A position-and-form-based distance and its application in geographical analysis

May 22, 2024

Abstract

Position and form are basis characterizations of objects in space. However, in the distance-based geographical analysis of objects, conceptualization of distance is usually based on some representative point like centroid of objects and therefore suffers a significant loss of information associated with their forms. In this paper, we propose a position-and-from-based distance to explicitly take into these basis characterizations. For substantiation, its significance is demonstrated with respect to methodology and application. In methodology, we show the form effect on existing geographical analyses based on only position-based distance and show all point-distance-based analyses and relevant methods could be generalized for study of objects with forms, using the pattern analysis by L statistic as an example. In application, we demonstrate the proposed distance could successfully solve the matching problem between the same object in the OpenStreetMap and the correspondent standard reference data in a small region (to show where it is). Such matching problem cannot be perfectly handled by the traditional methods utilizing incomplete position and form information. The newly-proposed distance is applicable to more real-life cases when the object forms have to be considered. The proposed position-and-form-based distance and the associated methods could give us a new perspective on the conceptualization of distance. Actually, it can also be further extended to include other object attributes. Therefore, it is an ideal notion of distance that can fully reveal the multi-facet nature of geographical relations. The proposed research will advance the frontier of theoretical and applied research in geography where distance plays an important role. (248/250 words)

1 Instruction

Space in general and distance in particular are the two most central concepts in geography (Gatrell, 1983). To conceptualize space, it essentially needs some objects composing a set. Then, distance can be defined as a relation, usually qualified as a numeric metric, between a pair of objects that satisfies three conditions, namely i) positive definiteness, ii) symmetry, and iii) triangle inequality. In a more general sense, distance can be considered as a measure of dissimilarity of objects. A set with a defined relation, such as distance, between its elements, like objects, creates a space (Gatrell, 1983). Geographical analysis often involves the distance between objects in space. For example, examinations of the spatial pattern of objects and their relationships, particularly those based on the notion of distance, have been the common concern in geography (Boots and Getis, 1988). The distance-based methods, e.g., G function, F function, K function and the corresponding L statistic (Baddeley et al., 2015; Boots and Getis, 1988; Diggle, 2003), have been powerful tools for the examination of spatial patterns and relationships. Spatial statistics in general and spatial auto-correlation in particular, including variogram (Cressie, 2015), Moran's I (Moran, 1950), and Geary's C (Geary, 1954), often include distance as a key factor in the analysis (see a special issue of *Geographical Analysis* (Griffith, 2009)). The famous and yet controversial "first law of geography" (Tobler, 1970) is a living testimony to the importance of distance and spatial auto-correlation, in our understanding of relationships among objects in space. It has generated tremendous research interest and debates in both theory and applications since its inception (see a special issue of *the Annals of the Association of American Geographers* "On Tobler's First Law of Geography" in 2004). In geographically weighted regression (GWR), distance and spatial auto-correlation are again specifically taken into consideration in the estimation of regression coefficients when local relationships exist (Fotheringham et al., 1998).

One of the most common forms of geographical objects is point in space because any objects can be abstracted as points given a sufficiently large scale of observation (Boots and Getis, 1988). Consequently, distance has traditionally been specified between two points. Although different definitions of distance have been proposed in geography over the years, such as the time distance (Clark, 1977; Muller, 1978), economic distance (Lowe and Moryadas, 1975), cognitive distance (Golledge et al., 1969; Lowrey, 1973), and social distance (Hall, 1959, 1966; Laumann, 1966, 1973), the Minkowski p-metrics, especially the Euclidean distance for $p = 2$, is the most common distance used in geographical studies (Miller and Wentz, 2003). However, if the scale of observation is not large sufficiently, spatial objects may only be generally conceptualized lines and areal units/zones/regions/-polygons of various forms (used interchangeably in our discussion) (Boots and Getis, 1988; Gatrell, 1983; Lloyd, 2014). Unfortunately, the specification of distance in geography has little taken into consideration the exact forms of the objects. By the argument of Alexander (1964), the form of an object actually affects its function. City size and shape, for example, is a tell-tale story about the function of a city, e.g., infrastructure development and public service (Batty and Longley, 1994; Batty, 2007, 2008, 2013). Also, it is argued that the form and function of buildings are closely related (Nasar et al., 2005). Therefore, the effect of forms of the objects (points, lines and polygons) on the distance between them at a certain observational scale needs to be investigated. If the impact of form cannot be ignored at a given scale of observation, we have to incorporate object forms into consideration when we study the distance between them. As the matter of face, the study of such form-related distance is also a central research problem in computer vision (Loncaric, 1998; Veltkamp and Hagedoorn, 2001).

In particular, distance between a pair of objects with at least one not a point is often transformed into the distance between points because of the simplicity in calculation. It can be treated as a generalization of the point distance, exemplified by the distance between centroids (Miller and Wentz, 2003), the minimum value of distance between turning points of boundaries of two areal objects (Arkin et al., 1989; Latecki and Lakamper, 2000), difference in the two shortest inner distances between two landmark points in each areal unit (Ling and Jacobs, 2007), and the minimum (Peuquet, 1992) or maximum distance (Okabe and Miller, 1996) between a pair of points separately belonging to two objects. Taking the whole object into consideration, the distance between two lines has been measured by the epsilon band method (Skidmore and Turner, 1992), the Hausdorff method (Abbas et al., 1995), the simple buffer method (Goodchild and Hunter, 1997), and the double buffer method (Tveite, 1999). Distance between areal units can be further specified into two general types: The geometry-based distance, if only areal shapes are considered; and the intensity-based distance, if there is extra information inside the areal units (Veltkamp and Hagedoorn, 2001). In terms of the geometry-based distance, it can be defined on the basis of the time series extracted from the boundaries of the areal units. In the extracted time series, the Fourier spectrum (Zahn and Roskies, 1972), bending energy (Young et al., 1974), and dynamics (Kashyap and Chellappa, 1981) have been employed to define the distance. In addition, syntactic analysis (Fu, 1974), multi-scale feature (Witkin, 1984), approximation (Pavlidis, 2012), and decomposition (Liu and Srinath, 1990) of boundaries can also enable us to quantify the distance in the shape-based sense. Utilizing the whole areas, it has been introduced into the study of area distance the Fréchet (Alt and Godau, 1995) and Hausdorff (Huttenlocher et al., 1993) distance, shape matrices method (Davis, 1986), and the overlap area (Mount et al., 1996) and symmetric difference (Alt et al., 1996). Involving the intensity in areas, the area distance can be calculated by the moment-based (Prokop and Reeves, 1992), morphological (Haralick et al., 1987), and fractal method (Pentland, 1984).

Although there are quite a number of classical definitions on distance between lines or areas, the point distance still plays an important role in geographical studies, especially in the calculation of distance in spatial analysis and geographic information system (GIS) which has been seemingly oblivious to the potential weakness in our negligence of the impact of form on distance-based studies (Miller and Wentz, 2003). This might be partially due to our lack of theories and methods for measuring and analyzing geographical relationships with respect to geometric forms, particularly in relation to the notion of distance (Miller and Wentz, 2003).

In this study, we explicitly incorporate form into our notion of distance. To facilitate our discussion, we will limit forms to the two-dimensional space, the most common dimension in geographical studies. With this new notion of distance, we intend to show the effect of the object form and to demonstrate the applicability of the proposed notion. This paper is then organized as follows: In Section 2 a new notion of distance explicitly incorporating the object form is proposed at conceptual level; In Section 3 the effect of the object form on the traditional distance-based geographical analysis is shown by a simple pattern analysis of a generated example using the L statistic. The feasibility of generalization of these analyses and relevant methods from points to forms is also demonstrated in this section; In Section 4 the applicability of the proposed distance notion is justified by showing how to solve a real-life problem in a VGI study; In Section 5 the paper ends with a summary and a bright outlook for further generalizations and more possible applications of the proposed distance notion with the object form incorporated.

2 Position-and-Form Distance

As argued, the most basic characterization of an object is its position and form. Therefore, a position-and-form-based distance for geographical analysis should satisfy some conditions under which relationships of objects in space can be appropriately characterized. The position and form are the most basic dimensions of an object to which further dimensions/attributes can be augmented. It is noteworthy that a point in space takes on a value of zero for its form because it has no geometric structures.

To simplify our investigation and explicate the effects, we target at objects with only position and form in our analysis, i.e., no other attributes about an object will be involved in the consideration of the notion of the position-and-form-based distance. However, it should be noted that the proposed distance concept can be extended to the general situation in which objects are multi-facet, characterized by other attributes. In other words, we specify distance with respect to only these two attributes in our analysis, i.e. the position-based and form-based distance, but with possibility of further generalization for more applicable to more general cases. With this premise, the position-and-form-based distance should be defined with respect to the combination of these two basic attributes with which their individual effects can also be separately examined.

Referring to Fig. 1, for illustration, we can calculate the position-and-form-based distance between two given objects X_1 and X_2 as below. Step I, the centroids of objects X_1 and X_2, C_{X_1} and C_{X_2} respectively, are first extracted to account for the information about their positions. Although different definitions of distance have been proposed in geography over the years, such as the time distance (Clark, 1977; Muller, 1978), economic distance (Lowe and Moryadas, 1975), cognitive distance (Golledge et al., 1969; Lowrey, 1973), and social distance (Hall, 1959, 1966; Laumann, 1966, 1973), the Minkowski p-metrics, especially the Euclidean distance for $p = 2$, is the most common distance used in geographical studies (Miller and Wentz, 2003). Therefore, the Minkowski p-metric, particularly the Euclidean distance, between them is used to measure distance between two centroids. This is the first component: the position-based distance $d_p(X_1, X_2)$.

With respect to the form-based distance, it should also strictly satisfy conditions of distance i)–iii) in the metric space, i.e., positive definiteness, symmetry, and triangle inequality, for the purpose of combination with $d_p(X_1, X_2)$. To exclude the duplicate consideration of the effect of position when form is considered, two objects are shifted to let their centroids C_{X_1} and C_{X_2} coincide and then the form-based distance is calculated in Step II in Fig. 1.

In addition to conditions i)–iii), the form-based distance, $d_f(X_1, X_2)$, is expected to satisfy the following conditions:

iv) $d_f(X_1, X_2)$ should be a unified metric suitable for any forms including points, lines, and polygons;

v) $d_f(X_1, X_2)$ should be a metric directly qualifying the difference between two forms without transforming them to other type of forms, such as from polygons to lines or points;

vi) $d_f(X_1, X_2)$ should remain the same if the relative relationship does not change, e.g., both of them simultaneously rotate by a certain degree or distorted in the same manner;

vii) $d_f(X_1, X_2)$ should be sensitive to any slight changes in forms of the objects.

It is noteworthy that condition vii), stipulating sensitivity, is in contrast to the robustness to noise. It is understandable because distinguishing noise from the essential difference in form is very difficult. To capture any slight change in the form-based distance, sensitivity instead of robustness should be a more appropriate expected condition. If some changes can be certainly attributed to noise, we can then set a threshold to filter these noise-caused changes out. On the basis of conditions i)–vii), we select the Hausdorff distance, which has been demonstrated as an effective distance to capture any slight difference between two objects in arbitrary forms (Feder, 1988; Falconer, 1990), as an appropriate form-based distance. We would like to emphasize here that we select the Huasdorff distance because it is an appropriate distance satisfying the seven conditions but not implying that it is the best one of all form distances. As illustrated in Step III in Fig. 1, the Hausdorff distance can be simply obtained as: first create buffers of the object X_1 and mark the minimum buffer size δ_{X_1} for those that completely cover the object X_2; and δ_{X_2} can be similarly obtained; then the Hausdorff distance equals $\max\{\delta_{X_1}, \delta_{X_2}\}$ and is the to be employed form-based distance $d_f(X_1, X_2)$.

After obtaining the position-based $d_p(X_1, X_2)$ and the form-based distance $d_f(X_1, X_2)$, it is necessary to combine them to construct the position-and-form-based distance $d(X_1, X_2)$. Leung et al. (2013) proposed an effective way to combine a position-based distance and another attribute-based distance to form an intrinsic distance for the study of spatial relationships. They prove that the combined measure is also a metric distance in the strict sense, i.e., satisfying the distance conditions i)–iii). Along similar line of reasoning, we can define the position-and-form-based distance $d(X_1, X_2)$ in the combination of $d_p(X_1, X_2)$ and $d_f(X_1, X_2)$ as

$$d(X_1, X_2) = cd_p(X_1, X_2) + e^{cd_f(X_1, X_2)} - 1, \tag{1}$$

with an empirical penalty parameter c, referring to Step IV in Fig. 1. If $d_f(X_1, X_2)$ is short, then $d(X_1, X_2)$ is approximately $cd_p(X_1, X_2) + cd_f(X_1, X_2)$. We re-emphasize in here that $d(X_1, X_2)$ is a distance in the strict sense, i.e., satisfying the distance conditions i)–iii). With such combination of the position and form distance, we can examine the composite effect of position and form on a relationship, and we can individually examine the effect of each component also. Similarly, more attributes of an object can be easily incorporated into the position-and-form-based distance by adding one more attribute distance $d_a(X_1, X_2)$, defined as the difference in the corresponding attribute, to $d(X_1, X_2)$. And more detailed information about the combination of two different distances can be referred to Leung et al. (2013).

3 Effect of the Object Form

Given the positive-definite property of $d_f(X_1, X_2)$, it is easy to observe the effect of the form-based distance on increasing $d(X_1, X_2)$. In this section, we examine the form effect on one of the fundamental distance-based methods in geographical analysis, i.e., L statistic. As an illustrative example, we employ the L statistic to investigate the form effect on the spatial pattern consisted of lines, the most simple example of forms.

Point pattern analysis is a basic research topic in geographical analysis (Boots and Getis, 1988). Three basic point patterns are usually considered, namely the complete spatial randomness (CSR), clustered, and regular pattern (Boots and Getis, 1988). The CSR, a result of a homogeneous Poisson point process, is taken as a benchmark. Points that distribute together tend to take on the clustered pattern, whereas those which are more spread out tend to take on a regular pattern. If points follow the CSR pattern like the example generated in a area $[0, 1] \times [0, 1]$ in R^2 as exhibited in left panel of Fig. 2, the expected number of point-to-point distance less than h divided by the intensity is $K(h) = \pi h^2$, and the corresponding $L(h)$ equals $\sqrt{K(h)/\pi} - h$ with a value zero (Boots and Getis, 1988; Lloyd, 2014). The clustered and regular pattern should respectively have L statistic significantly larger and smaller than that of the CSR pattern. The confidence interval for the significance test can be determined by the Monte Carlo experiment with surrogates generated by the same homogeneous Poisson point process. For each h, the corresponding quantiles of $L(h)$ determines the confidence interval. In this study, we use the 95% confidence interval, represented by the red dashed envelope in Fig. 3. $L(h)$ of the generated points is around zero and always locate in the confidence interval for all $h \in [0, 0.4]$ (Fig. 3), indicative of a consistent result, i.e., not significantly different from the CSR point patterns.

Then, a line with length 1/32 is introduced for each point with the line centroid locating at the exact position of the corresponding point and the line orientation independently randomly assigned (right panel of Fig. 2). We replace the centroid-to-centroid distance by the position-and-form-based distance with $c = 1$ and recalculate the corresponding $L(h)$. To set c at 1, we can examine how the form-based distance impacts the position-and-form-based distance by directly comparing to the position-based distance. It can be observed that the recalculated $L(h)$ now falls into the 95% confidence interval of the CSR point pattern, i.e., in the red dashed envelop in Fig. 3, only if $h \geq h_0 \approx 0.17$. Analogous to the CSR point pattern, it is reasonable to consider the patterns of lines with their centroids and orientations independently and randomly generated as CSR. In this sense, the conclusion with respect to the line pattern based on the centroid pattern analysis would be misleading. Furthermore, it seems that the difference between $L(h)$ on the basis of centroid distance and position-and-form-based distance almost unvaries if $h > 0.1$ (the estimated trend is only -0.00049). As detailized in Fig. 4, such difference indeed unchanges if $h > 0.1$ with the mean value about -0.01. Therefore, we define this mean value an effect indicator to quantify the effect of lines on the pattern analysis. In another sense, if we investigate the pattern based only on the position information, then this effect indicator should be added to correct $L(h)$ for form pattern.

Here, we would like to highlight the importance of h_0 for the pattern analysis using L statistic that h_0 can actually indicate when the effect of lines can be ignored. Essentially, h_0 give a measure of sufficient large scale of observation, over which lines could be extracted as points for the pattern analysis without drawing misleading conclusions. For scales below h_0, the effect of lines has to be taken into consideration. Therefore, the new confidence interval for the CSR line patterns should be determined by $L(h)$ based on the position-and-form-based distance. In Fig. 4, this new confidence interval is represented by a blue dashed envelope, based on which the consistent conclusion on the generated line pattern could be drawn for all scales h, i.e., not significantly different from the CSR line patterns. In addition, the mean $L(h)$ of these generated CSR line patterns is calculated, which should correspond to the zero-value horizontal line, indicative of the expected $L(h)$ of a prefect CSR pattern. As presented in Fig. 3, the mean $L(h)$ for the CSR line pattern is always below that for the CSR point pattern.

In this section, we demonstrate the effect of the object form on traditional distance-based geographical analysis only by a simple case, i.e., the pattern analysis using L statistic. Similarly, the effect of form on all other distance-based methods and analyses for points, such as G function, F function, K function, and spatial auto-correlation analysis including Moran's I, Geary's C, and variogram, could be investigated. Furthermore, all these distance-based methods and analyses can be generalized for forms, simply replacing the point distance by our proposed position-and-form distance. In addition, GWR is a useful tool for exploring the local variation of spatial structures over space (Fotheringham et al., 1998; Leung et al., 2000a,b; Leung, 2010). In case studies, GWR is often applied to some regional data. However, the weight is usually determined according to the centroid-to-centroid distance. It is of interest to explore how the form can impact on the GWR local coefficients under the proposed position-and-form-based distance.

4 Volunteered Geographic Information

With the development of private mobile sensors and Web 2.0, the geographic researchers are more interested in an alternative source of data, i.e., Volunteered Geographic Information (VGI), because of its free and enrich content (Goodchild, 2007). However, the mechanism of VGI also raises the researchers' concern about the data quality, because no certification or expert knowledge of the data contributor is required (Flanagin and Metzger, 2008; Kounadi, 2009; Goodchild and Li, 2012).

As one of the most representative projects of VGI, OpenStreetMap (OSM) has been developed to a free and online editable map of the world in the past decade, where the researchers could freely access and download the data under an open data licence (Haklay and Weber, 2008). Meanwhile, many researchers are interested in assessment of the OSM data quality from the extrinsic and intrinsic aspect (Haklay et al., 2010; Senaratne et al., 2017). For the extrinsic quality assessment, the researchers always compare the OSM dataset with a standard reference dataset and in such cases, map matching is an essential pre-process for data comparison. The majority of the currently existing literature focus on the network matching (e.g. road network), whereas there are only few approaches for the area objects matching (Ruiz-Lendínez et al., 2017). Huh et al. (2013) developed a method to detect the point pairs corresponding to polygon object pairs, using a string matching method based on a confidence region model of a line segment. Fan et al. (2014) defined the correspondence among building footprints of OSM and ATKIS existent if the intersected area goes over 30 percent of the minimum footprints area. However, these methods cannot produce a high accurate result in the complicated high-density building region, because the OSM building footprints data is generally produced by manually digitalizing the Bing map images. Such digitalization may introduce the the distortion-caused random offset to the OSM data. In these cases, the overlapping area is inconvincible to accurately define the correspondence.

In this section, the proposed distance is introduced to the matching case and the matching result will be compared with the overlapping area method result. The experiment data is a subset of OSM building data in kowloon (the highlight region in Fig. 5), and the iC1000 Digital Land Boundary Map (Hong Kong SAR Authority digital map) is introduced as reference data for comparsion. For the simplicity, the OSM data and the reference data will be termed as OSM building and STD buidling repectively, and their centroid of gravity will be termed as OSM centroid and STD centroid repectively. Fig. 6 is the details of the experiment region, and the matching pair is manually labeled as O_i and S_i firstly to show the true correspondence.

Fig. 6 shows a general matching error when we employ the matching method proposed by Fan et al. (2014). $\mathrm{O}_0, \mathrm{O}_1, \mathrm{O}_2, \mathrm{O}_3$ and O_7 have the similar shape with $\mathrm{S}_0, \mathrm{S}_1, \mathrm{S}_2, \mathrm{S}_3$ and S_7 respectively, but there are some different offsets between each corresponding pair. Such offset decreases their overlapping below the corresponding standard, i.e., the defined 30 percent of the minimum footprint area. However, each of these pairs of footprints obviously represent the same ground truth. Besides, O_6 should be matched to S_6 according to the position and shape. However, it has a bigger intersection with S_7 thus would be matched to S_7.

As aforementioned, position and form are the most characterisation of objects in space. We attempt to correctly match the OSM footprints to their correspondent standard reference objects based on the position and form information. On one hand, we utilize their position-based information for match of the OSM footprints. The matching result is almost prefect expect that O_6 is miss-matched to S_7 (see Tab. 1). On the other hand, we use the form-form distance. Each OSM footprint matches to the correct standard reference footprint except O_2 and O_{12}, both of them matching to the wrong standard reference footprints S_1 and S_8 respectively. In addition, all of O_1, O_3, O_{10} and O_{11} match to more than one standard reference footprints, but the true correspondent reference footprint is included. As the matter of fact, the overlapping information is also determined by their position and form but may be too sensitive to offset of the OSM objects. However, even if the overlapping area is too small to meet the matching standard, the correctly matched pairs of the OSM and STD object may still have the small distance showing their similarity larger than other pairs.

The situation in the case asks to consider the matching from position and form together because the position-position distance could provides correct matching pair for the matching error in the form-form distance and vice versa. For example, position-based information miss-match O_6 to S_7, but form-based information match O_6 to S_6 correctly. In this way, the position-based and form-based information could be the additional information to each other for a more accurate matching result. It means the position-and-form distance proposed in the paper could provides a useful measure to characterize the distance between each pair of the footprints in OSM building and STD building. In the proposed distance, the paremeter c is defined as 0.1 because the position imformation and form information are both important to characterize the distance in this case. Actually, Tab. 1 shows that the position-and-form distance could capture all the correspondence in Fig. 6 correctly.

5 Conclusions

Geographical analysis often involves the distance between objects in space. The most basic characterization of an object is its position and form. Distance between objects is a typical relation in the metric space or a measure of dissimilarity of objects. Based on the positions of two objects, we can measure their distance. Such distance is essentially a position-based measure of how far away one object is from the other. Conventionally, the representative point of an object is its centroid. Distance between the centroids has been extensively used to analyze spatial relationships of objects according to their positions in space. However, such distance measure does not consider the form of an object which is also a basic characterization of an object in space and usually bears significant information about its function. Therefore, distance based solely on the positions of the representative points of objects suffers from a significant loss of information associated with their forms. Hence, an ideal distance for distance-based geographical analysis is the one that takes both position and form of the objects as its constituent components. This is exactly what the existing notions of distance fail to capture. Without considering both position and form, the relationships that we are unravelling may be incomplete or dubious.

In this study, we have proposed a notion of distance explicitly incorporating information about both position and form of objects. At conceptual level, we formulate a position-and-form-based distance combined the position-distance and form-distance as its constituent components. With such proposed distance notion, we can examine the composite and individual effect of position and form on a relationship. We conceptually demonstrate the correctness of the position-and-form-based distance. Similarly, the object attributes other than form could be further integrated into proposed distance notion.

In methodology, we show the form effect on existing geographical analyses based on only position-based distance and how to generalize these analyses and relevant methods for study of objects with forms based on the proposed position-and-form-based distance, using the pattern analysis by L statistic as an example.

In application, we utilize the position-and-form-based distance to perfectly solve a matching problem of the OSM objects to the STD objects, which cannot be handled by the traditional methods based on the overlapping areas. Actually, the position-and-form-based distance is a potential tool for more real-life case studies. For example, the variation in economic attributes such as the gross domestic product of cities an important research topic (Huang and Leung, 2002). Although each city is an areal unit surrounded by its administrative boundary, the distance employed to determine the regression coefficients in GWR is usually between the city centroids (Huang and Leung, 2002). It is of interest to explore the dependence of interested variables on the form of cities can impact on the GWR local coefficients using the proposed position-and-form-based distance. Another research topic is the relationship between the form and function for cities and buildings, which has been recognized in relevant fields (Alexander, 1964; Batty and Longley, 1994; Batty, 2007, 2008, 2013; Nasar et al., 2005). It is possible to investigate whether the similarity in function well corresponds to the similarity in form measured by the form-based distance. In addition, a road network can be treated as the skeleton of a city. Its spatial structure is closely related to its service efficiency. Essentially, the spatial structure is based on the form of lines. In the dimension of space, spatial analyses based on the position-and-form-based distance between roads can provide a new perspective to connect form and function. In terms of time, the evaluation of a road network can be done by detecting it changes along with time using the position-and-form-based distance.

In summary, the proposed position-and-form-based distance and the associated methods could give us a new perspective on the conceptualization of distance. Actually, it can also be further extended to include other object attributes into the measure. Therefore, it is an ideal notion of distance that can fully reveal the multi-facet nature of geographical relations. The proposed research will advance the frontier of theoretical and applied research in geography where distance plays an important role.

References

Abbas, I., P. Grussenmeyer, and P. Hottier (1995). Contrôle de la planimétrie dune base de données vectorielle: une nouvelle méthode basée sur la distance de hausdorff: la méthode du contrôle linéaire. *Bull. SFPT 1*(137), 6–11.

Alexander, C. (1964). *Notes on the synthesis of form*. Harvard University Press.

Alt, H., U. Fuchs, G. Rote, and G. Weber (1996). Matching convex shapes with respect to the symmetric difference. In *European Symposium on Algorithms*, pp. 320–333. Springer.

Alt, H. and M. Godau (1995). Computing the Fréchet distance between two polygonal curves. *Int. J. Comput. Geom. Appl. 5*(01n02), 75–91.

Arkin, E. M., L. P. Chew, D. P. Huttenlocher, K. Kedem, and J. S. Mitchell (1989). An efficiently computable metric for comparing polygonal shapes. Technical report, DTIC Document.

Baddeley, A., E. Rubak, and R. Turner (2015). *Spatial point patterns: Methodology and applications with R*. CRC Press.

Batty, M. (2007). *Cities and complexity: Understanding cities with cellular automata, agent-based models, and fractals*. The MIT press.

Batty, M. (2008). The size, scale, and shape of cities. *Science 319*(5864), 769–771.

Batty, M. (2013). *The new science of cities*. Mit Press.

Batty, M. and P. A. Longley (1994). *Fractal cities: A geometry of form and function*. Academic Press.

Boots, B. N. and A. Getis (1988). *Point pattern analysis*. California: Sage Publications.

Clark, J. W. (1977). Time-distance transformations of transportation networks. *Geogr. Anal. 9*(2), 195–205.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Davis, L. (1986). Two-dimensional shape representation. In I. Young and K. S. Fu (Eds.), *Handbook of Pattern Recognition and Image Processing*, pp. 233–245. New York: Academic Press.

Diggle, P. J. (2003). *Statistical analysis of spatial point patterns* (2nd ed.). London: Arnold.

Falconer, K. (1990). *Fractal geometry: mathematical foundations and applications*. Wiley New York.

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

Feder, J. (1988). *Fractals*. Plenum Press New York.

Flanagin, A. J. and M. J. Metzger (2008). The credibility of volunteered geographic information. *GeoJournal 72*(3-4), 137–148.

Fotheringham, A. S., M. E. Charlton, and C. Brunsdon (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environ. Plan. A 30*(11), 1905–1927.

Fu, K. S. (1974). *Syntactic methods in pattern recognition*. Elsevier.

Gatrell, A. C. (1983). *Distance and space: A geographical perspective*. Oxford University Press.

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *Inc. Stat. 5*(3), 115–146.

Golledge, R. G., R. Briggs, and D. Demko (1969). The configuration of distances in

intra-urban space. *Proc. Assoc. Am. Geogr. 1*(9), 60–66.
Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal 69*(4), 211–221.
Goodchild, M. F. and G. J. Hunter (1997). A simple positional accuracy measure for linear features. *Int. J. Geogr. Inf. Sci. 11*(3), 299–306.
Goodchild, M. F. and L. Li (2012). Assuring the quality of volunteered geographic information. *Spatial statistics 1*, 110–120.
Griffith, D. A. (2009). Celebrating 40 years of scientific impacts by Cliff and Ord. *Geogr. Anal. 41*(4), 343–345.
Haklay, M., S. Basiouka, V. Antoniou, and A. Ather (2010). How many volunteers does it take to map an area well? the validity of linus law to volunteered geographic information. *The Cartographic Journal 47*(4), 315–322.
Haklay, M. and P. Weber (2008). Openstreetmap: User-generated street maps. *IEEE Pervasive Computing 7*(4), 12–18.
Hall, E. T. (1959). *The silent language*, Volume 3. Connecticut: Premier Books.
Hall, E. T. (1966). *The hidden dimension* . New York: Doubleday & Co.
Haralick, R. M., S. R. Sternberg, and X. Zhuang (1987). Image analysis using mathematical morphology. *IEEE Trans. Pattern Anal. Mach. Intell. 9*(4), 532–550.
Huang, Y. and Y. Leung (2002). Analysing regional industrialisation in jiangsu province using geographically weighted regression. *J. Geogr. Syst. 4*(2), 233–249.
Huh, Y., S. Yang, C. Ga, K. Yu, and W. Shi (2013). Line segment confidence region-based string matching method for map conflation. *ISPRS journal of photogrammetry and remote sensing 78*, 69–84.
Huttenlocher, D. P., G. A. Klanderman, and W. J. Rucklidge (1993). Comparing images using the hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell. 15*(9), 850–863.
Kashyap, R. and R. Chellappa (1981). Stochastic models for closed boundary analysis: Representation and reconstruction. *IEEE T. Inform. Theory. 27*(5), 627–637.
Kounadi, O. (2009). Assessing the quality of openstreetmap data. *Msc geographical information science, University College of London Department of Civil, Environmental And Geomatic Engineering*.
Latecki, L. J. and R. Lakamper (2000). Shape similarity measure based on correspondence of visual parts. *IEEE Trans. Pattern Anal. Mach. Intell. 22*(10), 1185–1190.
Laumann, E. O. (1966). *Prestige and association in an urban community: An analysis of an urban stratification system.* Bobbs-Merrill Company.
Laumann, E. O. (1973). *Bonds of pluralism: The form and substance of urban social networks.* New York: Wiley-Interscience.
Leung, Y. (2010). *Knowledge discovery in spatial data.* Springer.
Leung, Y., C.-L. Mei, and W.-X. Zhang (2000a). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environ. Plan. A 32*(1), 9–32.
Leung, Y., C.-L. Mei, and W.-X. Zhang (2000b). Testing for spatial autocorrelation among the residuals of the geographically weighted regression. *Environ. Plan. A 32*(5), 871–890.
Leung, Y., D. Meng, and Z. Xu (2013). Evaluation of a spatial relationship by the concept of intrinsic spatial distance. *Geogr. Anal. 45*(4), 380–400.
Ling, H. and D. W. Jacobs (2007). Shape classification using the inner-distance. *IEEE Trans. Pattern Anal. Mach. Intell. 29*(2), 286–299.
Liu, H.-C. and M. D. Srinath (1990). Partial shape classification using contour matching

in distance transformation. *IEEE Trans. Pattern Anal. Mach. Intell. 12*(11), 1072–1079.

Lloyd, C. D. (2014). *Exploring spatial scale in geography.* John Wiley & Sons.

Loncaric, S. (1998). A survey of shape analysis techniques. *Pattern Recognit. 31*(8), 983–1001.

Lowe, J. C. and S. Moryadas (1975). *The geography of movement.* London: Houghton Mifflin.

Lowrey, R. A. (1973). A method for analyzing distance concepts of urban residents. In R. M. Downs and D. Stea (Eds.), *Image and environment*, pp. 338–360. Chicago: Aldine Press.

Miller, H. J. and E. A. Wentz (2003). Representation and spatial analysis in geographic information systems. *Ann. Assoc. Am. Geogr. 93*(3), 574–594.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika 37*(1/2), 17–23.

Mount, D. M., R. Silverman, and A. Y. Wu (1996). On the area of overlap of translated polygons. *Comput. Vis. Image Underst. 64*(1), 53–61.

Muller, J.-C. (1978). The mapping of travel time in edmonton, alberta. *Can. Geogr. 22*(3), 195–210.

Nasar, J. L., A. E. Stamps III, and K. Hanyu (2005). Form and function in public buildings. *J. Environ. Psychol. 25*(2), 159–165.

Okabe, A. and H. J. Miller (1996). Exact computational methods for calculating distances between objects in a cartographic database. *Cartogr. Geogr. Inf. Syst. 23*(4), 180–195.

Pavlidis, T. (2012). *Algorithms for graphics and image processing.* Springer Science & Business Media.

Pentland, A. P. (1984). Fractal-based description of natural scenes. *IEEE Trans. Pattern Anal. Mach. Intell. 6*(6), 661–674.

Peuquet, D. J. (1992). An algorithm for calculating minimum euclidean distance between two geographic features. *Comput. Geosci. 18*(8), 989–1001.

Prokop, R. J. and A. P. Reeves (1992). A survey of moment-based techniques for unoccluded object representation and recognition. *Graph. Models Image Process. 54*(5), 438–460.

Ruiz-Lendínez, J. J., M. A. Ureña-Cámara, and F. J. Ariza-López (2017). A polygon and point-based approach to matching geospatial features. *ISPRS International Journal of Geo-Information 6*(12), 399.

Senaratne, H., A. Mobasheri, A. L. Ali, C. Capineri, and M. Haklay (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science 31*(1), 139–167.

Skidmore, A. and B. Turner (1992). Map accuracy assessment using line intersect sampling. *Photogramm. Eng. Remote Sens. 58*(10), 1453–1457.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Econ. Geogr. 46*(sup1), 234–240.

Tveite, H. (1999). An accuracy assessment method for geographical line data sets based on buffering. *Int. J. Geogr. Inf. Sci. 13*(1), 27–47.

Veltkamp, R. C. and M. Hagedoorn (2001). State of the art in shape matching. In *Principles of visual information retrieval*, pp. 87–119. Springer.

Witkin, A. (1984). Scale-space filtering: A new approach to multi-scale description. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'84.*, Volume 9, pp. 150–153. IEEE.

Young, I. T., J. E. Walker, and J. E. Bowie (1974). An analysis technique for biological shape. I. *Inf. Control 25*(4), 357–370.

Zahn, C. T. and R. Z. Roskies (1972). Fourier descriptors for plane closed curves. *IEEE T. Comput. 100*(3), 269–281.

Fig 1: Illustrative procedure for the calculation of the position-and-form-based distance between two objects X_1 and X_2.

Fig 2: Illustrative example of a point pattern (left panel) and the corresponding line pattern with their centroids locating on the points (right panel).

Fig 3: Illustrative example to show the form effect on the pattern analysis using lines as an example.

Fig 4: Effect of the form of lines on the pattern analysis with respect to L statistic quantified by an effect indicator about -0.01, defined as the mean difference between $L(h)$ on the basis of centroid-to-centroid distance and position-and-form-based distance for $h > 0.1$.

Fig 5: The position of experiment data in Kowloon OSM footprints data where the buildings have been highlighted.

Fig 6: The subset of OSM building footprints data (OSM building) and the correspondent standard reference data (STD building) with their related centroid of gravity, OSM centroid and STD centroid.

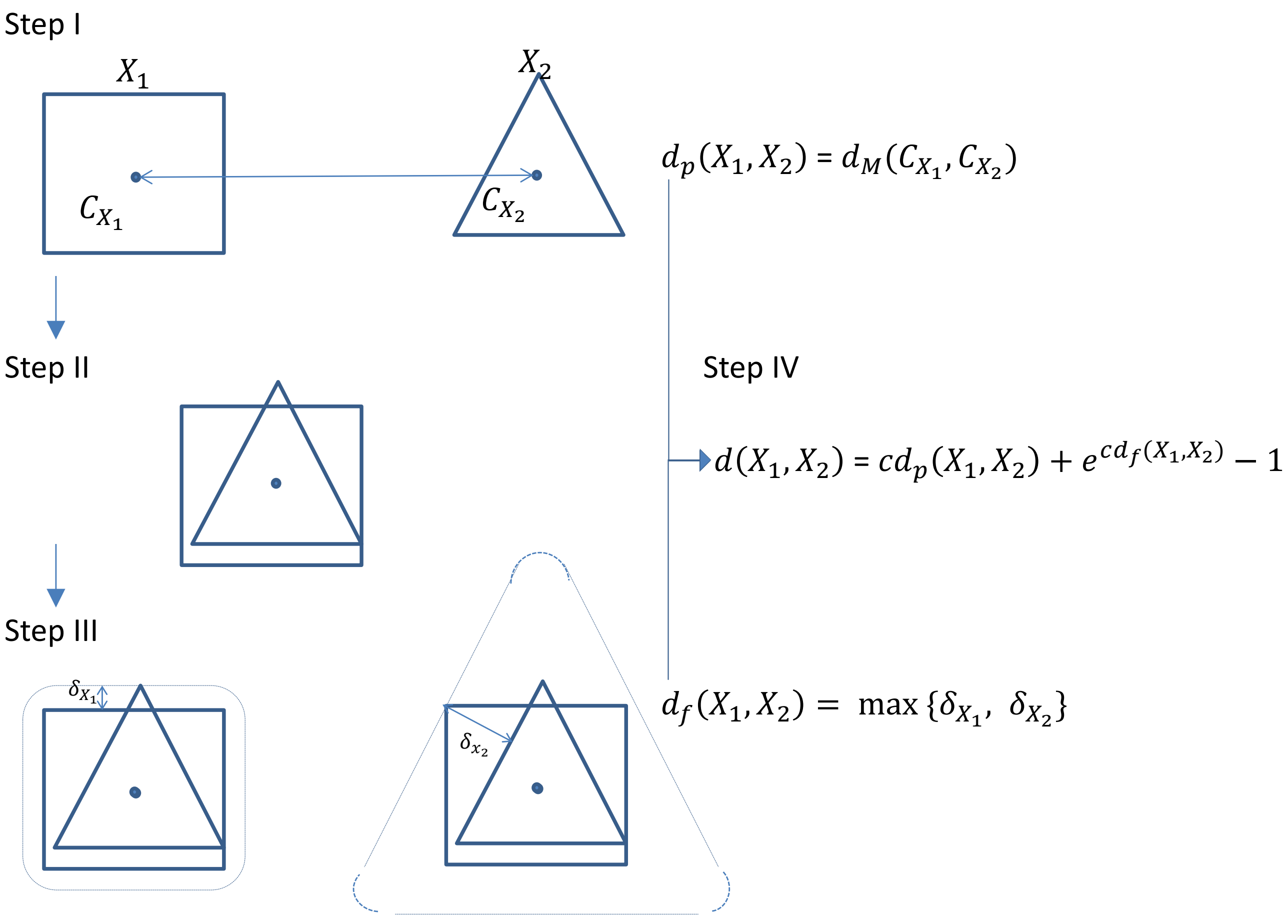

Figure 1:

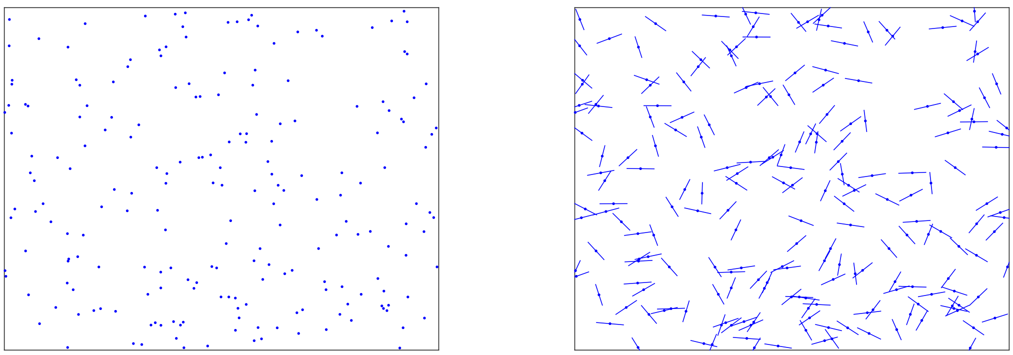

Figure 2:

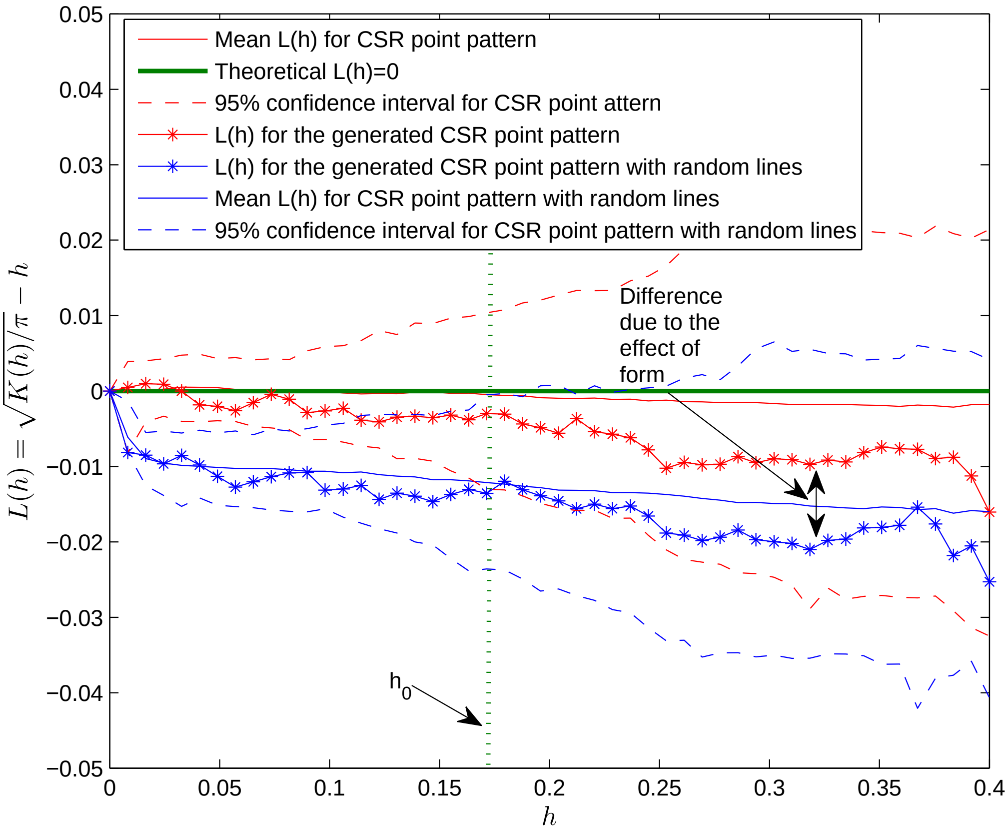

Figure 3:

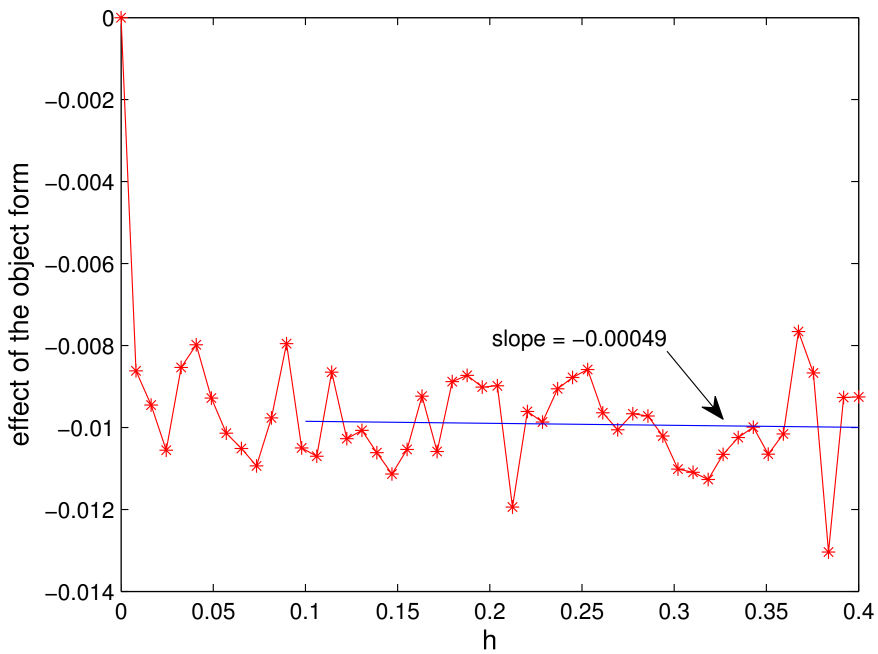

Figure 4:

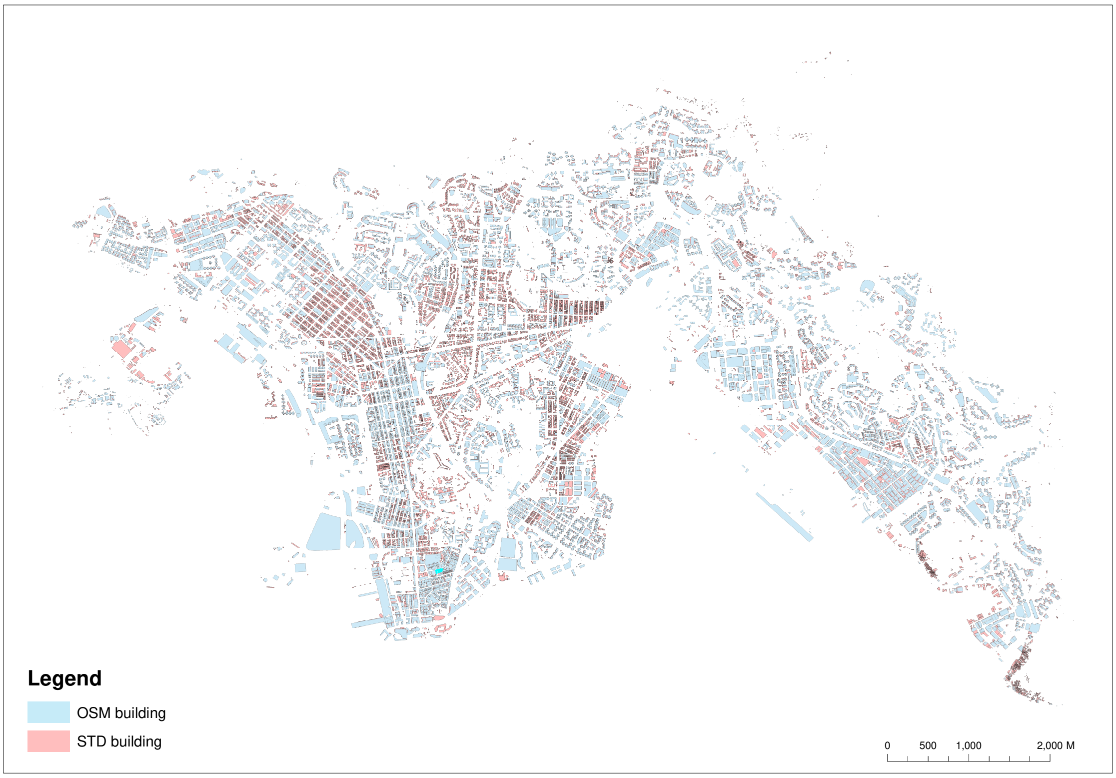

Figure 5:

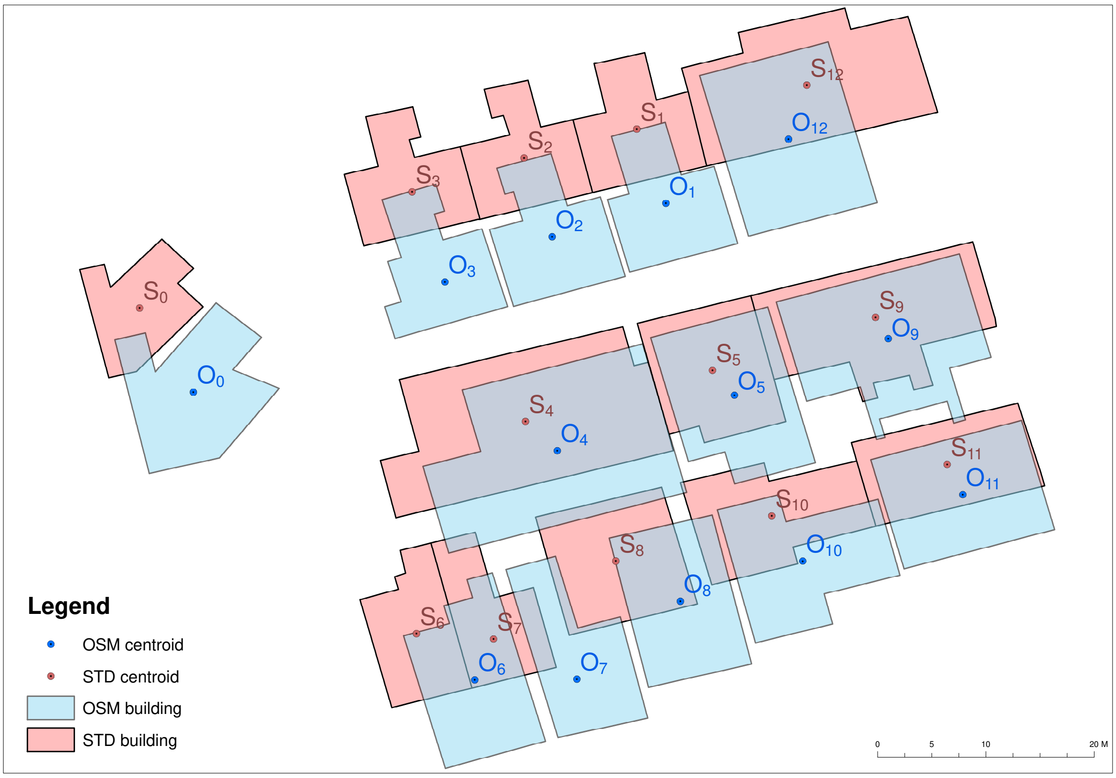

Figure 6:

Table 1: The matching results of three kinds of distance, the position-position distance, the form-form distance and the position-form distance.(Note: the bold number is the minimum value of each row which means the OSM object on the row is closest to the STD object on the column of the minimum value based on the specific distance.)

The position-position distance

	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
O_0	**9.02**	44.97	35.55	26.55	28.49	44.38	29.31	32.80	39.30	58.68	50.70	64.73	59.55
O_1	45.98	**7.27**	12.80	22.05	23.43	15.92	45.05	41.03	33.27	20.79	30.23	34.04	16.24
O_2	35.85	12.28	**7.64**	13.19	17.20	18.43	38.40	35.51	30.40	28.63	31.87	39.76	25.87
O_3	25.68	21.34	12.57	**8.09**	16.07	25.37	33.78	32.57	31.06	37.75	36.63	47.21	35.81
O_4	38.03	30.44	27.15	27.68	**3.82**	15.21	20.71	16.39	11.34	29.86	19.29	33.34	39.90
O_5	51.44	25.91	28.31	33.97	18.00	**2.96**	34.94	29.21	18.34	14.05	11.58	19.28	29.26
O_6	44.68	52.65	48.32	46.07	24.22	35.04	6.56	**5.92**	16.30	47.86	29.54	44.98	61.75
O_7	49.11	48.82	46.08	45.74	21.98	28.79	13.71	**8.03**	9.37	40.38	21.15	36.32	56.18
O_8	53.52	43.69	42.99	44.92	21.20	21.47	22.73	16.27	**6.63**	31.04	11.09	26.05	48.80
O_9	64.00	28.87	35.26	43.37	31.90	15.27	48.61	42.59	31.01	**2.22**	19.14	12.67	24.37
O_{10}	61.26	42.28	44.13	48.39	26.94	19.20	33.65	27.15	15.95	23.33	**4.92**	15.23	43.89
O_{11}	72.37	43.70	48.64	55.32	37.93	24.23	48.37	41.87	30.23	17.95	16.40	**3.06**	40.03
O_{12}	57.56	12.96	22.63	32.71	34.38	22.28	55.58	50.92	41.61	18.06	34.78	32.95	**5.22**

The form-form distance

	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
O_0	**2.13**	3.24	3.56	3.30	7.49	2.31	3.87	4.37	2.39	4.54	4.34	3.49	4.80
O_1	3.16	**1.03**	1.29	1.05	8.43	3.04	2.87	1.56	3.54	5.99	3.90	3.77	5.92
O_2	3.30	**1.11**	1.36	1.12	8.20	3.00	3.04	1.81	3.26	5.52	3.51	3.46	5.54
O_3	3.30	**1.30**	1.39	1.32	8.21	2.98	3.00	2.05	3.32	6.65	3.96	4.37	6.69
O_4	7.55	7.22	7.28	7.22	**3.08**	6.77	7.76	8.41	6.07	3.86	5.29	4.99	3.73
O_5	5.44	3.60	3.88	3.64	7.81	**1.69**	3.21	4.11	2.31	4.91	4.59	3.96	5.10
O_6	4.41	2.96	3.08	3.05	8.42	3.34	**1.61**	3.57	3.56	5.68	5.58	5.20	5.84
O_7	5.01	3.04	3.09	2.95	9.56	3.44	4.71	**1.38**	3.64	6.42	4.36	5.28	6.77
O_8	4.55	2.50	2.75	2.55	8.36	2.27	3.05	3.91	**1.57**	5.38	3.96	3.17	5.72
O_9	6.09	4.97	4.87	4.95	5.28	3.27	4.90	5.76	3.37	**2.19**	5.54	4.97	3.79
O_{10}	3.30	3.84	4.15	3.89	7.14	2.85	4.39	4.06	2.22	4.04	**1.52**	2.45	4.37
O_{11}	4.06	3.51	3.46	3.52	6.45	1.92	3.46	4.36	**1.17**	3.24	2.27	1.41	3.59
O_{12}	5.68	4.07	4.30	4.07	7.35	3.17	3.89	4.86	**2.22**	4.21	4.71	3.86	4.57

The position-form distance

	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
O_0	**1.14**	4.88	3.98	3.05	3.96	4.70	3.40	3.83	4.20	6.44	5.61	6.89	6.57
O_1	4.97	**0.84**	1.42	2.32	3.67	1.95	4.84	4.27	3.75	2.90	3.50	3.86	2.43
O_2	3.98	1.35	**0.91**	1.44	2.99	2.19	4.19	3.75	3.43	3.60	3.61	4.39	3.33
O_3	2.96	2.27	1.41	**0.95**	2.88	2.88	3.73	3.48	3.50	4.72	4.15	5.27	4.53
O_4	4.93	4.10	3.79	3.83	**0.74**	2.49	3.24	2.96	1.97	3.46	2.63	3.98	4.44
O_5	5.87	3.03	3.30	3.84	2.98	**0.48**	3.87	3.43	2.09	2.04	1.74	2.41	3.59
O_6	5.02	5.61	5.19	4.96	3.74	3.90	**0.83**	1.02	2.06	5.55	3.70	5.18	6.97
O_7	5.56	5.24	4.97	4.92	3.80	3.29	1.97	**0.95**	1.38	4.94	2.66	4.33	6.59
O_8	5.93	4.65	4.62	4.78	3.43	2.40	2.63	2.11	**0.83**	3.82	1.59	2.98	5.65
O_9	7.24	3.53	4.15	4.98	3.89	1.91	5.49	5.04	3.50	**0.47**	2.65	1.91	2.90
O_{10}	6.52	4.70	4.93	5.31	3.74	2.25	3.92	3.22	1.84	2.83	**0.66**	1.80	4.94
O_{11}	7.74	4.79	5.28	5.95	4.70	2.64	5.25	4.73	3.15	2.18	1.90	**0.46**	4.43
O_{12}	6.52	1.80	2.80	3.77	4.52	2.60	6.03	5.72	4.41	2.33	4.08	3.77	**1.10**