International Cartographic Association
Association Cartographique Internationale

A Novel GIS-based Polygon Shape Similarity Measure Applied to OSM Building Footprints

Milad Moradi [a,*], Stéphane Roche [a], Mir Abolfazl Mostafavi [a]

[a] *Centre for Research in Geospatial Data and Intelligence, Laval University, Quebec, QC G1V 0A6, Canada, milad.moradi.1@ulaval.ca, stephane.roche@scg.ulaval.ca, mir-abolfazl.mostafavi@scg.ulaval.ca*

* Corresponding author

Abstract: Assessing the similarity between polygonal shapes is a fundamental problem in geographic information science (GIS) with applications in spatial data quality assessment, feature matching, and cartographic generalization. This paper introduces a novel and computationally efficient shape similarity measure tailored for comparing building footprints in OpenStreetMap (OSM). Unlike traditional methods that rely on complex transformations such as Fourier descriptors or graph-based techniques, our approach is based on the average boundary distance between two polygons after applying translation and rotation corrections. This method is both easy to implement and computationally light, making it suitable for large-scale applications. The proposed measure demonstrates strong alignment with human perception of shape similarity. However, a notable limitation is that it tends to produce similarity values predominantly within the range of 70% to 100%. This behaviour arises because the measure emphasizes overall shape alignment while overlooking finer local discrepancies. As a result, subtle deviations, such as missing details or minor geometric distortions, may not significantly impact the computed similarity score. Despite this drawback, the method remains a practical and efficient alternative for evaluating shape similarity in large spatial datasets, particularly where computational simplicity and scalability are prioritized. Future works can explore potential refinements to enhance sensitivity to local shape variations while maintaining computational efficiency.

Keywords: Shape Similarity, OpenStreetMap, Geometry Similarity, Polygon Comparison, Buildings Footprint

1. Introduction

With the advent of Web 2.0 technologies, citizens gained the ability to share data over the internet (Goodchild, 2007). This development gave rise to diverse projects that rely on the active contributions of individuals acting as sensors to collect and share data with different motivations (Goodchild, 2007; Lotfian, Ingensand and Brovelli, 2020, 2021). OpenStreetMap (OSM) is one of the most prominent examples of such projects. However, since any contributor, regardless of certification or formal knowledge in geomatics, can edit OSM data, the quality of its datasets has become a significant challenge of research (Fan *et al.*, 2014; Arsanjani, Mooney and Zipf, 2015; Törnros *et al.*, 2015; Brovelli and Zamboni, 2018; Moradi, Roche and Mostafavi, 2022).

In OSM, buildings are added with the tag 'building = yes' to the project. Numerous measures have been developed to assess the quality of OSM building data across various dimensions of spatial data quality, including completeness (Hecht, Kunze and Hahmann, 2013; Fan *et al.*, 2014; Herfort *et al.*, 2023; Moradi, Roche and Mostafavi, 2023b; Ullah *et al.*, 2023), positional accuracy (Fan *et al.*, 2014; Brovelli *et al.*, 2016; Brovelli and Zamboni, 2018; KÜÇÜK and ANBAROĞLU, 2020), temporal accuracy (Hecht, Kunze and Hahmann, 2013; Moradi, Roche and Mostafavi, 2023b), shape accuracy (Fan *et al.*, 2014; Fan, Zhao and Li, 2021; Xu *et al.*, 2021; Ďuračiová, 2023), and semantic accuracy (Nowak Da Costa, 2016; Basaraner, 2020). Among these, measures designed to evaluate shape accuracy are particularly complex, as assessing the degree of similarity between two polygons still presents a challenging problem. In addition to spatial data quality, shape similarity is used for feature matching, classification, spatial inquiry, and cartographic generalization (Xu *et al.*, 2017; Lu *et al.*, 2024).

This paper provides a comprehensive review of existing methods for assessing shape similarity in GIS, with a specific focus on their application to OSM building footprints. Subsequently, it introduces a novel and easy-to-implement method to evaluate the similarity of polygons, offering new insights into the degree to which two shapes align.

The remainder of this paper is structured as follows: Section 2 reviews related works, Section 3 presents the fundamentals of shape similarity, Section 4 details the proposed method, Section 5 discusses the case study and results, and Section 6 concludes the study with future directions.

2. Related Works

Shape accuracy is the degree of similarity between shape *A* and shape *B*. A measure for polygonal shape accuracy is defined as a cost function $d(A, B)$ that is associated with the two polygons and quantifies the degree to which

Advances in Cartography and GIScience of the International Cartographic Association, 5, 22, 2025.
32nd International Cartographic Conference (ICC 2025), 17–22 August 2025, Vancouver, Canada. This contribution underwent double-blind peer review based on the full paper. https://doi.org/10.5194/ica-adv-5-22-2025 |

the two shapes are dissimilar (Arkin *et al.*, 1991). In other words, this measure should describe the similarity of two polygon boundaries using a single number and it should be consistent with human cognition of the similarity of the two shapes (Ai *et al.*, 2013).

Arkin *et al.*, (1991) proposed that if we represent the boundary of two polygons as a series of lengths and angles (for each edge), we can then use turning function $\Theta_A(s)$ to quantify their similarity. The turning function $\Theta_A(s)$ measures the angle of counterclockwise tangent as a function of arc-length, measures from a point on the boundary of A (Arkin *et al.*, 1991). By using turning function each polygon is represented as a series of horizontal line segments that the Y of each of them is equal to the tangent of the edge that they represent and their length is the length of that edge (Arkin *et al.*, 1991). Finally, $d(A, B)$ is the distance between these two representations of polygons.

Ai *et al.*, (2013) proposed a shape similarity method based on Fourier descriptor. It represents the boundary of a polygon shape as a periodic function, then, the distance between normalized Fourier coefficients is used as a measure of shape similarity. This method captures the main shape characteristics and ignores the details of the two shapes (Ai *et al.*, 2013).

Since complex spatial features are stored as multipolygons, Xu *et al.*, (2021) proposed a method based one similarities in shape and distribution of polygons that is capable of measuring the similarity between multipolygons. This method uses position graph to denote the distribution of subpolygons.

Since most of shape similarity methods are developed for simple polygons, Xu *et al.*, (2017) proposed a method capable of measuring shape similarity between complex polygons where there are several holes in the polygon. This method uses angles and distances to represent a polygon with its possible holes, then, it uses position graphs and Fourier transformation to measure the similarity between them (Xu *et al.*, 2017). The strength of this method is that when holes are represented in angles and distances, this representation is invariant under polygon translation, rotation and change of its scale. However, this method is computationally heavy.

Fan, Zhao and Li, (2021) converted polygons into grid representation in which the contour feature is represented as a multiscale statistic feature. This method, unlike previous methods, does not measure the distance between two representations of polygons (Fourier transformation or turning function representations). Instead, it defines the similarity as the correlation between textures extracted by shape features (Fan, Zhao and Li, 2021). This method showed a better accuracy than turning function and Fourier descriptor methods (Fan, Zhao and Li, 2021).

Fréchet distance is proposed by Shahbaz, (2013) as an effective similarity measure for spatial representation of features. This method unlike Hausdorff distance, takes the ordering of the points along the curves. This feature makes this shape similarity measure suitable for GIS applications where sequence of traversal matters, such as time-series analysis (Shahbaz, 2013). This measure is based on the minimum distance needed to traverse the two shapes without backtracking (Shahbaz, 2013). This method is robust against noise.

One of the most recently proposed methods of shape similarity, introduced by Lu *et al.*, (2024), is based on graph edit distance. This method first constructs a graph for each building contour, then measures a cost function based on the number of substitutions and deletions required to transform Graph 1 into Graph 2 (Lu *et al.*, 2024). Finally, this cost function quantifies the dissimilarity between the two polygons (Lu *et al.*, 2024).

In the context of OSM building footprint shape accuracy analysis, several research works have been done. Turning function method is widely used by research works (Mooney, Corcoran and Winstanley, 2010; Fan *et al.*, 2014; Müller, Iosifescu and Hurni, 2015; Hung, Kalantari and Rajabifard, 2016; Zhou *et al.*, 2018) to measure the shape accuracy of OpenStreetMap buildings footprints.

3. Fundamentals of Shape Similarity

If A and B are two polygonal shapes in the plane, then their similarity can be computed using a cost function $d(A, B)$. Arkin *et al.*, (1991) argued that such a measure should satisfy the following four properties to be consistent with human cognition:

- $d(A, B) \geq 0$ for all A and B.
- $d(A, B) = 0$ only and only if $A=B$.
- $d(A, B) = d(B, A)$. (symmetry)
- $d(A, B) + d(B, C) \geq d(A, C)$. (triangle inequality)

Additionally, since we want that this measure only represents the dissimilarity of the two shapes (boundaries of shapes), $d(A, B)$ should be invariant under translation, rotation and change of scale (Arkin *et al.*, 1991; Xu *et al.*, 2017). More importantly, this measure should align with human intuitive judgment about the dissimilarity of the two shapes. The proposed measure adheres to these properties.

4. The Proposed Shape Similarity Measure

The proposed method measures shape similarity based on the average boundary distance between two polygons after applying corrections to eliminate the effects of translation, rotation, and scale differences between corresponding polygons in the two datasets. The approximate average distance between corresponding points on the polygon boundaries is then estimated using the area enclosed between the two boundaries. Figure 1 illustrates the steps involved in feature matching, geometric corrections, and the main shape similarity algorithm.

4.1 Preprocessing

Before applying any shape similarity algorithm, it is necessary to perform feature matching between the OSM building footprints and the reference footprints. While feature matching is not the focus of this research, it can be performed by measuring the area overlap between polygons (Fan *et al.*, 2014), calculating the distance between their centroids (Hecht, Kunze and Hahmann, 2013), or using more accurate algorithms (Moradi, Roche and Mostafavi, 2023a). Since the shape similarity measure should be invariant to rotation, translation, and scale, corrections must first be applied to eliminate the effects of these transformations. However, in the case of OSM building footprints, scale mismatch is not a concern (Moradi, Roche and Mostafavi, 2023b), and therefore, no correction for scale is required.

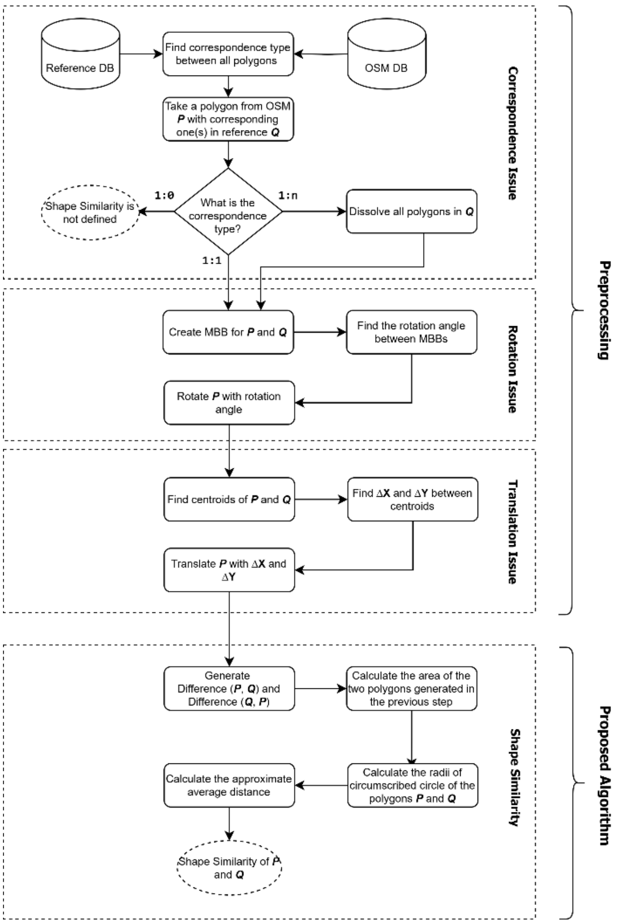


Figure 1. Workflow of the preprocessing steps and the proposed shape similarity algorithm.

4.1.1 *Feature matching*

To determine which reference polygon should be compared to each OSM polygon, a feature matching algorithm must be applied. If the correspondence type is 1:0, it indicates that the OSM building has no equivalent in the reference dataset, and therefore, the shape similarity is undefined in this case. For 1:1 correspondence, where there is exactly one polygon in OSM and one in the reference, the necessary corrections can be applied, and the shape similarity measure can be calculated.

In cases where one polygon in OSM corresponds to multiple polygons in the reference dataset (1:n), the reference polygons must be dissolved into a single polygon before proceeding with the process. For the many-to-many (m:n) case, both sets of polygons must be dissolved into a single polygon before continuing because the proposed algorithm in its current form is not capable of measuring shape similarity between multipolygon geometries.

4.1.2 *Rotation correction*

Since OSM polygons are often generalized or may contain errors compared to the reference polygons, and there can be mismatches in the number of edges or vertices, calculating the rotation angle directly is challenging. To address this, the rotation angle is calculated using the minimum bounding boxes (MBBs) of the two polygons rather than the polygons themselves. This approach reduces the impact of digitization errors on the calculated angle.

The rotation angle is defined as the angle between the two longer lines connecting the midpoints of opposite edges in the MBBs (see Figure 1). Once the rotation angle is determined, a rotation correction is applied to the OSM polygon to eliminate the impact of rotation on shape dissimilarities. Specifically, all vertices of the OSM polygon are rotated by θ degrees using the following formula:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} \quad (1)$$

Before applying the rotation correction, the centroid of the polygon should be translated to the origin. After the rotation, the polygon should be translated back to its original centroid location.

4.1.3 *Translation correction*

Since digitizing aerial images is one of the most common methods of data production in OSM, there are some systematic errors associated with OSM data. One common issue is that OSM contributors often digitize the roof outlines of buildings instead of their actual footprints. A potential error arises when the roof outline is displaced far from the building's footprint, particularly when the building is far from the centre of the aerial image. A translation correction should be applied to the vertices of the OSM polygon using as:

$$\begin{bmatrix} {x_i}' \\ {y_i}' \end{bmatrix} = \begin{bmatrix} x_{OSM\ Centroid} - x_{Ref.\ Centroid} \\ y_{OSM\ Centroid} - y_{Ref.\ Centroid} \end{bmatrix} + \begin{bmatrix} x_i \\ y_i \end{bmatrix} \quad (2)$$

where x_i, y_i are the coordinates of i-th vertex of OSM polygon and $x_{OSM\ Centroid}$, $x_{Ref.\ Centroid}$ are the x coordinates of the centroid of the OSM polygon and the centroid of its corresponding reference polygon, respectively.

Figure 2 illustrates an example of an OSM building and its corresponding reference polygon, highlighting the need for rotation and translation corrections.

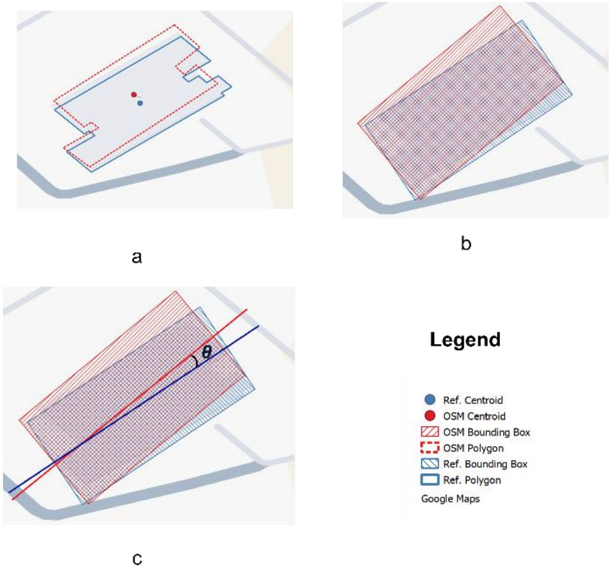


Figure 2. (a) The OSM polygon with its corresponding reference polygons. (b) The minimum bounding boxes of the two polygons. (c) The angle between the two bounding boxes.

4.2 Proposed Algorithm

If the two shapes are identical, we expect that after applying rotation, translation, and scale corrections, the boundaries of both polygons will overlap perfectly. In this ideal scenario, shape dissimilarity is zero. However, in real-world cases, the two boundaries typically do not align perfectly.

Here we propose a method where the average distance between the two boundaries, after applying the corrections, serves as an indicator of shape dissimilarity. However, computing this average distance is challenging. One might assume that simply generating sample points along the boundary of the OSM polygon and calculating their distances to the boundary of the reference polygon would be sufficient. However, our implementation of this method in our study area revealed that it can overlook significant shape dissimilarities in certain cases (see Figure 3).

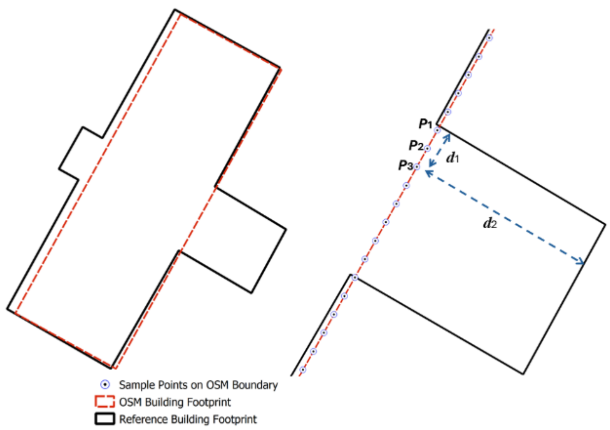


Figure 3. Potential error in average distance calculation when generating sample points on the OSM boundary.

The issue arises because the average distance should be computed between corresponding points on the two polygons, rather than the nearest point on the other polygon's boundary. Figure 3 illustrates this issue. P_3 is one of the sample points generated on the boundary of the OSM polygon. If we calculate its distance to the reference polygon, d_1 represents the shortest distance between P_3 and the reference polygon. However, to align with human perception of shape dissimilarity, the distance between corresponding points on the two polygons (d_2) should be considered instead.

On the other hand, finding corresponding points on the boundaries of the two polygons can be challenging, as OSM shapes may be highly generalized or differ significantly from the reference polygons.

To address this issue, we propose using the area between the two polygons to approximate the average distance between their boundaries.

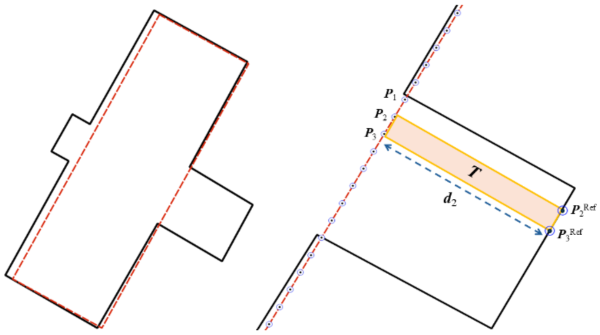


Figure 4. Average distance calculation based on the area of trapezoid *T*.

Let P_2^{Ref} and P_3^{Ref} be the corresponding points for P_2 and P_3 on the reference polygon. The line connecting P_2 and P_3, and the line connecting P_2^{Ref} and P_3^{Ref}, are not necessarily parallel. However, for approximation purposes, we assume that these two lines are parallel. In this case, T forms a trapezoid, and d_2 represents its height. d_2 can then be calculated as:

$$d_2 \approx \frac{Area\ of\ T}{\frac{Dist_{P_2-P_3}+Dist_{P_2^{Ref}-P_3^{Ref}}}{2}} \qquad (3)$$

where $Dist_{P_2-P_3}$ represents the distance between P_2 and P_3 along the perimeter of the OSM polygon, and $Dist_{P_2^{Ref}-P_3^{Ref}}$ represents the corresponding distance along the perimeter of the reference polygon. Thus, the average distance between the corresponding points of the two polygons is approximately equal to the area between their boundaries divided by the average of their perimeters.

$$Average\ Distance \approx \frac{Area\ between\ the\ boundaries}{\frac{Perimeter_{OSM}+Perimeter_{Ref}}{2}} \qquad (4)$$

Advances in Cartography and GIScience of the International Cartographic Association, 5, 22, 2025.
32nd International Cartographic Conference (ICC 2025), 17–22 August 2025, Vancouver, Canada. This contribution underwent double-blind peer review based on the full paper. https://doi.org/10.5194/ica-adv-5-22-2025

where the area between the boundaries is defined as the sum of the area of the OSM polygon that lies outside the reference polygon and the area of the reference polygon that lies outside the OSM polygon. In GIS applications, the difference operation generates a polygon representing the portion of the first polygon that does not overlap with the second polygon. Thus, the average distance can be expressed as:

$$Av.Dist. \approx \frac{Area\ (Diff_{Ref,OSM}) + Area\ (Diff_{OSM,Ref})}{\frac{Perimeter_{OSM} + Perimeter_{Ref}}{2}} \quad (5)$$

Figure 5 illustrates the difference between the OSM and reference polygons, as well as the difference between the reference polygon and the OSM polygon. In this example, the area of the OSM polygon extending beyond the reference polygon is larger, as most parts of the OSM polygon extend beyond the reference polygon.

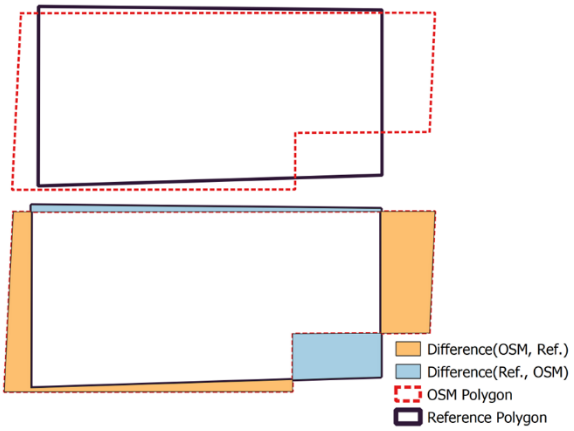


Figure 5. Area between the OSM building footprint and corresponding reference footprint.

The average distance is always greater than zero and indicates the degree of shape dissimilarity between the two polygons. The shape similarity is then calculated using the following formula:

$$shape\ similarity = 1 - \frac{average\ distance}{max\ radius_{OSM,Ref}} \quad (6)$$

where max radius is the largest radius among the circumscribed circles of the reference polygon and the OSM polygon (see Figure 6).

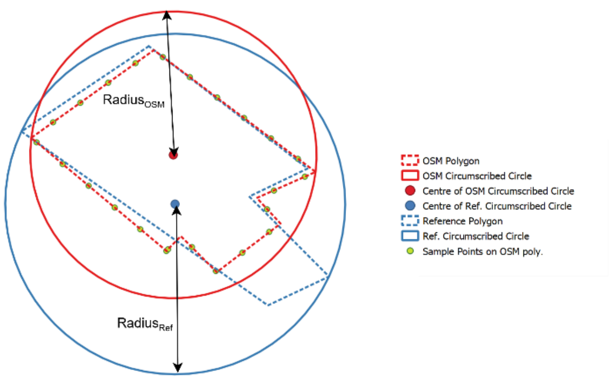


Figure 6. Sample points on OSM polygon, and OSM and reference circumscribed circles.

This measure ensures that similarity values range between 0 and 1, where 1 indicates identical shapes, and values closer to 0 represent greater dissimilarity.

5. Case Study

The proposed algorithm is implemented in python and is used to measure the shape accuracy of OSM buildings in Quebec City.

5.1 Study area

Quebec City, the capital of the province of Quebec in Canada, is home to nearly 500,000 residents. According to our reference data, the study area includes 171,648 buildings. The study area is defined as a rectangular region with the following coordinates: Left: −71.5, Bottom: 46.7, Right: −71.0, Top: 46.9.

5.2 Data Description and Preprocessing

The OSM data was downloaded as of January 1, 2025, using the OSMnx Python package for Quebec City, based on the coordinates mentioned in the previous section. The reference data was obtained from the open data portal of the Government of Canada: https://open.canada.ca/data/en/dataset/be7053a8-7122-4514-91a2-5a8f5a60b341

The completeness of the OSM data, based on the number of buildings, is 43.49%, while the completeness based on the total area of building footprints in the two datasets is 57.14%. This discrepancy arises because the average area of OSM building footprints is 295.95 m², compared to 225.27 m² for the reference building footprints. Larger buildings tend to attract more attention from OSM contributors compared to smaller buildings, which could explain this difference.

Figure 7 illustrates the percentage distribution of building footprint areas in the two datasets. The figure shows that in the OSM dataset, there is a higher percentage of buildings with footprints larger than 1000 m². However, for buildings with footprints smaller than 500 m², the reference dataset contains a higher percentage. This pattern supports the observation that larger buildings are more frequently digitized in OSM, while smaller buildings may be underrepresented.

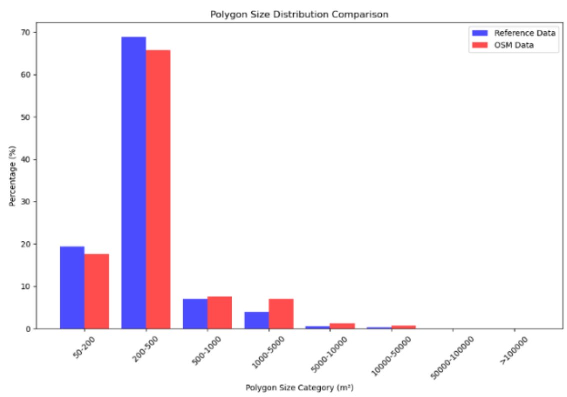


Figure 7. Percentage distribution of building footprint areas in the two datasets.

Advances in Cartography and GIScience of the International Cartographic Association, 5, 22, 2025.
32nd International Cartographic Conference (ICC 2025), 17–22 August 2025, Vancouver, Canada. This contribution underwent double-blind peer review based on the full paper. https://doi.org/10.5194/ica-adv-5-22-2025

Feature matching was performed between the two datasets. In most cases (approximately 90%), the buildings have a 1:1 correspondence, which simplifies the comparison of the two shapes. In instances where multiple reference buildings correspond to a single building in OSM, the reference polygons are dissolved into a single polygon to ensure the shape comparison remains feasible.

The translation correction was applied to OSM polygons. The average displacement in the X-direction is 1.06 m, and in the Y-direction is 1.55 m. The scatter plot of the centroid displacements is shown in Figure 8. Figure 8 indicates that most of the OSM centroids are on the left side of the centroids of the reference polygons.

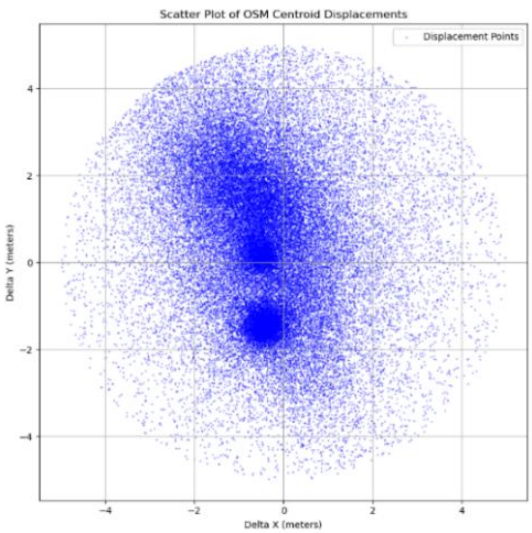


Figure 8. Scatter plot of the centroid displacement of the OSM buildings in comparison to the reference dataset.

After applying the necessary corrections, the shape similarity algorithm was implemented using the GeoPandas Python package. The implementation is straightforward, as all required operations (e.g., centroid calculation, difference computation, minimum bounding box, etc.) are built-in functionalities of GeoPandas and other Python libraries.

5.3 Results and Discussion

The shape accuracy values for Quebec City building footprints are mostly between 0.7 and 1, as the proposed method is not sensitive to minor dissimilarities between the two shapes. Figure 9 illustrates the histogram of shape similarity values.

Based on Figure 9, it is evident that in most cases, the proposed method calculates a relatively high similarity value. While this can be considered a disadvantage, it is an expected behaviour for a measure based on average distance, as it is not highly sensitive to small details.

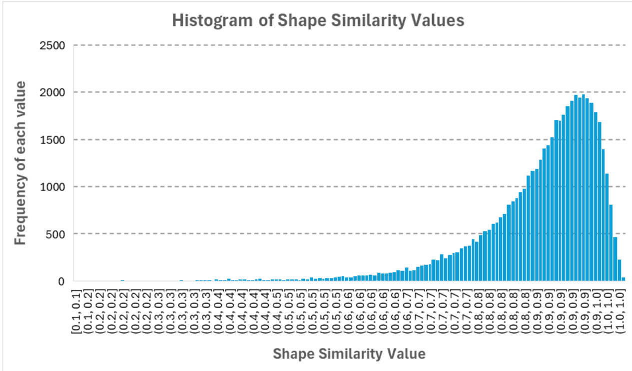


Figure 9. Histogram of the values of the proposed shape similarity measure.

We benchmarked the proposed shape similarity measure against two well-established measures: Elliptic Fourier Descriptors (EFD) and the Fréchet-distance shape similarity measure.

- Elliptic Fourier Descriptors (EFD).

We extracted the Fourier coefficients of each polygon contour with the **“pyefd”** library (https://github.com/hbldh/pyefd). Using the first $n = 30$ harmonics of the normalised boundary, we formed a coefficient vector for every shape and quantified dissimilarity as the Euclidean distance between the two vectors.

- Fréchet distance

The **“shape-similarity”** library (https://github.com/nelsonwenner/shape-similarity) computes the continuous Fréchet distance between the ordered boundary point sequences of the two polygons. This distance is then normalised by the larger of the two perimeters, yielding a score in [0,1] where 0 indicates identical shapes and 1 denotes maximal dissimilarity. Since the Fréchet measure is a dissimilarity (smaller values denote greater similarity), we converted it to a similarity score by computing 1 – (Fréchet distance). This aligns with the scale of the proposed method.

Both reference methods apply the same pre-processing steps used in the proposed approach: translation to a common centroid, rotation to the principal axis, and isotropic scaling. It ensures that any differences in the resulting scores arise solely from the similarity formula itself rather than from other factors.

Table 1 presents selected OSM and reference polygons to illustrate the compatibility of the proposed method with human cognition and perception of shape similarity.

In Table 1, in the first and second examples, the OSM polygon is a generalized version of the reference polygon, with many details that have not been digitized. As a result, the shape similarity based on the proposed method is below 90%.

Fourier Descriptors and Fréchet Distance produced much lower similarity degrees in comparison to the proposed method.

In the third example, the shape similarity is 99.6%, which aligns with human perception of how similar these two shapes are. The two other methods also generated shape similarity values around 97% and 99.4%, respectively.

In the fourth example, only one part of the reference shape is missing in OSM. Since the proposed method is not sensitive to small dissimilarities, the calculated shape similarity is 97%. Fréchet Distance is much more sensitive to big shape dissimilarities. The proposed method is based on the average distance and is not greatly affected by small shape mismatches.

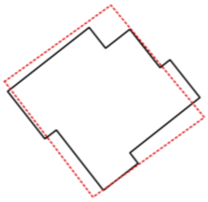
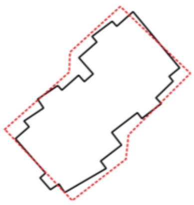
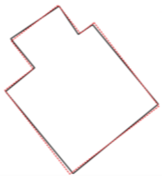
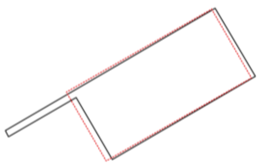
id	OSM and Reference shapes	Shape Similarity		
		Proposed Method	Fourier Descriptors	Fréchet Distance
1		0.89	0.69	0.58
2		0.84	0.66	0.48
3		0.996	0.977	0.994
4		0.97	0.87	0.73

Table 1. Examples of the OSM polygon (red) and corresponding reference polygon (black) with their shape similarity value.

6. Conclusions

This paper introduced a novel shape similarity measure for comparing polygonal geometries, particularly applied to OpenStreetMap (OSM) building footprints. The proposed method is based on the average boundary distance between two polygons, making it both computationally efficient and straightforward to implement. Unlike more complex shape similarity techniques, which require advanced transformations or graph-based methods, the simplicity of the proposed approach makes it practical for large-scale spatial datasets.

One key advantage of the proposed measure is its alignment with human perception of shape similarity. The results indicate that the measure effectively captures the overall similarity between building footprints, making it a useful tool for applications such as data quality assessment, feature matching, and cartographic generalization. Additionally, since scale and rotation are not major factors in OSM buildings, the main transformation required is translation correction, further simplifying its application in this context.

Another advantage of the proposed method is its reliance on familiar GIS primitives (centroid position, boundary distance, and enclosed area), whereas Fourier-based techniques depend on less intuitive frequency-domain coefficient distances.

However, the method has some limitations. First, most shape similarity values fall between 0.7 and 1.0, making it less effective in differentiating varying degrees of similarity. This is because the measure is based on average boundary distance, which tends to overlook small differences in shape details. Second, the method is not designed to handle complex polygons with holes or multiple parts. In this study, this issue was addressed by dissolving polygons into single geometries, but more sophisticated solutions could be explored in future research.

 https://doi.org/10.5194/ica-adv-5-22-2025 |

For future improvements, modifications could be made to enhance the measure's sensitivity to small shape dissimilarities without sacrificing computational efficiency. Additionally, researchers could explore alternative lightweight similarity measures that remain easy to implement while capturing finer shape variations. Despite these limitations, the proposed method offers a practical and scalable solution for shape similarity analysis in GIS applications, particularly when working with large datasets such as OSM building footprints.

7. References

Ai, T. *et al.* (2013) A shape analysis and template matching of building features by the Fourier transform method, *Computers, Environment and Urban Systems*. Elsevier Ltd, 41, pp. 219–233. doi: 10.1016/j.compenvurbsys.2013.07.002.

Arkin, E. M. *et al.* (1991) An Efficiently Computable Metric for Comparing Polygonal Shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 209–216. doi: 10.1109/34.75509.

Arsanjani, J. J., Mooney, P. and Zipf, A. (2015) Quality Assessment of the Contributed Land Use Information from OpenStreetMap Versus Authoritative Datasets, in *OpenStreetMap in GIScience*. Springer, pp. 37–58. doi: 10.1007/978-3-319-14280-7.

Basaraner, M. (2020) Geometric and semantic quality assessments of building features in OpenStreetMap for some areas of Istanbul, *Polish Cartographical Review*, 52(3), pp. 94–107. doi: 10.2478/pcr-2020-0010.

Brovelli, M. A. *et al.* (2016) Positional accuracy assessment of the openstreetmap buildings layer through automatic homologous pairs detection: The method and a case study, in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, pp. 615–620. doi: 10.5194/isprsarchives-XLI-B2-615-2016.

Brovelli, M. and Zamboni, G. (2018) A New Method for the Assessment of Spatial Accuracy and Completeness of OpenStreetMap Building Footprints, *ISPRS International Journal of Geo-Information*, 7(8), p. 289. doi: 10.3390/ijgi7080289.

Ďuračiová, R. (2023) An Aggregated Shape Similarity Index: A Case Study of Comparing the Footprints of OpenStreetMap and INSPIRE Buildings, *ISPRS International Journal of Geo-Information*, 12(12). doi: 10.3390/ijgi12120495.

Fan, H. *et al.* (2014) Quality assessment for building footprints data on OpenStreetMap, *International Journal of Geographical Information Science*, 28(4), pp. 700–719. doi: 10.1080/13658816.2013.867495.

Fan, H., Zhao, Z. and Li, W. (2021) Towards measuring shape similarity of polygons based on multiscale features and grid context descriptors, *ISPRS International Journal of Geo-Information*, 10(5). doi: 10.3390/ijgi10050279.

Goodchild, M. F. (2007) Citizens as sensors: The world of volunteered geography, *GeoJournal*, 69(4), pp. 211–221. doi: 10.1007/s10708-007-9111-y.

Hecht, R., Kunze, C. and Hahmann, S. (2013) Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time, *ISPRS International Journal of Geo-Information*, 2(4), pp. 1066–1091. doi: 10.3390/ijgi2041066.

Herfort, B. *et al.* (2023) A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap, *Nature Communications*. Springer US, 14(1), pp. 1–14. doi: 10.1038/s41467-023-39698-6.

Hung, K.-C., Kalantari, M. and Rajabifard, A. (2016) Assessing the quality of building footprints on OpenStreetMap: a case study in Taiwan, in *Smart World*, p. 237.

KÜÇÜK, K. and ANBAROĞLU, B. (2020) Spatial Accuracy Assessment of Buildings in Openstreetmap, *International Journal of Environment and Geoinformatics*, 7(2), pp. 205–212. doi: 10.30897/ijegeo.710508.

Lotfian, M., Ingensand, J. and Brovelli, M. A. (2020) A framework for classifying participant motivation that considers the typology of citizen science projects, *ISPRS International Journal of Geo-Information*, 9(12). doi: 10.3390/ijgi9120704.

Lotfian, M., Ingensand, J. and Brovelli, M. A. (2021) The partnership of citizen science and machine learning: Benefits, risks and future challenges for engagement, data collection and data quality, *Sustainability (Switzerland)*, 13(14). doi: 10.3390/su13148087.

Lu, X. *et al.* (2024) Measuring the similarity between shapes of buildings using graph edit distance, *International Journal of Digital Earth*, 17(1). doi: 10.1080/17538947.2024.2310749.

Mooney, P., Corcoran, P. and Winstanley, A. C. (2010) Towards quality metrics for OpenStreetMap, in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, p. 514. doi: 10.1145/1869790.1869875.

Moradi, M., Roche, S. and Mostafavi, M. A. (2022) Exploring five indicators for the quality of OpenStreetMap road networks: a case study of Québec, Canada, *Geomatica*, 31(April), pp. 1–31. doi: 10.1139/geomat-2021-0012.

Moradi, M., Roche, S. and Mostafavi, M. A. (2023a) A Novel Feature Matching Method for Matching OpenStreetMap Buildings with Those of Reference Dataset, in *International Symposium on Web and Wireless Geographical Information Systems*. Cham: Springer Nature Switzerland, pp. 139–152.

Moradi, M., Roche, S. and Mostafavi, M. A. (2023b) Evaluating OSM Building Footprint Data Quality in Québec Province, Canada from 2018 to 2023: A Comparative Study, *Geomatics*, 3(4), pp. 541–562. doi: 10.3390/geomatics3040029.

Müller, F., Iosifescu, I. and Hurni, L. (2015) Assessment and Visualization of OSM Building Footprint Quality, in *Proceedings of the 27th International Cartographic*

Conference. Rio de Janeiro.

Nowak Da Costa, J. (2016) Towards Building Data Semantic Similarity Analysis: OpenStreetMap and the Polish Database of Topographic Objects, *Proceedings - 2016 Baltic Geodetic Congress (Geomatics), BGC Geomatics 2016*. IEEE, pp. 269–275. doi: 10.1109/BGC.Geomatics.2016.55.

Shahbaz, K. (2013) Applied Similarity Problems Using Frechet Distance, (June). Available at: http://arxiv.org/abs/1307.6628.

Törnros, T. *et al.* (2015) Uncertainties of completeness measures in openstreetmap – A case study for buildings in a medium-sized German city, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(3), pp. 353–357. doi: 10.5194/isprsannals-II-3-W5-353-2015.

Ullah, T. *et al.* (2023) Assessing Completeness of OpenStreetMap Building Footprints Using MapSwipe, *ISPRS International Journal of Geo-Information*, 12(4). doi: 10.3390/ijgi12040143.

Xu, Y. *et al.* (2017) Shape similarity measurement model for holed polygons based on position graphs and Fourier descriptors, *International Journal of Geographical Information Science*, 31(2), pp. 253–279. doi: 10.1080/13658816.2016.1192637.

Xu, Y. *et al.* (2021) Measuring the similarity between multipolygons using convex hulls and position graphs, *International Journal of Geographical Information Science*. Taylor & Francis, 35(5), pp. 847–868. doi: 10.1080/13658816.2020.1800016.

Zhou, X. *et al.* (2018) Change detection for building footprints with different levels of detail using combined shape and pattern analysis, *ISPRS International Journal of Geo-Information*, 7(10). doi: 10.3390/ijgi7100406.

 |