Matching assessment of road network objects of volunteered geographic information.

Dr. Maythm Al-Bakri

Email : m.m.s.albakri@gmail.com
Zuhair Yahya Sfoog
Department of Surveying
College of Engineering
University of Baghdad
E-mail: eng_zuhair_70@yahoo.com

ABSTRACT:-

Recently new concepts such as free data or Volunteered Geographic Information (VGI) emerged on Web 2.0 technologies. OpenStreetMap (OSM) is one of the most representative projects of this trend**.** Geospatial data from different source often has variable accuracy levels due to different data collection methods; therefore the most concerning problem with (OSM) is its unknown quality, [1].

This study aims to develop a specific tool which can analyze and assess the possibility matching of OSM road features with reference dataset using Matlab programming language. This tool applied on two different study areas in Iraq (Baghdad and Karbala), in order to verify if the OSM data has the same quality in both study areas. This program, in general, consists of three parts to assess OSM data accuracy: input data, measured and analysis, output results. The output of Matlab program has been represented as graphs. These graphs showed the number of roads during different periods such as each half meter or one meter for length and every half degree for directions, and so on .The results of the compared datasets for two case studies give the large number of roads during the first period. This indicates that the differences between compared datasets were small. The results showed that the case study of Baghdad was more accurate than the case study of holy Karbala.

1. Introduction

Recent development of geospatial data collection technologies and the growth of the World Wide Web (WWW) for different applications have led to a massive increase in the amount of geospatial data. Nowadays, there are large amounts of geospatial data have been distributed and shared by means of the Internet.

Map users become able to publish maps online, quickly and without need for distribution or printing costs, [3]. There are wide assortments of geospatial data sources available on the Internet such as the Google Map service, the OpenStreetMap (OSM) project, the Flickr service, the interactive Wikimapia website, Yahoo imagery and others, [5].

The (OSM) project is a collaborative project like Wikipedia born in 2004 in England. The aim of OSM is to create and provide free geographic data. The project aims to compensate the lack of free data because

geographic data, even freely available, are provided with licenses restricting the use of information and the creativity according to project leaders.

Data stored in OSM by contributors of the project are modeled and stored in tagged geometric primitives. The question of the quality of such data should be asked, [6]. Furthermore there are several websites that are built upon the crowd-sourced movement and are also free and open-sourced. The advantages of these websites are that they can update existing mapping databases and also enable the mapping of the developing world that is not complete so far.

One such example is OpenStreetMap; a free editable map of the world made by volunteers. The OSM project has many Positives such as: anyone can easily obtain vector data for free, users can improve and update the map and that is very useful especially for the case of poorly mapped regions and also users can create specialist maps or make other innovative uses. However there is a concern about the quality of VGI data because the volunteers that contribute lack sufficient cartographic training needed to create a good representation of geospatial information. Also the quality of the data cannot be precisely evaluated, [7].

Many researchers have been conducted on VGI data which are fundamentally focus on quality analysis. Fairbairn and Al-Bakri, [4] studied VGI quality in UK and Iraq by comparing OSM data with authoritative large-scale datasets, such as Ordnance Survey(OS) data in a first case study and General Directorate for Survey (GDS)data in second case study, for assessing positional and shape matching. They used different methods to compute positional discrepancy based on principles of NSSDA and directional statistics to select and analyze tested points. Line geometric similarity measurement based on buffering techniques and statistics. The authors developed MATLAB Graphical User Interfaces to present, interpret and analyze the results .The output shown that the OSM data does not match the formal data in each of the study area.

A recent study by Koukoletsos et al., [8] evaluated VGI data completeness for several areas in UK. The tests were carried out by comparing OS data, as the reference dataset, with the OSM project based on multi-stage approach for an automated matching procedure. The method was applied on geometric and attribute (road name and type) quality. The results indicated that the data complete in urban area and sparser in the rural area. In addition, Koukoletsos et al., [8] shown that the matching method may not be so efficient When conducted a comparison of the various road networks.

Zielstra and Zipf, [9] investigated the completeness of OSM data in

Germany by comparing it with the TeleAtlas data. This work extended the studies from England by Haklay (2008) and Ather (2009). The results showed that the geospatial data has been continues growth to freely available compared to open source spatial data in the past few years. Furthermore, the results indicated that there is still a very strong heterogeneity of OSM data in terms of completeness.

2. Datasets

2.1 The OSM data

In this paper the OSM data was downloaded as shapefile format from Geofabrik service (http://geofabrik.de.com/) for the whole Iraq then, only the road datasets were extracted for the two study area (Baghdad and Karbala) as shown in Figs 1 and2.

The OSM data has been collected and uploaded by volunteers. Anyone Consist from nine layers (Source: geofabrik.com) as shown in Fig. 3.

2.2 Reference data (Ref.)

The dataset used as the reference dataset was from the General Director of Surveying (GDS) / Iraq. The dataset was aerial image with resolution of 0. 1 meter .The GDS can provide: geodetic – geophysical products, aerial photographs, cartographic products, boundary surveying, digital geographical data, services to the public organizations who has registered on the OSM website can upload data and contribute to the OSM project. The layers available in the OSM website are : highway, barrier, cycle way, track type, waterway, railway, airway ,aerial way ,power, manmade, leisure, amenity, shop, tourism, historic, land use, military, natural, route, boundary, sport, abutters, accessories, properties, restrictions, name, references, places, addresses and annotation .

The geographic datum used in OSM is the WGS-84 datum, and all OSM data are projected on UTM projection, [10]. There are various websites that can be used to download OSM data such as Cloud Made (route calculation) or GeoFabrik (which provides OSM data as a shape files for example, the OSM data of Iraq

and the citizens. The projection of this dataset is Universal Transfer Mercator (UTM) projection, zone 38N and the spheroid is World Geodetic System 84 (WGS 84). Digitizing procedures allow the user to record details of what has been captured, including feature names/types, registration positions, line thicknesses or color. Hence, the roads datasets have been digitized in order to compare with OSM road datasets.

Fig. 1 OSM data for Baghdad: road layer only Baghdad case study [12].

Fig. 2 The OSM data for the Holly Karbala city (tested dataset) [12].

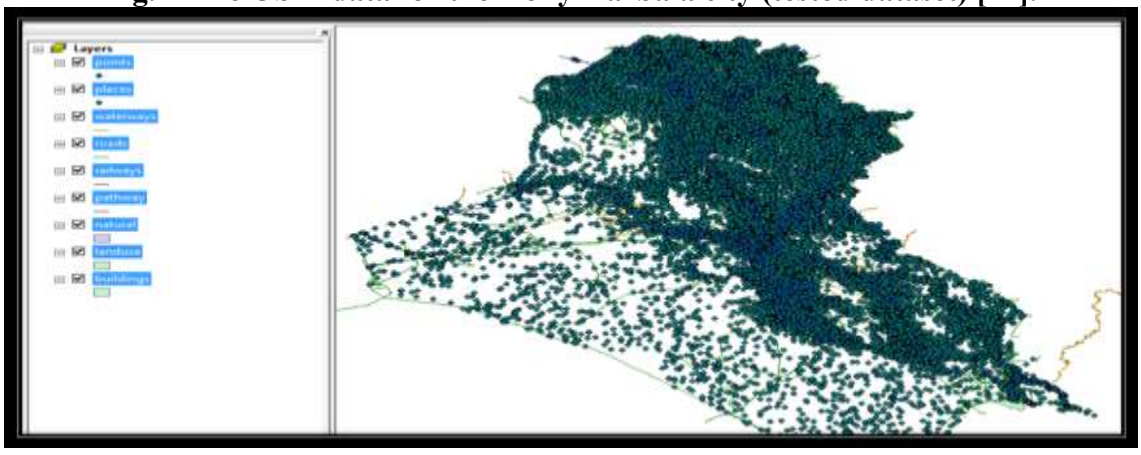


Fig. 3 illustrates the shape file of the OSM data sets (Iraq data) [12].

3. The study area

Two study areas have been chosen to assess the quality of OSM road network data, in order to investigate if the OSM data has the same accuracy in both study areas. They

are located on the center of Baghdad city and the holy Karbala city .The area of each is approximately 36 km². The study areas and their Directorate of Surveying / Iraq was used as a base for the comparison as shown in the Fig.s 4 and 5 , which properties represented in Table 1 .In addition to OSM data, reference data (aerial photograph) from General

shows the study areas in Baghdad and the holy city of Karbala.

Table 1. Study areas and their Properties [11].

Datasets	Upper left corner		Lower right corner		Number of roads	Area (km2)
	Easting (m)	Northing(m)	Easting(m)	Northing(m)		
Baghdad	446689.593	3693843.390	452769.347	3687840.206	2699	36.497
Karbala	404748.800	3610084.173	411354.147	3604081.510	1272	36.048

Fig. 4 The aerial image for Baghdad city (reference dataset) [13].

Fig. 5 The areal image for the Holy Karbala city (reference datasets) [13].

4. Methodology

The aim of the presented research is to compare the quality of OSM road linear features with the corresponding lines present in the reference data, [2]. The following methodology aims to assess the quality elements of OSM road network (tested dataset) against the aerial image dataset (reference dataset). The methodology focuses on the analysis of each element and on the comparison between the results of those.

4.1 Preparation of data

Two types of datasets have been used in this study: Formal datasets or reference datasets such as data from (GDS) and informal or VGI data such as OpenStreetMap (OSM). The preparation of the datasets is essential. The main reason for this choice is to compare the quality of OSM data in two different areas.

The reference dataset was obtained from the General Directorate for Surveying (GDS) as an aerial image with resolution (0.1) m. The tested datasets was the OSM data (only the road data sets). It was downloaded as a shape file format from GeoFabrik service (http://geofabrik.com/). This methodology consists of a sequential use of several tools available in ArcGIS.

4.2 Classifications of OSM roads

The shape file of OSM data consist from many types of roads layers, in this study (Iraq case study) OSM layers consist of nine layers as showed in Fig. (1). The road‘s layer contains a large number of roads, the classifications of OSM roads can be summarized as follows:

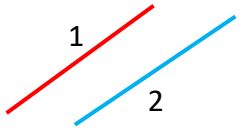
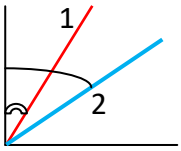
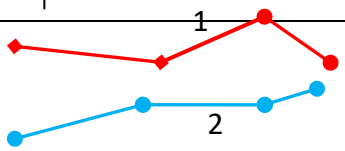
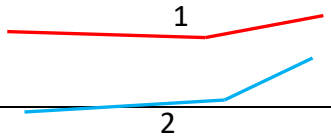
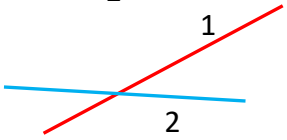
- Pedestrian : a pedestrian is a person traveling on foot, whether walking or running
- Primary: a major highway linking large towns, in developed countries normally with 2 lanes.
- Primary – link : roads which connect roadways to other roadways as part of an interchange
- Residential: roads accessing or around residential areas but which are not a classified or unclassified highways.
- Road : roads are used as local authority designations for routes within their area for administrative purposes
- Secondary: a road supplementing a main road, usually wide enough and suitable for two-way, all-weather traffic at moderate or slow speeds.
- Secondary –link: roads used to identify turning lanes connecting through lanes of highways to other roadways of all types.
- Service: generally for access to a building, service station, beach, campsite, industrial estate, business park, etc.
- Tertiary: roads connecting smaller settlements, and within large settlements for roads connecting local centers.
- Trunk: is a major road, usually connecting two or more cities, ports, airports and other places.

- Trunk –link: roads used to identify turning lanes connecting through lanes of highways to other roadways of all types.
- Unclassified: unclassified roads are local roads with no defined destination.

4.3 Conditions for datasets comparison

In this research six conditions have been adopted in order to evaluate the matching between the OSM and reference dataset as shown in table 2:

Table 2. Description and properties of the six conditions [11].

Condition	Description	Sketch
Length	The distance. How far from end to end or from one point to another.	
Direction	Defined by a horizontal angle between the line and a defined reference line called a meridian	
Shift in start point	Difference distance for start point of OSM road and the start point of Ref corresponding road	
Number of nodes	A single point with latitude and longitude.	
Number of segments	A segment is the distance between two nods.	
Intersections	Ref road intersect with same corresponding line on OSM.	
1: OSM 2: Ref.		

4.4 Program design

In order to achieve the main goal of this research, Matlab programming language was used to design a program that can analyze and asses the possibility of matching of OSM road features. Firstly, the data will be exported from ArcGIS to Excel file to save it as. txt file in order to call into the program through the use of specific code .

The account differences in lengths, directions, shifting in coordinates, number of nodes, number of

segments, and the number of intersections can be calculated .The Matlab program will calculate the roads numbers for this differences .The data for two case study, the procedures and processes of calculations for the logical steps followed in the development of this program can be explained in flowcharts below.

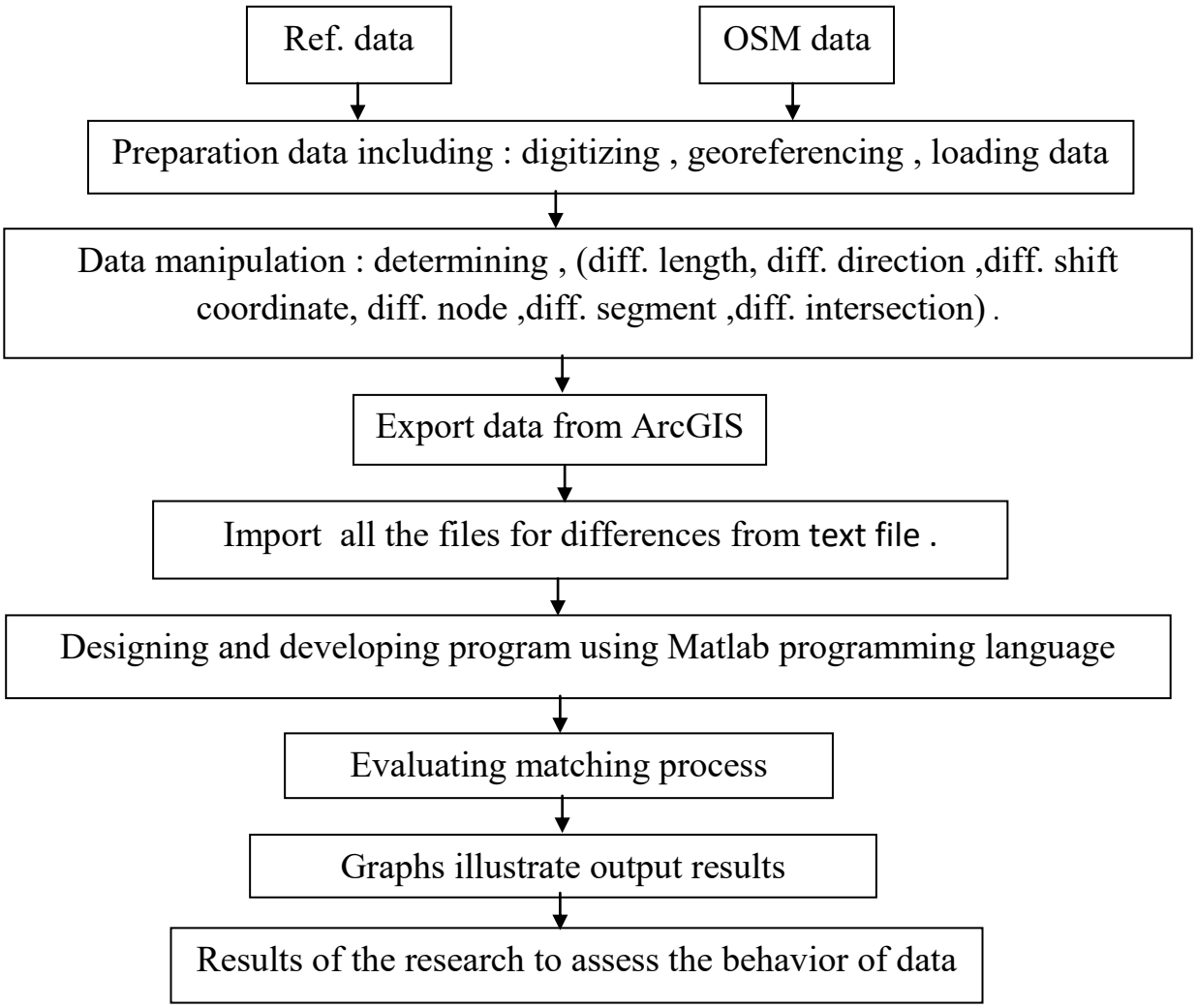


Fig. 6 The mechanism of the program [11].

5. Results and discussion

Fig. (7) illustrates the number of tested roads of (holy Karbala) when the differences in lengths are specified in a certain period of 1 m. The program gives (226) roads through the period bounded between zero and one meter ($\leq 0 \leq 1$) , while the program giving 166 roads to a period that are larger than one meter and smaller or equal than 2 meter ($< 1 \leq 2$) . For the period between two and three meter, the program gives 165 ($< 2 \leq 3$) . The process continues until they get less number of roads. The same procedure has been followed in Fig. (8). The number of roads when the differences in directions between zero and half degree (0 - 0.5) was (1102) roads,

while in the period between (0.5 – 1) the roads numbers was (42). Fig. (9) illustrates roads number when differences in shifting of start points was (≤ 0 <10) meter. For example for the period (0 -1) meter the number of roads was (289), and (1-2) the roads number was (272). Fig. (10) shows the numbers of roads when the differences in the number of intersections (0 , 1 , 2, , 3 , and 4) , when the number of intersections = 0, there is non-intersecting roads .The number of roads = (851) when the number of intersections = 0,while the roads give numbers =393 when the intersection = 1. Segments number illustrated in Fig. (11) for (≤ 0 ≤ 10), when the differences in segments number is zero, the number of roads was (374), but when the differences in segments = 1, the number of roads was (435). Same procedure was applied on Fig. (12) which calculates roads number when the differences in segments were (≤ 0 ≤ 10). For example, when the differences in the number of nodes = 0, the roads = 362, when dffi .= 1, the roads = 424 while the dff.=3, the roads = 231 and so on .

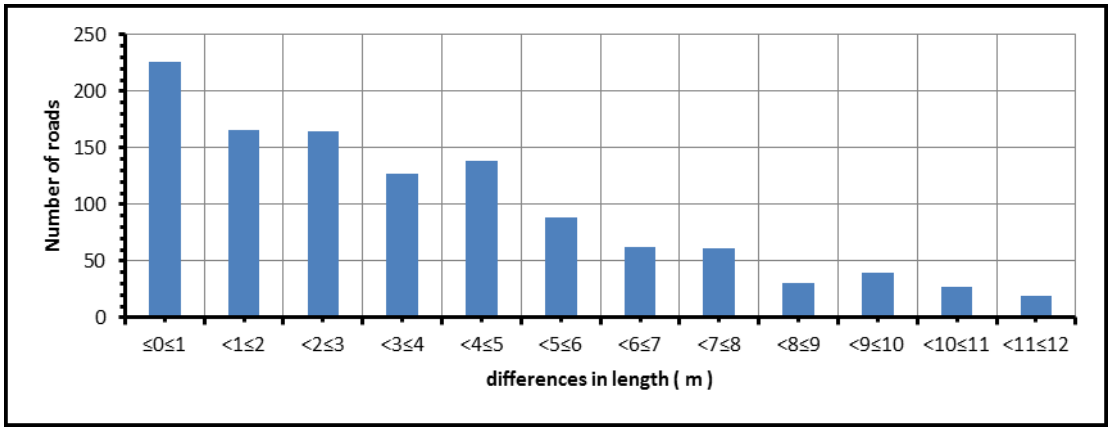


Fig. 7 illusterates the results of the length coparision (holy Karbala case study) [11].

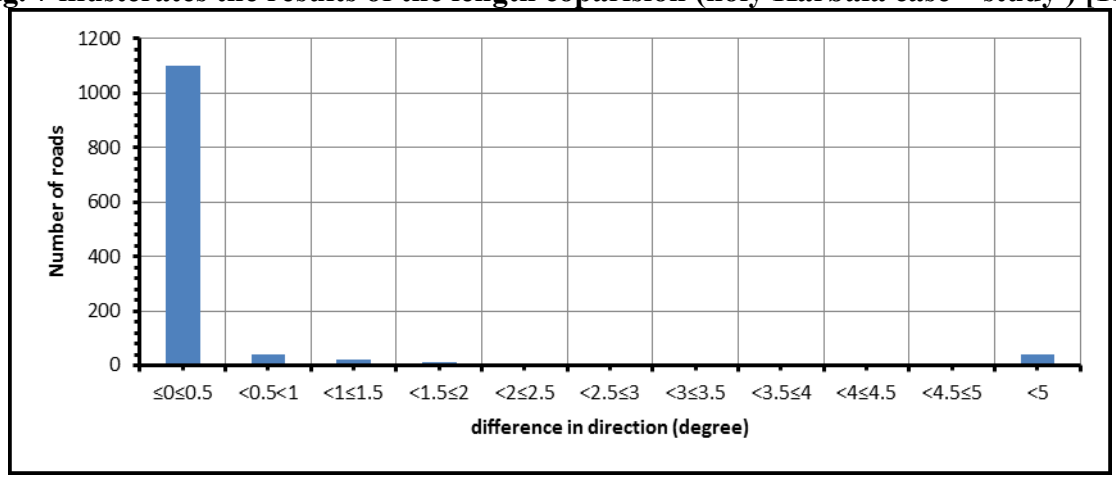


Fig. 8 illustrates the results of the direction comparison (holy Karbala case study) [11].

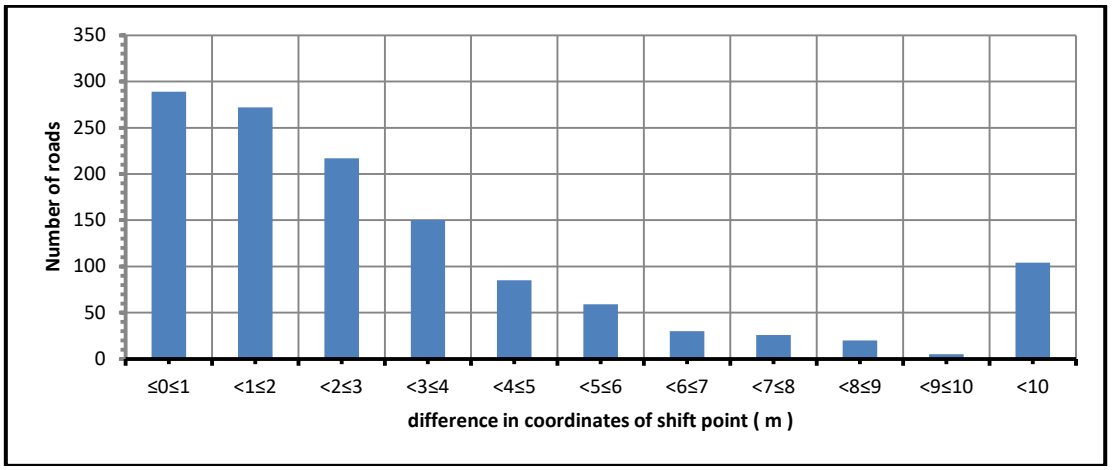


Fig. 9 illustrates the results of the comparison in shift of start point (holy Karbala case study) [11].

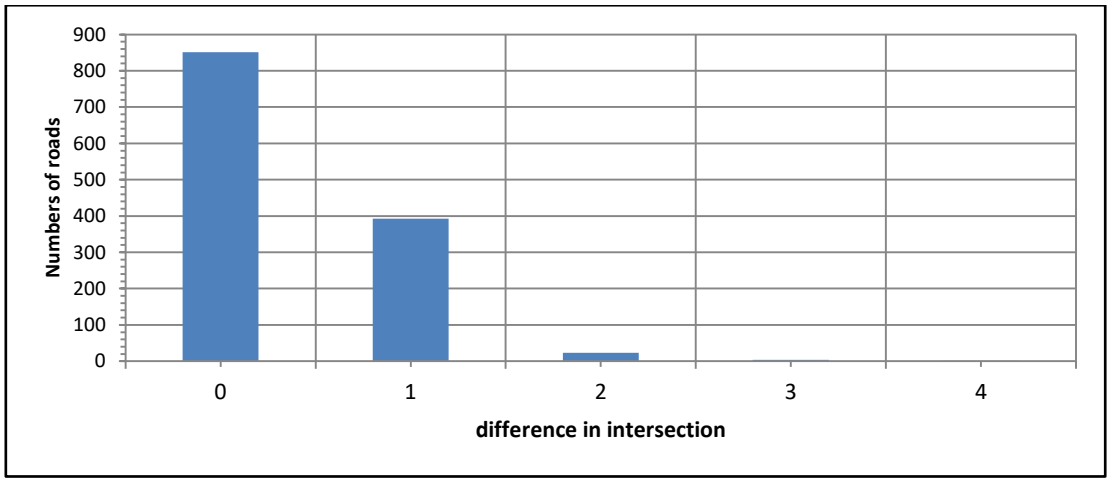


Fig. 10 illustrates the results of number of intersections comparison holy Karbala case study) [11].

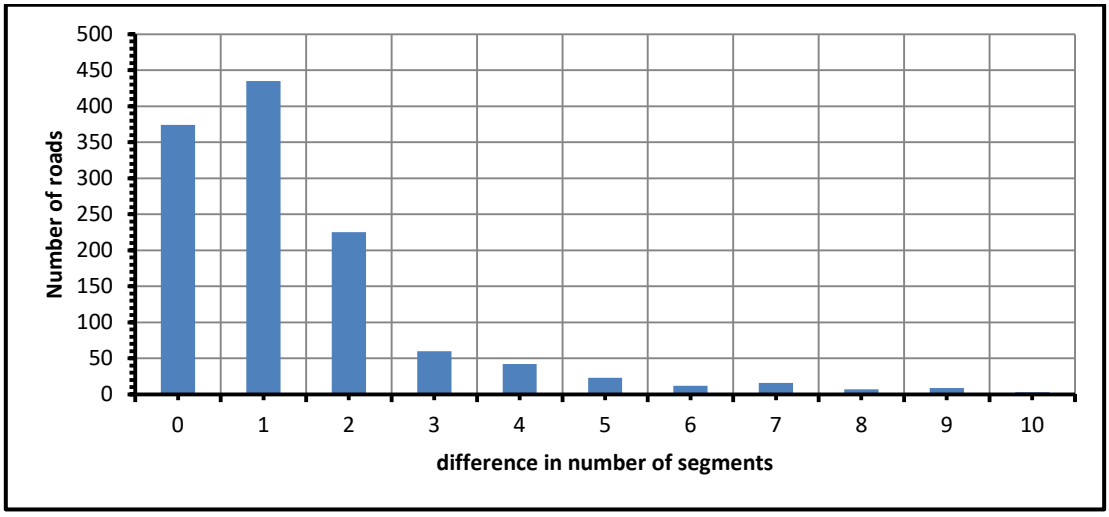


Fig. 11 illustrates the results of the number of segments comparison (holy Karbala case study) [11].

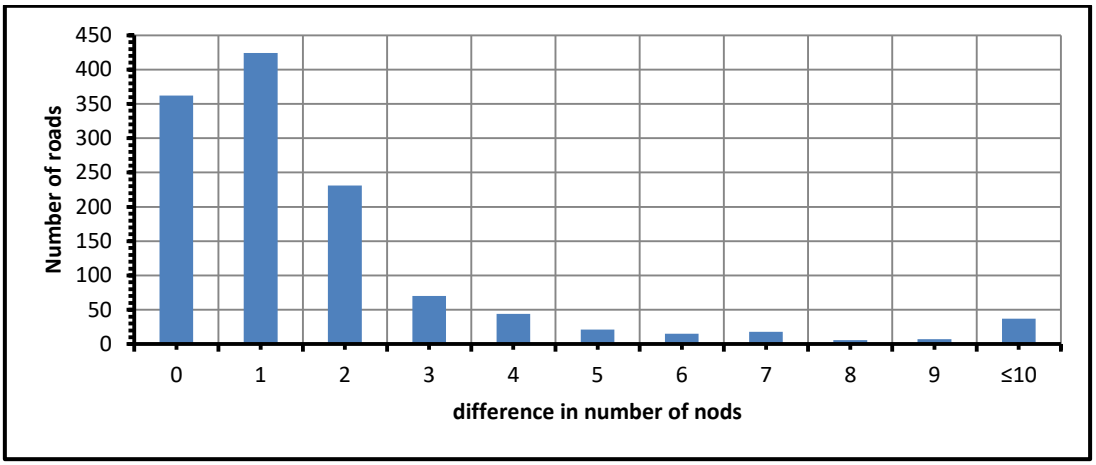


Fig. 12 illustrates the results of the number of nods comparison (holy Karbala case study) [11].

The same results and charts for Baghdad case study and same procedures can be followed for each condition to calculate the differences in (the lengths , the shifting in coordinates, the number of nodes, the numbers of segments, and the numbers of intersections) each one alone . For example the Fig. (13) illustrates the number of tested roads of Baghdad when the differences in lengths are specified in a certain period of 1 meter. The program gives (291) roads through the period bounded between zero and one meter (≤ 0 ≤ 1) ,while the program giving 252 roads a period that are larger than one meter and smaller or equal than 2 meter (<1 ≤ 2) . For the period between two and three meter, the program gives 215 (< 2 ≤3) .The process continues until they get last period the program gives 1376 (≤ 12). The same procedure has been followed in figur (14) . The number of roads when the differences in directions between zero and half degree (0 – 0.5) was (2450) roads , while in the peroid between (0.5 – 1) roads number (64). Fig. (15) illusteates roads number when differences in shifting of start point was (≤ 0 ≤ 10) meter. For example for the period (0 - 1) meter the number of roads was 401, and (1-2) the roads number was 499. Fig. (16) shows the number of roads when the differences in the number of intersections (0 , 1 , 2 , 3 , and 4) ,when the number of intersections = 0 , there is non- intersecting roads , in another meaning number of intersections = 0 give roads = 1605 , while the roads give numbers =1042 when the intersection = 1 ,while the program giving 48 when intersections= 2 . Number of roads illustrated in Fig. (17) when the differences in segments was (≤ 0 ≤ 10) . For example when the differences in segments number is zero, the number of roads was709, but when the differences in segments = 1, the number of roads (1069), while the roads give numbers =365 when the segments =3. Same

procedure was applied on Fig. (18) which calculates roads number when the differences in nods were (≤ 0 ≤ 10). For example, when the differences in number of nods= 0 , the roads =709 but when the difference = 1, the roads 294,while the differences =2 , the roads = 896 and so on .

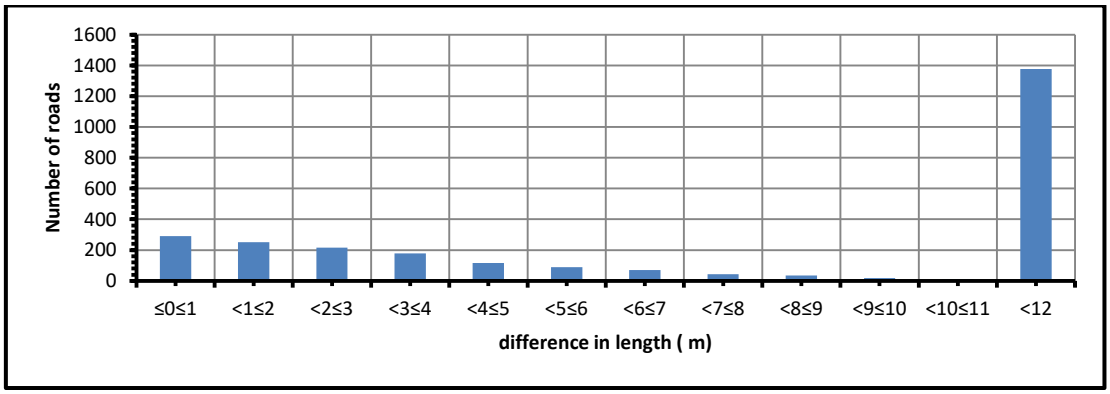


Fig. 13 illusterates the results of the length coparision (Baghdad case study) [11].

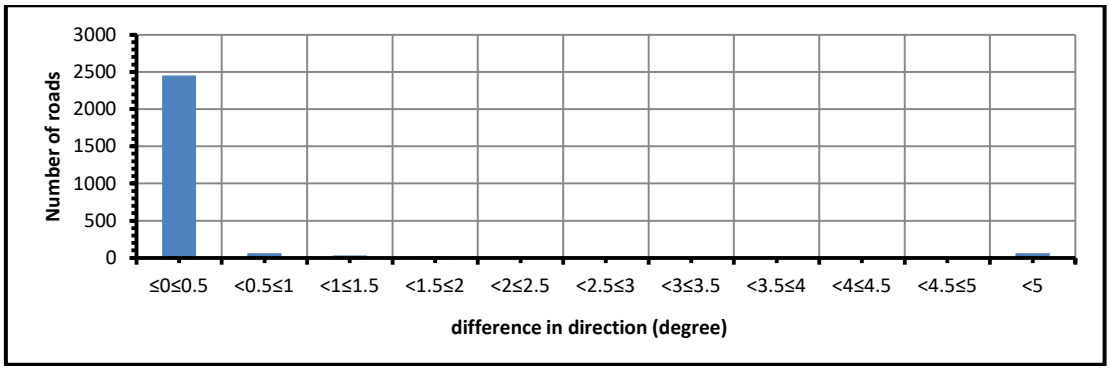


Fig. 14 illusterates the results of the dierctions coparision (Baghdad case study) [11].

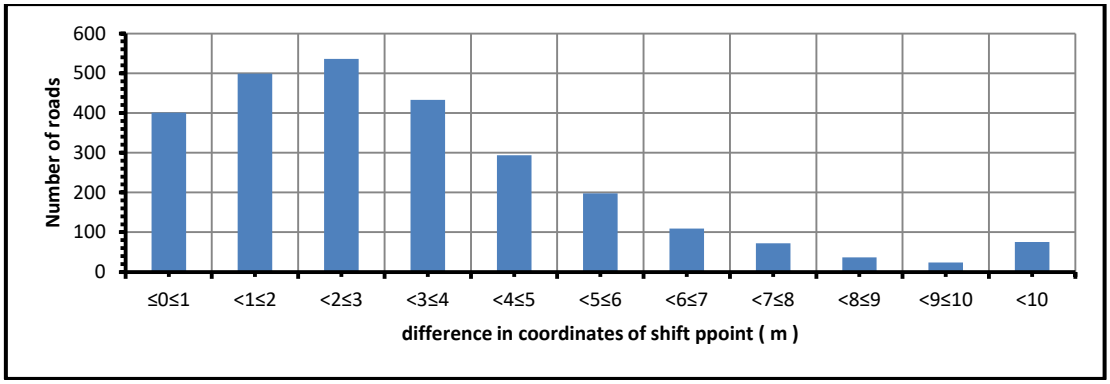


Fig. 15 illusterates the results of the comparision in shift of start point (Baghdad case study)[11].

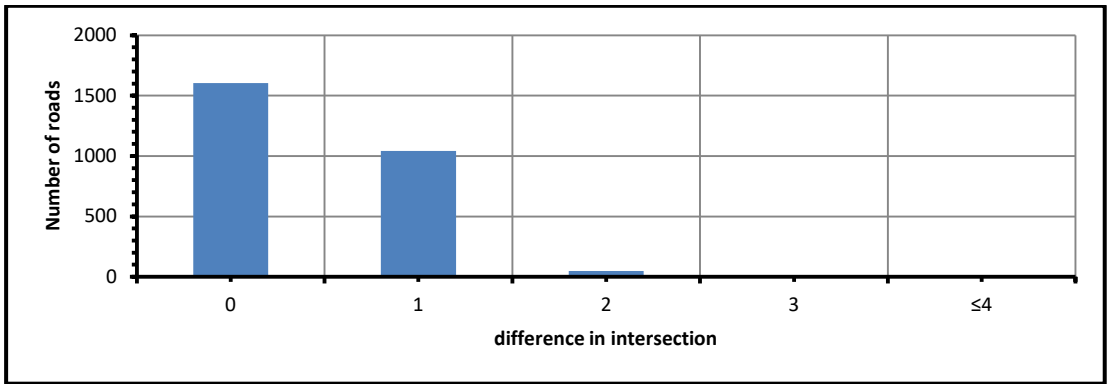


Fig. 16 illustrates the results of number of intersections coparision (Baghdad case study) [11].

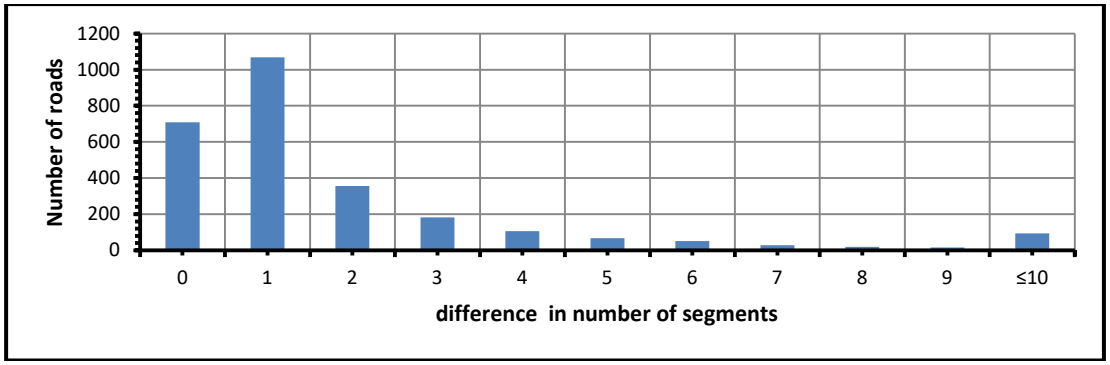


Fig. 17 illustrates the results of the number of segments comparison of (Baghdad case study) [11].

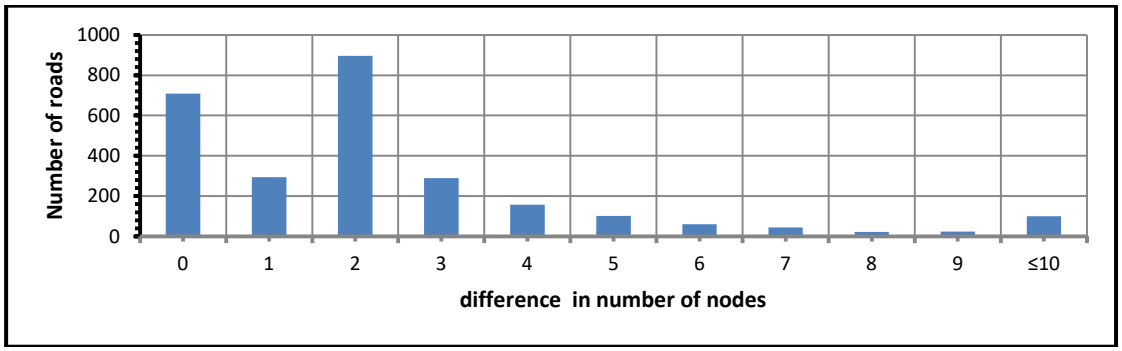


Fig. 18 illustrates the results of the number of nods comparison of (Baghdad case study) [11].

The number of road can be calculated for each of the six conditions, for each study area of the two case of the study. For example, to determine the numbers of the roads in which differences in the lengths or directions and other six conditions, certain specific period depending on the nature of these differences .The roads can calculate when the difference in (directions , shift start point , nodes ,segments ,intersections). As shown in the Fig.s above, which show the numbers of roads in areas of study in Baghdad and Holly Karbala. For example, Fig. (7) illustrates the numbers of roads of Holly Karbala when the differences in lengths are specific in a certain period of between (0-1) meter or between (1-2) meter for example .The program gives 226 through the period bounded between zero and one($\leq 0 \leq 1$) , while the program giving 166 roads to a period that are larger than one and

smaller or equal to 2 (< 1 ≤ 2) , The process continues for others period until they get less number of roads.

6. Conclusions

The aim of this study was to evaluate the quality of the VGI data. In particular the research assessed the matching of the OpenStreetMap (OSM) project. This study attempted to examine the matching of OSM data in Iraq. The study area is located on the center of Baghdad and the holy Karbala city and the area of each is approximately 36 km².

The methodology started by comparing the OSM data with the General Directorate of Surveying (GDS) data. The implementation of the methodology involved digitizing processes using Arc GIS software, reclassified classification of the roads, developing specific tool by Matlab programming language. In general, the results showed that OSM has very variable differences results which means that the utility of the data can be used with reliability for several cartographic purposes but not accurate business .The results showed that Baghdad data more accurate a certain degree.

There are some recommendations that could be drawn from this study. Firstly in relation to the OSM project it is clear that the map making guidelines should include detailed information about adequate mapping procedures so as to ensure fewer errors and avoid the homogeneity of the data. Also, possible future extension of this dissertation work would be to correlate these results with the mapping methodology that has being implied in each case; for example if it was done by using GPS tracks or by digitizing aerial imagery. If this kind of analysis reveals that a specific mapping technique gives more accurate results then it could be promoted by the OSM website that increasing the quality results of the project.

7. References

1. Al-Bakri, M. and D. Fairbairn,2010, Assessing the accuracy of " crowdsourced" data and its integration with official spatial data sets, in In Proceedings of Accuracy 2010: Ninth International Symposium on Spatial Accuracy Assessment in Natural Resuorces and Enviromental Sciences 2010, UK, 20-23 July 2010: Leicester. p. pp.317-320.
2. Brovelly, Minghini, Moliniri and Mooney, 2015, AFOOSS 4G-based procedure to compare OpenStreetMap and authoritative road network datasets.
3. Cartwright, W., 2008, Google maps and mobile devices: can just one generic design work? (1808-0936): p. PP.215-222.
4. Fairbairn, D. and M. Al-Bakri, 2013, Using Geometric Properties to Evaluate Possible Integration of Authoritative and Volunteered Geographic Information. ISPRS International Journal of Geo-Information. 2(2).
5. Goodchild, M.F., 2007, Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. International Journal of Spatial Data Infrastructures Research. 2: p. pp.24-32.

6. Girres, J.F. and G. Touya,2010, Quality assessment of the French OpenStreetMap dataset. Transactions in GIS. 14(4): p. 435-459.

7. Kounadi, O.,2009 , Assessing the quality of OpenStreetMap data. Msc geographical information science, University College of London Department of Civil, Environmental And Geomatic Engineering.

8. Koukoletsos, T., M. Haklay, and C. Ellul, 2012 ,Assessing Data Completeness of VGI through an Automated Matching Procedure for Linear Data.Transactions in GIS. 16(4): p. 477-498.

9. Zielstra, D. and A. Zipf, 2010 , A comparative study of proprietary geodata and volunteered geographic information for Germany, in 13th AGILE International Conference on Geographic Information Science20.

10. Haklay, M., 2010, How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. Environment and Planning B: Planning and Design. 37(4): p. 682-70.

11. Researcher: Zuhair Yahay, Surveying department, College of engineering, University of Baghdad.

12. (http://geofabrik.de.com)

13. The General Directorate for Surveying (GDS), 2010

تقييم أمكانية تكامل البيانات المجانية لشبكات الطرق

م.د. ميثم مطشر شرقي

جامعة بغداد/كلية الهندسة/قسم هندسة المساحة

زهير يحيى صفوك

جامعة بغداد/كلية الهندسة/قسم هندسة المساحة

الخلاصة

لقد ظهرت في الأونة الاخيرة على شبكة الانترنت (WWW) كميات كبيرة من البيانات الجغرافية المكانية المتاحة مثل Open Source) وغيرها, في اغلب الاحيان تطلق تسمية OpenStreetMap , Yahoo imagery ...Google Map, data) على هذه الخدمات . ان البيانات المكانية المنتجة من مصادرها المختلفة تكون عادة بمستويات دقة متباينة بسبب اختلاف طرق جمع هذه البيانات . وهذا بدوره قد لا يلبي متطلبات المستخدمين للتطبيقات مختلفة. هذا النوع من البيانات المكانية يسمى بالمعلومات الجغرافية للمتطوعين (VGI Volunteered Geographic Information .ان مشروع خارطة الشارع المفتوح (OSM) OpenStreetMap هو خير مثال على هذه البيانات . يحتوي ملف ال (OSM) على عدة طبقات (layers) متاحة للمستخدمين تشمل عوارض مختلفة , مثل الطرق , الانهار , البنايات وغيرها من الطبقات .

تهدف هذه الدراسة الى امكانية مطابقة البيانات المتعلقة بالعوارض الخطية (الطرق) في العراق ولمنطقتي دراسة في بغداد وكربلاء , حيث سيتم مقارنتها مع مجموعة بيانات مرجعية (بيانات ذات دقة أعلى) والتي تم الحصول عليها من الهيئة العامة للمساحة / وزارة الموارد المائية. تم التقييم بالاعتماد على عناصر مختلفة مثل مقارنة الاطوال , الاتجاهات , عدد النقاط والقطع المكونة للطرق وأحداثيات نقطة البداية . بأستخدام الحسابات الاحصائية مثل حساب الانحراف المعياري، المعدل،اعلى قيمة ، واقل قيمة للفروقات تمت المقارنة وحساب الدقة . في هذا البحث أيضا تم تطوير برنامج باستخدام ال (Matlab) يستطيع معالجة , تحليل و تقييم جودة العوارض الخطية (الطرق) . استنادا الى بيانات مطابقة النتائج , يتم تحديد امكانية استخدام بيانات (VGI) المتاحة كمصدر بديل للبيانات الرسمية . أظهرت النتائج الاولية وجود تطابق جيد ومقبول في نتائج منطقتي الدراسة .تم تمثيل النتائج برسومات بيانية , حيث تعطي هذه الرسومات أعداد الطرق خلال فترات مختلفة كأن تكون كل نصف متر أو متر بالنسبة للأطوال وكل نصف درجة بالنسبة للأتجاهات وهكذا .حيث بينت النتائج أن هناك عدد كبير من الطرق خلال الفترات الأولى ولكلا منطقتي الدراسة. وهذا يعطي مؤشر بأن الأختلافات بين مجاميع البيانات المقارنة يكون قليل . .ومن خلال دراسة وتحليل النتائج الاحصائية التي تم الحصول عليها يمكن القول بأن منطقة الدراسة في بغداد اكثر دقة من منطقة