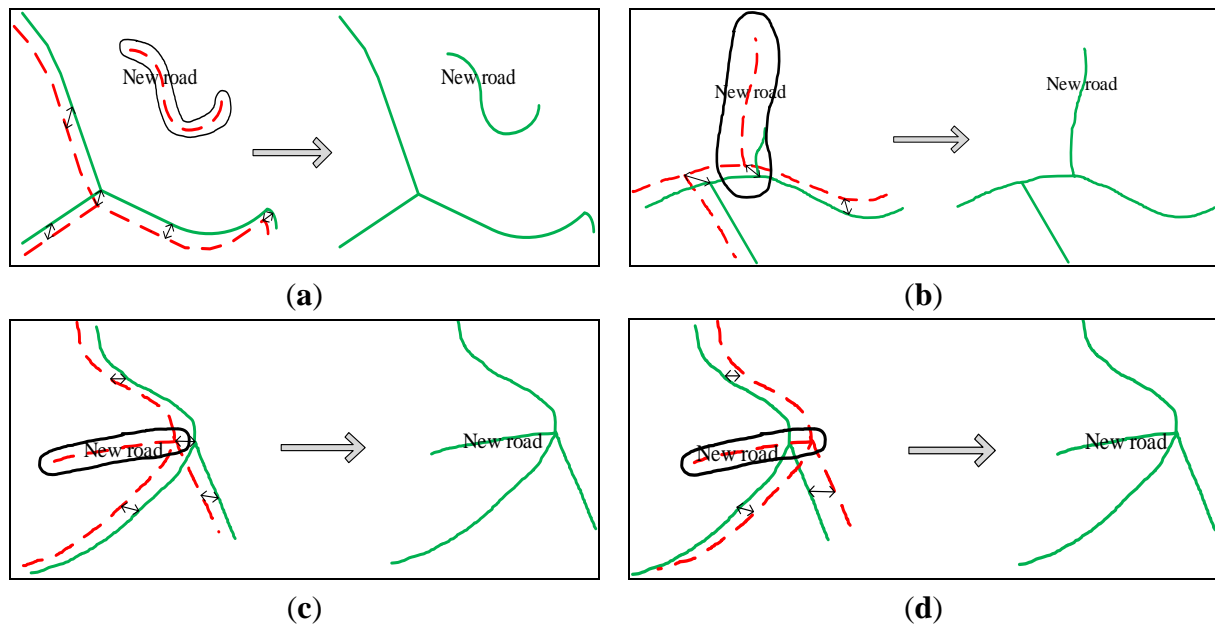


Figure 5. Scenarios of adding OSM (new) roads (red) into reference roads (green).
 (a) Isolated new road; (b) Removing insignificant reference road; (c) New road undershooting; (d) New road overshooting.

Besides geometry and topology, attributes of the reference roads need to be updated under certain semantic criteria. Since the attributes of the two datasets may not be consistent, one-to-one correspondence rules need to be established. Such fields as Shape, NAME and TYPE (or “KIND”) in the OSM roads are to be found for their best matches in the reference roads. When two fields in two datasets have the same name, they are matched directly. Otherwise, the fields will be matched manually to avoid matching mistakes. There are situations that the attributes of the final roads need to be determined by using the attributes of the OSM roads. For this purpose, similarity measures between attributes should be defined first. The attribute (string) similarity between the OSM and reference roads are calculated by a widely-used measure, the Levenshtein distance (or Edit distance) [34], which in our case is defined as the cost of transforming one string in the OSM road to another one in the reference road. The weights for insertion and deletion are set to 1 and the weight for substitution is set respectively to 0 for two identical words and 1 for two different words. For example, the Edit distance between the “ying yuan road” in the OSM road and the “da xue yuan road” in the reference road is calculated as follows:

Table 1. An example for calculating the Edit distance between two road names.

OSM Reference				
	Ying	Yuan	Road	
	0	1	2	3
Da	1	1	2	3
Xue	2	2	2	3
Yuan	3	3	2	3
Road	4	4	3	2

As shown in Table 1 the numerical element $D_{i,j}$, which indicates the Edit distance between the first i words in reference road name and the first j words in OSM road name, can be calculated by following expression:

$$D_{i,j} = \begin{cases} (\text{insertion cost}) \times i, & j=0 \\ (\text{deletion cost}) \times j, & i=0 \\ \min(D_{i-1,j-1} + \text{substitution cost}, D_{i,j-1} + \text{insertion cost}, D_{i-1,j} + \text{deletion cost}), & i > 0 \text{ \& } j > 0 \end{cases} \quad (4)$$

The Edit distance between “da xue yuan road” and “ying yuan road” is 2 (the number in the lower right corner). If the Edit distance is within the acceptable tolerance (<3), we adopt the attributes of the reference road and all of them are transferred to the final roads. If the Edit distance is beyond the tolerance, the final roads will use the attributes of the OSM roads. Sometimes, the OSM road attributes need to be changed to the type of corresponding attributes of the reference road. For example, the attribute of the OSM roads is ONEWAY and the value can be “yes” or “no”. If this value of the new road is yes, the value of LANNUM in the attributive table of the reference road map is set as 1 after the new road is added to the reference road.

3. Test Data

Navigation road maps are used as reference roads in this study. Such data were compiled based on the NFGIS road base maps, which were initially collected with VirtuoZo and JX4 (China made digital photogrammetric workstations) from aerial photographs of September 2003. The accuracies of the NFGIS road maps are respectively 5.0 m for planimetry and 1.5 m for elevation. Since 2003, routine field surveying was conducted by a navigation company (NavInfo) every year to keep the road data up-to-date so as to produce navigation road maps. The navigation road map used in this study was completed in August 2008 and covers an area of 8494 km² in the city of Wuhan. For evaluation and illustration purposes, we also selected five smaller subsets from this Wuhan dataset, with each being 5 km × 5 km. Their distribution is shown in Figure 6 below. To consider diverse road structures, subset 1 is chosen as a suburban area containing a few highways, whereas subset 2 is an old commercial zone. Both subsets 3 and 4 are new commercial zones and educational districts. Subset 5 is near the suburb and contains a part of an educational district.

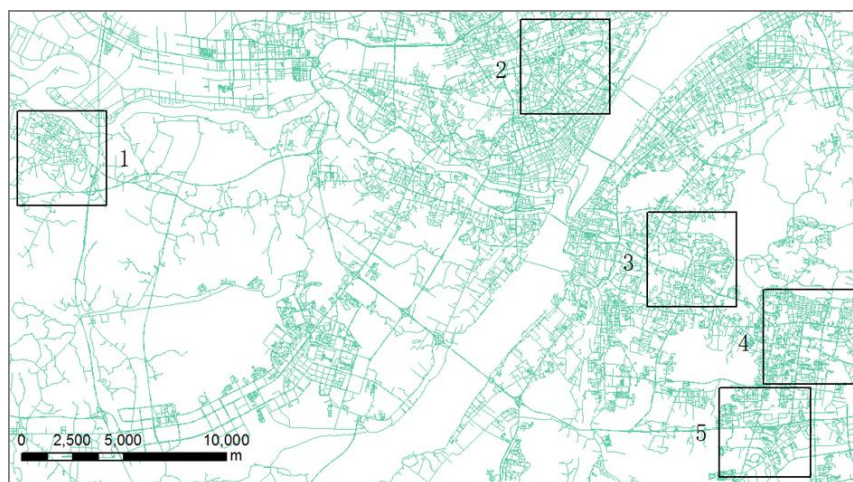


Figure 6. Reference roads of part of Wuhan city and its five 5 km × 5 km subsets.

As for the OSM road data, there are mainly four ways to acquire: (1) by JOSM (a desktop OSM Java editing tool), (2) by inputting URL composed of OSM API and the latitude and longitude of the target area, (3) by the Osmosis tool, a command line Java application, and (4) by directly downloading from the website of Cloudmade Corp, a business enterprise that offers free downloading of OSM road data in six formats. We used Cloudmade website to download the OSM road data within the test area in both OSM .xml and shapefile formats. However, the OSM road data, often produced by people without trained technical expertise and without a common quality requirement, may exhibit many topological errors. Some of these topological errors may influence the subsequent conflation procedure and should be corrected beforehand. In our experiments, four kinds of topological errors in OSM roads are observed: (a) redundancy, (b) self-intersection, (c) over-shoot, and (d) under-shoot. In Figure 7, the circled numbers designate the nodes and the line between nodes is a road record in the OSM road. Redundancy occurs when part of a line repeats itself. In Figure 7a, it can be seen that the line formed by node 3 and node 4 is the same as the one formed between node 5 and node 6. To correct it, the redundant line needs to be removed. Self-intersection means that part of a line intersects another part of the same line, as shown in Figure 7b. To correct, we split the lines of node 1–2 and node 3–4 at their intersection. Over-shoot is illustrated in Figure 7c, where the line of node 29–30 intersects the line of node 1–2. If the superfluous line segment at node 30 is within a tolerance, node 30 should be clipped to the line formed by nodes 1 and 2. Under-shoot in Figure 7d occurs when the line of node 29–30 fails to touch the line of node 1–2 within tolerance. We correct this error by extending the line of node 29–30 to meet the line of node 1–2.

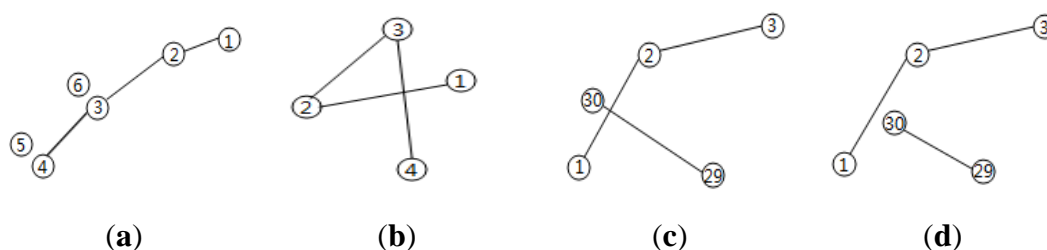


Figure 7. Some topological errors in OSM roads: (a) Redundancy; (b) Self-intersection; (c) Over-shoot; (d) Under-shoot.

The number of topological errors in the city of Wuhan OSM roads is summarized in Table 2. The results for the two data formats stay the same, which shows that the OSM .xml and shapefile data formats are topologically consistent. However, the attribute information in the .xml data format is more complete.

Table 2. Number of topologic errors in the city Wuhan OSM roads.

Redundancy	Self-Intersection	Over-Shoot	Under-Shoot	Total
86	47	56	72	261

4. Results and Discussion

This section will present and evaluate the results from the city of Wuhan dataset. It will address the selection of optimal buffers, and present the detected new roads and the updated reference roads.

4.1. Optimal Buffers

As described earlier, the optimal buffer radius is determined based on the detected new roads under the buffer radii ranging from 11 to 24 m. Figure 8 below displays the plots of the number of detected new roads in terms of the buffer radii. It is shown that the results experience a stable tendency in some intervals for every subset. Using the method presented in Section 2.2, the selected optimal buffer radii are listed in the second row in Table 3. The subsets 2, 3, and 4 have a relatively higher optimal buffer radius compared to the other two subsets because the commercial zones often have wider roads due to heavy traffic needs. As for subset 1, its highways contribute to the width of the roads in the subset. The subset 5 is distant from the downtown region and its optimal buffer radius is therefore the smallest. The optimal buffer radius for the big area of the city of Wuhan is influenced by the wide roads in the subsets 2, 3, and 4, and is determined as the same as that of those three subsets.

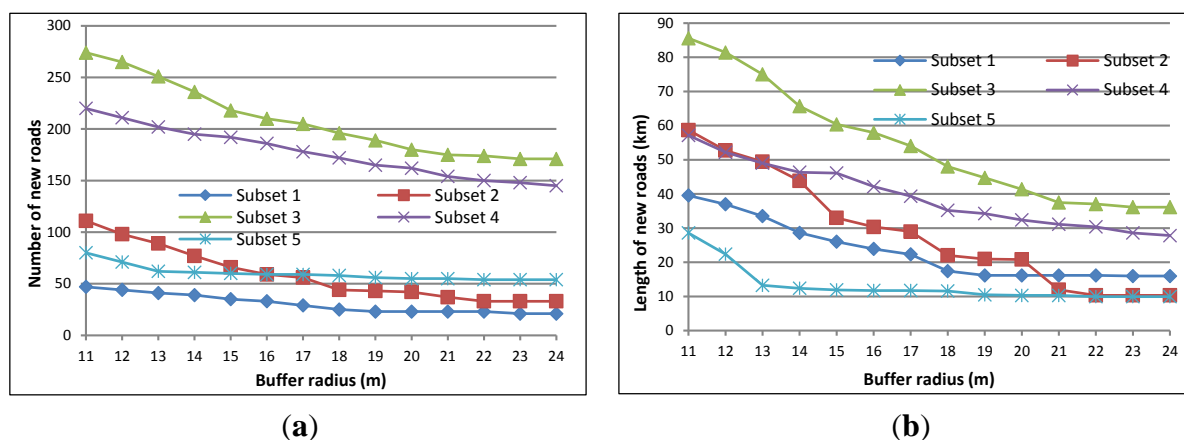


Figure 8. Plots of detected number of new roads (a) and their lengths (b) vs. the buffer radius for five 5 km × 5 km subsets in Wuhan.

4.2. New Roads Detected

New roads are those that do not correspond with the buffer zones of the reference roads. As an example, the results for subsets 1 and 3 are shown in Figure 9. The detected new roads are shown in blue in Figure 9. They are unique because there are no corresponding roads in the reference dataset (green). Also, the result for the city of Wuhan is shown in Figure 10 under the same color scheme. It shows that most of the OSM roads (red, Figure 10a) which have not been overlaid by reference roads (green) in Figure 10b found as new roads (blue) in Figure 10c. That means the new roads are successfully detected with our method.



Figure 9. Reference roads (green), OSM roads (red), and detected new roads (blue) for subsets 1 and 3 in Wuhan. (a) Reference atop OSM roads for subset 1; (b) Detected new roads for subset 1; (c) Reference atop OSM roads for subset 3; (d) Detected new roads for subset 3.

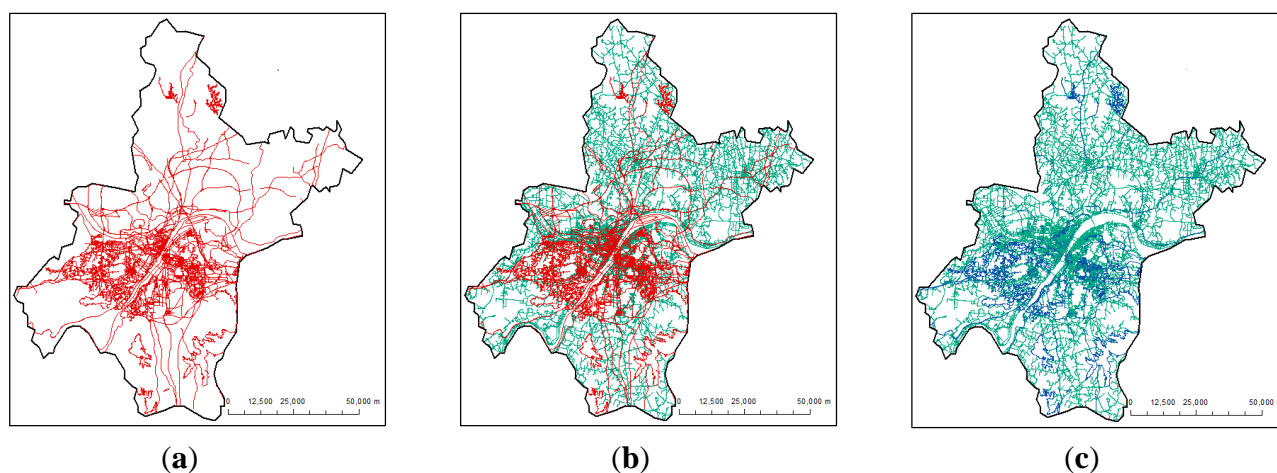


Figure 10. OSM roads (red), reference roads (green), and detected new roads (blue) for the Wuhan dataset. (a) OSM roads; (b) OSM atop the reference roads; (c) Detected new roads atop the reference roads.

To further evaluate the resultant new roads quantitatively, we manually select all the new roads in the five $5 \times 5 \text{ km}^2$ subsets and randomly choose 100 roads for the Wuhan dataset. They are then used as a reference to compare with the automated detected new roads. In our work, precision and recall are used to evaluate the results:

$$\text{Precision} = \frac{TPL}{TPL + FPL} \quad (5)$$

$$\text{Recall} = \frac{TPL}{TPL + FNL} \quad (6)$$

In the above two equations, TPL (true positive length) is the length of the correctly detected new roads, FPL (false positive length) is the length of the incorrectly detected new roads, and FNL (false negative length) is the length of the missed new roads. The precision measures the degree of accuracy with reference to the claimed new roads, and the recall represents the degree of accuracy with reference to the actual new roads. The outcomes are listed in Table 3. As it can be seen, our method demonstrates a high degree of validity with an average precision of 96.49% and an average recall of 97.63%.

Table 3. Selected optimal buffer radii, detected new roads and evaluations for the five subsets and the city of Wuhan. Note: 100 roads of the Wuhan dataset were randomly selected for evaluation.

Datasets	1	2	3	4	5	Wuhan
Optimal Buffer (m)	20.5	22.5	22.5	22.5	16.5	22.5
Correctly detected New Roads (km)	7.587	5.984	35.012	21.980	10.371	52.803
Wrongly Detected New Roads (km)	0.022	0.297	1.581	1.406	0.134	2.456
Missed New Roads (km)	0.135	0.090	0.589	1.117	0.322	0.808
Precision %	99.71	95.27	95.68	93.99	98.72	95.56
Recall %	98.25	98.52	98.35	95.16	96.99	98.49

4.3. The Updated Reference Roads

The detected new roads are finally merged into the reference roads in the study. As an example, Figure 11 shows the outcome of the subset 3 and subset 5. In terms of geometry, the new roads have been transformed to be nearer to the reference roads based on the geometric relations of the neighboring roads with which they intersect. The new roads (blue) show a small offset to the reference roads according to the OSM roads (dotted grey) to become the merged new road database (pink). In terms of topology, after the geometric transformation, the topology is updated if the distance is within tolerance.

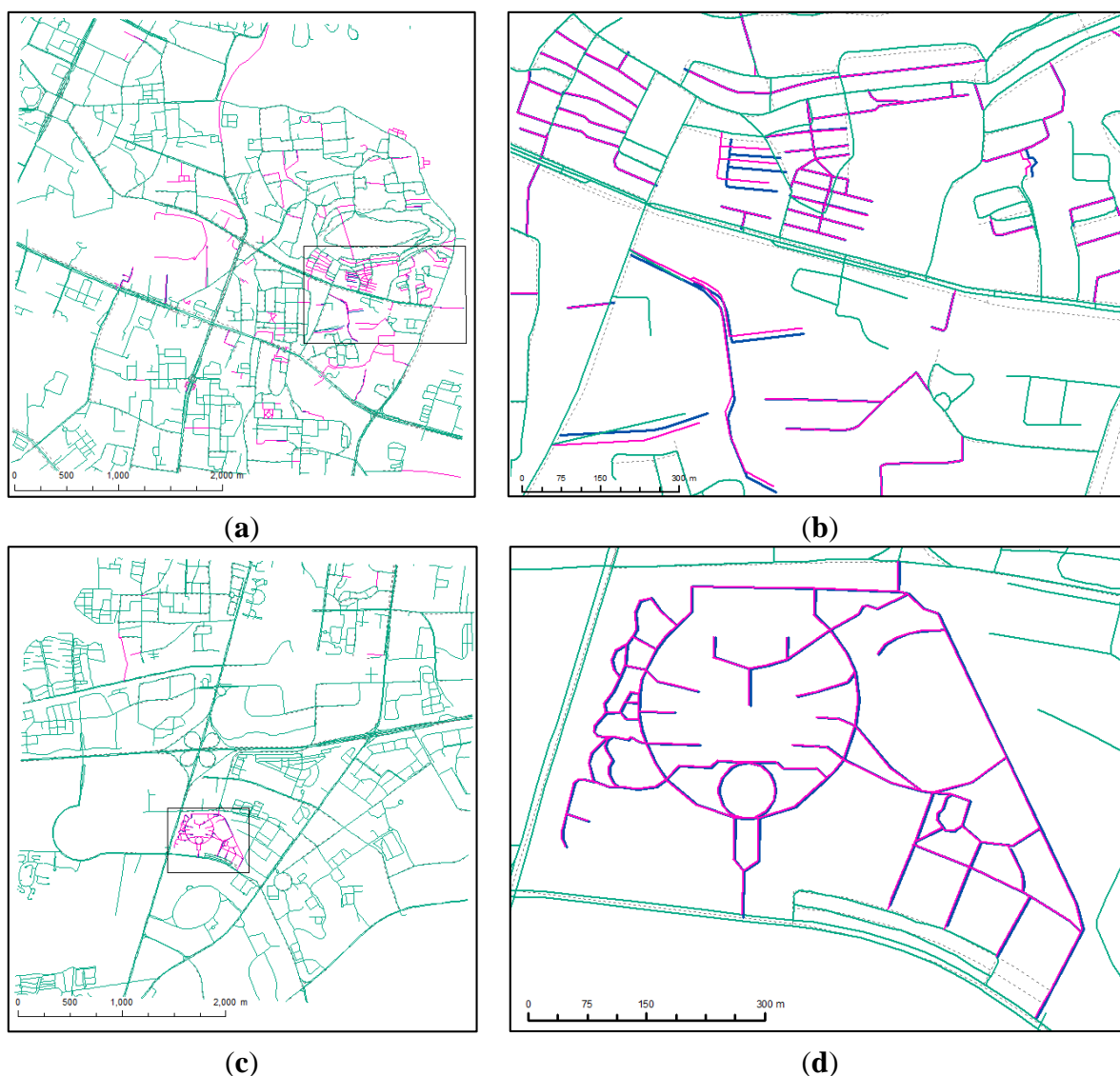


Figure 11. New roads added to the reference roads of subset 3 (a,b) and subset 5 (c,d). The right Figures b and d are an enlarged version of the boxed areas in the left Figures a and c, respectively. Reference roads are shown in green, new roads in blue, the merged new roads in pink, OSM roads in dotted gray.

In terms of attributes, the “Type” attribute of the OSM roads is added to the “Kind” attribute of the reference roads, whereas the “Name” field should be decided by its neighbors when necessary. In

Figure 12, the attributes of Shape, Kind, and Name, *etc.* for the new roads (highlighted items) are added.

FID	Shape*	MAPID	ID	KIND	Name	RoadLength
1795	Polyline	455463	45546304883	0814	zhuo dao quan bei road	39.556298
1796	Polyline	455463	45546306309	0601	zhuo dao quan bei road	151.532975
1797	Polyline	455463	45546306310	0601		92.170956
1798	Polyline	455463	45546304886	0814		107.645716
1799	Polyline	455463	45546304887	0814		89.984276
1800	Polyline	455463	45546304888	0601		40.361445
1801	Polyline	455463	45546304889	0601		78.967362
1802	Polyline	455463	45546304890	090a		79.320407
1803	Polyline	455463	45546304891	090a		319.84996
1804	Polyline	455463	45546304892	0804		8.682497
1805	Polyline	455463	45546304895	0814		53.481824
1806	Polyline	455463	45546304897	090a		111.881633
1807	Polyline	455462	45546304898	0601	dong san road	42.912438
1808	Polyline	455462	45546304899	0904		8.30538
1809	Polyline			secondary	shi pai ling road	28.446533
1810	Polyline			tertiary		119.843923
1811	Polyline			steps		8.890382
1812	Polyline			footway		30.191366

Figure 12. New OSM roads added to the attribute table of the updated reference roads (subset 3).

The length of new roads and their percentage to the length of the reference roads are listed in Table 4. Since subset 2 mostly consists of old commercial zones, there was only a small percentage (2.61%) of new roads developed from August 2008 to November 2011. On the other hand, subsets 3 and 4 are new commercial zones and educational districts. Both of them are rapidly developing centers as a result of the local government's policy in recent years. The new roads in those areas reach a highly distinctive 19.60% and 9.68%, respectively. Subsets 1 and 5 are located far away from the city center and they were not targeted for development during that time period. Only a few manufacturers were located there and the changes were moderate (5.62% and 5.89%). It may also be concluded that the percentage of new roads is influenced by the contributors to the OSM roads in specific areas. Subset 3, for example, has the highest new road rate (19.60%); and this may be partially due to the college students in that area, who are sensitive to new technology and more likely to upload and edit their data in a timely manner. For the big area of Wuhan, the detected OSM new roads can be as high as 11.96% of the reference roads, demonstrating the value of crowdsourcing geographic data.

Table 4. Length and percentage of the detected new OSM roads.

Datasets	1	2	3	4	5	Wuhan
New roads Length (km)	7.722	6.074	35.601	23.097	10.693	2008.644
$\frac{\text{New Road Length}}{\text{Reference Road Length}} \times 100\%$	5.62	2.61	19.60	9.68	5.89	11.96

5. Conclusion

The objective of the work is to find new roads, either newly mapped or built ones, in publically available OSM data for a given reference road map and create an updated road map. Current studies show that the OSM data should normally have a position accuracy of several meters, which provides a

reasonably good initial co-registration with the reference road map. Under this condition, the new roads are largely those that are not within zones of properly buffered reference roads. We present a progressive buffering method to determine the optimal buffer radius such that it yields a stable number of detected new roads. Tests with the city of Wuhan dataset of 8494 km² and a total of 16,783 km roads demonstrated the efficiency of such a technique. With this method, one can reach an average precision of 96.49% and an average recall of 97.63% for road change detection between the two datasets. The small variation of ~6% in quality across the tested five 5 km × 5 km datasets demonstrates the robustness and generality of the progressive buffering method. Moreover, it was shown that the OSM data can add 11.96% or a total of 2008.6 km new roads to the reference road map, which is quite a significant addition, demonstrating the value of OSM data when used properly. Since geometric, topologic, and semantic inconsistencies may exist between corresponding roads, some rules are necessary to define and transform the detected new roads into the reference road map to achieve a consistent conflation result. It is shown most detected new roads are located in new commercial zones and educational districts. This study demonstrates that the OSM data can be a very promising source for dynamic road data updates and change detection.

In spite of the great value of OSM data, there are still some limitations in road map updates. The completeness and quality of OSM roads still remain to be an issue. For example, among all cities in Hubei province, only Wuhan has detailed OSM data so far. For other cities, OSM data only includes major highways. Therefore, we believe the following aspects need improvement in the future. First, during the step of adding geometric information to the reference road database, fine matching algorithms using road nodes or intersections may be needed. Furthermore, another meaningful study could be undertaken in the conflation of the semantic information. It may be possible to improve the matching by considering the attributive information under a scheme of using levels of details. Secondly, it should be noted that OSM data cannot be used at this time to detect roads that no longer exist (*i.e.*, negative change) because there is no guarantee that OSM dataset itself includes all existing roads. In fact, OSM data often do not have complete coverage. Only when people truly digitized a road, either with reference to an image or through a GPS, can they record and share the road information. When OSM has a more complete coverage or can accommodate certain update annotations, determination of eliminated roads may become possible and should be added to the process to more realistically update the road database. Finally, the four types of topologic errors encountered in OSM data may not be inclusive and there is a need to carry out a more comprehensive study on the topologic errors in OSM data. Moreover, there are some topologic inconsistency between the added OSM roads and the reference roads in complicated areas, creating the need for sophisticated topologic checks and corrections in future endeavors.

We expect that OSM will attain greater recognition as a valuable data source soon because of its faster growth. In other words, it is quite possible that we may no longer assume that the official or commercial road database is superior in its position, themes, topology, and other aspatial information. Thus, how to reconstruct a new road network through data fusion may be a more pertinent topic in the future.

Acknowledgment

This work was partially sponsored by the National Natural Science Foundation of China under Grant No.61172175.

Author Contributions

L.X. designed the initial method, carried out the first tests and drafted the early version. C.L. modified and improved the method, and completed the tests and analyses. J.S. advised the study and contributed to writing in all phases of the work. X.H. advised the study and contributed to writing.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Haklay, M.; Singleton, A.; Parker, C. Web mapping 2.0: The neogeography of the Geoweb. *Geogr. Compass* **2008**, *2*, 2011–2039.
2. Goodchild, M.F. Citizens as sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221.
3. Goodchild, M.F.; Glennon, J.A. Crowdsourcing geographic information for disaster response: A research frontier. *Int. J. Digit. Earth* **2010**, *3*, 231–241.
4. Haklay, M.; Weber, P. Openstreetmap: User-generated street maps. *IEEE Pervasive Comput.* **2008**, *7*, 12–18.
5. Flanagan, A.J.; Metzger, M.J. The credibility of volunteered geographic information. *GeoJ.* **2008**, *72*, 137–148.
6. Ying, F.Y.; Corcoran, P.; Mooney, P.; Winstanley, A. How little is enough? Evaluation of user satisfaction with maps generated by a progressive transmission scheme for geospatial data. In Proceedings of 14th AGILE International Conference on Geographic Information Science, Utrecht, Netherlands, 18–22 April 2011.
7. Haklay, M. How good is OpenStreetMap information? A comparative study of OpenStreetMap and ordnance survey datasets for London and the rest of England. *Plan. Des.* **2010**, *37*, 682–703.
8. Ather, A. A Quality Analysis of OpenStreetMap Data. Master's Thesis, University College London, London, UK, 2009.
9. Girres, J.-F.; Touya, G. Quality assessment of the French OpenStreetMap dataset. *Trans. GIS* **2010**, *14*, 435–459.
10. Zielstra, D.; Zipf, A. Quantitative studies on the data quality of OpenStreetMap in Germany. In Proceedings of GIScience 2010: Sixth International Conference on Geographic Information Science, Zurich, Switzerland, 14–17 September 2010.
11. Ludwig, I.; Voss, A.; Krause-Traudes, M. A Comparison of the Street Networks of Navteq and OSM in Germany. In *Advancing Geoinformation Science for A Changing World*; Springer: Heidelberg, Germany, 2011; pp. 65–84.

12. Kounadi, O. Assessing the Quality of OpenStreetMap Data. Master's Thesis, University College of London, London, UK, 2009.
13. Helbich, M.; Amelunxen, C.; Neis, P.; Zipf, A. Comparative spatial analysis of positional accuracy of OpenStreetMap and proprietary geodata. In Proceedings of GI_Forum 2012: Geovisualization, Society and Learning, Salzburg, Germany, 4–6 July 2012. .
14. Fan, H.; Zipf, A.; Fu, Q.; Neis, P. Quality assessment for building footprints data on OpenStreetMap. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 700–719.
15. Mooney, P.; Corcoran, P.; Winstanley, A.C. Towards quality metrics for OpenStreetMap. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, New York, NY, USA, 3–5 November 2010.
16. Barron, C.; Neis, P.; Zipf, A. A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Trans. GIS* **2014**, *18*, 877–895.
17. Chen, S.Y. A Web-Based Accessibility Service OpenStreetMap Data. Master's Thesis, Shanghai Normal University, Shanghai, China, 2010.
18. Over, M.; Schilling, A.; Neubauer, S.; Zipf, A. Generating web-based 3D City Models from OpenStreetMap: The current situation in Germany. *Comput. Environ. Urban Syst.* **2010**, *34*, 496–507.
19. Mondzech, J.; Sester, M. Quality analysis of OpenStreetMap data based on application needs. *Int. J. Geogr. Inf. Geovisualization* **2011**, *46*, 115–125.
20. Pourabdollah, A.; Morley, J.; Feldman, S.; Jackson, M. Towards an authoritative OpenStreetMap: Conflating OSM and OS Opendata national maps' road network. *ISPRS Int. J. Geo Inf.* **2013**, *2*, 704–728.
21. Zhang, Y.F.; Yang, B.; Luan, X. Automated matching crowdsourcing road networks using probabilistic relaxation. In Proceedings of the 2012 XXII ISPRS Congress, Melbourne, ViC, Australia, 25 August–1 September 2012.
22. Neis, P.; Zielstra, D. Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. *Future Internet* **2014**, *6*, 76–106.
23. Cobb, M.A.; Chung, M.J.; Foley, H.; Petry, F.E.; Shaw, K.B.; Miller, H.V. A rule-based approach for the conflation of attributed vector data. *Geoinformatica* **1998**, *2*, 7–35.
24. Jiang, J.; CHEN, J. Some Consideration for Update of Fundamental Geoinformation Database. *Bull. Surv. Mapp.* **2000**, *5*, 1–3.
25. Rosen, B.; Saalfeld, A. Match criteria for automatic alignment. In Proceedings of Auto-Carto VII, Washington, DC, USA, 11–14 March 1985.
26. Volz, S. An iterative approach for matching multiple representations of street data. In Proceedings of 2006 ISPRS Workshop on Multiple Representation And Interoperability Of Spatial Data, Hannover, Germany, 22–24 February 2006.
27. Xiong, D.; Sperling, J. Semiautomated matching for network database integration. *ISPRS J. Photogramm. Remote. Sens.* **2004**, *59*, 35–46.
28. Zhang, M.; Meng, L. An iterative road-matching approach for the integration of postal data. *Comput. Environ. Urban Syst.* **2007**, *31*, 597–615.
29. Ruiz, J.J.; Ariza, F.J.; Ureña, M.A.; Blázquez, E.B. Digital map conflation: A review of the process and a proposal for classification. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 1439–1466.

30. Gabay, Y.; Doytsher, Y. Automatic adjustment of line maps. In Proceedings of GIS/LIS' 94 Annual Convention, Phoenix, AZ, USA, 1994.
31. Walter, V.; Fritsch, D. Matching spatial data sets: A statistical approach. *International Journal of Geographical Information Science* **1999**, *13*, 445–473.
32. Gösseln, G.V. A matching approach for the integration, change detection and adaptation of heterogeneous vector data sets. In Proceedings of XXII International Cartography Conference, Coruña, Spain, 9–16 July 2005.
33. Wiki.Roundabout, Available online, <https://en.wikipedia.org/wiki/Roundabout> (accessed on 20 May 2015).
34. Levenshtein, V.I. Binary codes capable of correcting deletions, insertions, and reversals. *Sov. Phys. Dokl.* **1966**, *10*, 707–710.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).