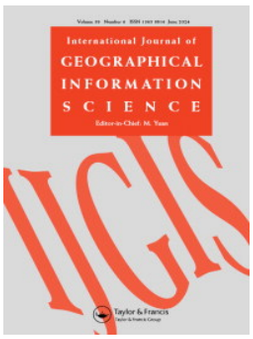


International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tgis20

# Location retrieval using qualitative place signatures of visible landmarks

Lijun Wei, Valérie Gouet-Brunet & Anthony G. Cohn

**To cite this article:** Lijun Wei, Valérie Gouet-Brunet & Anthony G. Cohn (10 May 2024): Location retrieval using qualitative place signatures of visible landmarks, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2024.2348736

**To link to this article:** https://doi.org/10.1080/13658816.2024.2348736

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2024.2348736

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

# Location retrieval using qualitative place signatures of visible landmarks

Lijun Wei[a], Valérie Gouet-Brunet[b] and Anthony G. Cohn[a,c,d]

[a]School of Computing, University of Leeds, United Kingdom; [b]LaSTIG, IGN-ENSG, Gustave Eiffel University, France; [c]Department of Computer Science and Technology, Tongji University, Shanghai, China; [d]The Alan Turing Institute, United Kingdom

**ABSTRACT**

Location retrieval based on visual information is to retrieve the location of an agent (e.g. human, robot) or the area they see by comparing their observations with a certain representation of the environment. Existing methods generally treat the problem as a content-based image retrieval problem and have demonstrated promising results in terms of localization accuracy. However, these methods are challenging to scale up due to the volume of reference data involved; and the image descriptions might not be easily understandable/communicable for humans to describe surroundings. Considering that humans often use less precise but easily produced qualitative spatial language and high-level semantic landmarks when describing an environment, a coarse-to-fine qualitative location retrieval method is proposed in this work to quickly narrow down the initial location of an agent by exploiting the available information in large-scale open data. This approach describes and indexes a location/place using the perceived qualitative spatial relations between ordered pairs of co-visible landmarks from the perspective of viewers, termed as '*qualitative place signatures*' (QPS). The usability and effectiveness of the proposed method were evaluated using openly available datasets, together with simulated observations by considering different types perception errors.

**ARTICLE HISTORY**
Received 26 July 2022
Accepted 24 April 2024

**KEYWORDS**
Place recognition; qualitative location; qualitative spatial relation; place descriptor; place signature

## 1. Introduction

The definition of a *location* or a *place* varies depending on the context and scale of applications. For instance, it can be a geographical name like Place Dauphine, Paris either with a crisp or a rough boundary (Bittner and Stell 2000). Alternatively, it can be an area (Kuipers 2000), a 2D/3D linestring or a zero-dimensional 2D/3D point on a map, or a point with the agent's pose attached in the field of Robotics (Irschara *et al.* 2009, Sattler *et al.* 2012, Kendall *et al.* 2015). The goal of location retrieval is to identify

**CONTACT** Lijun Wei villager5whu@gmail.com

Supplemental data for this article can be accessed online at https://doi.org/10.5518/1506.

the corresponding location and/or pose of an agent in an environment based on certain sensing modalities along with the existing knowledge of the environment. Satellite-based positioning systems such as GPS are the most widely used global positioning approaches. In situations where accurate satellite positioning information is not available, sensing modalities such as visual information have been explored for localization (Korrapati *et al.* 2012, Zamir *et al.* 2016, Piasco *et al.* 2018, Pion *et al.* 2020) or place recognition (Lowry *et al.* 2016, Ali-Bey *et al.* 2023). Existing methods often use low-level visual features such as 2D hand crafted/learnt corners, edges to represent geo-tagged images Durrant-Whyte and Bailey 2006, Jegou et al. 2010, Korrapati *et al.* 2012, Chen *et al.* 2017, Zhang *et al.* 2021, or 3D features from 3D point clouds thanks to the widespread use of LiDAR scanners (Uy and Lee 2018, Luo *et al.* 2024). Other solutions use more abstract semantic representations of the environment such as object categories (Lamon *et al.* 2001, 2003, Ardeshir *et al.* 2014, Li *et al.* 2014, Schlichting and Brenner 2014, Panphattarasap and Calway 2016, Zang *et al.* 2017, Hery *et al.* 2021). Compared to low-level features, semantic features are easier to communicate and may provide a more robust representation of the environment, particularly in the face of changing conditions such as occlusions and the change of seasons and viewpoints. While there are recent contributions focusing on addressing these challenging conditions (Piasco *et al.* 2021), they generally treat the localisation problem as a content-based image retrieval problem through learning distinctive image descriptors without considering the semantic meaning of the learnt descriptions.

In fact, humans often use a mental map of the environment to locate themselves and to navigate to a destination. We rely on high-level semantic objects and less precise but easily produced and understood spatial language to describe our surroundings and communicate our locations (Tversky 1993, Chen *et al.* 2013). In these descriptions, those semantic objects with known or relatively better known locations are selected as the *landmarks* (Sadalla *et al.* 1980) or *anchor points* (Couclelis *et al.* 1987) to define the locations of adjacent points. The descriptions of the spatial relationships between observed landmarks, or between an observer and their surrounding landmarks can help us understand the overall structure of the environment, which is especially useful for navigating in more open and/or less structured environments.

In this study, we propose an upstream approach to quickly narrow down the search area of an agent's initial location by exploiting the available information in large-scale open data, and describing and retrieving locations using *qualitative place signatures (QPS)*. *QPS* are defined as the perceived qualitative spatial relationships between ordered pairs of co-visible landmarks from a location, including the ordered sequence of landmark types, their relative orientations, and the qualitative angles in between. Note that even though individual landmarks might not be identifiable, when multiple co-visible landmarks form a place signature they might be used to identify a location. Based on this definition, a space division method is proposed to automatically divide the navigable space into distinct locations, or 'place cells' such that each cell is attached with a unique place signature. A coarse-to-fine location retrieval method is then used to efficiently identify the possible location(s) of a viewer based on their observations using approximate hashing. It should be noted that our strategy does not intend to replace other localization approaches, but rather to be complementary to quickly narrow down the initial search area.

### 1.1. Related work

A location can either be described using its spatial relations with respect to a landmark(s) in a fixed reference frame, for example, the 'the building is to the east of the central station'; or, using the perceived spatial relations between landmarks from the perspective of an agent, for example, 'I can see a church next to a tower in front of me'. These two modes form the foundation of human spatial mental models (Tversky 1993), and have both been studied for representing locations qualitatively in applications such as navigation (Wang *et al.* 2005, Fogliaroni *et al.* 2009) and spatial information queries (Yao and Thill 2006). More specifically,

- For methods with a fixed reference frame, disjoint points (Clementini *et al.* 1997), polygons, regions (Bittner and Stell 2000) or the combination of these geometries (Du *et al.* 2015) have been used to divide space into non-overlapping areas, enabling spatial inference using relevant direction, distance or topological relations (Cohn *et al.* 2014, Freksa *et al.* 2018). For example, Wang *et al.* (2005) proposed to describe the qualitative position of target objects using their cardinal directions, such as $\{E, W, S, N, NE, NW, SE, SW, O\}$ (Frank 1991, Egenhofer *et al.* 1999) in relation to their corresponding landmarks determined by a Voronoi model.
- For methods without a fixed reference frame, Levitt and Lawton (1990) proposed to divide the environment into regions using the lines connecting pairs of point landmarks such that the same order of landmarks can be perceived by agents from any locations within each region. This method was improved on by Schlieder (1993) to differentiate between adjacent regions where a same order of landmarks is observed by augmenting the order of landmarks with their complementary directions. Fogliaroni *et al.* (2009) proposed a similar approach although the decomposition of space was based on the extended convex landmarks instead of points. Places were also represented using qualitative distances to landmarks such as [*very close, close, medium, far, very far*] (Wagner *et al.* 2004) and nearness relations identified through data mining (Duckham and Worboys 2001).

Both of these methods generally assume landmarks can be correctly, completely, and uniquely identified by agents, the locations, or the *ids* of landmarks are known, and the initial global positions of agents are usually given if used in navigation applications. However, visual perceptions are prone to errors either due to the environmental or internal factors, and the initial position of agents may be unknown. Moreover, although various theoretical models were proposed to identify qualitative locations, the scales of existing experiments are generally small with limited number of landmarks. The scalability as well as the time complexity of these models were rarely investigated. Little work has been done in this area from the perspective of information retrieval. In this work, as shown in Figure 1, we propose to not use the explicit location of landmarks (Wang *et al.* 2005), but their relative locations and semantic information as such information is generally easy to capture compared to other more accurate measurement.

Among the large range of existing location retrieval techniques, the concept proposed by Weng *et al.* (2020) is the closest to our approach. However, (1) instead of sampling

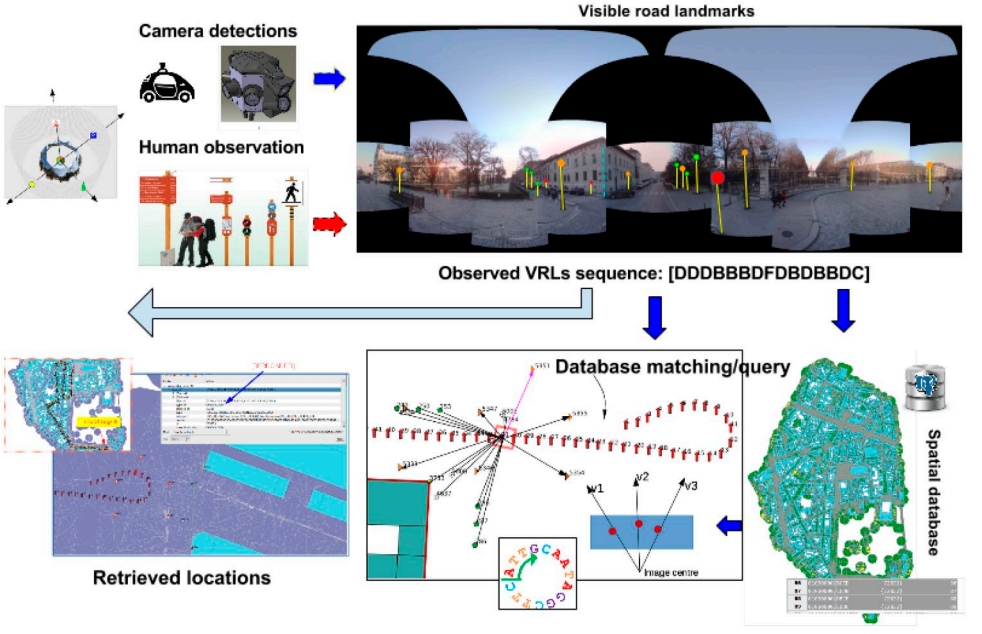


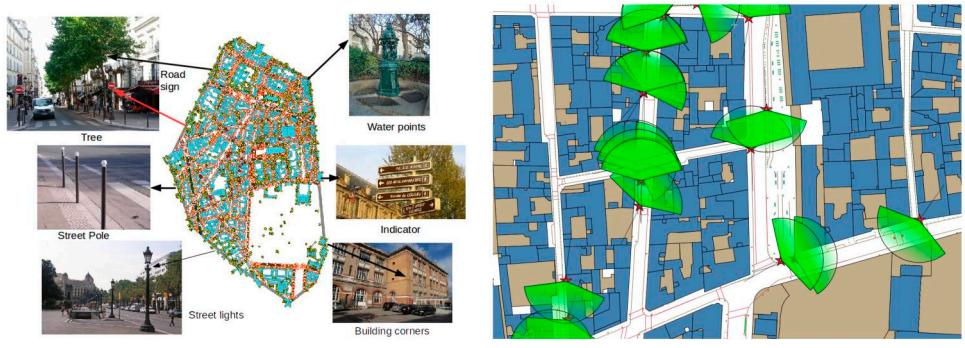

**Figure 1.** Demonstration of the proposed location retrieval method: given the perceived ordered sequences of visible road landmark (VRL) *types* and other qualitative spatial relations by users (or images), the location retrieval problem is to find the reference place cells with the most similar place signature to the observed one.

locations using 10x10m regular grids, we divide navigable space into distinct locations (i.e. place cells) following the definition of individual qualitative spatial relations, ensuring that consistent spatial relations can be observed by agents from anywhere inside each place cell; (2) instead of computing the direction of landmarks with respect to the *True North* which is not always feasible to judge, we consider the relative angles between the lines of sight of ordered landmark pairs; (3) we consider the possible occlusions of landmarks by other objects when creating place signatures; and (4) when comparing place signatures, instead of using distance measures under an exhaustive searching strategy, we propose a coarse-to-fine location retrieval method by using an approximate hashing technique to improve the retrieval efficiency. A detailed discussion on time complexity will be given in Sections 4.1 and 5.3.

In the remainder of this paper, the proposed qualitative place signature is presented in Section 3 and the location retrieval method in Section 4; experimental results are given in Section 5, followed by discussion in Section 6 and conclusion in Section 7.

## 2. Landmarks

In this work, landmarks are defined as distinctive, static, stable and easily recognisable objects in an environment, such as road signs (Soheilian *et al.* 2013) and street lights in an urban environment, and mountain ridges in rural areas. Examples of urban landmarks are shown in Figure 2(a).




(a) Different types of visible road landmarks (in Paris) (b) The visible zones of road landmarks.

**Figure 2.** Examples of urban landmarks and their visibility zones, which are not always full circles since some landmarks may have an inner front. For example, traffic lights usually have a designed inner front that are expected to be seen by road users, while a tree may not have a front or back. Pictures are from Google images by searching 'Paris'.

Though the appearance of certain landmarks may change in a regular or an irregular way, for example, most trees change colour and forms from spring to winter, their locations are mostly static and they can be easily identified by humans using descriptions with a strong semantic meaning that is informative and robust to visual changes. The location information of these landmarks is often readily available, either being automatically reconstructed from optic sensor data captured by mobile mapping systems, manually annotated by human surveyors, or sourced from various open data services, such as the *Ordnance Survey Roadside Asset Data Services* and *Find Open Data* in the United Kingdom, the *OpenDataParis* dataset in France, and crowd-sourced maps like *OpenStreetMap* (Rousell *et al.* 2015). In terms of volume, maps of landmarks are also representing the environment in a much more compact and refined way compared to other kinds of more commonly used data sources, such as images or LiDAR. Thanks to the fact that these open-sourced initiatives are now widespread and describing spatial regions on a large scale, it would be interesting to exploit their usage in applications like large-scale location retrieval.

In this work, each landmark $S_i$ has an attached set of attributes, written as:

$$S_i = (id, type, type\_id; centre(x, y), contour_{2d}, visible\,zone) \tag{1}$$

where *id* is the unique index of a landmark in a database that is usually unknown to agents; *type* is the category of a landmark (e.g. tree); *type_id* is a character encoding such category (e.g. '*J*' for *tree*); *centre(x,y)* captures the 2D coordinates of a landmark's centroid in a geographic system, $contour_{2d}$ captures its 2D extent, and *visible_zone* captures the area from where this landmark can be perceived. The *default* visibility zone is a circle with a varied radius for individual landmarks (Figure 2(b)), affected by factors such as the location and intrinsic direction of a landmark (if there is any), its size, height and visual salience (i.e. the perceptual quality which makes some items stand out from their neighbors), occlusion caused by other objects, observers' eyesight and height, and the weather and lighting conditions (e.g. day/night).

More detailed visual (e.g. color, shape, text, material), semantic (e.g. the type of a tree) or spatial attributes of landmarks can also be added into the above list to further

improve the discriminating ability of a landmark if such information cannot be inferred from the general 'type'. For a landmark $T$ with an intrinsic front/back, the circular visibility area can be separated into front and back half-circles and treated as belonging to two different landmarks. The orientation of such landmarks can either be collected by on-site survey, reconstructed from sensing data, or inferred from the landmark type, direction of nearby road networks, or existing standards on infrastructure installation. For example, traffic signs normally face oncoming traffic except that those indicating on-street parking controls are parallel to the edge of the carriageway, and some flag-type direction signs are pointing approximately in the direction to be taken. Methods for how to determine which landmarks are visible from a particular viewpoint will be given in Section 3.5 and detailed in the supplementary material.

## 3. Place signatures based on the qualitative spatial relations between visible landmarks

In this section, qualitative place signatures (*QPS*) are introduced to describe the spatial configuration of visible road landmarks from viewers' perspective, including their order of appearance, the relative orientations, and the qualitative angles between the lines of sight of ordered pairs of landmarks. Based on this definition, a study space can be divided into distinct reference place cells such that the same *QPS* can be observed by agents from anywhere within each cell. Then, given a viewer's new observation, their location can be retrieved by finding the place cell(s) with the best matched reference signatures.

In this section, the three types of qualitative spatial relations are respectively introduced in Section 3.1 to 3.4, followed by practical steps for creating and maintaining a reference database and discussions on the impact of landmarks uncertainty in Section 3.5.

### 3.1. The viewing order of visible landmarks on a panorama

As landmarks seen from a particular viewpoint appear as if they are overlaying on the surface of a sphere centered on the viewer's eyes (Figure 3) or on the image plane of a panoramic camera (Galton 1994), the first component of our proposed place signature is the ordered sequence of the types of visible landmarks seen from a location.

If we represent a viewer as $\sigma_i = (P_i, \nu_i)$, where $P_i$ is the 2D position of the viewer's centre (of eyes) in a coordinate system, and $\nu_i$ is an unit vector representing the viewer's viewing direction, as shown in Figure 3, the viewer's field of view (FOV) can be defined as the fan-shaped area centred at $P_i$ and oriented to $\nu_i$; **A** and **B** are the projections of two landmarks on a selected horizontal line (i.e., any line above the ground from the viewer's perspective) on the viewer's image plane. The projected interval of each landmark is defined by the extreme points of its projections, written as $I = (x_1, x_2)$, where $x_1$ and $x_2$ are real numbers and $x_1 \leq x_2$. When multiple landmarks are present, a viewer will be able to use the ordering relations of the projected intervals of these landmarks to describe their environment.

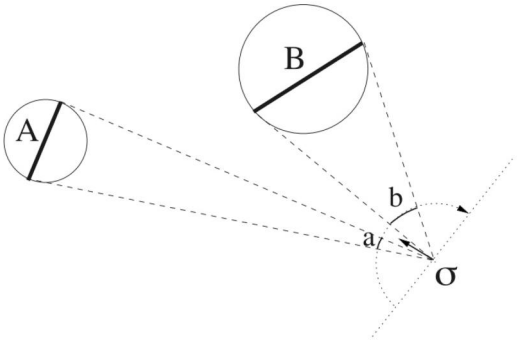



**Figure 3.** The projections of two landmarks A and B on the image plane of viewers from a viewpoint $\sigma$. The dashed lines represent the lines of sight (Ligozat et al. (2015)).

**The viewing order of two landmarks A, B**: assuming a viewer is facing towards the landmarks and looking from left-to-right and turning clockwise (or alternatively right-to-left in the anti-clockwise direction), the viewing order of these two landmarks can be decided using their extreme points $I_A(x_1^A, x_2^A)$ and $I_B(x_1^B, x_2^B)$ by following the rules below:

1. If $x_1^A < x_1^B$, then $A \rightarrow B$; otherwise, $B \rightarrow A$ : if the leftmost extremes of the two intervals are different, the one with a smaller starting point is considered appearing first.
2. If $((x_1^A = x_1^B)\ \ (|x_2^A - x_1^A| > |x_2^B - x_1^B|))$, then $A \rightarrow B$; otherwise, $B \rightarrow A$ : if the leftmost extremes of the two landmarks are the same, the one with a longer/wider interval is considered appearing first.
3. If $(x_1^A = x_1^B)\ \ (|x_2^A - x_1^A| = |x_2^B - x_1^B|)\ \ (d(A) < d(B))$, then $A \rightarrow B$; otherwise, $B \rightarrow A$ :: if the leftmost extremes of the two landmarks are the same, and the two intervals are of the same length, the landmark closer to the viewer is considered as appearing first (followed by the landmark behind if it is visible).

For point-like landmarks such as those infrastructure assets attached to the ground with a single pole, e.g. traffic lights, street lamps, trees, the locations of their poles on the horizontal line scanning across these assets can be used to decide the order of landmarks.

### *3.1.1. The relative positions of two landmarks on the panorama*

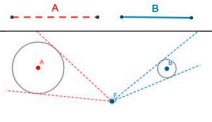
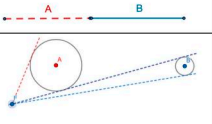
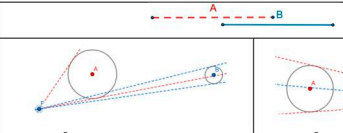
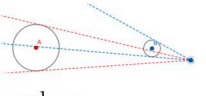
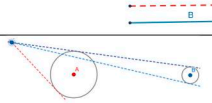
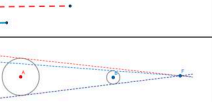
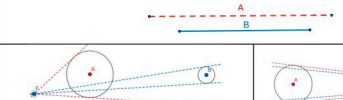
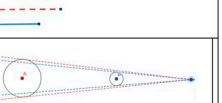
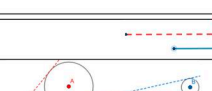
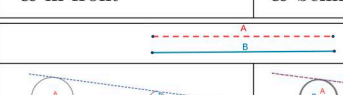
In addition to the ordering, as spatial occlusion can occur when a landmark appears before another with respect to a viewpoint, the *Interval Occlusion Calculus* (IOC) (Ligozat *et al.* 2015) is adapted in this work to further describe the relative positions of the projected intervals of ordered pairs of landmarks.

In terms of their ordering on a 0-360$^o$ panoramic coordinate system, when a best viewing direction(s) can be found such that all landmarks are located in a viewer's field of view, the left-extreme of the left-most landmark is set to zero in the 360 degree reference system; otherwise, the starting direction of the viewer is set as the zero-direction. In this work, landmarks with smaller left extremes are always

considered first; their relative positions on the panorama are represented using a reduced set of the 11 *IOC* relations as shown in Table 1, including five of the original *IOC* relations (Ligozat and Santos 2015) and six of the inverse *IOC* relations. These relations resemble *Allen's Interval Algebra* (Allen 1983, Du *et al.* 2016, Gatsoulis *et al.* 2016) except that the intervals are the projections on a panorama and the occlusion information is considered to encode the relative closeness of landmarks to viewers. In more detail:

- when there is no occlusion between two landmarks from a viewpoint, the first seen landmark could precede (***A p B***) or meet the following one (***A m B***).
- when there is partial occlusion between them, the first seen landmark could overlap and be in front of (***A*** $\boldsymbol{o^+}$ ***B***) or behind the other one (***A*** $\boldsymbol{o^-}$ ***B***).
- when the starting points of the two intervals are the same and the interval of one landmark (e.g. *A*) is longer than the other (e.g. *B*), *A* could be started by and in front of *B* (***A*** $\boldsymbol{si_*^+}$ ***B***), or behind *B* (***A*** $\boldsymbol{si^-}$ ***B***).
- when the two starting points are different, the landmark with a longer interval (e.g. *A*), could contain and be in front of *B* (***A*** $\boldsymbol{di_*^+}$ ***B***), or behind *B* (***A*** $\boldsymbol{di^-}$ ***B***); or be finished by and in front of *B* (***A*** $\boldsymbol{fi_*^+}$ ***B***) or behind *B* (***A*** $\boldsymbol{fi^-}$ ***B***).
- when the two intervals coincide, one landmark could coincide with and be in front of another, e.g. ***B*** $\boldsymbol{c_*^+}$ ***A***. Note since the front landmark is always considered first, the relation 'A coincides with B and B is in front of A' is not used in this work.

**Table 1.** The 13 Allen's Interval Calculus relations between an interval A and B (Allen 1983) are shown in the top row of each cell, and the corresponding adapted Interval Occlusion Calculus relations (Ligozat et al. 2015) are shown in the second row.

| Relation | Relation | Relation | Relation |
|---|---|---|---|
|  |  |  | |
|  precedes : $A\ p\ B$ |  meets : $A\ m\ B$ |  overlaps & in front : $A\ o^+\ B$ |  overlaps & behind : $A\ o^-\ B$ |
|  started by & in front : $A\ si_*^+\ B$ |  started by & behind : $A\ si^-\ B$ |  contains & in front : $A\ di_*^+\ B$ |  contains & behind : $A\ di^-\ B$ |
|  finished by & in front : $A\ fi_*^+\ B$ |  finished by & behind : $A\ fi^-\ B$ |  coincides with & in front : $B\ c_*^+\ A$ | $A\ eq\ B$ (not used) |

The symbol +, – encode the relative closeness of *A* and *B* to the viewer, i.e. + for in front of and – for behind.

- When two landmarks are equal to each other or the front one completely occludes the one behind, only one of them will be observed/considered in this work. Therefore, the equivalent relation *eq* is not used.

Note that the subscript $*$ in $A\{si^+_*,\ di^+_*,\ fi^+_*,\ c^+_*\}B$ means that for the landmark behind, i.e. *B*, it is only visible if tall enough as its bottom half is occluded by the landmarks in front; otherwise, it will be completely occluded and only the front landmark *A* will be observed. This constraint can be expressed with the projection $h_A$, $h_B$ of the two landmarks on the vertical axis using Allen's interval calculus (by adding a subscript $_a$ for each relation) as $(\neg[(h_A\ si_a\ h_B)\ \ (h_A\ di_a\ h_B)\ \ (h_A\ fi_a\ h_B)\ \ (h_A\ eq_a\ h_B)]$.)

Using the above definition of viewing order and the 11 modified *IOC* relations, there could be 18 types of relations between two co-visible landmarks depending on the viewpoint and location of viewers. As shown in Figure 4, 10 of the 18 relations starting from *A* are marked in blue, including $(A\{si^+_*, di^+_*, fi^+_*, o^+, m, p, o^-, fi^-, di^-, si^-\}B)$, and the other eight relations starting from *B* are marked in black, including $(B\{o^-, m, p, fi^+_*, di^+_*, si^+_*, c^+_*, o^+\}A)$. Note that *ApB*, *AmB*, *BpA*, and *BmA* are each appearing twice in different place cells.

An exemplar location $o_1$ (with viewing direction marked in pink) is given to illustrate how the relation $\langle A\ p\ B\rangle$ can be observed in the bottom place cell. If a viewer can provide their observed relation between the two landmarks, we can then roughly identify their located areas. Note that when the viewer is between the two landmarks, the landmarks can only be seen when the viewer turns around. Therefore, the viewed relation can either be $\langle A\ p\ B\rangle$ or $\langle B\ p\ A\rangle$ depending on the viewer's initial direction.

Moreover, it can be seen from Figure 4 that certain areas are spatially adjoining to each other while others are not. The corresponding *IOC* relations of adjoining areas are therefore conceptual neighbors as they can be directly transformed into one another without encountering any other types of relations (Freksa 1992) when a viewer starts moving. The neighborhood constraints between the *IOC* relations used in this work are shown in Figure 5. For example, the relation '*precedes' (p)* and '*meets' (m)* are neighbors because viewers can directly go from an area where '*A precedes B*' is

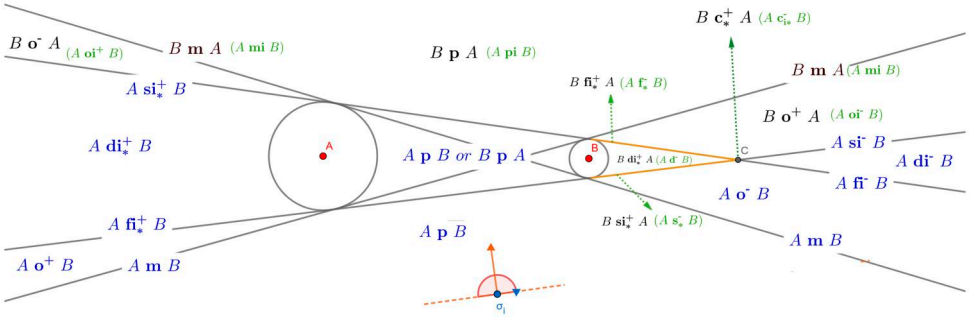



**Figure 4.** The space can be divided into individual areas where different *IOC* relations are observed between the projected intervals of landmark *A* and *B* (modified from (Ligozat et al. 2015)). For example, relation ***A p B*** would be observed in the area where the exemplar pose $\sigma_1$ is situated. As all such relations are given based on the left-to-right order from a viewer's perspective; the relations shown in green in the brackets are for illustration purpose only.

observed to an area where '*A meets B*' is observed. However, if a viewer wants to go from the area '*A precedes B*' to another one where '*A overlaps B*' is observed, they will have to go through the area/line where '*A meets B*' is observed. Areas with neighboring relations are in fact spatially connected as shown in Figure 4. By establishing the constraints on neighboring relations, we could possibly further reduce the viewer's locations or trajectory as they start moving and continuously report the observed relations between co-visible landmarks.

### 3.1.2. Abstraction of landmarks using points

As shapes can be described using points at various levels of abstraction (Freksa 1992), the above relations (shown in Figure 4) can be reduced based on point abstractions: (1) if one of the two landmarks is abstracted to a point, the 18 relations are reduced to eight, as shown in Figure 6(a); (2) if both landmarks are abstracted to points, the number of possible relations are further reduced to four as shown in Figure 6(b). When a viewer is on different sides of the line connecting the two points *AB* and facing towards the landmarks, we would expect them to observe the two landmarks in different orders, which is consistent with the observation made in Schlieder (1993) and Levitt and Lawton (1990). Note that with a reduced set of relations, we would only identify the viewer's location much more roughly compared to using the projected intervals as in Figure 4.

### 3.1.3. Viewing order of more than two landmarks on a panorama with IOC relations

When more than two landmarks ($n > 2$) are present, the study space could be divided using $n(n-1)/2$ lines connecting every two co-visible landmarks, or the extremities of landmarks when their extent is considered, such that landmarks are observed in different orders in each divided area.

For the simplistic situation when all landmarks are abstracted as points, as shown in Figure 7, the space could be divided into two parts using the convex hull of these landmarks in clockwise (or anti-clockwise) order. When viewers are outside or on the edge of the convex hull, they will be able to find an optimal viewing direction such

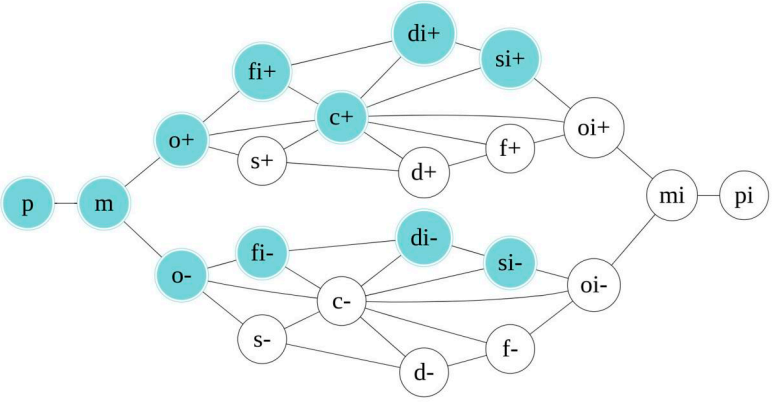
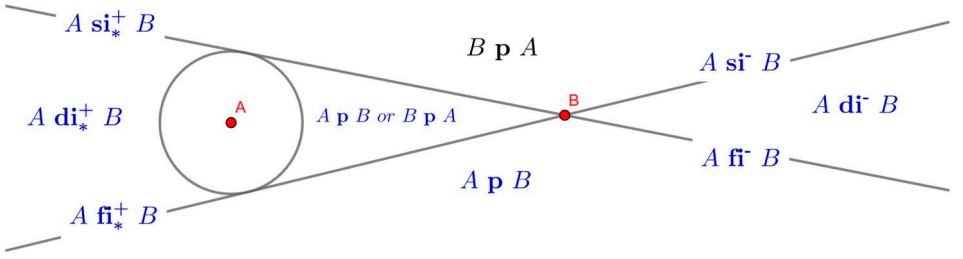
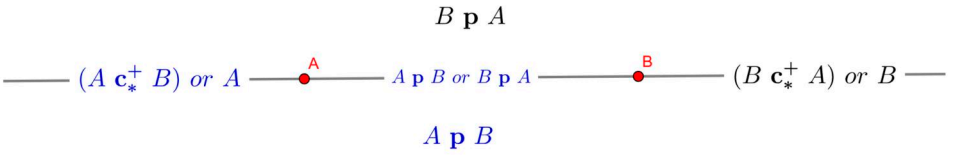



**Figure 5.** The neighborhood diagram between the used IOC relations (blue circles).




(a) The possible relations between a point landmark and a landmark with non-zero size.




(b) The possible relations between two point landmarks.

**Figure 6.** The possible relations between two landmarks using different levels of abstraction with points.

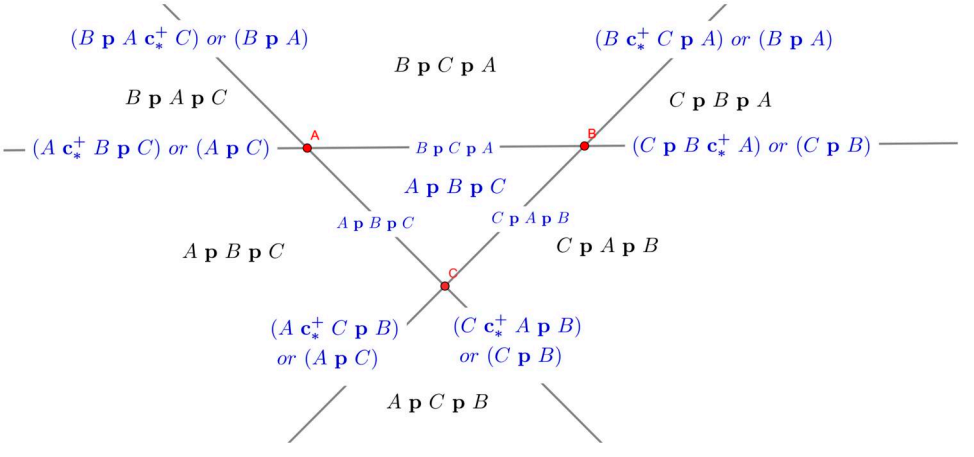



**Figure 7.** The possible relations of three co-visible landmarks. Note: in the area enclosed by the three landmarks, the observed order of landmarks depends on the initial viewing direction of viewers; while in other areas, we assume viewers have adjusted their orientations so that all landmarks are in their FOV.

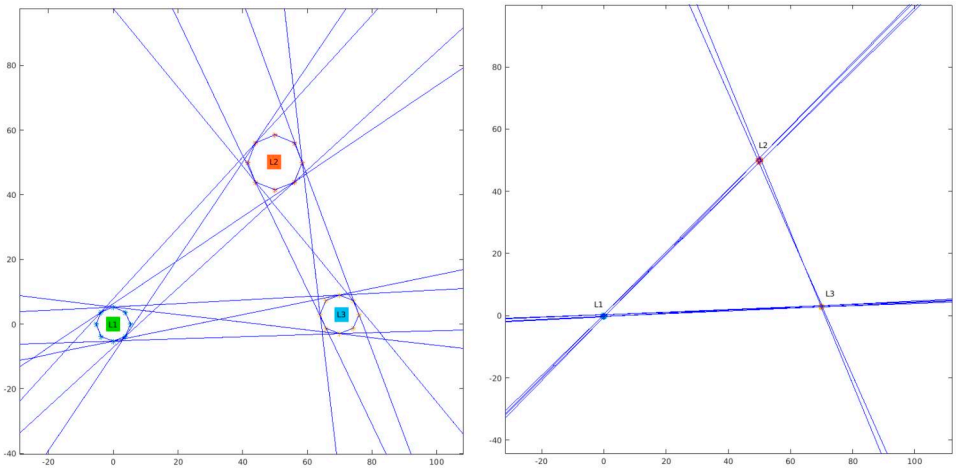
that all landmarks are located in their FOV. A unique viewing order of landmarks could then be identified for each of these areas. When viewers are inside the convex hull, they could start from any landmark so the observed order of these landmarks might be rotated, which means the observed sequences of landmarks are cyclically equal to each other, such as *ABC*, and *BCA*. In this work one such rotation sequence is created as a way to reference each area. Similarly, when the size of landmarks is considered, their relative *IOC* relations are identified by finding the left and right extremities of

each landmark from a viewpoint, and comparing the directions of corresponding tangent pairs $(\alpha_i^l, \alpha_i^r)$ in clockwise order.

Whether to consider the size of landmarks will depend on the data available and the specific application. For example, when the extent of landmarks is comparable to their distance apart, as shown in Figure 8(a), and the specific purpose of an application is to navigate robots for visual inspection of street infrastructure in urban environments (Peel *et al.* 2018), considering the extent of landmarks and the occlusion information can differentiate locations at a much finer level, especially for those areas close to the lines connecting the centroids of landmarks. On the contrary, if the extent of landmarks is relatively small compared to their distance apart such as shown in Figure 8(b), and the purpose of an application is to roughly identify the initial location of a viewer, using the point representation and dividing the space with lines connecting each point pairs would be enough, as areas near those lines would be small anyway.

Another question is whether to encode the full set of relations between every ordered pair of landmarks or only the relations between consecutively observed landmarks. For example, for the four scenarios shown in Figure 9, if we record the ordering of landmarks in each scenario based on their left extremities, we would observe $\langle ABC \rangle$ in all scenarios. A viewer will not be able to differentiate their locations based on this ambiguous description. However, if we store the sequential IOC relations of these landmarks based on their size and relative distance to the viewer, the observed place signatures would be $\langle Ao^+Bo^-C \rangle$ for the first two scenarios, $\langle Ao^+B\ p\ C \rangle$ for the third scenario, and $\langle A\ p\ B\ p\ C \rangle$ for the last scenario. These signatures are obviously more discriminative than the previous descriptions.

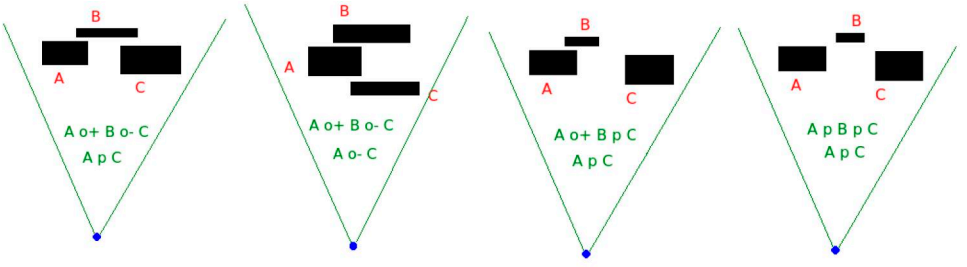
But still, the descriptions of the first two scenarios are the same while actually they are different: the relation between A and C are respectively $\langle A\ p\ C \rangle$ and $\langle Ao^-C \rangle$ in the




(a) The size of landmarks is comparable to their distance apart.

(b) The size of landmarks is relatively small.

**Figure 8.** Space division using landmarks with different distances apart. Note that the distances between landmarks in the two figures are the same but their sizes are reduced in the second figure.




(a) The same sequential relations are observed in both locations but not the first and last elements.

(b) Different sequential relations are observed in the first and second elements but if B is not observed, the two locations are indistinguishable.

**Figure 9.** Examples of different observed landmark configurations.

two scenarios. We can also see from the *composition table* (see supplementary material) that based on the relations between *A-B* and *B-C* there are four possible relations between *A-C*: $\{p, m, o^+, o^-\}$. Therefore, to achieve a better discriminative ability, ideally, it would be useful to encode the relative relations between all ordered pairs of landmarks. When landmarks are all points, only relation $p$ and $c^+$ are needed, and unique relations can be inferred from any combination of the two. Therefore, storing the relations between adjacent landmarks would be enough as no ambiguity will be caused by these relations.

In the following sections, two other components of the proposed *QPS* signature are introduced to further increase the 'resolution' of location description.

### 3.2. Adding relative orientations between ordered pairs of landmarks

The second component of the proposed place signature is the sequence of the relative orientations between ordered pairs of landmarks. The concept of relative orientations was originally proposed by Freksa (1992) to describe the location of a point with respect to two other points using the left/right and front/back dichotomies of 15 disjoint combinations of orientation relations.

For example, as shown in Figure 10 (left), objects in **area *1*** are on the *left-front* (LF) of point A with respect to vector $\overrightarrow{BA}$ and on the *right-back* (RB) of point B w.r.t. vector $\overrightarrow{AB}$; objects in area *2* are on the *left-neutral* (LN) of A and *right-back* of B; objects in area *3* are on the *left-back* (LB) of A and *right-back* of B; objects in area *4* are on the *left-back* of A and *right-neutral* (RN) of B; objects in area *5* are on the *left-back* of A and *right-front* (RF) of B; objects in area *6* are on the *straight-back* (SB) of A and *straight-front* (SF) of B; objects in area *7* are on location B and to the *straight-back* (SB) of A; objects in area *8* are on the *straight-back* of both A and B. Relations for the areas on the other side of the line AB (area 15, 14, 13, 12, 11, 10, 9) are similar to those in the area [1-7] except that *A* and *B* are switched.

Using the same division of space with the two perpendicular lines of line *AB*, one passing through *A* and the second through *B*, different groups of relative orientations can be observed by viewers *o* with respect to $(A, \overrightarrow{oA})$ and $(B, \overrightarrow{oB})$ when they are located in different areas. For example, as shown in Figure 10 (right), imagine a viewer is in area *1* and facing towards the two landmarks, they will observe *B* on the *right-*

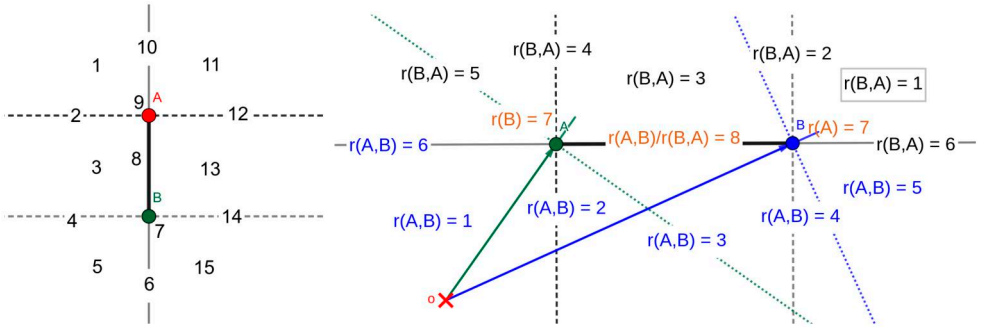



**Figure 10.** The left figure illustrates how the position of an object can be described by its relative relations to two points, A and B, when it is located within different areas, modified from (Freksa 1992); The right figure shows the subdivision of space from where different relative orientation relations can be observed between points A and B when a viewer is situated in different areas. (Note that the two figures may appear similar at first glance due to their shared use of dichotomies. However, it is important to note that they are fundamentally distinct: relations shown in the left figure describe the location of objects from a map-viewing perspective, while relations observed in the right figure are from an egocentric perspective.).

*front* of A by imagining a vector $\overrightarrow{oA}$ and its perpendicular line passing A; similarly, they will observe *A* on the *left-back (LB)* of B by imagining a vector $\overrightarrow{oB}$ and its perpendicular line passing B. Additionally, as landmarks are counted from left to right in this work and the nearer one is always considered first when multiple landmarks coincide, only eight of the original 15 orientation relations are needed in this work. More specifically, the observed relations of *B* with respect to $\overrightarrow{oA}$, and *A* with respect to $\overrightarrow{oB}$ from the eight indexed area in Figure 10 (right) are respectively: 1 (RF, LB), 2 (RN, LB), 3 (RB, LB), 4 (RB, LN), 5 (RB, LF), 6 (SF, SB), 7 (SB), and 8 (SB, SB). Four of them are on lines (2, 4, 6, 8), one on points (7), and the remaining three (1, 3, 5) are for regions.

### *3.2.1. Usage in practice*

Although it seems a bit tedious to define the exact relations observed in each area, in practice, after deciding the viewing order of two co-visible landmarks, i.e. *r(A,B)* or *r(B, A)*, a viewer will only need to select one index between 1 and 8 to describe their situated area. Or, to make the task even simpler, viewers will only need to describe whether they are situated *on the left*, *between*, or *on the right* of the two ordered parallel perpendicular lines by selecting from $\{1, 3, 5\}$. We will then be able to roughly differentiate their situated area.

When there are $n$ co-visible landmarks ($n \geq 2$), the space previously divided using $n(n-1)/2$ lines connecting every two co-visible landmarks (Figure 7) can be further divided using their perpendicular lines, totalling in $3n(n-1)/2$ dividing lines. In the case that the extent of landmarks are considered, the right extremity of the first seen landmark and the left extremity of the following landmark can be used to identify their relative orientations. The number of final areas will depend on the configuration of co-visible landmarks. For example, for the simple case shown in Figure 11 with three landmarks and line *AC* and being perpendicular to *BC*, the area marked in green where landmarks *ABC* are sequentially observed can be further divided into seven

areas (three regions and four on lines) each annotated with different combinations of relative orientation indices.

One advantage of using *relative orientation* relations is that they are able to distinguish between different qualitative distances. For example, as shown in Figure 12(a), a viewer on the left-hand side of the perpendicular line crossing the middle point of line segment will always be closer to the first seen landmark while a viewer on the other side will be closer to the following landmark. Another example is shown in Figure 12(b). Assuming a landmark of type *A* and one of type *B* are co-visible in three different locations *o*, the relative location between *A* and the viewer are the same in these scenarios, while the location of $B_2$, $B_3$ and $B_4$ are different, then it would be possible for us to differentiate between these locations as the indices of relative orientations between $A - B_2$, $A - B_3$ and $A - B_4$ are respectively *3 (RB-LB)*, *2 (RN-LB)*, and *1 (RF-LB)*; we can also infer that $B_2$ is the closest one to the viewer among the three. However, these relations do not resolve the orientation information more finely (Freksa 1992). For example, assume there is another location from where $A - B_1$ could be seen. The same configuration of relative orientations could be observed between $A - B_1$ and $A - B_2$, though the angles $\theta_1$ and $\theta_2$ between the lines of sight of the two landmarks are different. Therefore, in the next section, the angles between the lines of

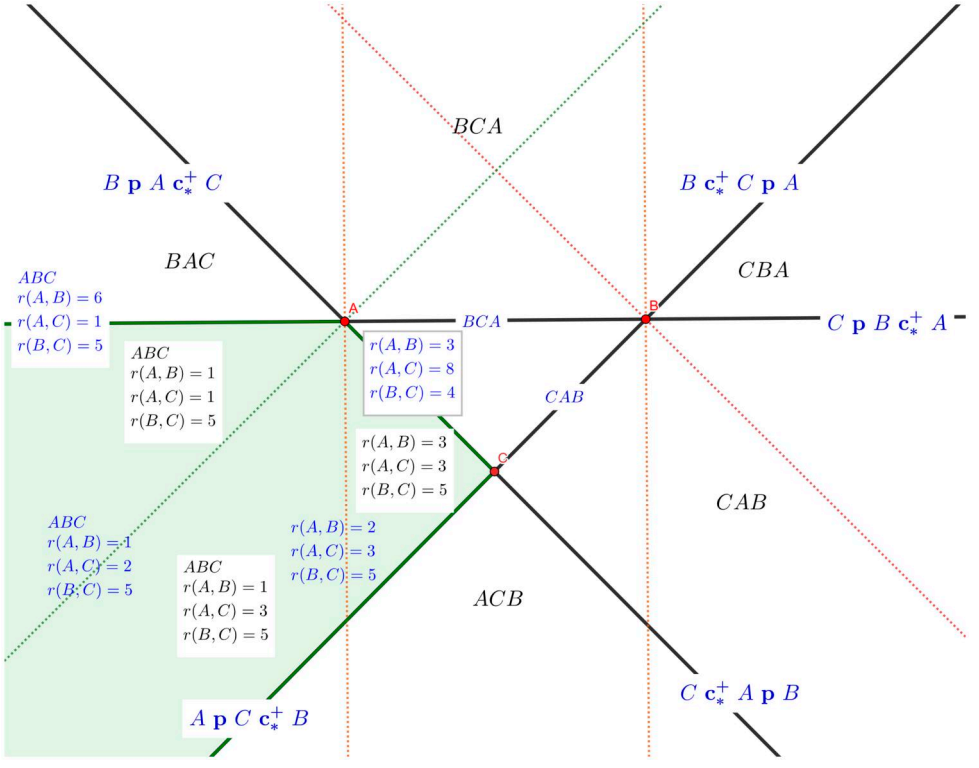
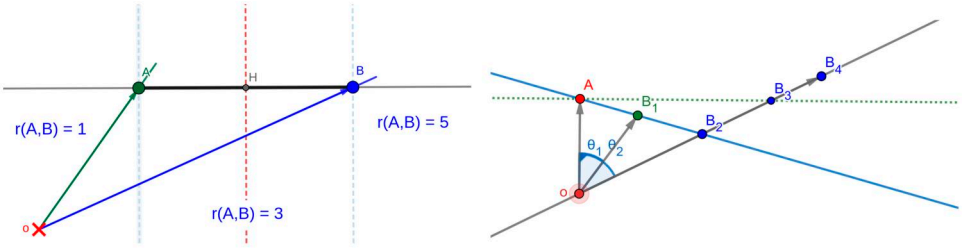



**Figure 11.** The refined space subdivision based on the ordering and relative orientation information of three co-visible point-like landmarks. Note that lines *AC, BC* are perpendicular in this simple scenario.




(a) Inferred relative closeness of two landmarks A and B.

(b) Differentiating between different qualitative distances.

**Figure 12.** (a) The space is divided by the perpendicular line crossing the middle point of line segment *AB*. A viewer on the left-hand side of the middle line will be closer to the first seen landmark while a viewer on the other side will be closer to the following landmark; (b) The relative orientations can distinguish between four locations from where $A - B_1$, $A - B_2$, $A - B_3$ and $A - B_4$ are respectively observed.

sight of ordered landmarks are introduced as the third component of the proposed place signature.

### *3.3. Adding qualitative angles between the lines of sight of ordered landmark pairs*

As previously suggested by Levitt and Lawton (1990), the set of locations from where a constant angle $o^o \leq \theta_{AB} \leq 180^o$ can be observed between the lines of sight of two landmarks is constrained to circular arcs in 2D space, which can be plotted as contour lines as shown in Figure 13 with the corresponding angles marked in black. For visualization purposes, the contour lines are plotted for every $5^o$ when angles are less than $65^o$, and every $10^o$ when angles are above. It can be seen from this figure that:

- when $\theta_{AB} = 0^o$ or $180^o$, the viewer must be co-linear with line *AB*. In this situation, the angular information $\theta_{AB}$ and the relative orientation *r* proposed in the Section 3.2 can be used to infer each other uniquely, i.e., $r(A, B) = 6) \leftrightarrow \theta_{AB} = 0^o$, $r(A, B) = 8 \Longleftrightarrow \theta_{AB} = 180^o$;
- when $\theta_{AB} = 90^o$, the corresponding contour line of $\theta_{AB}$ is the half-circle centered at the middle point of segment *AB*.
- when $90^o < \theta_{AB} \leq 180^o$, the corresponding contour lines are always between this half-circle and line *AB*, and between the two perpendicular lines of line *AB* passing through point *A* and *B*;

If viewers approach line *AB* by moving along one of its perpendicular lines that pass between A and B, then $\theta_{AB}$ will get continuously closer to $180^o$; but if they are outside the two perpendicular lines passing through A and B, then the observed angle can go larger, and then smaller, as one approaches the lines AB. If a viewer moves away from line *AB*, as the corresponding contour lines becomes quite large and the gap between two nearby arcs becomes wider, it suggests a higher ambiguity of viewers' possible location(s) between the two arcs.

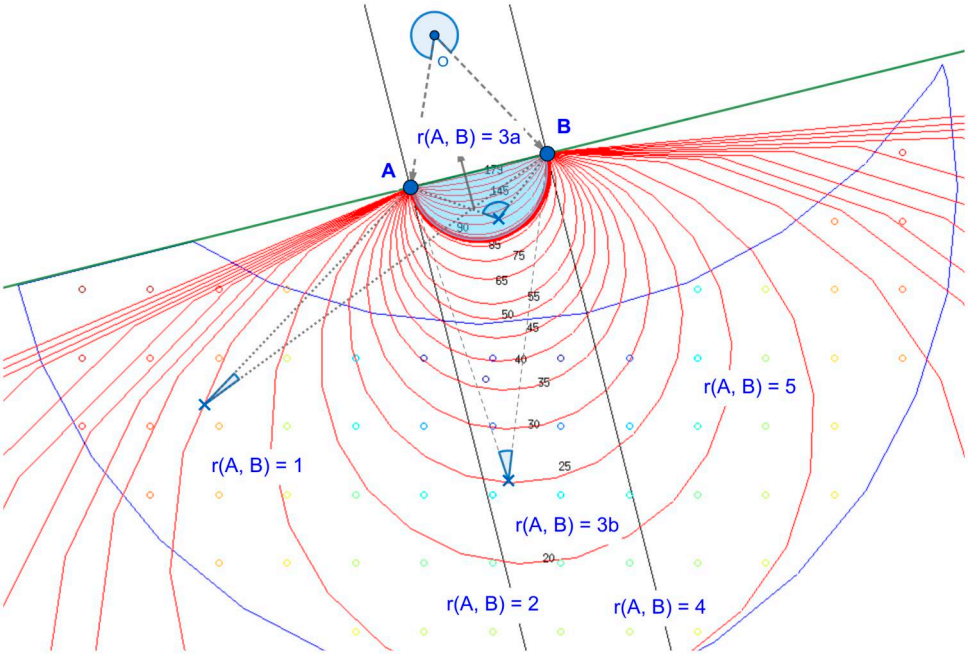



**Figure 13.** The locations from where constant angles $\theta_{AB}$ can be observed between two landmarks are plotted as red contour lines with the corresponding angles marked. Note that when the angle is closer to $180^o$, the contour lines get much denser so they are drawn for every $5^o$ (when $\theta \leq 55^o$) and $10^o$ (when $\theta \geq 65^o$) for illustration purpose. Note: the area outlined in blue is an exemplar place cell before further division.

Since humans are often not good at judging the exact value of an angle and it is not always feasible to use a separate measuring tool, qualitative angles between the lines of sight of landmarks are used in this work by judging approximately whether an angle is acute (i.e., $0^o < \theta < 90^o$), obtuse (i.e., $90^o < \theta < 180^o$) (Latecki *et al.* 1993), or right (i.e. $\theta = 90^o$). With this strategy, the region between the two perpendicular lines in Figure 13 (i.e. relative orientation $r(A, B) = 3$) is further divided as:

1. if the observed angle is ***obtuse*** (noted as 1), then the viewer must be in area 3*a*;
2. if the observed angle is ***acute*** (noted as 0) and their relative orientation is $r(A, B) = 3$, then the viewer must be in area 3*b*;
3. if the observed angle is exactly 90 degrees, then the viewer must be on the half-circle *AB* centered at the middle point of segment *AB*.

By combining the above three types of spatial relations described in Section 3.1 to 3.3, $4n(n-1)/2$ lines are used to divide the space if all ordered pairs are considered. This number is reduced to $4(n-1)$ if only adjacent pairs are considered. Each such divided area is called a *place cell* and has an associated place signature, consisting of three ordered sequences of qualitative relations between co-visible landmarks. For example, for the place cell (area outlined in blue) shown in Figure 13,

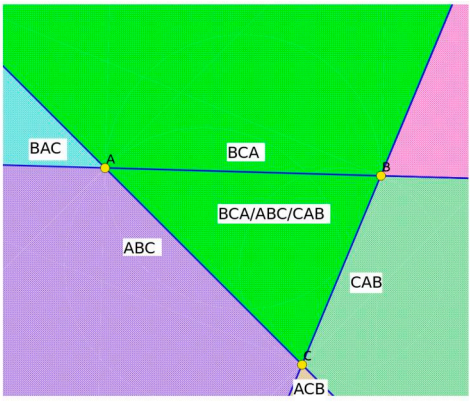
the left bottom area has a signature $S = (AB, p, 1, acute)$, where *(AB, p)* means that *A precedes B*, *r(A,B)=1* means that *B is on the right-front of A and A is on the left back of B*, and *'acute'* means that the observed relative angle between A and B is less than $90^o$.

### 3.4. Hedging place signatures in cyclic order

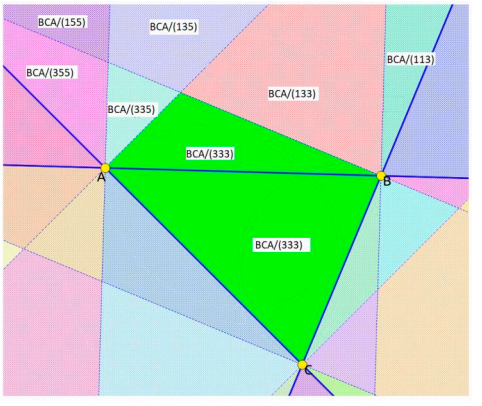
In certain locations, there exist a best viewing direction(s) such that all landmarks are located in the FOV, thus a unique place signature can be observed. In other locations, viewers may find landmarks are distributing around them and can start from any direction before turning clockwise, so the observed sequences of landmark types and relations could be rotated. For example, the sequences seen from the enclosed middle area in Figure 14(a) could be $\langle ABC \rangle$, $\langle BCA \rangle$, or $\langle CAB \rangle$, the same as from the adjacent regions. The starting element of these rotated sequences will only depend on the viewing direction not the exact location in the enclosed region. If we assume a viewer inside the enclosed area is facing *B*, then $\langle BCA \rangle$ will be observed, the same as the one from the adjacent top area marked in green.

To distinguish between the adjacent areas with same sequences of landmarks, Schlieder (1993) proposed to augment each landmark sequence with the complementary directions of all landmarks. In more detail, as illustrated in the Supplementary material, given a list of co-visible landmarks $P_1 \cdots P_n$ from a location *o*, the panorama is defined based on the $2n$ directed lines of $\overrightarrow{oP_i}$ and $\overrightarrow{P_io}$, and encoded as a sequence of upper-case letters (for the original landmarks) and lower-case letters (for the complementary directions of the original landmarks) in clockwise circular order. But this method can only distinguish between adjacent areas when a similar heading direction is assumed as a viewer moves across regions, which is reasonable for landmarks-guided robot navigation. But in our work, there is no constraint on viewers' starting directions and no movement between regions is strictly required, so the above method is not applicable.

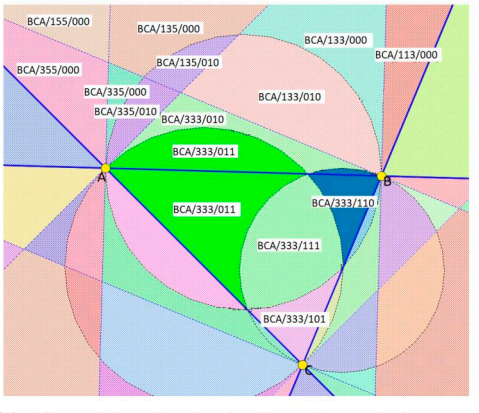
Instead, by considering the relative orientations of landmarks using perpendicular lines, the ambiguous non-enclosing area is first refined (Figure 14(b)) though the enclosed area is still indistinguishable (highlighted in green); then,by considering the qualitative angles between landmarks using half-circles, part of the enclosed area are separated (Figure 14(c)) except for those directly adjacent to the boundary (highlighted in bright green and dark blue). A simple solution suggested is to flag each region based on the distribution of the visible landmarks with respect to the viewer, which is considered as the last component of place signatures. For example, a region is flagged as $enclosed = 0$ if visible landmarks are all located on one side of the viewer (i.e. the clockwise angle between the first and last landmarks is less than $180^o$); and flagged as $enclosed = 1$ if visible landmarks are distributed around the viewer (i.e. the clockwise angle between the first and last landmarks is greater than $180^o$). It can be seen in Figure 14(d) that all ambiguous adjacent areas are separated with the suggested solution.



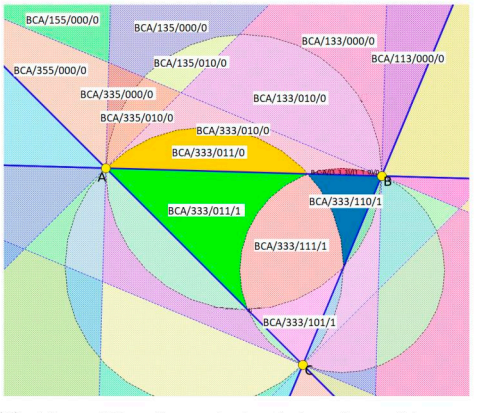

(a) The place cells created by the straight lines connecting the three landmarks. The enclosed area (green) has an observed sequence $\langle BCA \rangle$, i.e. one of the permutations of the landmark ordering.




(b) The place cells created by the straight lines and perpendicular lines of the three landmarks. The areas marked in green are inseparable with signature $\langle BCA, 333 \rangle$.




(c) After adding the circular lines, the enclosing and outside parts of the green and blue area are still inseparable.




(d) After adding the enclosing index, the ambiguous areas are separated.

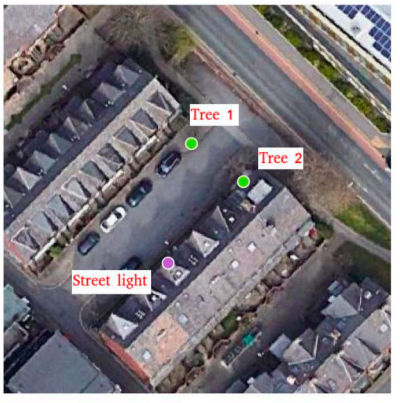
**Figure 14.** An example of place cells created using different combinations of dividing lines. Note the shown relative orientations and qualitative angles are given for landmark pairs in the order of 1–2, 1–3, and 2–3.

### *3.5. Practical steps for creating and managing a reference database of place signatures*

The section explains the steps for creating a reference database of place cells and signatures. Assume the location and attributes of a set of landmarks is givens for a study area, and the visibility range of each landmark has been selected, the *visibility areas* occluded by buildings are first removed based on the line-of-sight of each landmark (Figure 16(a)) (Algorithm is provided in the supplementary material); then, the intersections of all *visibility areas* are calculated to identify co-visible landmarks (a real world scenario for creating place cells is shown in Figure 15). After that,

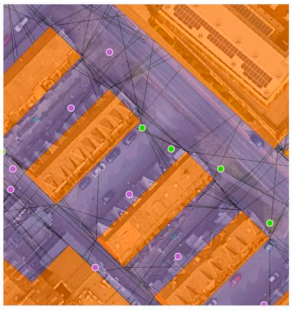
1. if landmarks are considered as *points*, each of the intersection areas is divided successively using the lines connecting each pair of co-visible landmarks, the two


(a) A street scene with three landmarks.

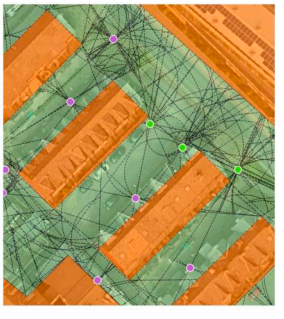



(b) Satellite view of the same area.

(c) Visible areas of individual landmarks

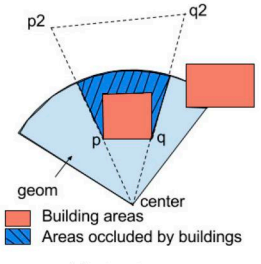


(d) Intersection of visible areas.

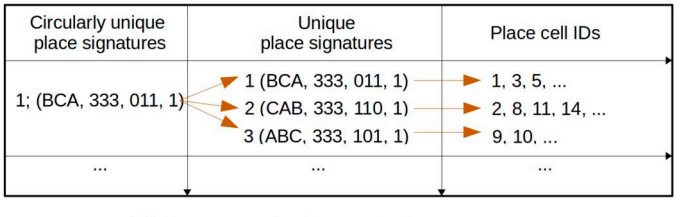


(e) Final areas.

**Figure 15.** A real-world scenario showing each step of the process of creating the place signatures (Google maps).




(a) Occlusion.




(b) Storing cyclically equal place signatures.

**Figure 16.** (a) removing the visibility area of landmarks occluded by buildings; (b) storing cyclically equal place signatures (any rotated version could be stored as a representative).

corresponding perpendicular lines of each connecting line, and the circles centered at the middle point of each connecting line segment.

2. if the *extent* of landmarks is considered, landmarks *A*, *B* on the X-Y plane are represented by their convex hulls and each of the intersection areas is divided using the extended lines of the upper-upper, lower-lower, upper-lower and lower-upper tangents of each pair of co-visible landmarks, the two perpendicular lines of line

*ab* (where *a, b* are the centroids of the two polygons) passing the two outer intersections $a_1b_1$, and circles centred at the middle points of line segment $a_2b_2$ (where $a_2b_2$ are the inner intersections of line *ab* with the two polygons).

In each of the resulting areas, the same ordering, relative orientation, and qualitative angles can be observed between landmark pairs, so a random point is chosen from each such area to compute the corresponding place signatures. Note when the size of landmarks is considered, the two tangents from the viewer to each landmark are used to identify the ordering relations between visible landmarks. Finally, unique place signatures are identified and place cells sharing the same place signatures are indexed for ease of information retrieval in the next stage.

Additionally, the unique place signatures that are rotations of each other are considered as cyclically equal, and stored using linked tables for future location retrieval. To check whether a sequence *s*1 is a rotation of *s*2, we first check whether they are of the same length; if so, we concatenate one sequence with itself then check whether it contains the other sequence. If so, they must be a rotation of each other. Note that all components of two place signatures need to be cyclically equal with corresponding shifts of the starting elements. One such example is shown in Figure 16(b).

### *3.5.1. Reference database management*

The created place cells and signatures can be managed in a relational database, as shown in Figure 17, using a table *Landmarks* storing the information of individual landmarks, a table *Place_cells* storing the information of place cells, a table *dividingLines* storing all lines used to divide the space with their type $\{SL, PL, CL\}$ attached, and a table *place_cells_relations* storing the adjacency relations between cells with link to the associated dividing line.

Based on the *type* of a dividing line, we can infer that:

1. if viewers walk across a ***Straight Line (SL)*** connecting two visible landmarks *A*, *B*, the observed orders of the two landmarks will be reversed in adjacent areas;

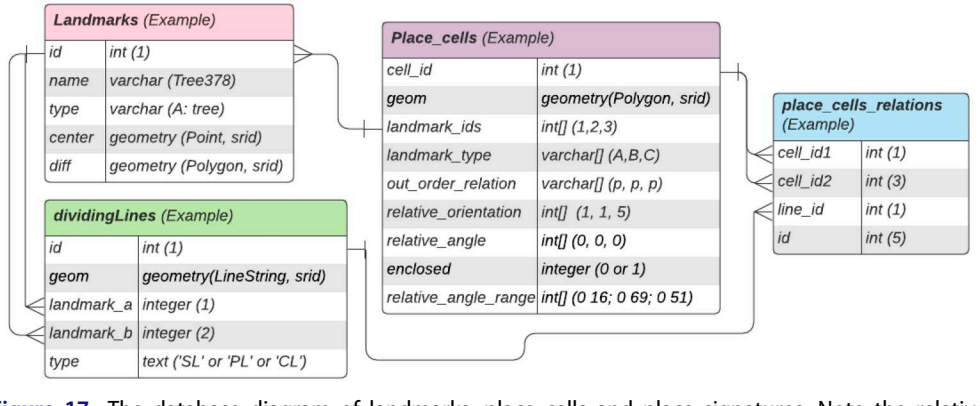



**Figure 17.** The database diagram of landmarks, place cells and place signatures. Note the relative relations are stored between landmark 1–2, 1–3, and 2–3 sequentially.

2. if viewers walk across one of the two ***Perpendicular Lines (PL)*** of the *Straight Line* connecting *A*, *B*, the observed *relative orientations* between the two landmarks will change between 1 and 3, or 3 and 5;
3. if viewers walk across a ***Circular Line (CL)***, the observed *qualitative angles* between the two corresponding landmarks will change between acute and obtuse;
4. if viewers walk across the visibility ***Boundary Line (BL)*** of a landmark, the landmark will be removed or added into the visible landmark list.

#### *3.5.2. Reference database updating to add new landmarks*

When a new landmark is added into the study area, only those place cells within or with intersections to the *visibility area* of the new landmark need to be examined. These areas will be further divided using the three types of dividing lines of the corresponding visible landmarks, and the place signatures of the updated place cells will be calculated accordingly.

#### *3.5.3. Impact of the uncertainty of reference landmark locations*

Due to the inaccuracy of GPS devices used in data capturing, errors in map digitizing, remote sensing surveys, etc, it is not uncommon to see uncertainty in GIS maps (Fisher 2005, Cheung *et al.* 2004),. In this work, we assume the semantic information of landmarks are well defined in the maps, only the uncertainty in landmark locations is discussed.

Assume the degree of uncertainty in landmarks location is much smaller than their visibility range, this uncertainty will primarily affect how the spatial relations discussed above will be observed by viewers rather than which landmarks will be observed. By modelling the location uncertainty of each landmark using independent multivariate Gaussian distributions on *X (Easting)* and *Y (Northing)* directions, and assuming the errors on both directions are uncorrelated, the uncertainty of any two co-visible landmarks $A(x_1, y_1)$ and $B(x_2, y_2)$ can be propagated to the three types of space dividing lines *SL, PL, CL* using first-order Taylor series propagation, as detailed in the supplementary material. As each type of dividing line corresponds to a certain type of qualitative spatial relation, we may expect viewers to observe a relation different to those stored in the database with a different likelihood which is depending on the viewers' location w.r.t. individual dividing line. Using this procedure, the likelihood of viewers observing an inconsistent spatial relation between any two co-visible landmarks in a reference place cell can be pre-defined.

#### *3.5.4. A simplified version of place signatures using relations between successively observed landmarks*

Assume there are *n* co-visible landmarks to a viewer, there would be $\frac{n(n-1)}{2}$ sets of relative relations between all ordered pairs of landmarks. For example, a signature with 35 visible landmarks will have 595 sets of relative relations. Though providing this full set of information would make a place signature more unique (as discussed in Section 3.1), it may be a great burden for humans to identify all such relations. Depending on the applications, it might be worth just providing the $(N-1)$ groups of relations

between successively observed landmarks. Note that with a database of complete place signatures generated (as shown in Figure 17), this step is equivalent to extracting the $k^{th}$ elements from the relation vectors, where

$$k = 1 - \frac{i(i-1)}{2} + (i-1)N, \;\; i = 1, \ldots, N-1; N \geq 2. \tag{2}$$

This simplified version will be used in the following location retrieval method.

## 4. A coarse-to-fine location retrieval method using visible landmarks based place signatures

To identify a viewer's location, they first need to report their observations by starting from the left-most landmark in their FOV, and continuously providing the types of following landmarks as well as their relative orientations and qualitative angles by turning clockwise until returning to the starting landmark. This step could potentially be automated if a camera(s) is used to capture the scene, though the information of ordering, occlusion as well as relative angles between 2D landmarks can be extracted from a single (panoramic) image, extracting the relative orientations between consecutive landmarks may need 3D cues.

In this work, we assume such a place signature $ps_j$ is ready for use and the reference database is complete or at least the landmarks in urban environment are modified gradually and the reference database is reviewed regularly such that the observed place signatures will be similar to those stored ones. Then, the location retrieval task turns to finding those place cell(s) in the database with the most similar signature $PS_i$ to the observed one, which is equivalent to finding those $PS_i$ with the smallest distance to the queried place signature, written as:

$$PS_i = \{\alpha : \alpha \in Database \;\; (\forall \beta \in Database : dist(\beta, \gamma) \geq dist(\alpha, \gamma))\} \tag{3}$$

where $\gamma$ is the observed signature, and $dist(,)$ measures the similarity of two signatures (see Section 4.1 and Section 4.2 below). The ideal situation is that viewers can observe all surrounding landmarks and their relations correctly so an exact match can be found in the database, but this is often not the case due to inaccurate perception and possible uncertainty in landmark locations. For example, a landmark could be occluded by cars or identified as a wrong type by viewers, or certain non-existent landmarks could be reported by mistake.

In the following sections, we will first discuss the pros and cons of a basic distance metric for real-time place signature matching in Section 4.1, then introduce a coarse-to-fine location retrieval method based on signature indexing in Section 4.2.

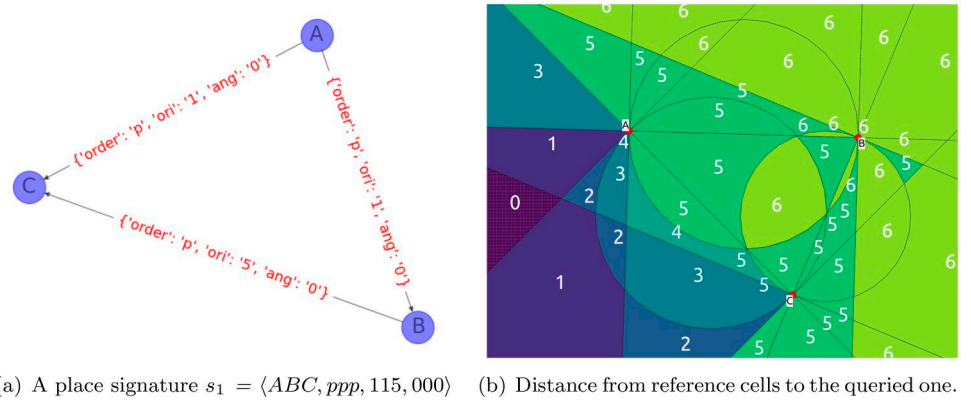
### 4.1. The pros and cons of a basic distance metric: edit distance

As each place signature is composed of three ordered sequences of landmark types, relative orientations and qualitative angles, the effect caused by perception errors are exactly the same as deleting, inserting or substituting characters in strings. *Edit Distance or Levenshtein distance* (Navarro 2001), a commonly used distance metric for string matching, would be a good candidate metric for comparing place signatures.

It measures the similarity of two strings by counting the minimum number of deletion, insertion, substitution or transposition (Damerau 1964) of characters required to transform the source string into the target one. When the complete set of relations between all ordered pairs of landmarks are considered, *graph edit distance* (Abu-Aisheh et al. 2015) can be used by representing landmarks as nodes with attached attributes (e.g. type) and representing ordering relations between landmarks as directional edges with other types of spatial relations attached. Then, finding the *edit distance* between two graphs is to find the minimum number of required edit operations on nodes and edges to transform one graph into another. For example, given an observed (and complete) place signature with three landmarks shown in Figure 18(a), its distances to all reference signatures are shown on the corresponding place cells in Figure 18(b). It can be seen that the queried place cell is correctly identified with a distance 0.

However, it is time-consuming to employ *edit distance* for real-time place signature matching, especially for large-size databases with long place signatures. As mentioned earlier, a simplified place signature with $n$ landmarks is attached with a sequence of $n$ landmark types, and two sequences of at least $(n-1)$ spatial relations between ordered landmark pairs. Given two place signatures with $n_1$ and $n_2$ landmarks, it will take quadratic time $O(3n_1n_2)$ to compare the two signatures of landmark types (and $O(2n_1(n_1-1)n_2(n_2-1)/2^2) = O(2(n_1n_2)^2)$ time to compare the qualitative relations). If there are $\mathbf{P}$ distinct place signatures in the reference database and the maximum number of landmarks in a place signature is $\bar{n}_2$, the total comparing time would be $O(3n_1\bar{n}_2 * \mathbf{P})$, which is linear to the size of the reference database, the length of the queried place signature, and the maximum length of reference place signatures.

For example, given a randomly selected place signature with 17 landmarks $\langle FFGGGGFGBGGGGGGJI \rangle$ from the *Leeds dataset* (will be detailed in Section 5.1), it took 0.04 seconds to calculate its *edit distance* to another randomly selected ignature with three landmarks, running on a laptop with Intel Core i7-7500U CPU @ 2.70 GHz processor using MatLAB R2021a. If we assume all reference place signatures contain an average of 35 landmarks and $\mathbf{P} = 1,178,445$ (i.e. numbers are from the *Leeds dataset*), it




(a) A place signature $s_1 = \langle ABC, ppp, 115, 000 \rangle$ (b) Distance from reference cells to the queried one.

**Figure 18.** An example of using *edit distance* to compare place signatures. It can be seen that the corresponding place cell of the signature shown in Figure (a) is correctly identified as the one with the smallest distance (dist = 0).

would take approximately $0.04 * 35/3 * 1,178,445$ seconds (around 152.76 hours) on the same machine to search through the whole database. In fact, it even took *316 seconds* to finish this procedure using parallel processing on a High Performance Computing facility with 12-cores. This time complexity makes *edit distance* impractical for real-time place signature matching. Therefore, it is important for us to investigate more efficient methods to quickly reduce the number of candidates such that more time-expensive yet more accurate methods could be used.

In the following sections, a coarse-to-fine location retrieval method is proposed by gradually reducing the number of candidates using *weighted MinHash*, *Jaccard distance of bags*, and *Edit distance* by considering the uncertainty in landmarks perception.

## 4.2. The proposed coarse-to-fine location retrieval method

### 4.2.1. Preparation step: representing place signatures as vectors of numbers using k-mers

To facilitate the use of other distance measures, the original place signatures are first mapped to vectors of numbers using K-shingles (k-grams) (Leskovec *et al.* 2014) (or k-mers in Bioinformatics (Arbitman *et al.* 2021)), which are substrings of length $k$ contained in a document or a sequence. Each component of a place signature is represented as a vector of k-mers by selecting a certain $k$ (or a combination of different $k$s). For example, given ten types of landmarks represented by $\{A, B, C, D, E, F, G, H, I, J\}$, there could be up to $10^1$ distinct 1-mers and $10^2$ 2-mers in a sequence of visible landmark types. The vector of 1-mer *term counts (tc)* in a sequence $\langle AFFJBAAAGBFF \rangle$ would be $\langle 4200041001 \rangle$, as shown in Table 2.

Other vector representations are also available (Table 2), such as the *frequency of terms (tf)* by dividing *tc*s with the total counts of terms appeared in a sequence, the *appearance of terms (ta)* by counting whether each term appears (1) or not (0) in a sequence, or the *binarized term counts (btc)* (Arbitman *et al.* 2021) by calculating the average count of all distinct k-mers in a sequence and keeping those elements with a count below the average as 0 and others as 1, etc. The representation using *ta* is selected in this work as it is shown in the experiments (Section 5) that it provides the best location retrieval performance compared to other representations.

Similarly, the sequence of relative orientations can be converted to vectors of numbers using $K_1 = \Sigma_{k \in [k_1, k_2]} 3^k$ terms of $\{1\ 2\ 3;\ 11\ 12\ 13\ 21\ 22\ 23\ 31\ 32\ 33; \ldots\}$; and the sequence of qualitative angles can be represented using $K_2 = \Sigma_{k \in [k_1, k_2]} 2^k$ terms of $\{0$

**Table 2.** An example of representing a sequence of landmark types $\langle AFFJBAAAGBFF \rangle$ as vectors of numbers.

| k = 1 | A | B | C | D | E | F | G | H | I | J | | | | | | | sum = 12, $f_{k=1} = \frac{12}{10^1}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| tc: | 4 | 2 | 0 | 0 | 0 | 4 | 1 | 0 | 0 | 1 | | | | | | | |
| tf | 0.33 | 0.17 | 0 | 0 | 0 | 0.33 | 0.08 | 0 | 0 | 0.08 | | | | | | | |
| ta | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | | | | | | | |
| btc: | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | | | | | | | |
| k = 2 | AA | … | AF | AG | … | BA | … | BF | … | FF | … | FJ | GB | … | JB | … | sum = 11, $f_{k=2} = \frac{11}{10^2}$ |
| tc | 2 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 2 | 0 | 1 | 1 | 0 | 1 | 0 | |
| tf | 0.18 | 0 | 0.09 | 0.09 | 0 | 0.09 | 0 | 0.09 | 0 | 0.18 | 0 | 0.09 | 0.09 | 0 | 0.09 | 0 | |
| ta | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | |
| btc | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | |

1; 00 01 10 11; … }, where $[k_1, k_2]$ set a range of values for $k$ between $k_1$ and $k_2$. Then, we can either concatenate the three vectors or use them sequentially for distance matching. Though using them sequentially can reduce instantaneous memory requirements, there is a potential that true positives could be filtered out in an earlier stage because it is unknown which part of the signature could be incorrectly observed. Therefore, the three vectors are concatenated as one vector for the following analysis.

After mapping each place signature into a vector of a fixed-length $K = K_1 + K_2 + K_3$, alternative distance metrics include *cosine distance* (Ballatore *et al.* 2013, Shahmirzadi *et al.* 2018, Steiger *et al.* 2016), *Hamming distance* (Arbitman *et al.* 2021), *Jaccard distance* (Leskovec *et al.* 2014), etc. The time complexity of calculating these distances between two vectors is $O(P*\mathrm{K})$ which is better than *Edit distance* but still linear to the size of distinct vectors. More explicitly, $O$(Edit distance) $\geq$ $O$(Jaccard bags) $\geq$ $O$(Jaccard Distance) $\geq$ $O$(Hamming distance) $\geq$ $O$(Cosine distance)). Detailed comparison will be given in Section 5.3.1.

As observations are subject to errors and an exact match may not exist in the database, it would be useful if 'similar' reference place signatures are placed together so we will only to examine the distances to these similar ones.

#### *4.2.2. Step 1: initial fast screening using locality sensitive hashing (LSH)*

Hashing is a technique that maps input data of different length to a fixed-length of values or keys to quickly identify a set of potential matches for a given query Leskovec *et al.* (2014). Being different from other Hashing methods to avoid hashing collision, *LSH* (or *approximate Hashing*) methods hash vectors such that similar ones are more likely to be hashed to same buckets and dissimilar ones into different ones (Leskovec *et al.* 2014, Marçais *et al.* 2019, Arbitman *et al.* 2021). One most used *LSH* method is *MinHash* which efficiently approximates Jaccard distance (Leskovec *et al.* 2014) by randomly generating $n$ hash functions to simulate the random permutations of term *ids*, then mapping each input vector to a vector of $n$ minimum hash values. This method can only be applied to unweighted vectors with binary values, such as vectors of term appearance (ta) (an example was given above in Table 2).

To take account of the exact number of terms appeared in each place signature, *Weighted MinHash* (Ioffe 2010, Shrivastava 2016) is used in this work to map vectors of term counts (tc) to vectors of $n_{hash}$ hash values such that the probability of drawing identical samples for a pair of inputs is identical to their Jaccard similarity. The algorithm is summarised in Algorithm 1.

**Algorithm 1:** Weighted MinHash (Ioffe 2010).

**procedure** w_MinHash($x$, $n_{hash}$)
% generate random hash variables
K: = length(x)
**for** $\mathtt{i} = 1$ to $\mathtt{n_{hash}}$ **do**
Sample $r_i$, $c_i \sim$ Gamma(2,1)
Sample $\beta_{ij} \sim$ Uniform(0,1)
**end for**
% calculate Hash values for each vector $x$
**for** $\mathtt{i} = 1$ to $\mathtt{n_{hash}}$ **do**

10: **for** iterate over $x_j$ s.t. $x_j > 0$ **do**
11: $t_j = floor\left(\frac{\log x_j}{r_{ij}} + \beta_{ij}\right)$
12: $y_j = \exp\left(r_{ij}(t_j - \beta_{ij})\right)$
13: $z_j = y_j * \exp\left(r_{ij}\right)$
14: $a_j = c_{ij}/z$
15: **end for**
16: $h^* = \min_j a_j$
17: hashPairs[i] $= (h^*, t_{h^*})$
18: **end for**
19: **return** `hashPairs`
20: **end procedure**

Using this method, the hash vectors of all reference place signatures are pre-computed offline and place signatures with the same hash vectors are placed in a same group, resulting in $P_I$ unique reference hash vectors, where $P_I \leq \mathbf{P}$. When an observation is provided by viewers, it is first hashed to a vector of $n_{hash}$ values using the same set of hash functions, then compared with all reference hash vectors by counting the number of positions with different hash values, noted as $n_0$. By setting a threshold to the proportion of different positions $t = \frac{n_0}{n_{hash}}, 0 \leq t \leq 1,$ those reference vectors with a proportion less than the threshold are considered as 'similar' candidates.

The comparison process requires $O(d)$ time to compute the hash values of a vector of $d$ non-zero values, and $O(P_I * n_{hash})$ time to compare with all reference hash vectors, which is linear to the number of hash functions $n_{hash}$ and the number of buckets $P_I$ (i.e. the number of unique vectors of hash values). Then, candidates in the 'similar' buckets can be further examined in the following steps.

Note that since *Weighted MinHash* is an approximation of *Jaccard distance* (which requires exhaustive searching through the whole database), detailed comparison of their retrieval performance in terms of query time and precision-recall will be given in the experiment Section 5.3.3. Note the recall and precision rate of exhaustive searching using *Edit distance* is not directly compared due to its impractical time complexity as discussed earlier in Section 4.1. Overall, *LSH* can significantly reduce the number of candidates in a fraction of the time compared to using *Jaccard distance*, and feeding this reduced set of candidates into the next stage with an exact matching metric will only take the corresponding proportion of the exhaustive searching time while maintaining the same precision rate. For example, the *LSH* approach can reduce the average number of candidates per query to half the size of the database at a recall rate 1 using nearly 10% of the searching time of using *Jaccard distance*. This suggests that by using at most $1/10 + \mathbf{1}/\mathbf{2} = 3/5$ of the exhaustive searching time, the same precision rate can be achieved with *LSH*. When a slightly smaller recall rate is considered, e.g. 0.97, the number of candidates can be reduced to 1/30 of the database size, which totaled in $1/10 + \mathbf{1}/\mathbf{30} = 2/15$ of the exhaustive search time.

#### *4.2.3. Step 2: Candidates refinement using jaccard distance of bags with an adaptive distance threshold*

The *Jaccard distance* of two sets $s_1$ and $s_2$ is one minus the Jaccard similarity of the two sets, defined as the ratio of the size of their intersection and the size of their

union (Leskovec *et al.* 2014), written as:

$$J(s_1, s_2) = 1 - \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} = 1 - \frac{\sum min(a_1, a_2)}{\sum (a_1 + a_2)} \tag{4}$$

When the number of terms appeared in each sequence is considered, the above Jaccard distance turns to *Jaccard distance of bags* (Jaccard bags): the intersection of two sets is the sum of the minimum number of each term appeared in the two sequences, and their union is the total number of terms (Leskovec *et al.* 2014). For example, the 1-mer tc vectors of two sequences '*ACBC*' and '*ADCA*' are respectively $a_1 = [1\ 1\ 2\ 0]$ and $a_2 = [2\ 0\ 1\ 1]$, and their Jaccard bags is $Jb(s_1, s_2) = 1 - \frac{\sum [1\ 0\ 1\ 0]}{\sum [3\ 1\ 3\ 1]} = 6/8$.

After calculating the Jaccard distance to all candidates found in the *weighted MinHash* step, the *k-nearest candidates* or those with a distance less than a threshold can be kept as the refined candidates. Although setting up a high distance threshold would allow severely deformed place signatures being corrected retrieved (recalled), it will inevitably bring in more false positives for those queries with smaller perception errors. Therefore, an adaptive threshold is chosen for each queried place signature by taking a fixed threshold $t_1$ (e.g. $t_1 = 0.59$), or the $l^{th}$ (e.g. $l = 50$) lowest distance $m(l)$ between the queried and reference candidates, whichever is smaller, $t_i = \min(t_1, m(l))$.

Through above procedures, the number of candidates can be quickly reduced to an acceptable level, enabling the usage of *Edit distance* to examine whether the number of candidates could be further reduced while retaining the same recall rate.

#### *4.2.3. Step 3: Further candidates refinement using Edit distance by considering the uncertainty in landmarks perception*

As each place signature is composed of three sequences, the *edit distance* between two place signatures $ps_1$, $ps_2$ is the weighted sum of the distances of the three components:

$$d(ps_1, ps_2) = w_1 * d(type_1, type_2) + w_2 * d(ro_1, ro_2) + w_3 * d(ra_1, ra_2) \tag{5}$$

All weighting factors are set equally as 1/3 in this work. The costs of *deletion, insertion* and *substitution* of elements were defined by considering the likelihoods of errors in landmarks perception. For example, we could imagine that it is often more likely for us to miss an existing landmark (e.g. due to occlusion) than to 'discover' a non-existent one if we assume the reference database is complete. Therefore, if a viewer reports a sequence $\langle ABC \rangle$ and there are two reference signatures $\langle ABCD \rangle$ and $\langle AB \rangle$, although the edit changes to them are both one, it would be more likely for the viewer to be in a place where $\langle ABCD \rangle$ should be observed by missing a 'D' than in a location where $\langle AB \rangle$ should be observed by adding a 'C'. This suggests that, generally, the edit cost of *deleting* a landmark should be lower than the cost of *inserting* a landmark. Similarly, we can imagine that once we observed a landmark, it would be very unlikely for us to misidentify its general type (given a limited number of options), at least more unlikely than missing the landmark. These observations can help us set up a global constraint on the edit cost of *deleting, inserting* and *substituting* landmarks in

this work:

$$Constraint\ 1: \quad C_{subs} \geq C_{ins} \gg C_{del} \tag{6}$$

In addition to the global constraint, it is worth noting that the costs associated with *deletion, insertion* and *substitution* changes in *Edit Distance* can vary depending on the attributes of landmarks. For example, imagine a situation that a viewer reports a sequence *ACD*, two candidate signatures '*ABCD*' ('bin' stands for bin) and '*AJCD*' ('J' stands for tree) would be ranked equally with one edit change of *deletion* for both candidates. But could '*ABCD*' be slightly more likely to be the correct match given that *bins* are usually much shorter than trees thus more likely to be occluded by other objects?

In general, the bigger a landmark is on viewers' retina or the image plane of cameras (which means taller, wider or closer), and more salient it is (depending on factors like colour, pattern, static/flashing etc.), the less likely it will be missed or being identified as a wrong type. Therefore, it should be assigned with higher change costs. This relation could be expressed as: $w_1^n \propto \{height,\ size,\ visual\ salience\}$. For example, the above example of *bin* and *trees* deletion can be expressed as: $w_1^{bin} C_{del} < w_1^{tree} C_{del}$. A more detailed discussion on the uncertainty in landmarks and place signature perception is given in the supplementary material.

After calculating the *edit distance* between a query place signature and all candidates, the final candidates are identified by choosing those with *k-smallest distances* to the queried one, or those with a distance below a threshold $t_e = b * len(ps_q)$, where $b$ is a proportion value (e.g. $b = [0\ 1/6\ 1/6\ 3/6\ 4/6\ 5/6\ 1]$) and $len(ps_q)$ is the length of the queried place signature.

#### 4.2.4. Assigning a probability to retrieved location hypotheses

After finding the final candidate(s) for each queried place signature, we can represent our estimate of the viewer's initial location with a probability density function distributed over the corresponding place cells of those candidate reference place signatures using a set of M particles $loc_t = (loc_t^1, ..., loc_t^M)$. Each particle contains the information of a candidate place cell and is considered as a hypothesis of the viewer's location. A uniform probability can be assigned to the viewer's possible locations inside each place cell w.r.t the area of the place cell, as $P(S_i) = 1/|X_k|$ where $X_k$ is the total volume/area of a place cell. This information can be used to further refine the viewer's location or trajectory as they start moving and continue providing their observations.

## 5. Experiments and evaluation

To evaluate the proposed location retrieval method, a reference database of place cells and place signature was created for the city centre of *Leeds* in the UK following the methods presented in Section 3 and 4; then, a set of place signatures were randomly selected and modified to simulate observations and used to evaluate the proposed location retrieval method by examining the query precision, recall and time complexity.

### 5.1. Creating a reference database of place cells and place signatures

The studied area in this work is situated in the city centre of *Leeds* in the UK, as shown in Figure 19. It covers an area of 6.47$km^2$, including 3.82$km^2$ *free* space after excluding those occupied by buildings. Semantic and spatial discrepancy were identified in landmark datasets sourced from *Find open data* and *OpenStreetMap (OSM)*, for example, *street lights* were stored in a table called *street_light* in *Find open data*, but labelled as *'highway' = 'street_lamp'* in the *OSM* map. As detailed alignment between

**Figure 19.** Overview of the *Leeds Dataset*. Top left) the location of the Leeds dataset in the UK. Top right) the studied area in Leeds city centre. A detailed view of the landmarks/place cells in the area outlined in blue is given in the bottom figure. Bottom) A zoomed view of the exemplar area divided by visibility boundary lines and lines connecting each visible landmark pair.

different datasets is out the scope of this work, only the data from *Find open data* was used. The landmarks include street lights, traffic signals, bins, trees, bus stops, etc. Although there is no limit of the number and type of landmarks being used in the proposed method, a total of *8,108* landmarks of ten types were used in this experiment, respectively represented using one character in ⟨*ABCDEFGHIJ*⟩, as summarized in Table 3.

Considering that the test scenario is in a dense urban environment, a 30-meter visibility range was selected from experience to construct the reference place signatures. The 'free' space was divided into *2,224,059* place cells of distinct place signatures with four components, including the observed ordered sequences of landmark types (symbols), relative orientations, qualitative angles, and whether the landmarks are distributed all around the viewer, or only on one side. Some statistics of the created dataset are given in Table 4.

For example, the average number of co-visible landmarks in a place cell is 35, and the maximum number is *168* due to close distribution of parallel street lights in certain areas. In the dataset under study, when only considering the categories of the ordered sequences of landmarks in the place signatures, the most frequently occurring signature or the one with the largest spatial coverage is ⟨GGG⟩. This indicates that a sequence of three streetlights is observable in the majority of areas, as streetlights comprise 56% of the landmarks in the collected dataset. However, after adding the relative orientations and angles between landmarks, the place cells are further divided, resulting in a reduction of the maximum coverage of a single signature from $0.23km^2$ to one third of that value ($0.08km^2$)".

Furthermore, as the size of individual place cell might be different, the count of place cells sharing the same signature may not well describe its potential location ambiguity. To better understand the discriminating ability of each place signature, we define their spatial coverage as the summed area of all place cells sharing this signature, noted as $sc^j$; and their spatial deviation as the standard deviation of the centroids $(c_x, c_y)$ of these place cells, written as $sd^j = \sqrt{var(c_x) + var(c_y)}$. Generally, place signatures with a small spatial coverage and deviation are relatively centralized, while those with a large deviation are Loosely distributed and viewers may observe them from many scatter locations. The maximum coverage of a single signature is $81,738.7m^2$ and the average is $1.71m^2$. The top-50 signatures with the largest coverage are displayed in Figures 20 and 21.

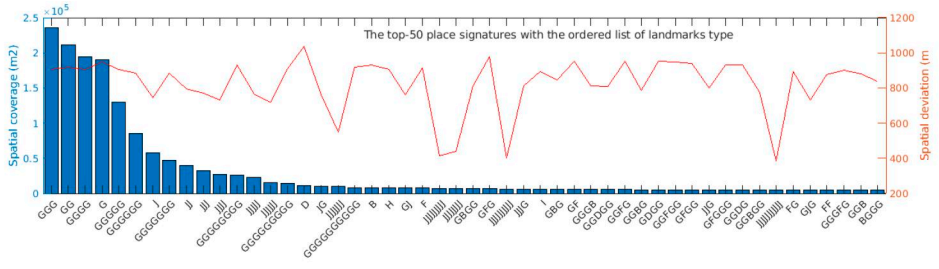
In the following sections, the overall performance of the proposed location retrieval method is evaluated step by step using randomly selected and modified place signatures from the above dataset.

**Table 3.** Summary of the ten types of landmarks used in the experiments.

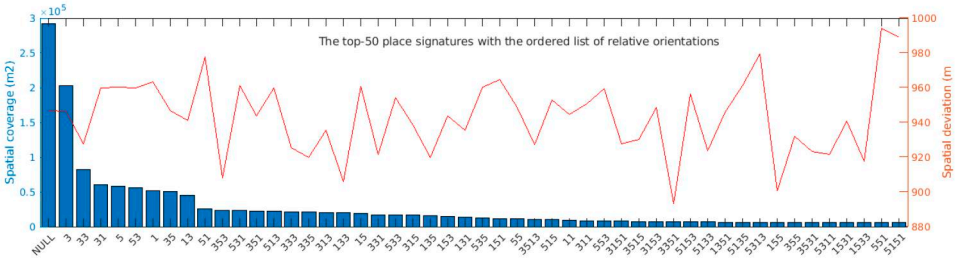
| *Landmark* | Symbol | Number (%) | *Landmark* | Symbol | Number (%) |
|---|---|---|---|---|---|
| *bicycle_parking* | A | 53 (0.65) | *road sign* | F | 888 (10.95) |
| *bin* | B | 348 (4.29) | *street light* | G | 4,566 (56.31) |
| *bollards* | C | 356 (4.39) | *toilets* | H | 3 (0.037) |
| *bus stop* | D | 245 (3.02) | *traffic signals* | I | 113 (1.39) |
| *memorial* | E | 2 (0.02) | *tree* | J | 1,534 (18.92) |
| Total | | | | | 8,108 (100) |

**Table 4.** This table shows the statistics of the created place cell and place signature database when applying different types of spatial relations on the *Leeds landmarks dataset*, including using ordered sequences of landmark *Symbols* (landmark type), *ro*(relative orientations between landmarks), *ra* (relative angles between landmarks), *symbols* + *ro* + *ra* (a combination of the three types of relations), or +*enclosed* (the three relations plus an extra element of landmarks enclosure).

| Attributes | *symbols* | *ro* | *ra* | *symbols* + *ro* + *ra* | + enclosed |
|---|---|---|---|---|---|
| *N. of distinct place signatures* | 1,178,445 | 1,916,974 | 9,011 | 2,224,059 | 2,232,311 |
| *Max. coverage of a single signature (km²)* | 0.2359 | 0.2032 | 0.2099 | 0.0817 | 0.0817 |
| *Avg. coverage of a single signature (m²)* | 3.23 | 1.98 | 422.86 | 1.71 | 1.71 |
| *Signature with the largest coverage* | ⟨*GGG*⟩ | ⟨3⟩ | ⟨0⟩ | ⟨*GG*, 3, 1⟩ | ⟨*GG*, 3, 1, 0⟩ |




(a) The top-50 place signatures when the ordered sequences of *landmark types* are used. Abbreviations for landmarks are: A (bicycle_parking), B (bin), C (bollards), D (bus stop), E (memorial), F (road sign), G (street light), H (toilets), I (traffic signals), and J (tree).



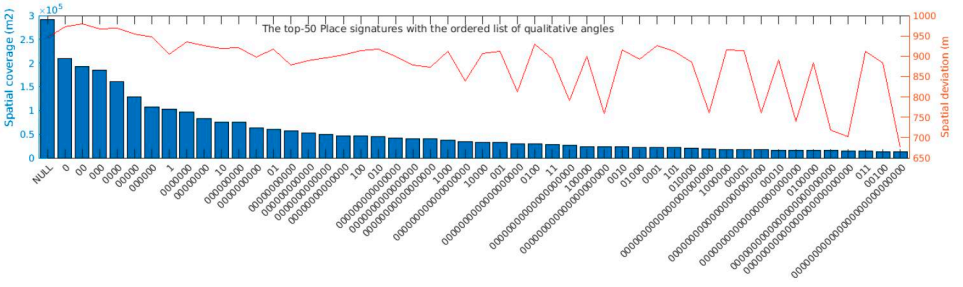

(b) The top-50 place signatures when *relative orientation*s are used. Note: when only one landmark is observed, the *relation orientation* is NULL.

**Figure 20.** The spatial coverage and spatial deviation of the top-50 place signatures with the largest coverage when individual component of place signatures is used for comparison (1).

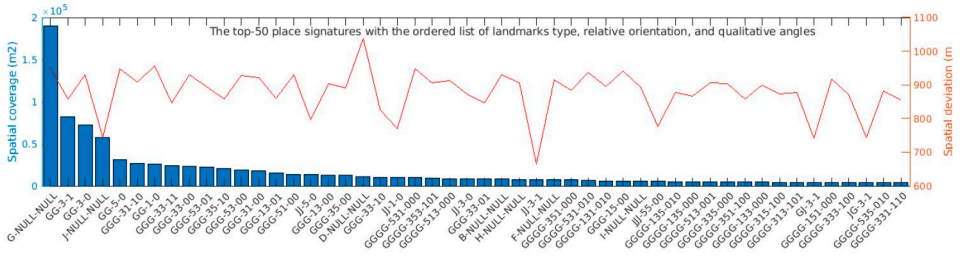
## 5.2. Evaluation criteria and observation simulation

### 5.2.1. Evaluation criteria for location retrieval

In this work, given the observed place signatures at *N* locations, a list of 'similar' reference signatures will be retrieved from the database. The *precision*, *recall* rates and *time complexity* of location retrieval methods are evaluated. A retrieved candidate is considered as a *True positive (TP)* if it is the correct correspondence of the queried signature, otherwise a *False Positive (FP)*. If none of the retrieved candidates contain the queried one, this query is considered as a *False Negative (FN)*. Then, the ***recall*** and ***precision*** rates are defined as below:




(a) The top-50 place signatures when *qualitative angles* are used, which could be '0' (acute angle), '1' (obtuse angle) or NULL when only one landmark is observed.




(b) The top-50 place signatures when all components are used.

**Figure 21.** The spatial coverage and spatial deviation of the top-50 place signatures with the largest coverage when individual component of place signatures is used for comparison (2).

$$Recall = \frac{\sum TP}{\sum(TP + FN)} = \frac{\sum TP}{N}, \; Precision = \frac{\sum(TP)}{\sum(TP + FP)} \tag{7}$$

At a given recall rate, the method with a higher precision rate would suggest that fewer irrelevant candidates (*FN*) are retrieved; while at a given precision rate, the method with a higher recall rate would suggest that more true positive candidates can be identified for the same amount of candidates.

### *5.2.2. Simulation of observations with errors*

The observed signatures by viewers are simulated by first randomly selecting $N = 1000$ place signatures from the reference database (or any other numbers or multiple sets of place signatures), then distorted to mimic different types of perception errors. To do this, we assume the probability of a landmark being deleted ($p_1$) varies between different types of landmarks, while the probability of a landmark being substituted ($p_2$) or being inserted ($p_3$) in each trial of observation are the same for all types of landmarks. For example, *bollards* are assigned a higher probability of $p_1$(0.3) compared to *traffic signals* (0.05) as they are often much shorter than traffic signals which are often have flashing signals attached. The assigned values of $p_1$ are given in Table 5, and $p_2$ and $p_3$ are set as 0.01 in the experiments. Note that these probabilities were chosen by authors based on the usual size, height and width of individual type of landmarks as the explicit information is not provided in the datasets. Future experiments will be needed to understand the actual probabilities in different environments/scenarios.

**Table 5.** The probability of different types of landmarks being *missed/deleted* in a trial of observation.

| *Landmark* | Symbol | $p_1$ | *Landmark* | Symbol | $p_1$ |
|---|---|---|---|---|---|
| *bicycle_parking* | A | 0.2 | *road sign* | F | 0.1 |
| *bin* | B | 0.2 | *street light* | G | 0.05 |
| *bollards* | C | 0.3 | *toilets* | H | 0.3 |
| *bus stop* | D | 0.1 | *traffic signals* | I | 0.05 |
| *memorial* | E | 0.2 | *tree* | J | 0.1 |

For each type of landmark in the selected place signatures, a random number $N_i^{sub}$, $N_i^{del}$ and $\overline{N_i}^{ins}$ are respectively generated from a binomial distribution $binornd(N_i, p_i)$ $(i = 2, 1, 3)$ to simulate how many landmarks of this type are to be substituted, missed, and inserted, where $N_i$ is the total number of this type of landmarks in the selected place signatures. Note *insertion* is simulated after substitution and deletion so the number of landmarks is recalculated as $\overline{N_i}$. Then, $N_i^{sub}$ and $N_i^{del}$ unique random integers are generated respectively between $[1, N_i]$ to simulate which of this type of landmarks are to be substituted and deleted; and $\overline{N_i}^{ins}$ numbers are generated between $[1, \overline{N_i}]$ to suggest where landmarks are to be inserted after. After that, each of the $N_i^{sub}$ landmarks are replaced by a randomly generated different landmark type; each of the $N_i^{del}$ landmarks as well as their related spatial relations are deleted from the original signatures; and landmarks and relevant relations are inserted after each of the $\overline{N_i}^{ins}$ landmarks with randomly generated numbers between $[1, 10]$, $\{1, 3, 5\}$ and $\{0, 1\}$. In this step, two of such spatial relations with respect to the previous and following landmarks also need to be inserted if a landmark is inserted after the first element and before the last element of a landmark sequence; otherwise, only one element of each such spatial relations needs to be inserted.

After these steps, the $N$ modified place signatures are used in all following experiments to evaluate the location retrieval performance.

## 5.3. Evaluation of the proposed location retrieval method

The proposed location retrieval method is evaluated step by step by first comparing *Jaccard bags* with other distance metrics, then comparing *Weighted MinHash* with another approximated Hashing method, following by evaluating the proposed adaptive distance metric and the contribution of using *Edit distance* by considering the uncertainty of landmarks perception.

### 5.3.1. Theoretical comparison of different distance metrics

Given a query place signature with $n_1$ landmarks and $K$ $(\sum_k 10^k + 3^k + 2^k)$ terms to represent the three components of a place signature, the following distance metrics are evaluated, including:

1. The *Edit distance* between original place signatures. It will take $O(\mathbf{P} * 3n_1\bar{n}_2)$ time to search across the whole database of $\mathbf{P}$ reference place signatures, where $\bar{n}_2$ is the maximum length of a reference place signature, where $n_1$ and $\bar{n}_2$ are the number and average number of landmarks in the query and reference place signatures;

2. The *cosine distance* between *term frequency* (tf), *term counts* (tc), or *tf-inverse document frequency* (tf-idf) vectors (Ballatore *et al.* 2013, Steiger *et al.* 2016, Shahmirzadi *et al.* 2018,). In *tf-idf,* the frequencies of terms appeared in a sequence are weighted by their *Inverse Document Frequency* in the corpus of sequences $idf(t) = 1 + log_{10}\frac{N}{n_t+1}$, where $N$ is the number of reference signatures and $n_t$ is the number of them containing the term $t$.

   Given an observed place signature, it is first represented as a vector $v_q$ of $K$ *tf* or *tf-idf* values by concatenating the three components, then compared with all reference *tf* or *tf-idf* vectors $v_p$ using cosine distance: $\cos(v_p, v_q) = 1 - \frac{v_p \cdot v_q}{\|v_p\|\|v_q\|}$. As the inverse document frequency $idf(t;P)$ of terms appeared in a corpus was pre-calculated and assumed to be consistent, we only need to calculate the term frequency in each queried signature. It will take $O((P_c + 2) * K)$ time (with time $O(P_c * K)$ for dot production and $O(2K)$ for calculating the norm of a queried vector) by using this measure where $P_c$ is the number of distinct *tf* (or *tf-idf, tc*) reference vectors and $K$ is the number of terms;
3. The *Hamming distance* between *term appearance* (ta) or *binarized term counts* vectors using logical exclusive function *xor* as both vectors are binary. It will take $O(P_h * 2K)$ time using this measure where $P_h$ is the number of distinct reference *Hamming* vectors;
4. The *Jaccard distance* between binary *term appearance* (ta) vectors using logical *xor* and *or* functions as: $J(s_1, s_2) = \frac{\sum xor(a_1, a_2)}{\sum or(a_1, a_2)}$. It will take $O(P_{ta} * 4K)$ time by using this measure where $P_{ta}$ is the number of distinct reference *ta* vectors.
5. The *Jaccard bags* between *tc* vectors (Section 4.2). It will take $O(P_{tc} * 4K)$ time by using this measure where $P_{tc}$ is the number of distinct reference vectors of *tc*.

As the length $n_2$ of original place signatures can be as high as 168 (as shown in Table 4), **P** is generally large, and the number of distinct term count vectors $P_{tc}$ is generally higher than the number of distinct binary vectors of term appearance $P_{ta}$, it can be expected that $O$(Edit distance) $\geq$ $O$(Jaccard bags) $\geq$ $O$(Jaccard Distance) $\geq$ $O$(Hamming distance) $\geq$ $O$(Cosine distance)).

### *5.3.2. Evaluation of the recall and precision rate of different distance measures*

To compare the performance of different distance measures, a threshold between $[0, 1]$ is selected for each distance measure other than *Jaccard bags*, whereas a threshold between $[0.5, 1]$ is selected. Reference place signatures with a distance below the selected threshold are considered as 'similar' candidates of a queried sample. Then, the recall rate, the average number of candidates per sample (which is equivalent to the inverse of *Precision* rate), and the average time per query using different distance measure are compared. At a same recall rate, a lower average number of candidates would suggest a better performance.

As shown in Figure 22 (left), the three methods using *Jaccard bags* with *tc*s (blue lines) gave the smallest number of candidates at a same recall rate, followed by the methods using *Jaccard distance* with term appearance (blue lines), cosine distance (green lines) and Hamming distance (magenta lines). With regards to the performance

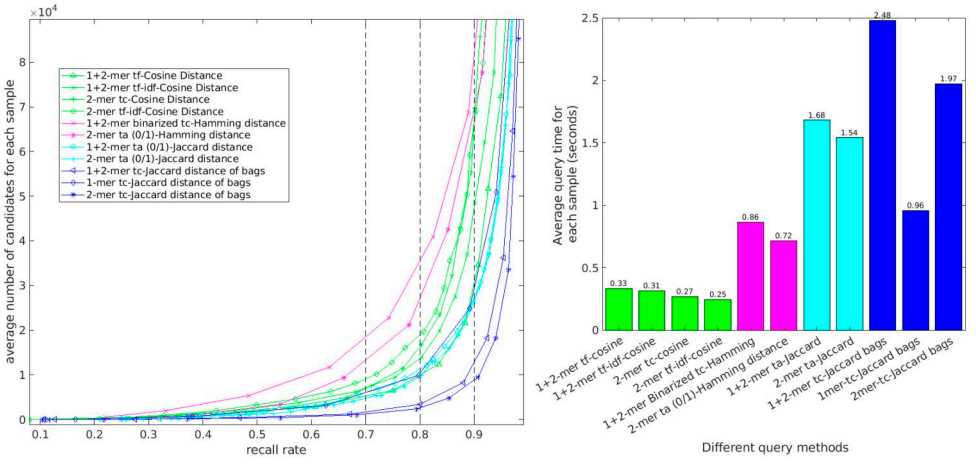



**Figure 22.** Comparison of multiple distance measures for location retrieval using 1000 randomly selected and distorted place signatures. (Left) The recall rate and the average number of candidates per quert; (right) the average query time per sample.

of different k-mers, generally, when the same distance measure is considered, the methods using *2-mer* suggest lower numbers of candidates than those using *1-mer*, or the combination of *1-mer* and *2-mers*. This may be because in some sense the ordering information between consecutive landmarks is kept in 2-mer terms but completely lost in 1-mer terms. Note that k-mers with $k > 2$ are not used in this experiment due to the high memory demands. For example, selecting $k = 3$ would convert each place signature into a $(10^3 + 3^3 + 2^3 = 1,035)$-dimensional vector, which requires approximately ten times the memory for storing all reference data compared to using $k = 2$ $(10^2 + 3^2 + 2^2 = 113)$.

While for the average time per query, as seen from the right part of Figure 22, both *Jaccard distance* and *Jaccard bags* take more time to query through the whole database (blue and blue bars) than other methods, which is consistent with our theoretical analysis in Section 5.3.1. Note *edit distance* is not directly compared at this stage due to its unpractical time complexity as discussed in Section 4. However, once the number of candidates is reduced to an acceptable level using other methods, its performance will be compared in a later section. All experiments were run on a PC with an *Intel® Core™ i7-7500U CPU @ 2.70 GHz* and one processor.

### *5.3.3. Evaluation of the locality sensitive Hashing methods on location retrieval*

As *Jaccard distance* and *Jaccard bags* based methods provide the best performance in recall-precision, experiments in this section evaluate whether locality-sensitive hashing techniques can reduce the initial query time while keeping the performance in location retrieval. To do this, the *tc* vectors of reference place signatures are mapped to *Weighted MinHash* vectors using $n$ random hash functions, and the *term appearance* vectors are mapped to unweighted **MinHash** vectors (Leskovec *et al.* 2014). Both methods require time $O(P_I * n)$ which is linear to the number of hash functions $n$ and the number $P_I$ of unique hash vectors. As there are generally more unique *tc* vectors than *term appearance* vectors, we would expect *Weighted MinHash* take slightly more

time than the non-weighted method if same numbers of hash functions are used. The recall rates and the average number of candidates per query of the two methods are shown in Figure 23 were drawn by choosing different thresholds for the proportion of different buckets between $[0, 1]$.

As shown in Figure 23 (right)), the query time per sample was significantly reduced to nearly ten percent of the original exhaustive searching time by using approximate hashing methods (green bars and red bars).

With regarding to their performance in identifying candidates, it can be seen from Figure 23 (left) that tests using *weighted MinHash* (blue/magenta lines) generally suggest lower numbers of candidates than *MinHash* (green lines) at the same level of recall rate. More specifically, the *weighted* method with 50 random functions on *2-mer tc* vectors (red dashed line with stars) gave the lowest number of candidates when the recall rate is below 0.97 while using the same method on *1 + 2-mer* terms (red solid line with triangles) gave the lowest number of candidates when the recall rate is above 0.97. This suggests that *2-mer tc*s are enough for most cases while the combination of *1-mer* and *2-mer* terms are needed for some extreme cases.

To keep as many true positives as possible, a large threshold is selected for *Weighted MinHash* method with *1 + 2-mer* terms, followed by *Jaccard* on *2-mer terms* (as seen in Figure 22) for location refinement. The average number of candidates was brought down to 1/30 of the size of the database at recall rate 0.97, and half the size of the database at recall rate 1. Then, feeding these candidates into the next exhaustive step will only take the corresponding proportion of the original computation time but retain the same levels of recall rate. For example, it can be seen from Figure 23 (left) that after combing these two steps, the curve of recall and average candidate numbers (shown as a magenta dashed line with squares) is almost the same as the exhaustive method using *Jaccard bags* on *2-mer* terms (blue line with stars), but the average query time per sample was reduced from 1.97*s* to 0.40*s*, shown as stacked red/blue bars in Figure 23 (right). Note that the average number of candidates still

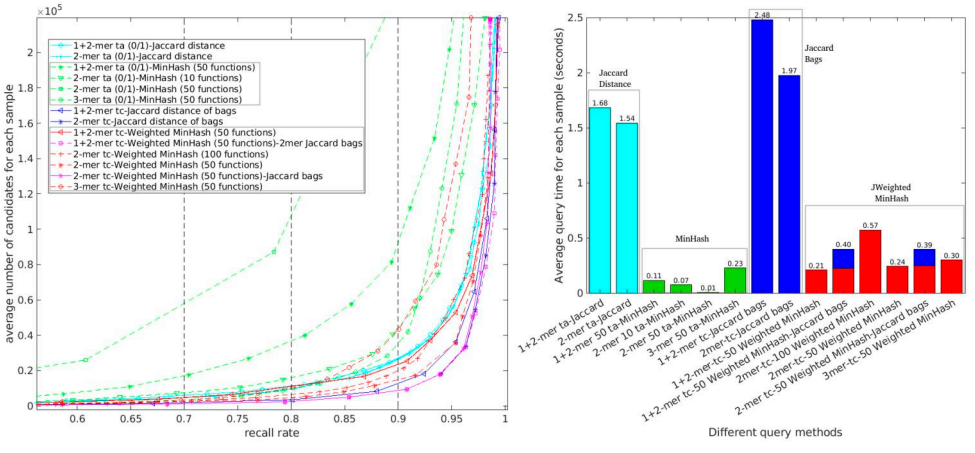



**Figure 23.** Comparison of the performance of Jaccard Distance, Jaccard bags, MinHash and Weighted MinHash in location retrieval using 1000 random place signatures, Left) The recall rate and the average number of candidates per query. (Right): the average query time per sample by using different methods.

increases sharply when the recall rate approaches 1. This is because the same distance threshold was used on *Jaccard Bags* for all queried samples and a large distance threshold will inevitably bring in more false positive for certain queries. In the next section, the proposed adaptive threshold on Jaccard bags is evaluated.

### 5.3.4. Evaluating the contribution of adaptive distance thresholds on Jaccard distance of bags

In this experiment, an adaptive distance threshold is chosen for Jaccard bags for each queried sample by taking the smaller value of a fixed threshold $t$ and the $l^{th}$ lowest distance $m(l)$ between the queried sample and all candidates. The results of using $l = 50$ and $l = 110$ with a list of fixed values of $t$ between 0.5 and 0.9 are shown in Figure 24 as dashed lines in blue. It can be seen that the average numbers of candidates are significantly reduced. For example, at a recall rate 0.9, the average number of candidates was reduced to around 3, 000 by using $t_i = \min([t, m(110)])$ on *Jaccard Bags*, which is only one-third of the number when using a fixed threshold (line in magenta), and one-eighth of the number when using *weighted MinHash* only (red dashed line).

### 5.3.5. Evaluating the contribution of edit distance by considering the uncertainty in landmarks perception

The candidates can be further refined using *edit distance*. The results by setting all costs as one (i.e. $C_{subs} = C_{ins} = 1$), and by considering the difference in landmarks perception with a higher substitution and insertion cost (i.e. $C_{subs} = C_{ins} = 5,\ C_{del} = 1$) are shown in Figure 25. It can be seen that whether a large or a small threshold is chosen for *Jaccard bags*, the corresponding maximum recall rates can all be achieved after adding the *edit distance* while the average numbers of final candidates are reduced, especially by setting different edit costs. For example, by using $\min(t, m(110))$, the average number of candidates was reduced from 7, 000 to 36 while keeping the maximum recall rate. Although it is more time-expensive using using $\min(t, m(50))$ as more candidates need to be examined by *edit distance*, the maximum recall rate is

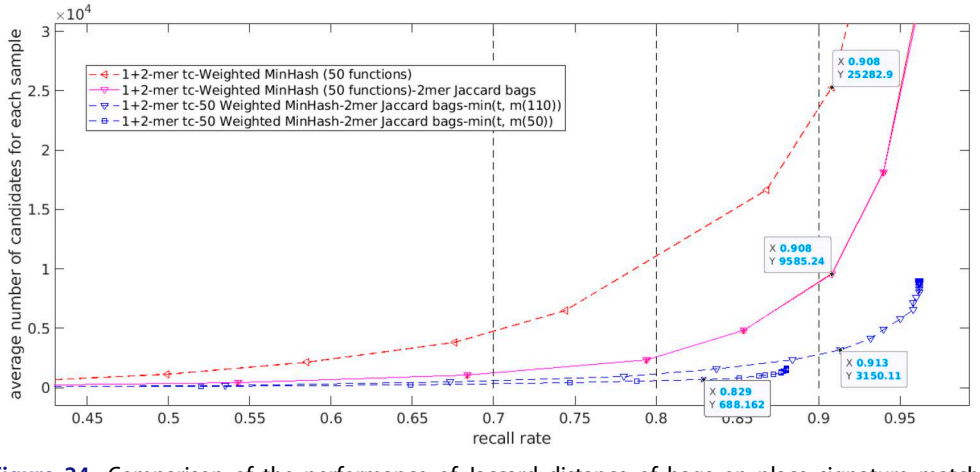



**Figure 24.** Comparison of the performance of Jaccard distance of bags on place signature matching by using or not using an adaptive distance threshold.

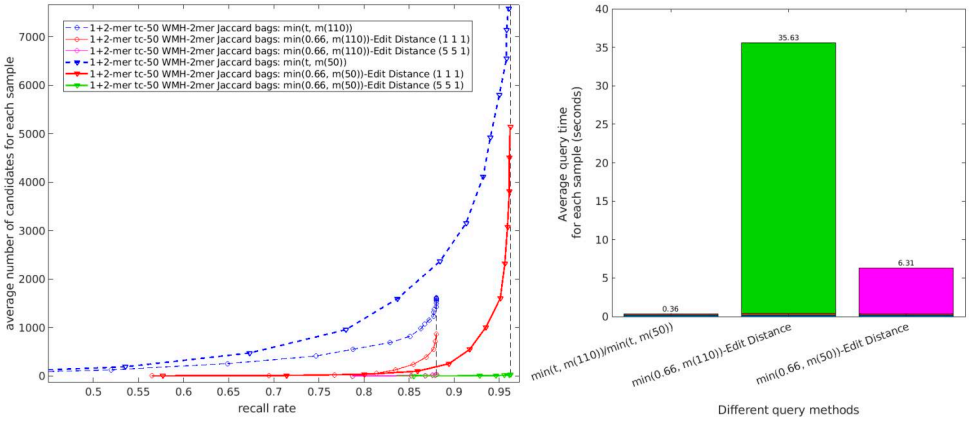



**Figure 25.** The recall rate and average number of final candidates after using *edit distance*. (Left) the results after adding the basic *edit distance* are shown as red lines, and the results by considering the difference in landmark perception are shown in magenta and green. (Right) the average query time per sample.

slightly higher. Therefore, depending on the required precision of specific applications and the available computing resources, a different *l* could be selected.

### *5.3.6. Comparison with other methods*

As the concept proposed by Weng *et al.* (2020) is close to our approach, the same *Leeds* dataset was tested on that method by comparing the performance on place cells generation and signature query. After sampling the study area as regular grids for every 10 meters on the *East-West* and *North-South* directions, 257x253 = 65, 021 point locations and their corresponding place signatures (i.e. sequences of visible landmarks and quantified angle indices w.r.t North) were created. The average coverage of the reference place signatures is $248m^2$ compared to $1.71m^2$ using our approach (Table 4). We then queried through this reference database using the same set of 1000 simulated viewer observations of randomly selected and modified place signatures. Note that as accurate measurement of angles are not assumed to be available in our work, only the sequences of landmark types were used for query. With the exhaustive searching method, it took about 45.9 seconds in average to calculate the edit distance between a query and all reference signatures. And the averaged query time was 0.36s (before applying edit distance) and 6.31s (after adding edit distance) in our approach by comparing all components of place signatures. Furthermore, only 4, 275 of the created reference signatures using Weng *et al.* (2020) were exactly the same as those 1178, 445 signatures encoded with our method (Table 4), this is partly because occlusions from buildings was not considered, and partly because the fixed-distance space division method only encode the observations from the center point of each 10x $10m^2$ square. It is difficult to quantify the difference between the observations from each center point and elsewhere inside each grid as this will depend on the number, type and distribution pattern of landmarks surrounding each individual place cell. Our approach provides a much more complete and accurate description of the environment. It also provides much richer information to help infer viewers' movement if they start crossing different types place cell dividing lines, as previously discussed in Section 3.5.

Concerning other vision-based place recognition methods relying on geo-referenced images or LiDAR datasets, we think that they are not directly comparable to our approach, first because they are challenging to be scaled up in terms of the availability of the reference datasets and of the complexity of the retrieval process through the visual features being used; second because although they may provide a more precise location or 6D pose, this is often achievable within a smaller search area; but this does make them complementary to our approach which can be exploited upstream to reduce the search area quickly and thus reducing the overall time complexity.

## 6. Discussion

### 6.1. Ability to handle similar scenes

In this work we proposed a location retrieval framework using urban landmarks, such as road signs, along with qualitative spatial relations to describe and retrieve locations. We acknowledge that this can be a challenge in the environment where we have a lot of similar scenes. However, this is a common problem in vision-based approaches, which is why in this work we propose to add the spatial configurations of visible landmarks to describe places, more specifically, the perceived qualitative spatial relations of visual landmarks in an egocentric reference frame. These relations are frequently used by humans for communicating places but are not well-studied in the literature for localisation. By including these spatial relations in a framework for location indexing, the search area of an agent's initial locations can be quickly narrowed down. The initial location(s) can then be refined using other cues, such as changes in observations when the agent moves or additional sensor input. It should be noted that our goal is not to solve the location-retrieval problem in one go, but rather as an upstream method by exploring the available information in large-scale open data. This approach has a range of potential applications where an agent's initial location is unknown or cannot be trusted, such as to re-estimate an agent's global location for loop closure in vision-based global mapping or in urban canyons, or to process the visual content from social networks or crowd-sourced data with unreliable geolocation metadata.

### 6.2. Point objects

In this work, urban objects are simplified as point objects for usage in a place signature. This is partly due to the availability of the reference data, and partly to simplify the approximate retrieval approach by enhancing the signature with semantic information instead of complex geometrical and visual attributes. It is possible to expand this semantic information with more detailed visual, semantic and spatial attributes if available as explained in the supplementary material. Note that when considering attributes such as 3D shapes and other characteristics for perception, more complicated features may be more suitable instead of only treating landmarks as points.

## 7. Conclusion

In this work, a qualitative place signature is proposed to describe locations using the perceived qualitative spatial relations between co-visible landmarks from viewers'

perspective. A framework is proposed to divide the space such that consistent place signatures can be observed inside each place cell; and a coarse-to-fine location retrieval method is proposed to identify viewers' possible location(s) by efficiently reducing the number of candidates using *weighted MinHash* and *Jaccard bags*, hypotheses refinement using *edit distance* by considering the uncertainty in landmarks perception. A reference database was created for the city of *Leeds* in the UK using openly available landmark datasets and observations were simulated to evaluate the proposed location retrieval method. The results suggest that by using *weighted MinHash* and *Jaccard bags* with *adaptive distance thresholds* for initial screening, the number of false positives can be significantly reduced to an acceptable level in less than a second; while incorporating the *edit distance* by considering the difference in perception error could further reduce the number of candidate locations by keeping the high recall rate. As the proposed approach only requires storing the configuration of high-level landmarks and utilising an approximate Hashing step for fast screening, it is easy to scale up thus can be exploited as an upstream approach for other location/pose refinement techniques. This technique could be used indoors, given a suitable database of landmarks and it will help in urban canyons where there are not enough satellites in view. For future work, we plan to create a more complete and coherent reference landmarks dataset by resolving the semantic and spatial discrepancy in different datasets using methods such as ontology alignment (Stoilos *et al.* 2005, Li *et al.* 2009) and geometry matching (Du *et al.* 2017), to test the proposed method in virtual reality environment by considering landmarks with extended sizes, and to extend the proposed scheme from single location retrieval to trajectory identification by considering the movements of agents.

## Author contributions

Lijun Wei was a research fellow at the University of Leeds and French Mapping Agency and contributed to the conceptualisation, methodology, analysis, validation, and writing-original draft preparation of this work. Valérie Gouet-Brunet is a senior researcher at the French Mapping Agency and Gustave Eiffel University and contributed to supervision and writing - review & editing. Anthony G. Cohn is Professor of Automated Reasoning at the University of Leeds and Foundation Models Lead at the Alan Turing Institute. He contributed to the methodology, funding acquisition, supervision, and writing - review & editing.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This research was partially supported by the Economic and Social Research Council (ESRC) under grant ES/W003473/1.

## Notes on contributors

***Lijun Wei***, Ph.D., is a senior advisor on data science and analytics at GHD (Gutteridge Haskins & Davey) with a particular focus on spatial data analytics and movement insights. Before her current role, she was a postdoctoral researcher at the French Mapping Agency (France) and the School of Computing, University of Leeds (UK), during which she was conducting researches on

image based localisation and AI-enhanced infrastructure asset management under the supervision of Dr. Valérie Gouet-Brunet and Prof. Anthony G. Cohn, respectively. She obtained her PhD degree from the University of Technology of Belfort-Montbéliard (France) on multi-sensor data fusion for vehicle localisation, and BSc degree from Wuhan University (China) on geographic information system and natural resource management.

***Valérie Gouet-Brunet***, Ph.D., has been a Research Director of the French Ministry of Ecology since 2012 at the LASTIG Lab. of the French mapping agency (IGN) and of Gustave Eiffel University in France. She is in charge of researches on the description by content, matching and indexing of large-scale and long-term multimedia collections, with a focus on their applications to cultural and natural heritage. She was the head of the MATIS laboratory (IGN) between 2014 and 2018, specialising in photogrammetry, computer vision and remote sensing. She obtained her PhD in Computer Vision in 2000 from the University of Montpellier (France) on colour image matching for intermediate view synthesis, and habilitation to direct research at the Pierre and Marie Curie University (France) in 2008 on content-based structuring of collections of still and animated images. She has supervised over fifty PhD students and researchers and participated in/coordinated over twenty projects funded by the French national ANR, FUI, European Council, industry, or international research funds. Currently, she is a member of the board of the European Association Time Machine Organisation, and the working group "Digital data" of the scientific site for the restoration of Notre-Dame de Paris.

***Anthony G. Cohn***, Ph.D., is a Professor of Automated Reasoning at the University of Leeds and Foundation Models Lead at the Alan Turing Institute. His PhD is from the University of Essex. He spent 10 years at the University of Warwick before moving to Leeds in 1990 where he founded a research group working on knowledge representation and reasoning with a particular focus on qualitative spatial/spatio-temporal reasoning, the best known being the well-cited region connection calculus (RCC) – the KR-92 paper describing RCC won the 2020 KR Test-of-Time award. He was awarded the 2021 Herbert A. Simon Prize for Advances in Cognitive Systems. He is the Editor-in-Chief of Spatial Cognition and Computation and has been Chairman/President of the UK AI Society SSAISB, the European Association for Artificial Intelligence (EurAI), KR Inc., the IJCAI Board of Trustees and was the Editor-in-Chief for Artificial Intelligence 2007–2014. He is the recipient of the 2015 IJCAI Donald E Walker Distinguished Service Award and the 2012 AAAI Distinguished Service Award. He is a Fellow of the Royal Academy of Engineering, and the Learned Society of Wales, and is also a Fellow of AAAI, AISB, EurAI, AAIA, the BCS, and the IET; he is also a Chartered Engineer.

## Data and codes availability statement

The data that support the findings of this study were derived from the open data resources as detailed in the supplementary material. The supplementary material to this manuscript can be found at https://doi.org/10.5518/1506. The derived data, codes, and instructions that support the findings of this study are available at https://doi.org/10.6084/m9.figshare.25680096.

## References

Abu-Aisheh, Z., *et al.*, 2015. An exact graph edit distance algorithm for solving pattern recognition problems. *In*: *International conference on pattern recognition applications and methods*, 271–278.

Ali-Bey, A., Chaib-Draa, B., and Giguere, P., 2023. MixVPR: Feature mixing for visual place recognition. *In*: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. Waikoloa, HI, USA: IEEE. https://www.computer.org/csdl/proceedings/wacv/2023/1KxUhhFgzlK.

Allen, J.F., 1983. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26 (11), 832–843.

Arbitman, G., *et al.*, 2021. Approximate Hashing for Bioinformatics. *In*: *CIAA 2021 - 25th international conference on implementation and application of automata*, Bremen, Germany, 1–12.

Ardeshir, S., *et al.*, 2014. GIS-Assisted landmark Detection and Geospatial localisation. *In*: *ECCV 2014*, 602–617.

Ballatore, A., Wilson, D.C., and Bertolotto, M., 2013. Computing the semantic similarity of geographic terms using volunteered lexical definitions. *International Journal of Geographical Information Science*, 27 (10), 2099–2118.

Bittner, T., and Stell, J.G., 2000. Approximate qualitative spatial reasoning. *Spatial Cognition and Computation*, 2 (4), 435–466.

Chen, J., *et al.*, 2013. A survey of qualitative spatial representations. *The Knowledge Engineering Review*, 30 (1), 106–136.

Chen, Z., *et al.*, 2017. Deep learning features at scale for visual place recognition. *In*: *International conference on robotics and automation*, 3223–3230.

Cheung, C.K., Shi, W.Z., and Zhou, X., 2004. A probability-based uncertainty model for point-in-polygon analysis in GIS. *GeoInformatica*, 8 (1), 71–98.

Clementini, E., Di Felice, P., and Hernández, D., 1997. Qualitative representation of positional information. *Artificial Intelligence*, 95 (2), 317–356.

Cohn, A.G., *et al.*, 2014. Reasoning about topological and cardinal direction relations between 2-dimensional spatial landmarks. *Journal of Artificial Intelligence Research*, 51, 493–532.

Couclelis, H., *et al.*, 1987. Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology*, 7 (2), 99–122.

Damerau, F.J., 1964. A technique for computer detection and correction of spelling errors. *Communications of the ACM*, 7 (3), 171–176.

Du, S., Shu, M., and Feng, C.C., 2016. Representation and discovery of building patterns: a three-level relational approach. *International Journal of Geographical Information Science*, 30 (6), 1161–1186.

Du, H., *et al.*, 2017. A Method for Matching Crowd-sourced and Authoritative Geospatial Data. *Transactions in GIS*, 21 (2), 406–427.

Du, S., Feng, C.C., and Guo, L., 2015. Integrative representation and inference of qualitative locations about points, lines, and polygons. *International Journal of Geographical Information Science*, 29 (6), 980–1006.

Duckham, M., and Worboys, M.F., 2001. Computational Structure in Three-Valued Nearness Relations. *In*: *International conference on spatial information theory: foundations of geographic information science (COSIT 2001)*, 76–91.

Durrant-Whyte, H., and Bailey, T., 2006. Simultaneous localisation and mapping: part I. *IEEE Robotics & Automation Magazine*, 13 (2), 99–110.

Egenhofer, M.J., *et al.*, 1999. Progress in computational method for representing geographical concepts. *International Journal of Geographical Information Science*, 13 (8), 775–796.

Fisher, P.F., 2005. Models of uncertainty in spatial data. *In*: *Geographical information systems: principles, techniques, management, and applications*, 191–205.

Fogliaroni, P., *et al.*, 2009. A Qualitative Approach to localisation and Navigation Based on Visibility Information. *In*: *Spatial information theory*. Springer: Berlin Heidelberg, 312–329.

Frank, A.U., 1991. Qualitative spatial reasoning with cardinal directions. *In*: *7th austrian conference on artificial intelligence*, 157–167.

Freksa, C., 1992. Using orientation information for qualitative spatial reasoning. *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, 54 (1-2), 162–178.

Freksa, C., 1992. Temporal reasoning based on semi-intervals. *Artificial Intelligence*, 54 (1-2), 199–227.

Freksa, C., van de Ven, J., and Wolter, D., 2018. Formal representation of qualitative direction. *International Journal of Geographical Information Science*, 32 (12), 2514–2534.

Galton, A., 1994. Lines of sight. *In*: *Proceedings of the seventh annual conference of AI and cognitive science*, 103–113.

Gatsoulis, Y., *et al.*, 2016. QSRlib: a software library for online acquisition of qualitative spatial relations from video. *In*: *29th international workshop on qualitative reasoning (QR16)*, 11 July, New York. https://ivi.fnwi.uva.nl/tcs/QRgroup/qr16/index.html.

Héry, E., Xu, P., and Bonnifait, P., 2021. Consistent decentralized cooperative localisation for autonomous vehicles using LiDAR, GNSS, and HD maps. *Journal of Field Robotics*, 38 (4), 552–571.

Ioffe, S., 2010. Improved Consistent Sampling, Weighted Minhash and L1 Sketching. *In*: *Proceedings of the 2010 IEEE international conference on data mining*, 246–255.

Irschara, A., *et al.*, 2009. From structure-from-motion point clouds to fast location recognition. *In*: *IEEE conference on computer vision and pattern recognition*, Miami, Florida, USA. IEEE.

Jacobs, L.F., and Schenk, F., 2003. Unpacking the cognitive map: The parallel map theory of hippocampal function. *Psychological Review*, 110 (2), 285–315.

Jegou, H., *et al.*, 2010. Aggregating local descriptors into a compact image representation. *In*: *IEEE conference on computer vision and pattern recognition (CVPR)*, San Francisco, CA, USA. IEEE.

Kendall, A., Grimes, M., and Cipolla, R., 2015. PoseNet: A convolutional network for real-time 6-DOF camera relocalisation. *In*: *IEEE international conference on computer vision (ICCV)*, 2938–2946.

Korrapati, H., *et al.*, 2012. Image sequence partitioning for outdoor mapping. *In*: *IEEE international conference on robotics and automation*, St. Paul, Minnesota, USA. IEEE, 1650–1655.

Kuipers, B., 2000. The spatial semantic hierarchy. *Artificial Intelligence*, 119 (1-2), 191–233.

Lamon, P., *et al.*, 2001. Deriving and matching image fingerprint sequences for mobile robot localisation. *In*: *Proceedings of the IEEE international conference on robotics and automation*, 1609–1614.

Lamon, P., *et al.*, 2003. Environmental modeling with fingerprint sequences for topological global localisation. *In*: *Proceedings of IEEE/RSJ international conference on intelligent robots and systems*, 3781–3786.

Latecki, L.J., *et al.*, 1993. Orientation and qualitative angle for spatial reasoning. *In*: *IJCAI*, 1544–1549.

Leskovec, J., Rajaraman, A., and Ullman, J.D., 2014. *Mining of massive datasets*. USA: Cambridge University Press, 73–129.

Levitt, T.S., and Lawton, D.T., 1990. Qualitative navigation for mobile robots. *Artificial Intelligence*, 44 (3), 305–360.

Li, A., Morariu, V.I., and Davis, L.S., 2014. Planar structure matching under projective uncertainty for geolocation. *In*: *European conference on computer vision*, 265–280.

Li, J., *et al.*, 2009. RiMOM: A dynamic multistrategy ontology alignment framework. *IEEE Transactions on Knowledge and Data Engineering*, 21 (8), 1218–1232.

Ligozat, G., and Santos, P., 2015. Spatial occlusion within an interval algebra. *In*: *AAAI spring symposium series*, 103–106.

Lowry, S., *et al.*, 2016. Visual place recognition: a survey. *IEEE Transactions on Robotics*, 32 (1), 1–19.

Luo, K., *et al.*, 2024. 3D point cloud-based place recognition: a survey. *Artificial Intelligence Review*, 57 (4), 44.

Marçais, G., *et al.*, 2019. Locality-sensitive hashing for the edit distance. *Bioinformatics (Oxford, England)*, 35 (14), i127–i135.

Navarro, G., 2001. A guided tour to approximate string matching. *ACM Computing Surveys*, 33 (1), 31–88.

Panphattarasap, P., and Calway, A., 2016. Visual place recognition using landmark distribution descriptors. *arXiv*.

Peel, H., *et al.*, 2018. Localisation of a mobile robot for bridge bearing inspection. *Automation in Construction*, 94, 244–256.

Piasco, N., *et al.*, 2018. A survey on visual-based localisation: on the benefit of heterogeneous data. *Pattern Recognition*, 74 (2), 90–109.

Piasco, N., *et al.*, 2021. Improving image description with auxiliary modality for visual localisation in challenging conditions. *International Journal of Computer Vision*, 129 (1), 185–202.

Pion, N., *et al.*, 2020. Benchmarking Image Retrieval for Visual localisation. *In*: *International conference on 3D vision (3DV)*, 483–494.

Rousell, A., *et al.*, 2015. Extraction of landmarks from OpenStreetMap for use in navigational instructions. *In*: *The 18th AGILE international conference on geographic information science*, Lisbon, Portugal.

Sadalla, E.K., Burroughs, W.J., and Staplin, L.J., 1980. Reference points in spatial cognition. *Journal of Experimental Psychology: Human Learning & Memory*, 6 (5), 516–528.

Sattler, T., *et al.*, 2012. Image retrieval for image-based localisation revisited. *In*: *British machine vision conference (BMVC)*, Newcastle, UK. The British Machine Vision Association (BMVA).

Schlichting, A., and Brenner, C., 2014. localisation using automotive laser scanners and local pattern matching. *In*: *Proceedings of IEEE intelligent vehicles symposium*, 414–419.

Schlieder, C., 1993. Representing visible locations for qualitative navigation. *In*: *Qualitative reasoning and decision technologies*, 523–532.

Shahmirzadi, O., Lugowski, A., and Younge, K., 2018. Text similarity in vector space models: a comparative study. *In*: *2019 18th IEEE international conference on machine learning and applications (ICMLA)*, Boca Raton, FL, USA. IEEE, 659–666.

Shrivastava, A., 2016. Simple and efficient weighted minwise hashing. *In*: *Advances in neural information processing systems*, vol. 29. Curran Associates, Inc., 1–9.

Soheilian, B., *et al.*, 2013. Generation of an integrated 3D city model with visual landmarks for autonomous navigation in dense urban areas. *In*: *IEEE intelligent vehicles symposium (IV)*, 304–309.

Steiger, E., Resch, B., and Zipf, A., 2016. Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. *International Journal of Geographical Information Science*, 30 (9), 1694–1716.

Stoilos, G., Stamou, G., and Kollias, S., 2005. A String Metric for Ontology Alignment. *In*: T*he semantic web – ISWC 2005*, vol. 3729. Springer, 624–637.

Tversky, B., 1993. Cognitive maps, cognitive collages, and spatial mental models. *In*: *European conference on spatial information theory COSIT 1993: spatial information theory A theoretical basis for GIS*, 14–24.

Uy, M., and Lee, G., 2018. PointNetVLAD: Deep point cloud based retrieval for large-scale place recognition. *In*: *IEEE/CVF conference on computer vision and pattern recognition*, 4470–4479.

Wagner, T., Schlieder, C., and Visser, U., 2004. An Extended Panorama: Efficient Qualitative Spatial Knowledge Representation for Highly Dynamic Environments. *In*: *Proceedings of the IJCAI-03 workshop on issues in designing physical agents for dynamic real-time environments: world modelling, planning, learning, and communicating*.

Wang, X., *et al.*, 2005. Landmark-based qualitative reference system. *In*: *Proceedings of 2005 IEEE international geoscience and remote sensing symposium*, 932–935.

Weng, L., Gouet-Brunet, V., and Soheilian, B., 2020. Semantic signatures for large-scale visual localisation. *Multimedia Tools and Applications*, 80 (15), 22347–22372.

Yao, X., and Thill, J.C., 2006. Spatial queries with qualitative locations in spatial information systems. *Computers, Environment and Urban Systems*, 30 (4), 485–502.

Zamir, A., *et al.*, 2016. Large-scale visual geo-localisation. *In*: *Advances in computer vision and pattern recognition*. Switzerland: Springer.

Zang, A., *et al.*, 2017. Accurate vehicle self-localisation in high definition map dataset. *In*: *Proceedings of the 1st ACM SIGSPATIAL workshop on high-precision maps and intelligent applications for autonomous vehicles*, New York, NY, USA, 1–8.

Zhang, X., Wang, L., and Su, Y., 2021. Visual place recognition: A survey from deep learning perspective. *Pattern Recognition*, 113, 107760.