Analysing Building Shapes Quality of Collaborative Mapping

Dr. Maythm Al-Bakri
Lecturer
Department of Surveying
College of Engineering
University of Baghdad
Baghdad - Iraq

Abstract

The very fast developments of web and data collection technologies have enabled non-experts to collect and disseminate geospatial datasets through web applications. This new type of spatial data is usually known as collaborative mapping or volunteered geographic information VGI. There are various countries around the world could benefit from collaborative mapping data because it is cost free data, easy to access and it provides more customised data. However, there is a concern about its quality because the data collectors may lack the sufficient experience and training about geospatial data production. Most previous studies which have outlined and analysed VGI quality focused on positional and linear features. The current research has been conducted to investigate the quality of another feature type such as polygons (buildings) of collaborative mapping data. Two different VGI data sources have been tested: Google Maps and WikiMapia services. The VGI data was compared with reference data extracted from high resolution aerial image which was provided from General Directorate of Surveying. The suggested methodology based on applying several metrics and methods such as surface distance method, compactness, elongation, and ratio of areas computation. The polygon shape accuracy was analysed by comparing conventional statistical values such as mean, median, standard deviation, minimum, and maximum. The results indicated that there is no big difference between the shape similarities of collaborative mapping polygons. Hence, it can be used for several applications such as spatial data infrastructures (SDI) and urban planning.

Key words: google maps; wikimapia; shape accuracy; polygons; collaborative mapping

1. Introduction

As a technology has evolved; the Internet permits people not only to gain information but also allows users to create new content and new knowledge. This includes community based websites which is usually called the crowdsource system. Crowdsource refers to the way that large number of distributed people can work on the same project in a very powerful manner, creating something where the whole is more than sum of the parts. The online encyclopaedia Wikipedia is a good example of Crowdsource technique.

Wikipedia is approach where in principle anybody with a computer can edit or submit content. In Wikipedia there is no peer review, and the material is subject to editing by any user [3]. The concept of the web as platform was also adopted by geographers to establish the volunteered geographic information (VGI) [9].

The VGI is collaborative web-based efforts to collect, produce and disseminate free geospatial data provided voluntarily by individuals. The VGI sites allow user to share their own contribution by defining locations where certain features exist. Nowadays, there are a wide variety of programs exist to display online maps on mobile phones. Amongst supported devices are nearly all phones that can run Java for mobiles, as well as platforms such as Android, the iPhone, Windows Mobile. The various programs distinguish themselves according to key features like if they use raster maps or vector maps, if they need an Internet connection or can be used offline, if they support address search or advanced features like routing, thus creating for a large number of different needs for viewing maps. Some examples of VGI phenomenon are WikiMapia, Google Maps Maker, OpenStreetMap (OSM) [15].

During the last decade, the amount of VGI data sources continues to grow on the Internet [6]. This may be due to the fact that in many developing countries, there is a need for free geospatial data because authoritative data are not existing or not complete. These countries lack financial or technical conditions to produce spatial data in digital formats. Therefore VGI can provide a rich source for spatial which can support a wide range of applications such as monitoring environment, disaster management, tourist services [7]. In addition, the authoritative data is expensive and map production is costly in most countries around the world [4]. The VGI data can be captured using handheld GPS-devices or any smart phones, and it can be also digitised from free aerial images. Furthermore, gaining benefits from free software. It becomes possible to use or download open source GIS software to produce and upload free vector datasets. However, due to the VGI data are produced by volunteers from different background, the quality of VGI data is not standardised and needs to be assessed [10].

Different approaches are being developed by spatial researchers to assess VGI quality. For example, [11] presented different techniques to measure the OSM building completeness in Germany. The comparison involved reference data from national mapping agencies. Their results showed that the completeness of OSM data in

Germany in the federal states of North Rhine-Westphalia were 25% and 15% in Saxony in November 2011. In a study undertaken by [5] analysed the OSM building footprints. The evaluation involved position, completeness, shape and semantic accuracy. The OSM data was compared to German Authority Topographic–Cartographic Information System as official datasets. The main results from the paper indicated that the semantic accuracy and completeness have high quality. Also, the shape of OSM building footprints is very close to official data. However there is a difference four meters in positional accuracy of OSM data. Most recently [6] assessed the completeness of OSM buildings data in UK. Three different study areas were tested: London, Leeds and Sheffield / UK. Their findings proved that the quality of OSM buildings is variable and inconsistent within UK cities.

In most of the existing studies, the OSM buildings were evaluated using official dataset. The study reported here compares the polygons (buildings) quality of other collaborative mapping datasets such as WikiMapia (Wiki) and Google Maps (GM).

2. Data Sources

Two different datasets from collaborative mapping projects have been tested: WikiMapia (Wiki) and Google Maps (GM) services. WikiMapia is a web site enabling collective annotation of geographic satellite imagery, and is representative of similar efforts such as Google Earth and mash-ups created with web application programming interface (API) to mapping services. It is allowing users to add information in the form of a note to any location on Earth. Although registration is not required to edit or add data to WikiMapia, over 2,500,000 users from around the world currently are registered [16]. All content uploaded by users becomes the intellectual property of WikiMapia and available for non-comercial use through WikiMapia API [15].

WikiMapia allows any contributor to add a tag (placemark) to any location by marking out a polygon around the location and then providing a default language, title, description and one or more categories. Location tagging is fully multi-lingual, meaning that there is no need to create separate tags for different languages. WikiMapia also enables users to add different map features such as buildings, roads, railroads and rivers. The interface allows specifying the size of features and providing a brief description and coordinates about them, as explain in Fig. 1 [15]. In this study, the coordinates of tested features are displayed and recorded manually from the main window of WikiMapia.

Google map is another geospatial dataset that has been investigated and analysed in this research. Google map is one of the webs mapping service which was launched in 2005. The main basic interface of Google Maps is incredibly intuitive and straightforward. Users can select a map, move it around, and zoom in and out to find particular area that, as shown in Fig. 2. Google Maps shows country, state, and county boundaries when available; regions and locations; country names. It also provides street-level and building maps. The Google Maps enable you to overlay lines and polygons objects onto a map through Google Map Maker service. This information can be used to provide additional detail about the location that looking for. The overlay features are specified in terms of their location within the Google Maps environment [1].

Additional information is usually displayed in an information window, which is a sort of speech bubble that is overlaid on the map. The information window usually appears in response to the user selecting a marker, either by clicking the marker in the map or by clicking a separate list of markers that is also displayed on the page. Google Maps also offer coordinates for places found in Google map [1]. The positional data coordinates can be extracted directly from the interface of Google Map service which is usually in World Geodetic System (WGS84).

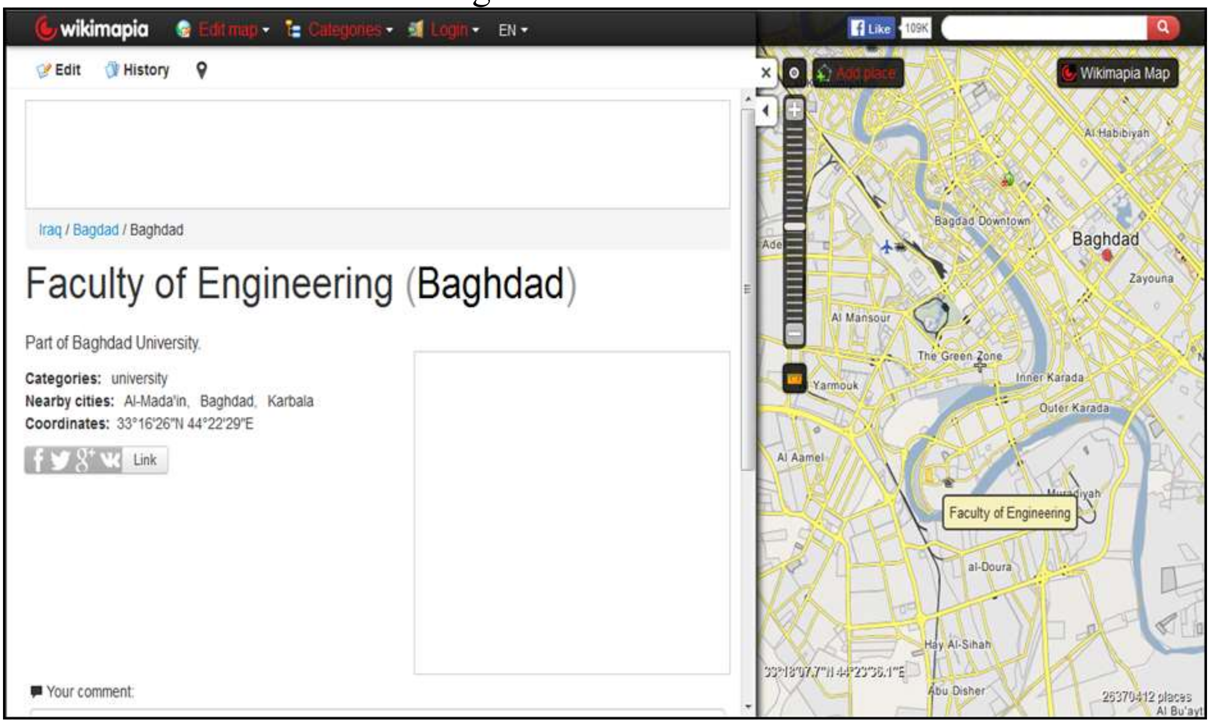


Fig 1. Screen shot of Wikimapia service, showing online information including map, place name, coordinates (http://wikimapia.org), Date: 20-07-2016.

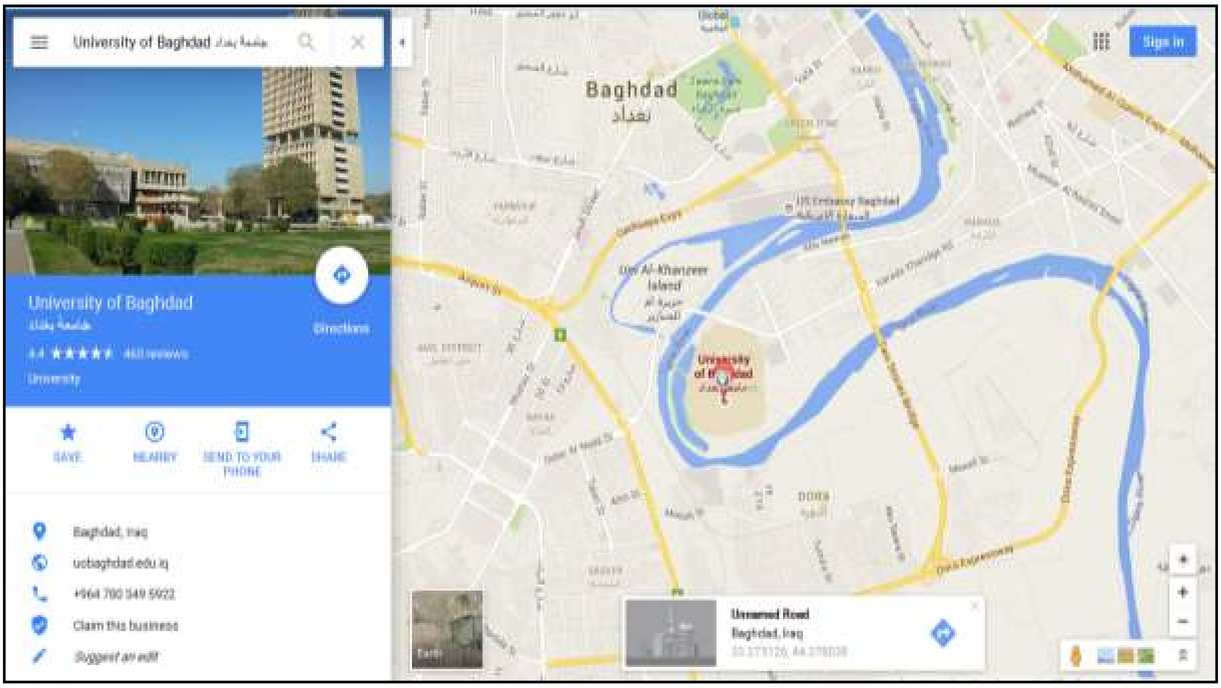


Fig 2. Screen shot of Google Maps showing online information including map, place name, coordinates, photo of place (https://www.google.com), Date: 20-07-2016.

In order to estimate the shape accuracy of tested data (WikiMapia and Google Maps), it was compared with self-generated dataset of higher accuracy. The digitized data will be used to create a definitive reference dataset (RD). The comparison will consist of an accurate vector dataset produced using high resolution (10 cm) aerial photo from general directorate of surveying. The compared datasets is composed of fifty polygons (buildings) were selected in the centre of Baghdad city. The intention is to compare the data quality of these datasets by applying the methods mentioned in section 3. Initially this can be done with visual comparison of derived maps, to get a general picture but the methods indicated below will be applied to construct a quantitative approach, identifying the strength of different datasets.

3. Methodology and Procedures

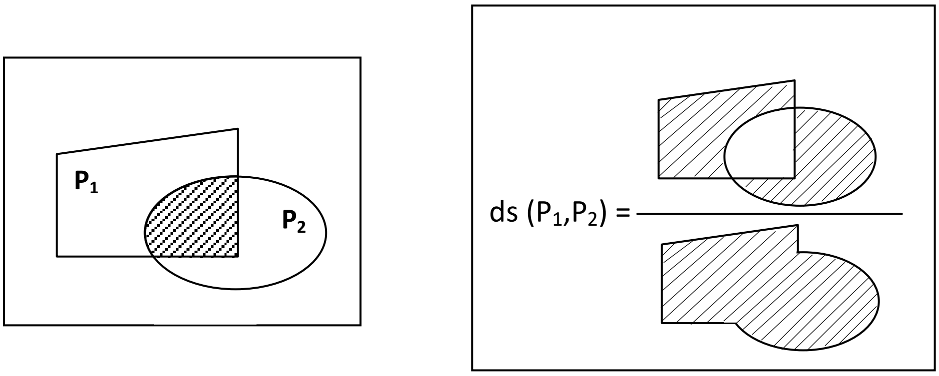
After presenting the main characteristics of the data sources in previous section, we now move to describe the overall procedures and methods which have been adopted in this study. Many methods for determining the similarity between polygons have been investigated and introduced. In the preliminary analysis, the function of the surface distance operation was applied as follows [8]:

$$ds = 1 - \frac{S(P_1 \cap P_2)}{S(P_1 \cup P_2)} \qquad (1)$$

Where: P_1 and P_2 are two different polygons.

ds: Surface distance value

In this method, the surface distance between polygons can be determined by dividing the intersection of polygons on union of them, as shown in equation 1. The value of surface distance is defined in the interval (0 and 1). The two polygons can be considered as similar, if the value of the distance is close to zero, while the two polygons are disjoint, if the distance is close to one. Fig. 3 shows the basic elements of the surface distance method [8].



(a) **(b)**

Fig 3. The surface distance method

The comparison of polygonal similarity also involved compactness measurements. The compactness can be defined as a numerical quantity representing the degree to which a shape is compact [12]. In GIScience, the shape compactness analysis has been a long standing for studying urban sprawl, city planning and describing the hydrological properties of drainage basins [2]. In literature, several methods for compactness measurement have been introduced; see for example [13]; [2]; [8]. The [13] method is adopted in this study because it can be considered as the most common method for measuring compactness. The compactness is determined as a ration between the area and perimeter of the polygon, as follows:

$$C = \frac{Area}{(0.282 \times perimeter)^2} \qquad (2)$$

Where: C is compactness value.

Elongation was also determined as another metric of shape quality measurement. It can be calculated by dividing the width and the length of the smallest rectangle that contain the polygon, Fig. 4. The values of elongation are in the range of (0-1). The value close to zero means that the shape is symmetric in all directions such as square while elongation value close to one denotes that the shape is elongated in one or more directions [14]. The equation for elongation can be expressed as:

$$E = 1 - \frac{W}{L} \quad (3)$$

Where:

W : is the width of the smallest rectangle containing the polygon.

L : is the length of the smallest rectangle containing the polygon.

E : is the elongation value.

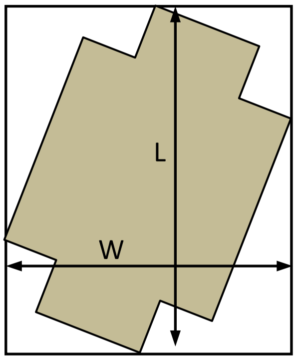


Fig 4. Elongation concept

Additional method for measuring similarity between polygons is the ratio of areas (R) computation. The ratio of area is computed as a ratio of the area of polygon in first dataset and the area of the same polygon in another dataset. The equation of this descriptor is given as follows:

$$R = \frac{A_1}{A_2} \quad (4)$$

Where:

A_1 : the area of polygon in one dataset.

A_2 : the area of the same polygon in another datasets.

R : the ratio of the area of polygon value.

The first step in the methodology consisted in the extraction the coordinates of required polygons from tested datasets. The extracted position information was imported by ArcGIS to generate point layers for showing building locations as represented in the three datasets: GM, Wiki and RD. The resulting points map was subsequently connected and joined with stratified closed polygons. In the following step of the analysis, the geometric properties of polygons, such as areas, perimeters, length of sides, and intersection of area, were determined. To be able to determine the shape similarity based on the geometric properties of tested datasets, MatLab programming language was employed to design a specialized program to perform this task. The designed program has the ability to compute and compare the values of surface distance method, compactness, elongation, and ratio of areas computation through statistical analysis. The methodology followed in this research can be shown in Fig. 5.

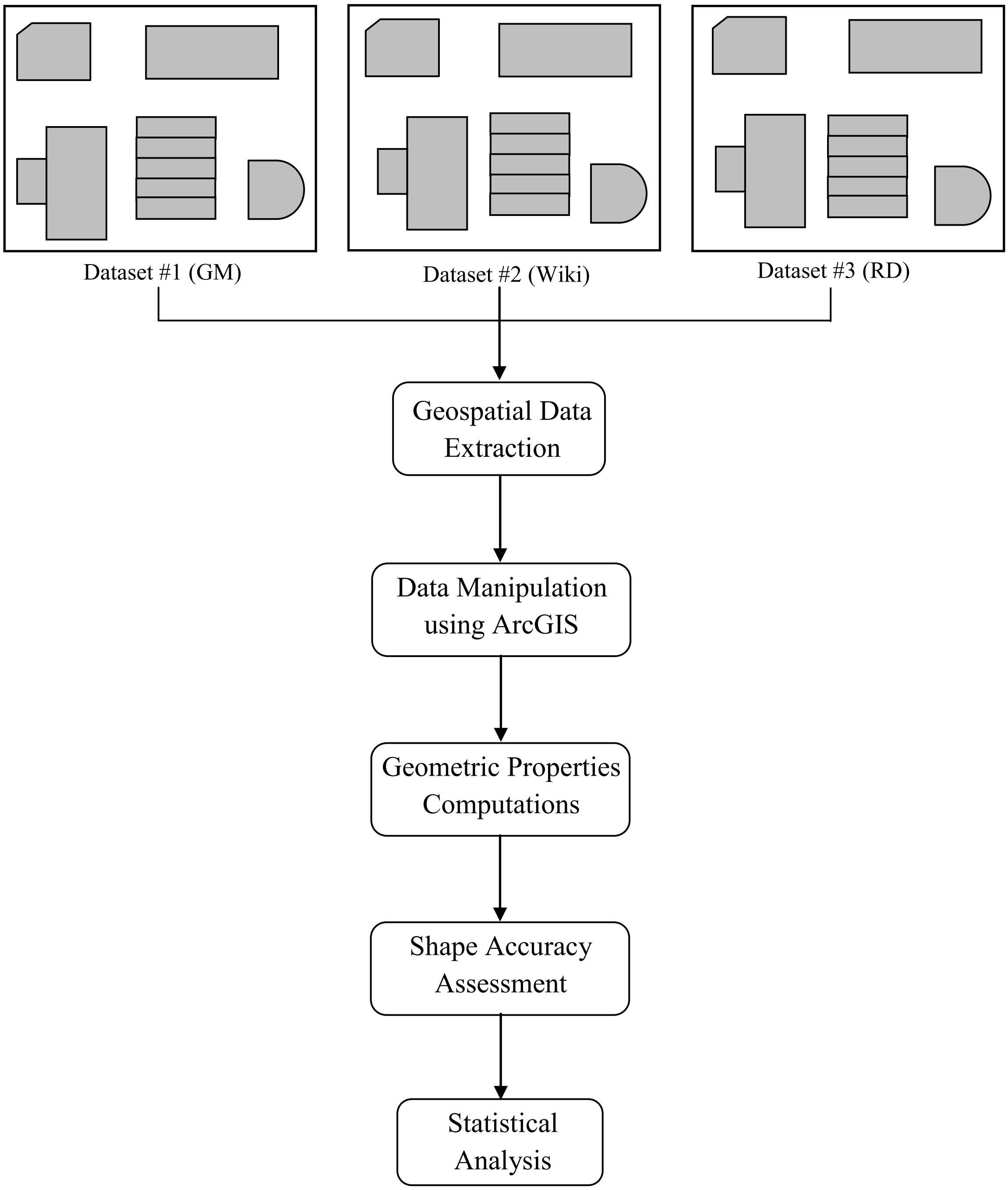


Fig 5. Flow diagram of the shape similarity process

4. Quality Assessment of Collaborative Mapping Buildings

Using the surface distance method described in section 3, the accuracy of tested buildings was computed. Figure 6 shows the distribution of the output of the surface distance method. It is clear that the compare data are close to each other, although the frequency of GM/RD data is a little more in the first value. The concentration peak is between 0.05

and 0.35, which corresponds small difference. This is also confirmed from the descriptive statistics in table 1. The mean values for the surface distance method of all tested data (i.e. buildings mapped from WikiMapia and Google Maps) are less than 0.5 which means that the differences between output values are small.

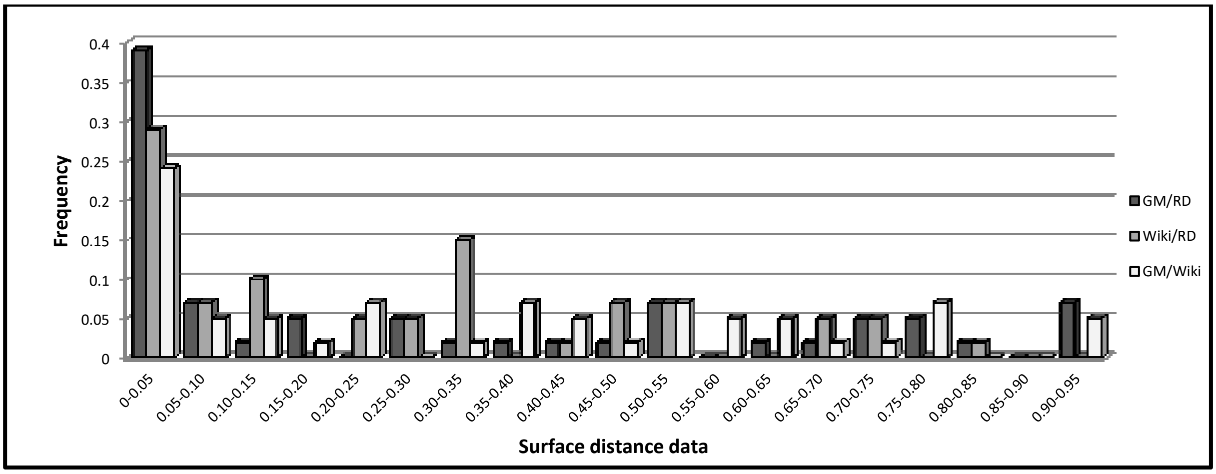


Fig 6. The distribution of the surface distance output for the tested datasets

Table 1. Descriptive statistics of surface distance method

	GM/RD	Wiki/RD	GM/Wiki
Max	0.9503	0.8124	0.9231
Min	0.0003	0.0020	0.0011
Mean	0.3320	0.2692	0.4800
Medean	0.1536	0.2211	0.3571
SD	0.3794	0.2471	0.5946

The compactness vales are determined for each polygon of tested dataset. More than fifty percent of RD data has compactness value less than 0.30, and it is almost similar to the distribution of GM and Wiki compactness values as shown in Fig. 7. The computed compactness is almost consistent among tested datasets. The uniformity is presented in table 2 by the low mean of 0.2564, 0.2515, and 0.2601 for the RD, GM, and Wiki respectively. The compactness standard deviations were computed for each dataset with a range of 0.0975 and 0.1105. Therefore, in general all tested datasets have a very similar compactness values because the buildings have regular shapes and compact.

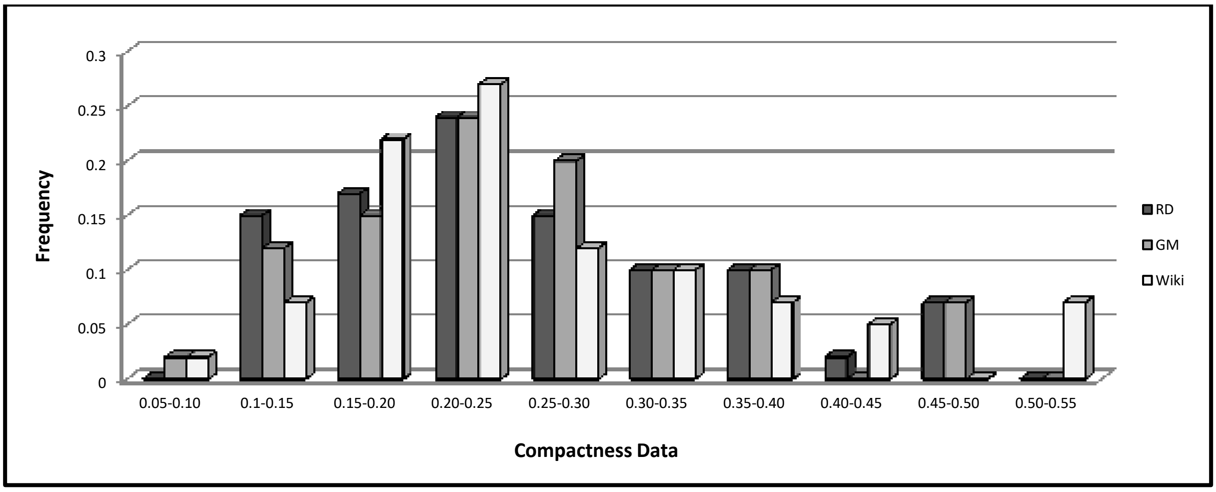


Fig 7. The distribution of the compactness output for the tested datasets

Table 2. Descriptive statistics of compactness method

	RD	GM	Wiki
Max	0.4686	0.4877	0.5358
Min	0.1072	0.0954	0.0847
Mean	0.2564	0.2515	0.2601
Medean	0.2287	0.2207	0.2346
SD	0.0975	0.0985	0.1105

As stated before, the elongation for each polygon was also determined and investigated. The elongation results of RD, Wiki and GM datasets are presented in table 3 and Fig. 8. The elongation standard deviations of the full sample (50 buildings) are 0.3013, 0.2784, and 0.2725 for RD, GM and Wiki respectively, with a range of 0.0169 to 0.9888. The data with lowest elongation mean is RD, and the highest is Wiki (0.4039 and 0.4340, respectively). The differences between GM and Wiki elongation are relatively small. The RD overall elongation values are slightly higher, which indicates that the RD polygons are a little elongated.

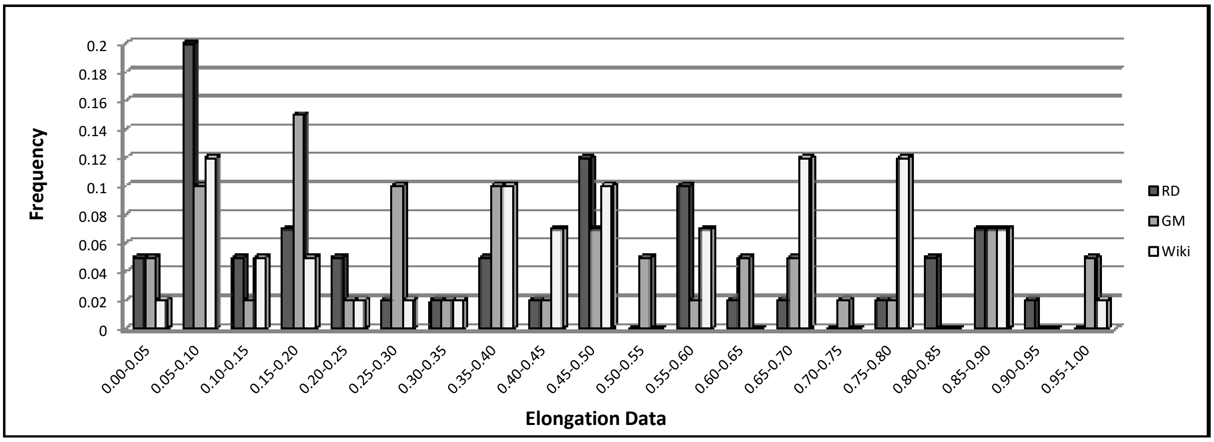


Fig 8. The distribution of the elongation output for the tested datasets

Table 3. Descriptive statistics of elongation method

	RD	GM	Wiki
Max	0.9564	0.9888	0.9658
Min	0.0427	0.0346	0.0169
Mean	0.4039	0.4112	0.4340
Medean	0.3902	0.3577	0.3568
SD	0.3013	0.2784	0.2725

Table. 4 and Fig. 9 compare the output results of the ratio of area method. It is clear from Fig. 9 that most ratios of area data are distributed between 0 and 1.8. This variety in values is reflected by the differences of standard deviations of 0.5007, 0.3463, and 0.7665 for GM/RD, Wiki/RD, and GM/Wiki respectively. Hence, it can be said that in general the areas of tested polygons tend to be different.

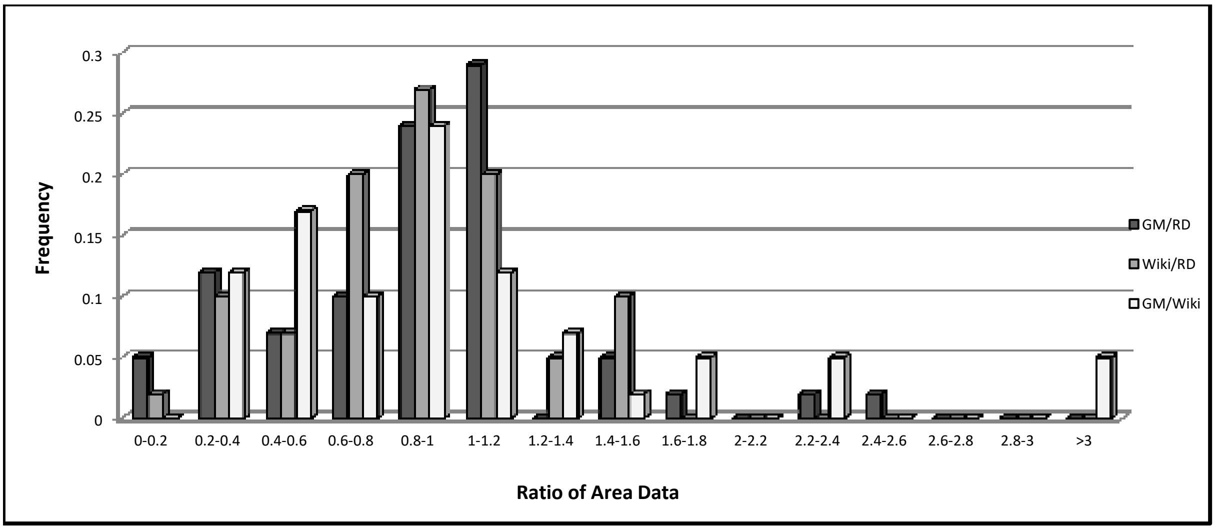


Fig 9. The distribution of the ratio of areas output for the tested datasets

Table 4. Descriptive statistics of the ratio of areas method

	GM/RD	Wiki/RD	GM/Wiki
Max	2.4503	1.5319	3.7131
Min	0.0676	0.1676	0.2355
Mean	0.9222	0.8774	1.0460
Medean	0.9900	0.9547	0.9564
SD	0.5007	0.3463	0.7665

5. Conclusion and Outlook

The VGI data or collaborative mapping is where volunteers, usually untrained, and they have different background and expertises, create spatial data based on web platform. Due to the lacking of standardization, the VGI quality is vary and heterogeneous from different data sources. Therefore, there is a need for better understanding of the quality of free geospatial data, in particular their accuracy. To the best of our knowledge, the majority of published research assessed the quality of OpenStreetMap (OSM) data as a common VGI data source. Here, this paper proposed a methodology for evaluating the quality of buildings on a further VGI services such as WikiMapia and Google Maps. Fifty polygons (buildings) were selected and examined in centre of Baghdad. The evaluation process involved several methods and techniques such as surface distance method, compactness, elongation, and ratio of area.

The findings showed that the polygons quality of different collaborative mapping datasets is almost similar, although there is a slight difference between statistical values of tested datasets. For instance, the mean values of compactness are 0.2564, 0.2515, and 0.2601 for RD, GM and Wiki data respectively. On the other hands, the mean values of surface distance method are 0.3320, 0.2692 and 0.4800 for GM/RD, Wiki/RD and GM/Wiki comparison respectively. Therefore, this kind of datasets can be used to enrich and improve each other. In addition, free geospatial data can also aid the completion of geospatial datasets in some areas around the world, especially in countries lacking full authoritative geospatial datasets. Collaborative mapping data can be utilised to develop spatial data infrastructure (SDI) or urban planning. Thus, policy and institutional issues could be addressed to demonstrate the advantages of the approach developed.

For further studies, the above methodology can be transferred to multi data sources such as Bing Maps, Yahoo Maps, etc. This can be done by the development of a prototypical application that allows for location, selection and assessed shape accuracy in the web environment. Thus, web services need to be included in the data flowline. Subsequently, refinements and feedback will be undertaken to the framework. Another possible area of future research would be to investigate the creation of a quality evaluation index which may assist to determine the extent to which dataset can be used for a range of purposes. Issues of the creation of a universal mathematically derived index need to be addressed as well.

References

[1] Brown, M., (2006), *Hacking Google Maps and Google Earth*, Wiley Publishing, Inc., Indianapolis, Indiana, USA.

[2] Chandra, S., Chhetri, P. and Corcoran, J., (2009). *Spatial patterns of urban compactness in Melbourne: an urban myth or a reality*. In: Ostendorf, B., Baldock, P., Bruce, D., Burdett, M. and Corcoran, P., eds. Proceedings of the Surveying and Spatial Sciences Institute Biennial International Conference, Adelaide, Australia, p.p. 231-242.

[3] Crampton, J., (2010), *Mapping A critical Introduction to Cartography and GIS*, John Wily & Sons Ltd, West Sussex, United Kingdom.

[4] Elwood, S., Goodchild, M. and Sui, D, (2012), *Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice*. Annals of the Association of Geographers 102(3). p.p.571-590.

[5] Fan, H., Zipf, A., Fu, Q., and Neis, P., (2014), *Quality assessment for building footprints data on OpenStreetMap*, International Journal of Geographical Information Science, 28 (4), 700-719.

[6] Fram, C, Chistopoulou, K, and Ellul, C, (2015), *Assessing the quality of OpenStreetMap building data and searching for a proxy variable to estimate OSM building data completeness*, GISRUK 2015 conference, Leeds, UK, p.p. 1-11.

[7] Genovese & Roche, (2010), *Potential of VGI as a resource for SDIs in the North/South context*, Geomatica, 64 (4), p.p. 439-450.

[8] Girres & Touya (2010). Quality assessment of the French OpenStreetMap dataset. Transactions in GIS 14(4). p.p. 435-459.

[9] Goodchild, M. F., (2007), *Citizens as Sensors: The World of*

Volunteered Geography, GeoJournal, 69, 211–221.

[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), PP. 682–703.

[11] Hecht R, Kunze C, Hahmann S,(2013), *Measuring completeness of building footprints in OpenStreetMap over space and time*, ISPRS Int J Geo-Information 2, p.p.1066-1091.

[12] Li, W., M. F. Goodchild, and R. L. Church. (2013). A*n Efficient Measure of Compactness for Two-Dimensional Shapes and Its Application in Regionalization Problems*. International Journal of Geographical Information Science 27(6), p.p. 1227–1250.

[13] Maceachren (1985). Compactness of geographic shape: Comparison and evaluation of measures. Geografiska Annaler 67(1). p.p 53-67.

[14] Mingqiang, Kidiyo & Joseph (2008). A survey of shape feature extraction techniques. Pattern Recognition p.p.43-90.

[15] Surhone, L., Tennoe, M., Henssonow, S., 2010, *Volunteered Geographic Information*, VDM Publishing house Ltd., Beau Bassin, Mauritius.

[16] Wikimapia Statistics, (2016), available at http://wikimapia.org/wiki/Wikimapia_Statistics [Accessed 10 April 2016]

تحليل جودة أشكال البنايات في الخرائط التعاونية

د. ميثم مطشر شرقي
مدرس
قسم هندسة المساحة /
كلية الهندسة - جامعة بغداد
بغداد – العراق

الخلاصة

لقد مكنت التطورات السريعة في تقنيات الحواسيب والانترنيت الغير المختصين من جمع ونشر البيانات الجغرافية على الانترنيت بكل سهولة وبدون تعقيد. ان هذا النوع الجديد من البيانات عادة مايعرف بالخرائط التعاونية او المعلومات المكانية التي ينتجها المتطوعون. هناك العديد من البلدان حول العالم ممكن ان تستفاد من هذه البيانات او المعلومات وذلك لكونها مجانية وتوفر المزيد من البيانات المكانية وحسب الطلب ، خصوصا في تلك الدول التي لاتتوفر فيها خرائط رسمية معتمدة ومنتجة من الموسسات المختصة. مع ذلك فان هناك قلق متزايد حول نوعية هذه المعلومات ودقتها لكون الاشخاص المتطوعين لجمع هذه المعلومات ونشرها قد تعوزهم الخبرة والاختصاص في هذا المجال. ركزت معظم الدراسات السابقة على دراسة وتحليل دقة العوارض النقطية والخطية لهذا النوع من البيانات. تهدف الدراسة الحالية الى تقييم جودة نوع اخر من العوارض وهي المضلعات (المباني) لنوعين من بيانات الخرائط التعاونية Google Maps , WikiMapia. لقد تمت مقارنة هذه البيانات مع بيانات مرجعية مستخلصة من صور جوية عالية الوضوح مجهزة من الهياة العامة للمساحة . ان المنهجية المقترحة لهذا البحث تعتمد بشكل اساسي على تطبيق عدد من الموديلات والاساليب الرياضية مثل ratio of areas ,elongation ,compactness ,surface distance method. لغرض تقييم وتحليل دقة المضلعات المطلوبة فقد تم حساب ومقارنة القيم الاحصائية التقليدية مثل الوسط الحسابي ، الوسيط ، الانحراف المعياري ، الحد الادنى ، الحد الاعلى للنتائج المستحصلة من تطبيق الموديلات الرياضية السابقة. أشارت نتائج هذه الدراسة الى وجود فروقات بسيطة بين دقة اشكال المضلعات (البنايات) المقارنة وعليه يمكن توظيف هذه البيانات في تطوير البنى التحتية للمعلومات المكانية او تطبيقات التخطيط الحضري والاقليمي.

الكلمات المفتاحيه:
خرائط الكوكل ، دقة الاشكال ، المضلعات ، الخرائط التعاونية .