Building-Level Comparison of Microsoft and Google Open Building Footprints Datasets

Jack Joseph Gonzales
Geospatial Science and Human Security Division, Oak Ridge National Laboratory, TN, USA

Abstract

Large-scale datasets of building footprints are a crucial source of information for a variety of efforts. In 2023, the general public benefits from open access to multiple sources of building footprints at the country scale or larger, such as those produced by Microsoft and Google. However, none of the available datasets have attained complete global coverage, and researchers and analysts may need to combine multiple sources to assemble a complete set of building footprints for their area of interest or choose between overlapping sources, requiring an understanding of the differences between different building sources. This paper presents a method to closely examine the quality of different building footprint sources by matching corresponding buildings across datasets, using building footprints in Ethiopia published by Microsoft and Google as an example set.

2012 ACM Subject Classification Computing methodologies

Keywords and phrases Open data, Building footprints, Data comparison

Digital Object Identifier 10.4230/LIPIcs.GIScience.2023.35

Category Short Paper

Acknowledgements The author gives thanks to Daniel Adams and Jessica Moehl for their thoughtful review and advice.

1 Introduction

Among many data resources characterizing the built environment, building footprints have proven to be extremely useful for a wide variety of purposes, from general public use mapping services like OpenStreetMap, to population modeling efforts such as WorldPop and LandScan [10, 1, 11]. At large scale, these building footprints are typically derived from satellite imagery via automated machine learning models, e.g. [14, 13, 12, 8, 5], or using volunteers to manually map out building footprints as in the case of OpenStreetMap [11].

Microsoft and Google have both released expansive datasets of building footprints for use by the general public, providing researchers and analysts with massive datasets covering multiple continents and growing. In addition to their 1.2 billion building dataset covering Europe, much of the Americas, Africa, and Asia, Microsoft has released several independent country-scale datasets, such as the 2018 dataset for the United States [8]. The Google Open Buildings dataset began with a near-complete mapping of buildings in Africa, and has since expanded to parts of Asia and the Americas to include 1.8 billion buildings [5]. Both

12th International Conference on Geographic Information Science (GIScience 2023).
Editors: Roger Beecham, Jed A. Long, Dianna Smith, Qunshan Zhao, and Sarah Wise; Article No. 35; pp. 35:1–35:6
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Microsoft and Google identify buildings using convolutional neural network-based semantic segmentation models to classify pixels in high-resolution satellite imagery as building or non-building, and then generate building footprint polygons from the positively classified pixels [8, 5, 12].

Other large-scale datasets exist as well, such as EUBUCCO v0.1, which aggregates and harmonizes data from 50 sources to build a dataset of over 200 million buildings for the European Union [9]. OpenStreetMap utilizes a vast number of volunteer analysts to manually map out buildings, providing a good alternative to machine learning-based datasets, albeit very labor intensive to develop and ensure quality, and often lacking in completeness [2, 15, 3, 6].

While all these datasets provide an excellent data resource, they vary in quality and completeness, sometimes requiring multiple sources to be used to completely cover an area of interest. In order to effectively use and integrate data from different sources, effort must be made to understand and account for systemic differences between building footprints from each source. This study presents a framework for comparing one dataset against the other based on matching building footprints from Microsoft and Google.

2 Methods

Study area

Two small areas of interest (AOIs) were selected: one from a densely built urban area and another from a low-density rural area. The urban AOI is in the eastern part of Addis Ababa, Ethiopia's capital city, covering roughly 108 hectares and including a good representation of building types found throughout the city. The rural AOI is located in the Amhara region, about 175 kilometers northeast of Addis Ababa, and is dominated by agricultural land with small villages and clusters of buildings scattered about. Examples of the settlement patterns in the AOIs can be seen in the imagery in Figure 1. Like many areas in the world, these AOIs are relatively data poor, with little to no data available other than machine-generated datasets. These two contrasting areas were chosen to evaluate the datasets in a variety of conditions, since settlement patterns heavily differ between urbanized and rural areas, placing different demands on building extraction models. Although small, these AOIs provide a good proof of concept in anticipation of larger-scale comparison efforts.

Data

Building footprints data were sourced from Microsoft's Global Building Footprints and Google's Open Buildings datasets. In addition to footprint geometry, Google provides a confidence value with each footprint, along with guidelines on suggested confidence thresholds to achieve 80%, 85%, or 90% precision. This confidence value allows Google to include many more geometries in their data, many of which may be false detections that can be filtered out using the prescribed confidence thresholds, especially in areas where natural building materials are common and buildings can often be confused with rocks and other landscape features [12]. For this study, we only used those geometries that meet the 90% precision confidence threshold. Microsoft does not report confidence values, but reports that their data achieves 94.4% precision in Africa. Microsoft and Google both report roughly 70% recall [8, 5].

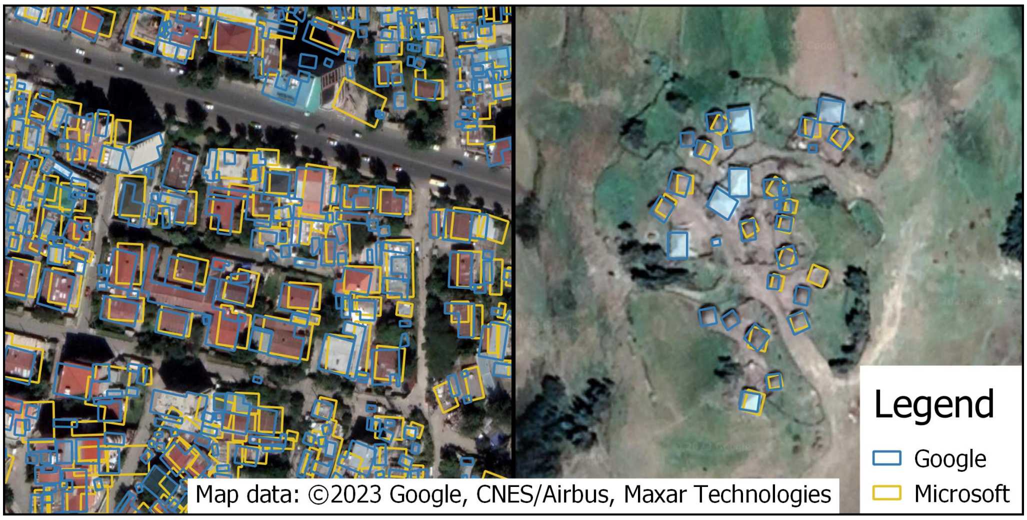

Figure 1 Typical settlement patterns and building footprints in the two study areas, with the urban area on the left and the rural area on the right, overlaid on Google Maps satellite imagery [4].

Comparison

This study seeks to compare matching building footprints from both Microsoft and Google. As such, the initial step is to pair each footprint in one dataset to the foorprint(s) that represent the same building in the opposite dataset. For each building in one dataset, matches were identified by identifying all footprints in the opposite dataset that overlap by at least 30% of the area of the smaller geometry, using a similar minimum overlap threshold to Fan et al. 2014 [3]. Individual building footprints may have multiple matches, especially in dense urban areas, where the Microsoft and Google models may disagree on where to divide buildings that are adjacent or have complex, disjointed roofs.

Matched buildings were compared based on area differences and the number of matches found in the other dataset. The number of matches describes the semantic similarity of building detection, or the models' agreement on how to divide complex and adjacent buildings, and can be expressed as a ratio of the number of building footprints in one dataset to the number of corresponding footprints in the other. Possible semantic similarity ratios include 1:1 similarity, where a building matches with exactly one footprint footprint in the other dataset, 1:0 if there is no match, 1:n if one building has multiple matches, m:1 if multiple buildings match one building, or m:n, where multiple buildings match with multiple other buildings [3]. In this study, only 1:1 and 1:n similarity ratios were considered, as other ratios demand a more complex analysis beyond the scope of a short paper, but are important to a complete and thorough examination of the differences between these two sources.

Area comparison is straightforward, taking the median area difference of corresponding footprints between the two datasets, as well as the percentage of buildings with a statistically significant difference from their counterpart in the opposite dataset. A threshold of 1.96 deviations from the median was used to identify values significantly different from the median. Median absolute deviation (MAD) was used to quantify data dispersion as it provides a much more intuitive description of data deviation than the traditional standard deviation [7].

In addition to metrics describing matched building footprints, aggregated statistics describing total number of buildings, percentage of buildings with at least one match, and total and average building area were used to further compare datasets and contextualize statistics of matched buildings.

3 Results

Aggregated statistics

Aggregated statistics shown in Table 1 reveal differing trends for the urban and rural study areas. In the urban area, Microsoft produced fewer, larger building footprints with greater total area, while Google produced more, smaller footprints covering less total area. In both datasets, the majority of buildings had at least one matching footprint.

In the rural area, Microsoft produced far fewer footprints than Google, totalling just 58% of the total area of Google. However, less than half of Google's buildings had a match in Microsoft, whereas 72.5% of Microsoft's buildings had a match. In addition, both produced similar sized footprints on average.

Table 1 Aggregated statistics of each sample set in both urban and rural study areas.

Aggregated Statistics				
Dataset	Total buildings	Percent matched	Total area (ha)	Mean building area (m^2)
Microsoft (Urban)	2,628	66.67	35.03	133.28
Google (Urban)	3,194	72.94	21.68	68.86
Microsoft (Rural)	1,942	72.50	7.02	36.13
Google (Rural)	3,094	46.19	12.11	39.13

Matched building statistics

In the urban area, Google buildings tended to be smaller than their matches in the Microsoft dataset, with a very high MAD, and very few buildings with more than one match, while Microsoft buildings had a higher average number of matches. Both datasets contained similar percentages of buildings with an area significantly different from the median difference.

In the rural area both Microsoft and Google had very similar results, with few buildings matching with more than one other, and Microsoft buildings running slightly smaller than their Google counterparts. MAD for both were nearly identical and far lower than in the urban area. Similar to the urban area, Microsoft buildings had a slightly higher percentage of buildings with a significant area difference.

Table 2 Statistics comparing buildings with their matched counterparts in the opposite dataset.

Matched Area Statistics				
Dataset	Median area difference (m^2)	Area Difference MAD (m^2)	Percent significant difference	Mean number of matches
Microsoft (Urban)	15.96	42.13	15.60	1.39
Google (Urban)	-62.47	131.86	10.48	1.06
Microsoft (Rural)	-2.32	6.65	16.48	1.02
Google (Rural)	2.03	6.55	10.63	1.00

4 Discussion and Conclusion

In the rural study area, matched buildings are remarkably similar, however the aggregated statistics show that Google detected far more buildings, and thus greater total building area. Although many of these buildings have no match in the Microsoft dataset, both datasets report at least 90% precision and roughly 70% recall, indicating that this discrepancy is most likely predominantly due to different imagery dates and new construction, allowing Google to detect buildings that simply did not exist in the imagery used by Microsoft [5, 8]. This is supported by inspection of Google and Bing satellite maps, with Google imagery appearing to be more recent.

In the urban study area, matched area differences in both datasets show large dispersion, likely due to difficulty in matching the correct buildings with one another. Correctly matching buildings becomes very difficult where imagery is misaligned or models disagree on where to divide and separate buildings. This can be seen on the left side of Figure 1, where overlapping footprints are often very different, as opposed to the rural area on the left where they are very similar. Microsoft tends to generate larger footprints that may encapsulate multiple buildings under a single footprint, while Google tends to break buildings up into smaller polygons, potentially dividing a single complex building into multiple parts. This led Microsoft to generate a larger total building area with fewer buildings, which can be seen Table 1. This discrepancy in polygonization also leads to poor matching results, as small Google footprints may match with a large Microsoft footprint that may completely envelope several Google buildings, leading to the large area difference and dispersion shown in Table 2.

Conclusions

By examining individual building footprints, one can gain a much more in depth understanding of the differences between two data sources that both seek to describe building footprints. This study demonstrates a framework for evaluating differences between two similar sets of polygons, which is crucial for integrating data from multiple sources. It is important to note that neither of these datasets can be considered absolute truth, and rather than determine accuracy, this workflow is designed to characterize differences to assist analysts in integrating or choosing between multiple available data sources. Analysis shows that in the rural area, the Microsoft and Google datasets are very similar where they are able to detect the same buildings, but it is likely that differences in imagery dates result in Google containing additional recently constructed buildings [5, 8]. Differences in the urban area are not likely due to imagery differences, but rather how the models define and separate buildings, as well as difficulty in matching footprints in dense urban areas.

This paper shows an effective method for comparing buildings datasets based on matched footprints in less dense areas, but a more refined matching strategy is needed for an appropriate building-level comparison in highly dense urban areas with complex building patterns. Goals for future work include further development and improvements on the building matching strategy, scaling to larger areas such as regions or countries, and incorporating other building morphology characteristics in addition to area to gain a better understanding of how these different sources characterize the same buildings.

References

1 Gianluca Boo, Edith Darin, Douglas R Leasure, Claire A Dooley, Heather R Chamberlain, Attila N Lázár, Kevin Tschirhart, Cyrus Sinai, Nicole A Hoff, Trevon Fuller, et al. High-resolution population estimation using household survey data and building footprints. *Nature communications*, 13(1):1330, 2022.

2 Maria Antonia Brovelli and Giorgio Zamboni. A new method for the assessment of spatial accuracy and completeness of openstreetmap building footprints. *ISPRS International Journal of Geo-Information*, 7(8):289, 2018.

3 Hongchao Fan, Alexander Zipf, Qing Fu, and Pascal Neis. Quality assessment for building footprints data on openstreetmap. *International Journal of Geographical Information Science*, 28(4):700–719, 2014. doi:10.1080/13658816.2013.867495.

4 Google. Google maps, 2023. URL: https://google.com/maps.

5 Google. Google open buildings, 2023. URL: https://sites.research.google/open-buildings/.

6 Robert Hecht, Carola Kunze, and Stefan Hahmann. Measuring completeness of building footprints in openstreetmap over space and time. *ISPRS International Journal of Geo-Information*, 2(4):1066–1091, 2013.

7 Christophe Leys, Christophe Ley, Olivier Klein, Philippe Bernard, and Laurent Licata. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4):764–766, 2013. doi:10.1016/j.jesp.2013.03.013.

8 Microsoft. Microsoft global ml footprints, 2023. URL: https://github.com/microsoft/GlobalMLBuildingFootprints.

9 Nikola Milojevic-Dupont, Felix Wagner, Florian Nachtigall, Jiawei Hu, Geza Boi Brüser, Marius Zumwald, Filip Biljecki, Niko Heeren, Lynn H. Kaack, Peter-Paul Pichler, and Felix Creutzig. EUBUCCO v0.1: European building stock characteristics in a common and open database for 200+ million individual buildings. *Scientific Data*, 10(1):147, March 2023. doi:10.1038/s41597-023-02040-2.

10 J. J. Moehl, E. M. Weber, and J. J. McKee. A vector analytical framework for population modeling. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVI-4/W2-2021:103–108, 2021. doi:10.5194/isprs-archives-XLVI-4-W2-2021-103-2021.

11 OpenStreetMap contributors. Open Street Map retrieved from https://planet.osm.org . https://www.openstreetmap.org, 2023.

12 Wojciech Sirko, Sergii Kashubin, Marvin Ritter, Abigail Annkah, Yasser Salah Eddine Bouchareb, Yann Dauphin, Daniel Keysers, Maxim Neumann, Moustapha Cisse, and John Quinn. Continental-Scale Building Detection from High Resolution Satellite Imagery, July 2021. Number: arXiv:2107.12283 arXiv:2107.12283 [cs]. URL: http://arxiv.org/abs/2107.12283.

13 Benjamin Swan, Melanie Laverdiere, H. Lexie Yang, and Amy Rose. Iterative self-organizing SCEne-LEvel sampling (ISOSCELES) for large-scale building extraction. *GIScience & Remote Sensing*, 59(1):1–16, December 2022. doi:10.1080/15481603.2021.2006433.

14 Hsiuhan Lexie Yang, Jiangye Yuan, Dalton Lunga, Melanie Laverdiere, Amy Rose, and Budhendra Bhaduri. Building Extraction at Scale Using Convolutional Neural Network: Mapping of the United States. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(8):2600–2614, August 2018. doi:10.1109/JSTARS.2018.2835377.

15 Qi Zhou, Yuheng Zhang, Ke Chang, and Maria Antonia Brovelli. Assessing osm building completeness for almost 13,000 cities globally. *International Journal of Digital Earth*, 15(1):2400–2421, 2022.