Spatial Analytics for the Identification of Salient Urban Areas

Zomer, Lara-Britt; Duives, Dorine; Cats, Oded; Hoogendoorn, Serge

Publication date
2017

Document Version
Final published version

Published in
CUPUM 2017

Citation (APA)
Zomer, L.-B., Duives, D., Cats, O., & Hoogendoorn, S. (2017). Spatial Analytics for the Identification of Salient Urban Areas. In *CUPUM 2017: 15th International Conference on Computers in Urban Planning and Urban Management*

Spatial Analytics for the Identification of Salient Urban Areas

Lara-Britt Zomer, Dorine Duives, Oded Cats and Serge Hoogendoorn

Abstract

Spatial urban route knowledge consists of the internalized representation of a sequence of actions to be performed at certain locations, cued by wayfinding landmarks. Determining the location of distinctive landmarks is thus important in research on route choice, urban cognition, and travel information. Currently, most approaches to identify landmarks require vast data collection efforts. To overcome these demands, this study proposes a spatial analytic method able to handle open-source datasets to identify urban wayfinding landmarks as salient urban areas.

The method consists of five steps based on data management, grouping analysis, and cluster and outlier analysis. Determinants to identify salient urban areas are building volume, surface, height, building year, and the number of buildings in a 100 square meters grid-cell.

Findings have been applied to identify differences in distribution of clustering and dispersion between local and global salient urban areas using the Gini coefficient, based on an open-source GIS dataset on the built environment of Amsterdam.

L.B. Zomer (Corresponding author) • D.C. Duives • O. Cats • S.P. Hoogendoorn
Department of Transport and Planning, Delft University of Technology, Delft 2628CN Delft, NL

L.B. Zomer
Email: l.zomer@tudelft.nl

D.C. Duives
Email: d.c.duives@tudelft.nl

O.Cats
Email: o.cats@tudelft.nl

S.P. Hoogendoorn
Email: s.p.hoogendoorn@tudelft.nl

1. Introduction

City users, to some extent, rely on memorized urban route knowledge to decide how to move from one place to the next. To this end, spatial urban route knowledge can be viewed as remembered sequences of landmarks, that, combined with directional actions support users to navigate across town. Following Lynch (1960) and Appleyard (1970), landmarks are defined as salient geographic objects, points, or polygons of buildings that structure the internal representation of a city (Richter and Winter 2014).

Over the last two decades, different approaches to identify and integrate landmarks have been developed, as can be noticed, e.g. in route descriptions. As such approaches require large-scale, detailed, diverse datasets, and correspondingly demanding data collection methods (Richter and Winter 2014), today, knowledge on the effects of urban landmark distribution on wayfinding behavior remains limited.

This study aims to contribute to methodology with an approach to handle open-source data. To do so, the concept of aggregate urban landmarks, coined as salient urban areas, is introduced. Salient urban areas possess noticeable characteristics that make them distinct from their surroundings. From a theoretical perspective, a landmark is salient (distinct) in relation to its immediate surrounding or context at large. Salient urban areas are considered unique, either because of dissimilarities to their (local) area, and/or else, because of characteristics considered similar in comparison to other (global) areas. Presumably, the more distinctive a landmark or area, the easier it will be to memorize and incorporate this saliency into the spatial route knowledge to be drawn upon in future. Therefore, salient urban areas are hypothesized to be important to structure spatial knowledge in long-term memory (Couclelis et al. 1987; Sadalla et al. 1980; Montello 1997).

Any method to identify salient landmarks has to be applicable in large-scale environments presenting unequal distributed data. Using open-source data on Amsterdam's urban structure, this study examines whether a spatial analysis approach is useful to identify salient urban areas. First, determinants in urban environments will be defined. As previous studies focused on identifying and integrating landmarks as salient buildings, metrics for salient urban areas can be inferred. Next, to allow for a systematic analysis, a cellular grid (100 square meters per grid-cell) is projected covering the case study area in Amsterdam. Last, grid-cells' determinants are spatially analyzed to identify the characteristics of local hotspots and global clusters of salient urban areas.

Section 2, synthesizing prior studies, offers insights into the identification of urban landmarks and methods to conduct research on landmark identification in relation to wayfinding behavior. Section 3 elaborates on the research approach and methodology. Results, presented in section 4, are categorized into three subsections, starting with a descriptive analysis of the determinants used for the Amsterdam case study. Next, both the findings on the identification of salient urban areas and analyses on the (spatial) distribution of identified landmarks will be put forward. Section 5 summarizes the conclusions and provides recommendations for further research.

2. Defining Urban Landmarks

In this paper, the literature review on the influence of urban structures on city users' wayfinding behavior focuses on landmark identification and urban typologies. To this end, section 2.1 presents first insights derived from cognitive sciences regarding landmark identification, and, next, in section 2.2 urban morphology techniques to distinguish urban typologies are put forward.

2.1 Identifying landmarks

The concept of landmarks originates from Lynch's research (1960) in which five elements according to which cities are perceived, comprise paths, nodes, landmarks, edges, and areas. Appleyard (1970) combines landmarks, being both objects in space and internal representations, with the notion of salience and hypothesizes the more unique a building is, the more likely it will be incorporated into survey knowledge.

Based on memorized buildings in one's "home town", Appleyard identifies significant determinants, both for local (neighborhood) and global comparison (across city areas) based on memorized buildings in a "home town". Using the correlation between property and frequency of recall, the author distinguishes three properties: form (contour, building volume, visual attributes of the façade), semantics (intensity and uniqueness of use), and structural (location and structure of environment). Resulting from Hillier's and Hanson's space syntax theory (1984), a fourth property, visibility (frequency of being in-sight and proximity to a vantage point), has been added (Morello and Ratti 2009). Working on the isovists idea regarding visible sights, Morello and Ratti argue urban environments will be legible due to their location-based visibility. Richter and Winter (2014) hold a building's total salience to become stronger as its distinctiveness on more categories increases.

It appears, whereas in urban planning, landmarks appear firmly grounded concepts, their appliance to large-scale environments is cumbersome, particularly, when buildings are unequally distributed. Based on Lynch, regarding their identification, generally, landmarks are analyzed as geo-referenced points or buildings. Although, resulting from social data, using pictures, new approaches to identify landmarks from (geo-referenced) user-generated data are being developed (Duckam et al 2010; Richter 2007), an aggregated, cellular approach is still lacking.

2.2 Landmarks determinants based on Urban Morphologies

Landmark identification frameworks appear intricate to apply to spatial experiences (Stevens 2004). This may be one reason why systematic research on how people learn and comprehend novel urban environments – i.e., how people organize, group, differentiate and catalogue their perceptions while moving across town- remains limited. Urban morphology aims to understand spatial structures and patterns, e.g., physical layouts of urban environments, but its underpinning methods have not been applied to identify salient landmarks in urban environments.

Levels of analysis in urban morphology range from regional to continuous points. Main objects of interest are building blocks, followed by neighborhoods. As can be noted in Table 1, in urban morphology, determinants may be quantified along different scales, and, moreover, depending on particular research goals, decisions as to what determinants to include, may vary. The earliest methods distinguish typologies (urban atmospheres) based on conceptual differences (Lynch 1960; Conzen 1960; Duany 2002; ABF Research 2003). From 2000, different methods have been used, such as plotting against two axes (Marshall 2005; Berghauser-Pont and Haupt 2010), or by applying a characteristic, to some extent (Morello and Ratti 2009; van Nes et al 2012; Oliveira and Medeiros 2016).

We conclude that although various methods and techniques have been developed to identify landmarks in relation to wayfinding behavior, little is known on how the distribution of landmarks in large-scale urban environments actually effect wayfinding behavior. From literature, shape turns out to be a consistent indicator in both landmarks and urban morphology, and, therefore, urban grid-cell landmarks will be identified using aspects of shape.

Scale	Determinants
Regional	City size Density (FSI, GSI) Proximity of services Land use mixture Building period
Plot or neighborhood	Town plan Land use pattern Composition of network hierarchy and directionality Configuration of intersection and connectivity of network Betweenness centrality Building density (FSI, GSI)
Street	Average (pedestrian) flow
Building block	Density and volume of the built environment (spacematrix) Land use mixture Building form pattern
Grid	Spatial integration of axial lines Building densities (FSI, GSI) Land use mixture
Continuous	Accessibility of network Ground Space Index (GSI) Building year Mixed building usage

Table 1. Landmark determinants in urban morphology

3. Research Approach and Methodology

This section, first, introduces the research approach, followed by an explanation of data processing procedures and spatial analyses using ArcGIS to analyze salient urban areas. Last, the case study area and cleaning processes on open-source data will be discussed.

3.1 Research approach

Spatial route knowledge on a city can be conceived of as the cognitive level of route choices, consisting of memorized (orders of) landmarks. It is hypothesized for landmarks characterized by more noticeable local or global (dis)similarities to be easier to memorize, and, thus, to be more probable to become part of the cognitive level of route choices. Following Lynch, in order to be distinct from its nearby surroundings, a landmark is to be strongly dissimilar from local buildings. Likewise, clusters (neighborhoods) can be considered distinct when there is a strong global similarity in terms of continuity and delineation of space. As form turns out a consistent characteristic, both regarding landmarks and urban morphology, urban grid-cells are identified by contour (Ground Space Index – GSI), volume (Floor Space Index – FSI and number of floors – L), and visual attributes of the façade (building year).

3.2 Data processing procedure and analysis

As concluded in section 2, due to approaches requiring vast data collection efforts, knowledge on the effects of the distribution of landmarks in large-scale urban environments on wayfinding behavior remains limited. To try and fill this gap to some extent, we propose a spatial analysis method to translate detailed disaggregate data of large-scale urban environments into meaningful and computationally efficient aggregate data. Below, five steps comprising the spatial analysis method are introduced. In 3.3.1, these methodical steps will be applied to the case study.

Step 1. Create map layers from data.

- 1.1 Create grid-cells using a fishnet that superimposes the area of interest.
- 1.2 Assign available data to grid-cells.

Step 2. Iterative grouping analysis to identify how the determinants relate to different urban morphologies.

Step 3. Cluster and outlier analysis based on Anselin Local Moran's I using the determinants of interest as input fields (Anselin 1995).

- 3.1 An inverse distance squared is used because nearby neighboring grid-cells have a much larger influence than grid-cells further away.

Step 4. Unite map layers of cluster and outlier results of relevant determinants.

Step 5. Create maps of urban salient areas.

- 5.1 Local urban salient areas: Cumulative summation of all low negative z-scores indicate statistically significant spatial outliers of a high value surrounded by low values (HL) and a low value surrounded by high values (LH).
- 5.2 Global urban salient areas: Cumulative summation of all high positive z-scores that indicate statistically significant clusters of high values (HH) and low values (LL).

3.2.1 Discussion of the elements of the spatial analytic method

Upon the creation of grid-cells, it has to be ensured such cells are of large enough size to contain at least one feature, and small enough to allow for variety within urban plots. Also, the size of grid-cells should be suitable for further analysis. Furthermore, dependent on available data, the specific combination of spatial joints, intersections, dissolves and unions to be used to transform the data to grid-cells will have to be decided.

Grouping analysis is used as an exploratory analysis to reveal underlying structures of the determinants of interest to be used to identify clusters of distinct urban areas with similar physical characteristics (Jain 2009).

Cluster and outlier analysis is applied in many domains, such as economics and geography to identify concentrations of values and outliers that explain (behavioral) patterns (Anselin 1995). This analysis is often preferred over hotspot analysis based on the Getis-Ord Gi*, as it also identifies statistically significant spatial outliers, which are expected to be the most important aggregate urban landmarks.

3.3 Amsterdam as a Case Study

Next, the case study area is presented as well as the operational choices needed to apply the spatial analytic method to the case study.

Founded in the 13th century, Amsterdam is situated along the river "het IJ" and the Amstel delta. Following several expansion periods, Amsterdam's residential area covers over 165 km^2. The open-source GIS data provided by the City of Amsterdam to analyze the urban structure of Amsterdam can be downloaded at http://maps.amsterdam.nl/open_geodata. This dataset consists of 471.580 BAG (key registers of addresses and buildings) address points that include attributes concerning building year and usage surface, and 17.791 polygon shapes of GBKA building typology (large scale topography), by which the surface area can be calculated to represent the Ground Space Index (GSI). Due to missing and incorrect geo-coded data, 23.532 (5%) points of the BAG addresses have been excluded because they did not include a building year, and 17.021 (4%) have been excluded because they did not include a usage surface. Regarding the polygons of buildings from GBKA, 13.437 (75%) polygons containing a BAG address point together with a building year and/or usage surface have been included.

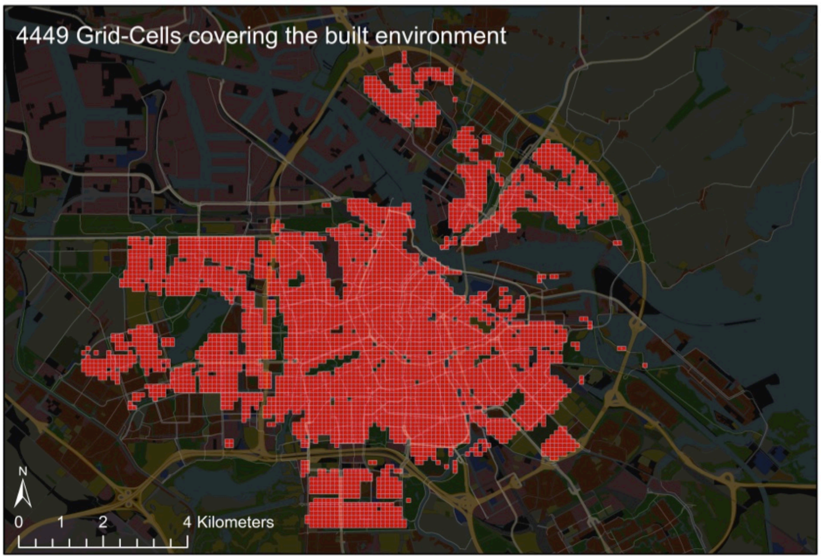

Fig 1. Case study Amsterdam, and the main road network. In red: 100 square meters grid-cells with one or more built environment features.

3.3.1 Applying five steps of the spatial analytic method to Amsterdam

Step 1. To ensure that most cells are large enough to contain at least one feature, and, simultaneously, small enough to allow for diversity within urban plots, the area of the grid-cells is set at 100 square meters. After data cleaning and overlaying open-source point data on BAG addresses with the polygon shapes of corresponding buildings, Floor Space Index (FSI), Ground Space Index (GSI), and number of floors (L) are calculated. As a direct spatial joint of many-to-one does not exist in ArcGIS, several steps (intersections, spatial joints and dissolves) are necessary to link the processed open-source data sets to the grid-cells. The final step in data processing combines all processed layers to one, which can be used for spatial analyses. This resulted in "aggregated" data for 4449 grid-cells, each containing at least one building, and a maximum of 73 buildings, see Figure 1.

Step 2. In deciding which determinant to include, a first selection is made based on the semantic meaning in relation to the concept of aggregate urban landmarks. Secondly, meaningful determinants should have at least one determinant where similar values are dispersed, and one determinant where similar values are clustered (referring to local and global salient urban areas). Table 2 gives an overview of these intermediate steps as well as 12 determinants that are hypothesized to describe the spatial pattern of salient urban areas. N/A indicates the determinant is either not applicable or there is no assignment; Low or High indicate whether low or high values are spatially dispersed (local) or clustered (global). The blank fields indicate a random distribution of the values of meaningful determinants.

		Count	Sum	Min	Max	Range	Std
Local	*Building Year*	N/A	N/A		N/A	High	High
	Building Volume (FSI)	N/A		High			
	Contour Surface (GSI)	N/A		High			High
	Levels	N/A	N/A	N/A	High		
	Number of buildings per grid-cell		N/A	N/A	N/A	N/A	N/A
Global	*Building Year*	N/A	N/A	Low	N/A		Low
	Building Volume (FSI)	N/A	Both				Low
	Contour Surface (GSI)	N/A	Both				Low
	Levels	N/A	N/A				Low
	Number of buildings per grid-cell	Both	N/A	N/A	N/A	N/A	N/A

Table 2. Urban landmark grid-cells identification metrics.

Grouping analysis aims to explore spatial patterns and identify the reliability of the 12 determinants hypothesized to describe these patterns. When performing this grouping analysis, each determinant has a Rho^2, describing the extent of discrimination amongst determinants. Because there is no ground truth about either the determinants or the identification of groups, a suitable determinant is defined as a determinant with a low range of Rho^2 for different number of groups. Furthermore, in grouping analyses no spatial constraint is used; features are partitioned using a k-means algorithm to minimize differences amongst features in a group, over all groups. Multiple iterations have been performed to identify suitable combinations of determinants to overcome the limitations of the greedy heuristic.

Step 3. Based on Anselin Local Moran's I statistic (Anselin 1995), cluster and outlier analysis identify statistically significant hot and cold spots and spatial outliers. Incremental spatial autocorrelation analyses provide insight into the maximum spatial autocorrelation. However, for many determinants the distance band turned out too high to ensure that no feature exceeds 1000 neighbors, which results with memory errors. Therefore, the fixed distance band was set at 700 meters. All grid-cells within the distance band are weighted equally.

Step 4-5. Final maps of urban salient areas can be created when the results of the cluster and outlier analysis are combined with "union". The total level of salience of a grid-cell is the cumulative score of significant values. Significant values of low negative z-scores of suitable determinants are summed to represent local salient urban areas. Significant values of high positive z-scores of suitable determinants are summed to represent global salient urban areas.

4. Results

Section 4, first, discusses descriptive statistics, followed by the identification of local and global salient urban areas. Finally, in the last part, a possible application of the spatial analysis approach is discussed, aiming to investigate the spatial distribution of salient urban areas with the Gini coefficient.

4.1 Descriptive statistics on the case study Amsterdam

Descriptive statistics regarding the case study Amsterdam are shown in table 3. The oldest buildings in the dataset date stem from 1300, and the average age of buildings within a grid-cell is 35, with a maximum of 709 years. On average the built volume of a grid-cell is 5145m^3, with a maximum of 110,288 m^3. If the surface of the buildings would be 10,000 this would correspond to 10 floors. On average the surface of buildings cover almost 20% of grid-cells. Within a grid-cell, the average smallest surface equals 343m^2, whereas the maximum equals 10,000m^2. The highest building level within a grid-cell reaches almost 23 floors, while the average building level is below 3 floors. The average value is lower than expected for an urban area like Amsterdam, and probably, results from an incomplete dataset. The average number of buildings within a grid-cell is just over 6, with a maximum of 73.

Figure 2 shows the values for rho^2 found for different grouping analyses. The figure indicates that regardless of the number of groups, building year determinants are most consistent, and the age of buildings within a grid-cell always scores the highest rho^2. The remaining four characteristics (FSI, GSI, number of floors and number of buildings) gain more consistency when 7 to 14 groups are created.

	Determinant	Mean	Std. Dev.	Min	Max
Building year	**Oldest** (min)	1912	78.91	1300	2016
	Newest (max)	1947	33.50	1600	2016
	Range	35.86	81.56	0	709.00
	Std. Dev.	13.61	29.03	0	251.25
FSI	**Average**	5145.64	4,523.64	0	110,288
	Smallest (min)	879.27	2,675.77	0	110,288
	Largest (max)	2521.82	3,120.71	0	110,288
	Std. Dev.	737.02	1,229.74	0	51,736.39
GSI	**Average**	1830.73	1,242.90	0	10,000
	Smallest (min)	343.73	722.34	0	10,000
	Largest (max)	911.78	769.25	0	10,000
	Std. Dev.	256.48	311.93	0	3,971.81
Building level	Average	2.76	1.38	0	22.86
	Lowest (min)	2.04	1.45	0	22.86
	Highest (max)	3.60	2.04	0	22.86
	Std. Dev.	0.58	0.74	0	16.16
Number of buildings	**Count**	6.28	7.97	1	73

Table 3. Descriptive statistics on determinants for case study Amsterdam. Bold: determinants of interest

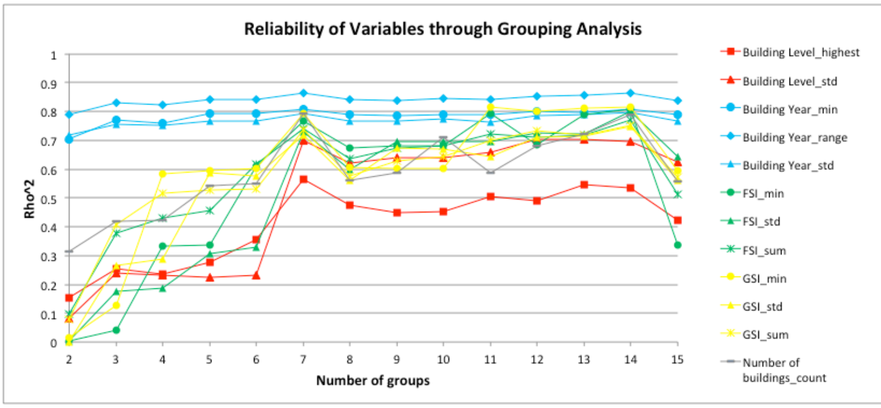

Fig 2. Reliability of determinants: grouping analysis.

4.2 Identification of Local and Global Salient Urban Areas

This section presents the results of step 5 on the identification of local and global salient urban areas.

Regarding case study Amsterdam and parameter settings, as visualized in figure 3A, 494 local salient urban areas are distinguished, covering 11% of the built environment. Highest level local salient urban areas arerepresented by pink grid-cells and comprise, amongst others, Amsterdam Central Station and the Rijksmuseum. From the distribution within figure 3A, it may be expected local salient urban areas cluster more within the historical city center and many local salient urban areas are located near (intersections of) the bicycle street network. Subsequent analysis shows that neighborhood percentages indeed deviate from the city average, e.g., the historical city center has a coverage percentage of 16%, while prewar extension plans like Plan Zuid yield coverage percentages of 10%, whereas percentages for urban extensions during the 1960's, such as Westelijke Tuinsteden, are just above coverage 9%. The Gini coefficient is used to determine how local salient urban areas cluster near (intersections of) the bicycle network as will be elaborated on in 4.3.

Regarding case study Amsterdam and parameter settings, as visualized in Figure 3B, there are 3284 global salient urban areas covering 74% of the built environment. Highest level global salient urban areas are represented by bright pink grid-cells and are central locations, such as Dam Square, Damrak, and the Nieuwmarkt. The images A to C in figure 4 indicate the historical city center, as a neighborhood, contains highest global salience (95% of the grid-cells have salience levels of 1 or higher). Just like the case regarding local salient urban areas, there seems to be a variation amongst urban expansion plans. For example, 70% of Westelijke Tuinsteden have statistically significant clusters of similar urban characteristics, while Plan Zuid reaches a coverage percentage of 58%. Furthermore, from the detailed figure 4B of Plan Zuid it can be seen that global salient urban areas follow the major axial streets.

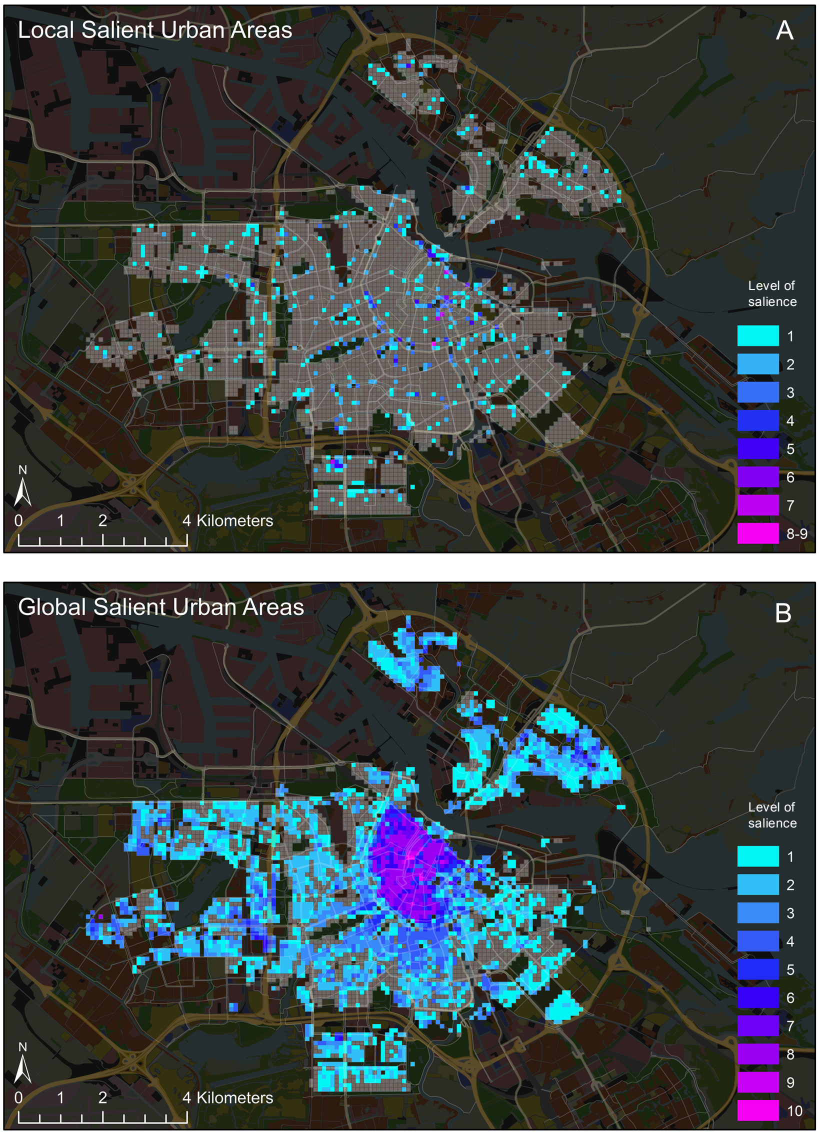

Fig 3A-B. Identification of Salient Urban Areas in Amsterdam.

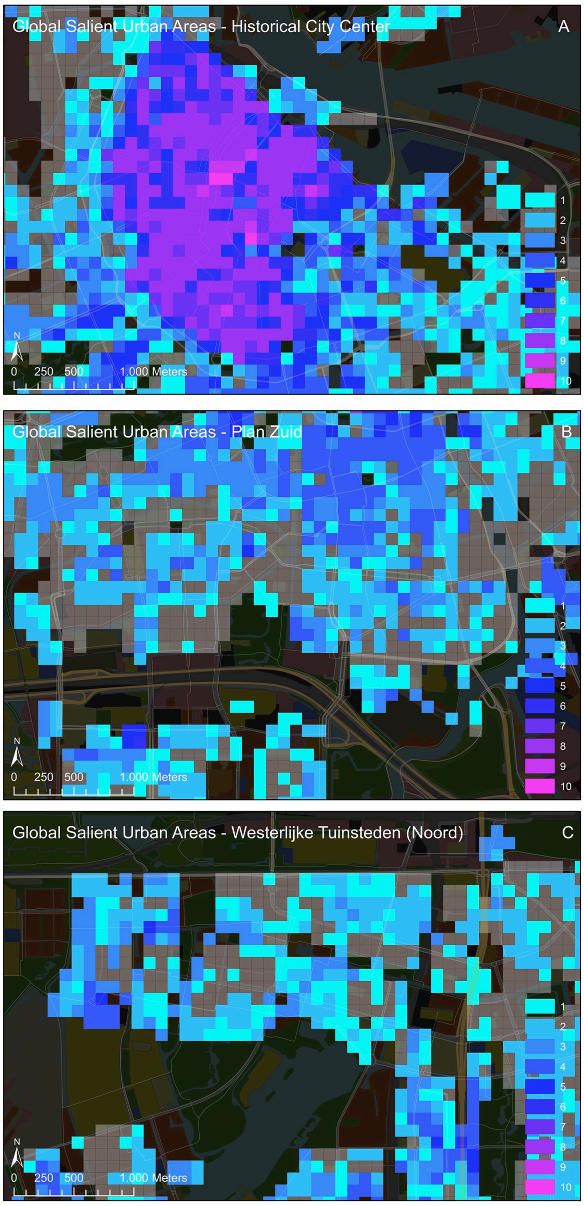

Fig 4A-C. Detail images of neighborhoods and global salient urban areas

4.3 Spatial Distribution of Salient Urban Areas

The last part of this section uses the Gini coefficient as a comparative measure of dispersion relative to salient urban areas within Amsterdam. This analysis is preferred over the multi-distance spatial cluster analysis because it is scale dependent (Tsai 2005). The ratio analyses are used to measure the inequality of the distribution of salient urban landmarks in Amsterdam, based on 1.) the extent to which an urban area is salient, and 2.) the number of salient urban areas within a certain distance field of a salient urban area. For example, a distance field of 300 meters represents 8 grid-cells surrounding a salient urban area. The Gini coefficient can range between 0 and 1, with 0 representing perfect equality, and 1 representing perfect inequality of the distribution of salient urban area in Amsterdam. Brown's formula has been used to calculate the Gini coefficients shown in figures 5A and B.

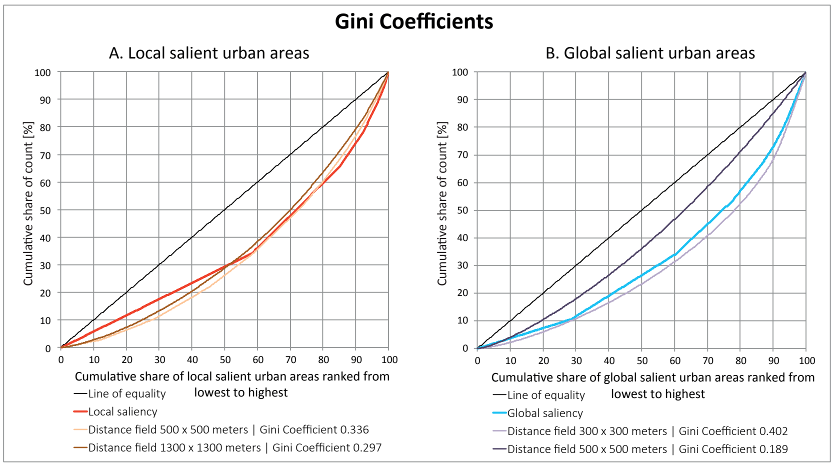

Fig 5A-B. Gini coefficients of local and global salient urban areas

The Gini-coefficient of saliency of local (and global) salient urban areas is 0.30 (0.35), meaning that saliency is distributed rather equally over all salient urban areas. Figures 5A-B show that 58% of the local salient urban areas (28% of global salient urban areas) have only one salient determinant. These percentages correspond to 34% (local), and 10% (global) of the cumulative salience. Both figures 5A-B also indicate the 10% highest levels of salient urban areas correspond to 25% (local), and 29% (global), of the cumulative salience. The Gini coefficient representing the number of salient urban areas within a certain distance range fluctuates between 0.25 and 0.35 (local), and 0.19 to 0.23 (global), depending on the distance field.

In line with previous statements (4.2), this means saliency of local salient urban areas to be slightly more equally distributed compared to global salient urban areas. As to distance fields, more variation is found. The number of local salient urban areas within a distance field, is least equal at 500 square meters, and most equal at 1700 square meters. On the other hand, the number of global salient urban areas, within a distance field, is most equal at 500 square meters, and least equal at 300 square meters.

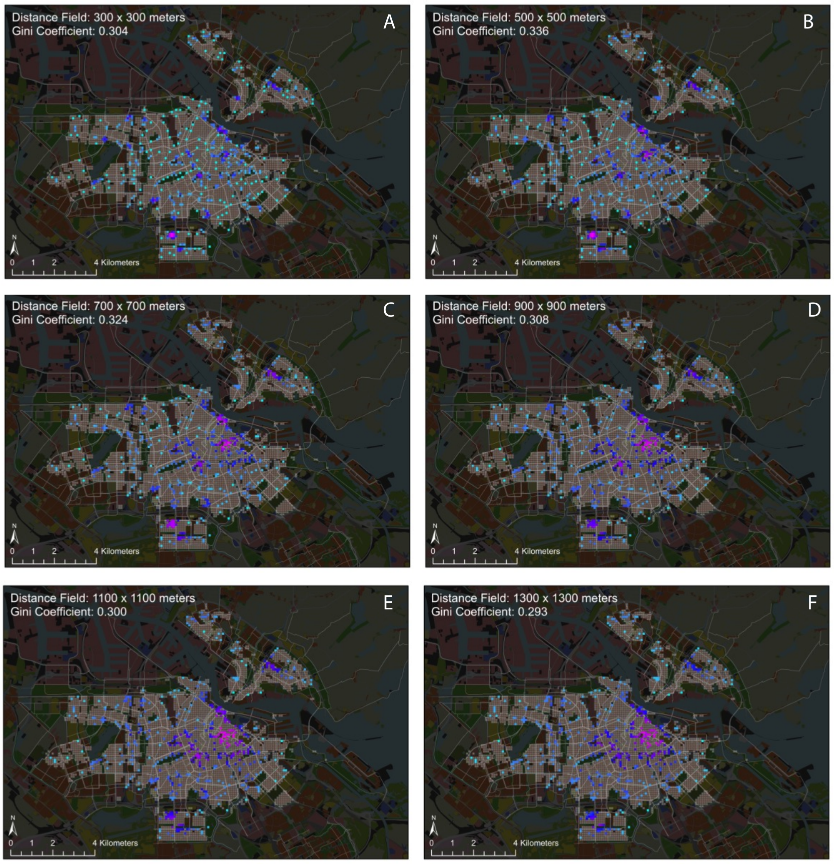

Fig 6A-F. Spatial distribution of Gini coefficients for different distance fields.

In ArcGIS the number of local salient urban areas surrounding one local salient urban area can be visualized for different distance fields used to compute the Gini coefficient. Insights from these maps are complementary to the Gini coefficient, as the latter does not explain how salience is spatially distributed. Figure 6A, for example, shows high values are concentrated around Vrije Universiteit van Amsterdam in the South, containing the smallest distance field of 300 square meters. Medium to high values are concentrated around larger public squares, such as, Central Station and Museumplein. Also, local salient urban areas with lower levels of salience appear to be located along major axial streets.

Figure 6B shows spatial distribution changes according to different distance fields. E.g. a distance field of 500 square meters shows a concentration near Mr. Visserplein. Moreover, it becomes clear, more local salient urban landmarks with relative more local salient urban landmarks are distinguished within the proximity of 250 meters, such as around Vondelpark. By increasing distance fields, local salient urban areas within and bordering the historical city center gain higher percentages, meaning, it is more likely to encounter more local salient urban area when moving across the historical city center. Hence, routes across the historical city center are expected to be easier to memorize and structure in long-term memory.

5. Conclusion and Recommendations

Landmarks are assumed to support wayfinding behavior in urban environments. Determining the location of distinctive landmarks is thus important for investigating route choice processes, structures of urban cognition, and travel information. However, currently most research approaches in this field require highly demanding data collection efforts. To overcome these demands, this study proposes an approach to handle open-source data.

The proposed method combines insights from cognitive sciences and spatial analytics from urban morphologies to identify aggregated local and global urban landmarks based on salient characteristics. The method consists of five steps based on data management, grouping analysis, and cluster and outlier analysis. Results have been applied to identify the differences in distribution of cluster and dispersion between local and global salient urban areas using the Gini coefficient, based on an open-source GIS dataset on the built environment of Amsterdam.

Implications of identifying salient urban areas can provide new insights to analyze how wayfinding landmarks structure environmental knowledge and investigate influences on wayfinding strategies. This environmental knowledge (configuration of landmarks) is assumed to become available when also knowledge has been memorized about the general interrelationships between landmarks (Hirtle and Hudson 1991). If people use these wayfinding landmarks as part of the wayfinding strategy, this is expected to be observable in their route choice behavior. For example it could be more likely to take a detour if more wayfinding landmarks will be passed.

Improved insights can potentially complement navigation apps, physical route signage, and urban planning

More research is needed to verify the parameter settings of this case study, and investigate other determinants. It is expected that digital elevation maps (AHN) or Lidar data will be a better indicator for building level. Further expansion of the determinants can also include traffic intensities, network characteristics, individual movement patterns using GPS and visibility using isovists, and functionalities. To improve validity it would be of interest to investigate to what extent the grouping analyses mimics the way people classify urban typologies through stated preference studies.

Acknowledgement

We thank the anonymous reviewers for the valuable comments and suggestions. The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 Framework Programme for Research and Innovation. It is established by the Scientific Council of the ERC Grant Agreement n. 669792 (Allegro).

References

ABF Research. (2003). Achtergrond informatie Woonmilieu-typologie.

Anselin, L. "Local Indicators of Spatial Association—LISA," Geographical Analysis 27(2): 93–115, 1995.

Appleyard, D. (1970). Styles and methods of structuring a city. *Environment and behavior*, *2*(1), 100-117.

Berghauser-Pont, M. Y., & Haupt, P. (2010). *Spacematrix: space, density and urban form*. Rotterdam: NAi Publishers.

Conzen, M. R. G. (1960). Alnwick, Northumberland: a study in town-plan analysis. *Transactions and Papers (Institute of British Geographers)*, (27), iii-122.

Couclelis, H., Golledge, R. G., Gale, N., & Tobler, W. (1987). Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology*, *7*(2), 99-122.

Duany, A., & Talen, E. (2002). Transect planning. *Journal of the American Planning Association*, *68*(3), 245-266.

Duckham, M., Winter, S., & Robinson, M. (2010). Including landmarks in routing instructions. *Journal of Location Based Services*, *4*(1), 28-52.

Hillier, B., & Hanson, J. (1984). The social logic of space, 1984. *Cambridge: Press syndicate of the University of Cambridge*.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*, *31*(8), 651-666.

Klippel, A. (2003, September). Wayfinding choremes. In *International Conference on Spatial Information Theory* (pp. 301-315). Springer Berlin Heidelberg.

Marshall, S. (2004). *Streets and patterns*. Routledge.

Montello, D. R. (1997). The perception and cognition of environmental distance: Direct sources of information. In *International Conference on Spatial Information Theory* (pp. 297-311). Springer Berlin

Morello, E., & Ratti, C. (2009). A digital image of the city: 3D isovists in Lynch's urban analysis. *Environment and Planning B: Planning and Design*, *36*(5), 837-853.

Van Nes, A., Berghauser Pont, M., & Mashhoodi, B. (2012). Combination of Space syntax with spacematrix and the mixed use index: The Rotterdam South test case. In *8th International Space Syntax Symposium, Santiago de Chile, Jan. 3-6, 2012*. PUC, Santiago, Chili.

Oliveira, V. (2013). Morpho, a methodology for assessing urban form. *Urban Morpohology* 17:149-161.

Oliveira, V. and Medeiros, V. (2016) Morpho: Combining Morphological Measures. *Environment and Planning B: Planning and Design*.

Richter, K. F. (2007). A uniform handling of different landmark types in route directions. In *International Conference on Spatial Information Theory* (pp. 373-389). Springer Berlin Heidelberg.

Richter, K. F.; Winter, S. (2014): Landmarks: GIScience for Intelligent Services. Springer, Cham, Switzerland.

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

Stevens, Q.. "*The Shape of Urban Experience: a Re-Evaluation of Lynch's Five Elements*." In *Evaluation in Progress - Strategies for Environmental Research and Implementation*. IAPS. Vienna, Austria: Österreichischer Kunst- und Kulturverlag, 2004.

Tsai, Y. H. (2005). Quantifying urban form: compactness versus' sprawl'. *Urban studies*, *42*(1), 141-161.

Xia, J. (2007). Modelling the spatial-temporal movement of tourists. *School of Mathematical and Geospatial Sciences*, (February), 1–202.