MIT Open Access Articles

Quantifying Memories: Mapping Urban Perception

Massachusetts Institute of Technology

Quantifying memories: Mapping urban perception

Quantifying memories: Mapping urban perception

Yuji Yoshimura[a,b]**, Shan He**[c]**, Gary Hack**[d]**, Takehiko Nagakura**[e]**, Carlo Ratti**[b]

[a] Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan

[b] SENSEable City Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

[c] Uber Technologies Inc, 1455 Market Street, San Francisco, CA 94103-1355, USA;

[d] Department of Urban Studies and Planning, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

[e] Department of Architecture, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Email: yyyoshimura@cd.t.u-tokyo.ac.jp, heshan0131@gmail.com, gahack@mit.edu, takehiko@mit.edu, ratti@mit.edu

Abstract

This paper discusses the relationship between the spatial structure of the built environment and people's memory of the city as derived from their perceptual knowledge. We explore how spatial comprehension is influenced by the spatial layout pattern in urban settings and individuals' daily activities. In doing so, we seek to determine whether better spatial knowledge is a function of the legibility of the city and of temporal factors, particularly the amount of time spent in a place. For this purpose, we created a web-based visual survey in the form of a geo-guessing game. The participants were asked to guess the locations of random street views within a familiar neighborhood by placing a pin on a map. This system enabled us to measure how well they remember different urban images on the basis of two indicators of spatial familiarity: location identification and visual recognition. Thus, the resulting datasets are quantitatively different from those collected manually by traditional techniques. By analyzing the combination of the quantitative and qualitative datasets, our proposed methodology can clarify previously unknown aspects of the cognitive role in exploring the built environment and on hidden patterns embedded in the relationship between the city's spatial elements and people's mental maps.

Keywords
City image, cognitive mapping, urban computing, visual perception

1. Introduction

This paper discusses the relationship between the spatial structure of the built environment and people's memory of the city as derived from their perceptual knowledge. We explore how spatial comprehension is influenced by the spatial layout pattern in urban settings and individuals' daily activities. In doing so, we seek to determine whether better spatial knowledge is a function of the legibility of the city and of temporal factors, particularly the amount of time spent in a place. For this purpose, we created a web-based visual survey in the form of a geo-guessing game.

Kevin Lynch argues that people navigate in a familiar urban environment using mental maps, or representations of spatial information stored in our minds [1]. A readily imageable urban environment has a high probability of evoking a strong image among various observers. Such an environment would have districts, nodes, landmarks, or pathways that are easily identifiable and can be grouped into a coherent overall pattern.

Our mental maps consist primarily comprises the representation of spatial relationships and some map-like qualities [2]. They register the topological relationships of spaces rather than their absolute coordinates and distances, which are often not useful in interpreting the real environment. This is because our mind tends to simplify visual patterns by reducing complex spaces to a simple collection of basic

shapes [3]. Consequently, mental maps may blur matters of distance and direction but treat topological relations with great clarity [4]. People's spatial knowledge is a complex collection of varyingly perceived items, qualities, and events, the sum of which constitutes in effect a multi-modal representation of the city [5].

Although urban scholars have acknowledged the importance of such mental maps for several decades, the relationships between individual memory and physical elements of the built environment have rarely been analyzed quantitatively. This lack of research knowledge derives largely from the difficulty of collecting the relevant data and the scarcity of robust tools for conducting the analysis. Traditional data collection has relied on expensive and time-consuming manual data collection via in-home or telephone surveys, impairing the accuracy and timeliness of the information delivered (see [6] for a general review of these methods). As a result, previous literature has had to rely on manually constructed forms of data such as hand-drawn sketch maps, along with responses to interviews and questionnaires [1, 2]. The sketch maps provide some general measures of spatial cognition, such as the relative location of places, their shapes, or even perceived distances between places, but the results of sketch maps are difficult to quantify and compare. Furthermore, variances in sketch maps may be attributable to differences in individuals' drawing abilities. Supplementing maps with interviews requires researchers to spend considerable time interacting with participants, and the translation of images into words may be unreliable.

To address these research challenges, we used a web-based survey for data collection, measuring the memories of participants. Google's geo-tagged street views provide numerous urban images, and contemporary web technology enables crowd-sourced data collection. By combining those technologies, we created a web-based visual survey in the form of a geo-guessing game. We asked participants to guess the locations of randomly presented street views from a familiar neighborhood by placing a pin on a map. This method can measure how well people remember various urban images on the basis of two spatial familiarity indicators: location identification and visual recognition. Further, we collected demographic and background information on each participant, including age, past experience of the study area, and frequency of visits. We then combined the datasets in an attempt to externalize participants' memories to better understand the relationship behind the city's structure and the development of people's memory.

This paper is structured as follows. Section 2 provides a literature review and describes the analytical methodologies used in previous studies along with how our methodology relates to this previous work. Section 3 describes in detail the methodology, analytical framework of our paper, and datasets used. Section 4 presents and discusses the results; we then make comparisons with prior work in section 5 and draw conclusions in section 6.

2. Related literature

One major challenge that researchers face when investigating people's mental images of a city is how to externalize an individual's spatial knowledge of a familiar environment. Although several kinds of analytical frameworks have been proposed, the previous research can be classified into three groups.

The first group of studies has employed traditional manual techniques to obtain qualitative data [1, 2, 5, 7–10]. The most common practice is to have subjects produce hand-drawn sketch maps of a specific urban area, based on their recollection. Other methods include oral interviews, questionnaires, and cognition tasks. Lynch [1] asked interview participants to draw sketch maps, collecting 100 samples from each of the three cities. Appleyard [5] conducted interviews and observations and then attempted to correlate urban spatial perception with human perceptual and cognitive processes. Golledge and Spector [7] interviewed 151 residents of a local area and used the data to elicit information concerning the spatial structure of participants' mental maps. Evans [2] employed interviews to investigate the correlations between length of residency, where people work or reside, social class, income, and degree of familiarity with the environment. However, all these methodologies are laborious, tedious, costly, and time-consuming, forcing researchers to base their conclusions on relatively small samples.

The second group has relied on the recent development of computational technologies along with quantitative analysis, collecting relevant data on a large scale and directly studying the relationship between the physical appearance of cities and human behavior [11–13]. Salesses et al. [12] used Google Street View (GSV) images to estimate differences between human perceptions of various urban areas. They employed a crowd-sourcing data collection methodology to rate the perceptions by pairwise comparison of geo-tagged images from four cities. Quercia et al. [11] proposed measuring people's

ability to recognize the urban environment through an online crowd-sourcing platform. In their study, when a participant entered a website, 10 randomly selected GSV photos of local scenes appeared, asking a participant to guess the nearest subway station, borough, or region based on the photo displayed. This approach enabled Quercia et al. [11] to estimate people's mental maps, resulting in a measure of people's recognizability of the city and to create psychological maps with large-scale datasets. These datasets were compared with socioeconomic data to discover the correlation between poor recognizability and social problems.

Isola et al. [14] dealt with the "memorability" of visual information. They measured this trait by analyzing photographs derived from a web-gaming platform, which presented a series of unique photos with repeated photos interspersed, making it possible to systematically quantify how people remember images.

The third group of studies has employed computer vision techniques to analyze people's sense of place [15–22]. Naik et al. [16] trained computer vision algorithms to predict ranked scores with regard to the perceived safety of a street. Similarly, Dubey et al. [15] trained a deep convolutional neural network (DCNN) with human-labeled data from Place Pulse 2.0 to explore the relationship between the visual features of the built environment and humans' sense of place. Zhang et al. [17] analyzed the perceived urban elements that create our sense of place through the combination of training a DCNN with Place Pulse 2.0 datasets and the segmentation analysis of images. Zhou et al. [18] applied a DCNN to explore cities' identities and measured their similarities. The trained DCNN estimated seven high-level attributes of the city's spatial form and its functionality using attribute analysis of geo-tagged images.

These methods can assess people's collective sense of a city, but may not be adequate to capture spatial knowledge or how it develops over time. Since people's spatial knowledge and memory of a place are acquired largely through repeated encounters with the environment, disregarding individual's habits of actually experiencing the urban structure will produce unreliable data. Furthermore, any differences in the subjects' accumulated knowledge and potential biases in perception are typically not considered. For example, a subject's spatial knowledge is partly a function of temporal factors, such as the amount of time spend living in a place or the number of visits made. If these potential confounding factors are not controlled, it is difficult to use the resulting dataset for a reliable analysis of the image of a city.

Our present work is inspired by Quercia et al. [11], who evaluated inhabitants' familiarity with a neighborhood through web-based visual surveys in the form of a guessing games. Our study participants were shown randomly selected GSV photos from a familiar neighborhood and asked to guess what location was shown in each photo by pointing to a map. Thus, people's familiarity with a specific image was assessed by two related indicators: whether they could recall it by the visual image and whether they could locate it accurately on a map. In addition, we collected background information on participants so that we could classify them by their relationship to the study area (e.g., whether they lived in the area or their previous visiting experience). Furthermore, we considered the direction from which each photo was taken, because the landscape and people's recognition of it change greatly depending on the direction from which one looks at an urban environment.

The paper's key contribution is combining the quantitative and qualitative datasets to more clearly understand how participants develop their memory of a city. Our system enables us to capture comprehensive details about every corner of the urban environment. Our methodology avoids conducting lengthy interviews with selected subjects, while the website's carefully designed user interface makes it possible to perform required actions without giving face-to-face guidance. Using online visual surveys exponentially increases the number of participants available and makes collecting qualitative data easier.

3. Methodology and analytical framework

As noted, our methodology combines quantitative and qualitative data collection. Questionnaires of a traditional sort provide qualitative data (age, residency status, frequency of visits to the study area); the online visual survey using the geo-tagged Street View images gives us quantitative data. Combining these two distinct, complementary data sources enables us to uncover people's memory of the city and how it relates to their perceptual knowledge.

3.1 Tool design

To evaluate the respondents' familiarity with different places in the urban area, we designed a web-based visual survey called UrbanExplorer, which functions as a game by which people explore a familiar urban neighborhood. UrbanExplorer uses thousands of geo-tagged GSV images. GSV offers virtually exhaustive street-level views of locations throughout thousands of cities in over 80 countries, taken using similar photographic equipment. Moreover, all GSV images are georeferenced; thus, this extremely large, standardized data source offers a cost-effective way to scan and study the visual structures of cities. These images have been used to quantify urban perception and safety [15, 16], map urban demographics [23], develop socioeconomic indicators [19], assess species' habitats [24], and measure urban tree cover and greenspace [25]. GSV has one other important advantage for our purposes: it shares a perspective and spatial resolution comparable with how humans see, experience, and collects visual information on the outdoor urban environment.
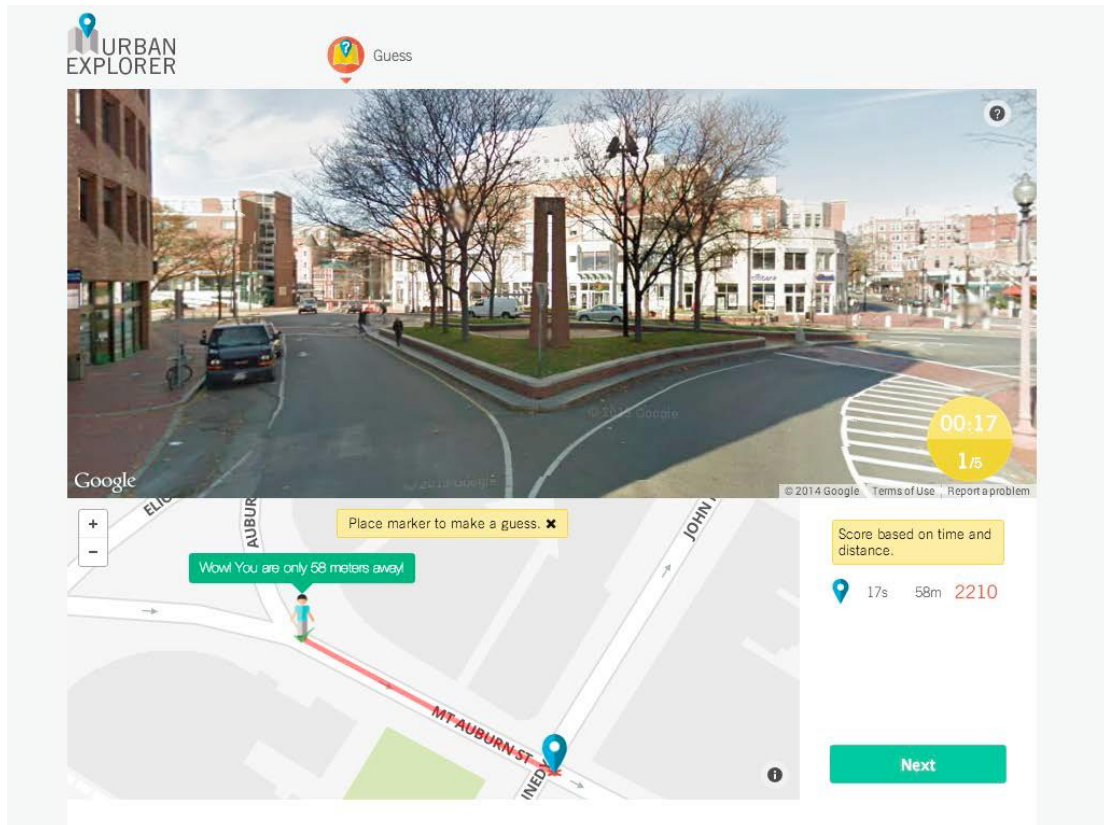

Figure 1. The user interface of UrbanExplorer

Figure 1 shows the user interface of UrbanExplorer. In the upper part of the screen, the randomly selected GSV is displayed; the lower portion shows a map. The participant is asked to guess the photo location by pointing to a map. Once the participant has made a guess, the real location of the GSV is revealed, along with the time required to make a decision (see Figure 1, bottom right). All guesses and associated information are stored in the database with the following attributes:

Stored information	Description of information
Place ID	Each street view is given a unique ID for easy retrieval
Time spent to guess	The time is calculated from the moment when a photo is shown to the moment when the person clicks the guess button
Location of street view	Longitude and latitude of the street view
Location of guess	Longitude and latitude of the guessed location
Distance to answer	The distance between the guessed and correct locations is calculated
User information	Information on the user, obtained from the demographic and background questions
Time of creation	Time when the guess was made

Score	Score is a function of distance and time spent, based on a scale of 1 to 2,500

Table 1. Attributes associated with each guess

3.2 Questionnaires

We collected relevant background and demographic information on each participant. To do this, UrbanExplorer asked each participant to answer three questions after they completed their guesses on the photos. The first question requested their past visit experience of the study area, with the following answer choices: (1) live/lived in, (2) work/worked in, (3) live/lived and work/worked in, (4) visited, and (5) have never been to. The second question asked participants to give their age as either (1) under 25, (2) 25 to 34, (3) 35 to 44, (4) 45 to 54, or (5) older than 55. The third question offered three choices with regard to the frequency of visits to the study area: (1) fewer than 3 times, (2) 3 to 10 times, and (3) more than 10 times.

This additional information enables us to associate people's memory of the city with their age and previous experience of the location and, more importantly, help us observe how people's memory develops, depending on the time spent in a location.

3.3 The selection of the study area and GSV images

We chose the Harvard Square neighborhood in Cambridge, Massachusetts, USA for our study area. There are primarily three reasons for this selection: First, Cambridge was one of the cities that Kevin Lynch studied in his seminal work on how people use visual cues to organize mental maps of cities. Thus, using the same location permits us to compare our results with his. Second, Harvard Square is heavily visited by both tourists and local residents. Due to its adjacency to Harvard University, the faculty staff, student, and the business people working at the start-ups, among others, commutes to the area daily. Actually, the Harvard Square MBTA stop is a major transportation hub that links the local subway, buses, and taxis, transporting thousands of people to and from the area every day. This environment permits us to measure the urban familiarity by the different residency statuses of the people. Finally, the road system in this area is complex and irregular. Few streets are in a grid arrangement, and most of the major intersections are not perpendicular. This urban morphology provides us to study whether the irregular and curved small streets could also enhance the people's familiarity of the district, i.e., the relationship between the spatial layout of the built environment as the legibility of the city and people's memory of the city.

Figure 2. (a) The location of all street views selected around Harvard Square; (b) sample images of the node; (c) sample images of the link.

Figure 2(a) presents the study area, showing the location of all collected street views. We retrieved the geo-tagged GSV images along with the street network of the study area from the GSV interface. Figure 2(b) presents sample images of one node, and Figure 2(c) shows the sample images of a link. This study focuses on the distinction between the node and link because people usually refer to the nodes as key spatial reference points or landmarks, according to Lynch's [1] five elements and Golledge's [7] anchor point theory, which proposes that the spaces in our mind are formed as a linked-node structure. However, this theory is primarily concerned with each individual's spatial knowledge and might not apply to the collective mental maps of the city. Golledge [26] argued that landmarks acted as anchor points for organizing other spatial information into a pattern. The complex urban forms are thus stored in our memory in the form of a linked-node configuration. The process of acquiring spatial knowledge involves continuously adding new nodes to the existing node–link framework [27].

For each node, three to four images were selected, looking at the location from different points of view. For each link, two images were selected, looking from different directions. In this phase of the selection of the images for each node and link, we manually checked them one by one and chose some of the most adequate ones. The selection was conducted based on the advice from the technical staff, who are familiar with this area, because the image of each node and link should be the representative ones for each location, which display the characteristics of its places. Finally, through these processes, we collected 190 geo-tagged GSV images around Harvard Square and 50 GSV images of other popular places in Cambridge. UrbanExplorer made random selections from among these 240 images and showed them to each participant. No image could be shown more than once to the same participant.

For quantifying the overall level of familiarity with each location, we computed the score based on the system as follows: The score is the sum of the percentage of guesses within 50 meters of the location plus the percentage of guesses between 50 to 100 meters multiplied by 0.75, plus the percentage of guesses made between 100 to 150 meters multiplied by 0.5. Specifically, we used the following formula:

$$S = (x_1 \char`\^ 1 + x_2 \char`\^ 0.75 + x_3 \char`\^ 0.5)/x \tag{1}$$

where x_1 expresses the number of guesses within 50 meters of the correct location, x_2 represents the number of guesses between 50 and 100 meters from the location, x_3 indicates the number of guesses between 100 and 150 meters from the location, and x represents the total number of valid guesses. Thus, the overall familiarity of the place can be inferred from the collective actions and responses of all players. The possible scores range from 0 to 1; an image would receive a score of 1 if all guesses about its location were within 50 meters. Since the location of each photo is already known, the scoring system is linked to each photo through our algorithm, resulting in automatic mapping and visualization.

4. Results

This section presents the obtained results for our study and describes the basic statistics of our data in terms of participants and their guesses. Moreover, it presents the most and least familiar places, calculated by our scoring system and shows the spatial distribution of the people's urban familiarity. The comparison between the node and link helps us understand the potential cause and visual cues for enhancing people's spatial familiarity. Finally, how people's urban familiarity develops, depending on the people's past visit experience to the study area in also discussed.

4.1 Descriptive statistics of the participants and guesses

We collected data from 394 respondents. A total of 4,216 guesses were made, of which 3,617 were made by people who also completed the survey. On average (excluding the guesses not associated with a completed survey), each image received 15 guesses. Most of the participants were students or faculty members. Overall, 68.0% of the participants stated that they currently or previously lived and/or worked in Cambridge, 79.4% of the subjects claimed to have visited the study area more than 10 times, and 15.5% reported 3 to 10 visits; 84.0% of the subjects were under age 35.

4.2 Most familiar and least familiar places

Figure 3. (a) The 15 most recognized GSVs; (b) the 15 least recognized GSVs

Figure 3(a) shows the 15 most recognized GSVs and their scores. Numbers in yellow circles indicate nodes and numbers in blue circles represent links. The score, which each image obtained, is based on formula (1) that we explained in the previous section. The places receiving the highest scores were primarily public spaces at the center of the area, such as Harvard Square, Brattle Square, and Winthrop Square. Certain stores at the center of Harvard Square (e.g., the Harvard Co-op) and along Massachusetts Avenue also received high scores. In fact, the most recognized place was the Qdoba Mexican grill along Massachusetts Avenue, with a score of 0.847. The links with the highest scores were streets within one block of the neighborhood center (e.g., JFK Street and Massachusetts Avenue).
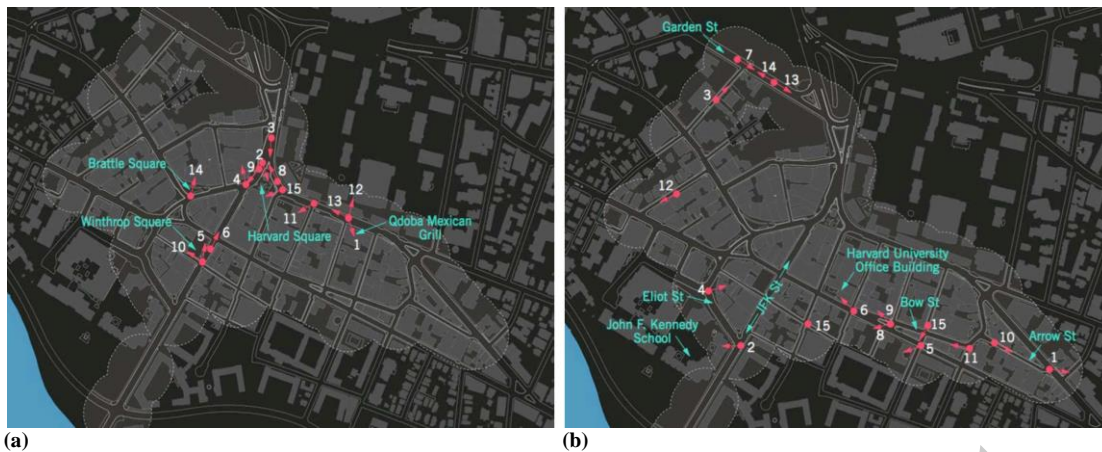
(a)   **(b)**

Figure 4. (a) The locations of the 15 most recognized GSVs; (b) the locations of the 15 least recognized GSVs

Figure 4(a) presents the mapping of the locations of the 15 most recognized GSVs. The geographic distribution of those places indicates that the degree of interaction and the proximity to the center are highly correlated with the probability of being recognized. Moreover, it is clear that the irregularity of urban structure does not prevent the formation of strong mental images. Most of the highly recognized places are in the center of the study area, where the road system is extremely complex; the shapes of these "squares" are actually triangular and the roads passing through them are curved. This finding seems to contradict our intuition that a clear urban structure is more likely to evoke strong mental images. However, a legible urban structure does not necessarily require a regular city grid. Rather, patterns of use trump geometric regularity, and even distorted streets can have a high level of imageability [1].

In contrast, the places with the lowest scores are primarily school or residential buildings with no eye-catching signs or distinct features and streets with minimal activity (e.g., Garden Street, Bow Street, and Arrow Street). The least recognized places are all quite distant from the center of the study area, as shown in Figure 4(b).

Some other specific findings are worth mentioning. For example, the second least recognized GSV was the John F. Kennedy School of Government building, located at the intersection of JFK Street and Eliot Street. JFK Street is the busiest street in Harvard Square, with heavy traffic and frequent congestion. The fact that a building at one busy intersection is unfamiliar to people indicates that frequently passing by a place does not necessarily evoke strong images. We speculate that institutional structures engage fewer passers-by than commercial areas do.

The sixth least recognized GSV image is the back of the Harvard University Office Building (Smith Campus Center), which takes up the entire block. However, the eleventh most recognized GSV shows the front of the same building, where shops and large red windows dominate the three-story podium in the center of the view. This observation suggests that tall structures in a dense environment do not necessarily evoke strong mental images, because pedestrians are more likely to look at features that are visible at eye level.

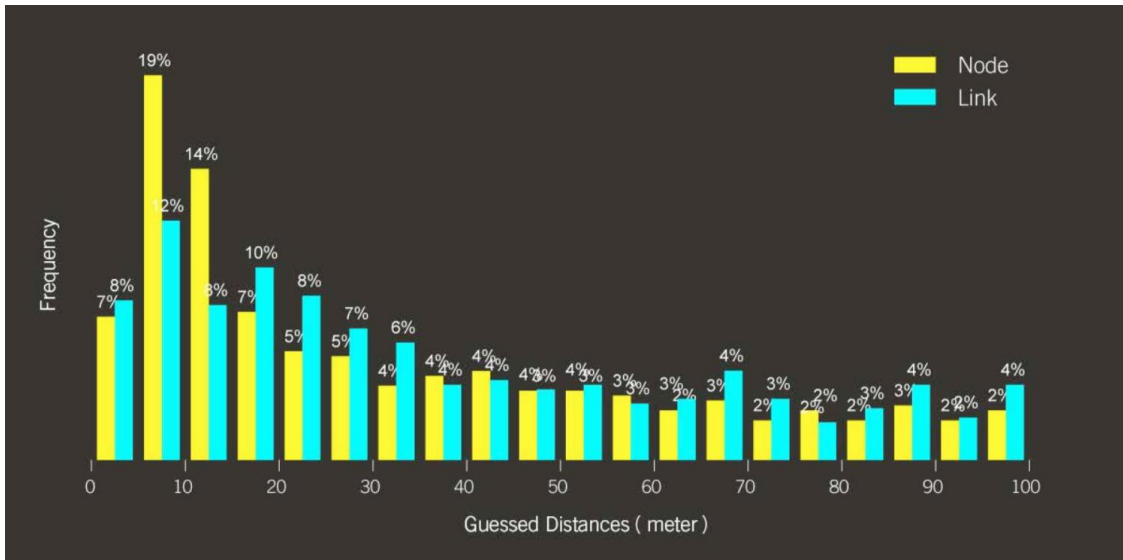**4.3 Differences in accuracy of guesses by nodes and links**

Figure 5. Distribution of guessed distances

Figure 5 presents the distribution of the distances between each guess made by a participant and the true location. Of all the answers that were within 100 meters of the correct location, if the actual site was a mode, almost half of the answers were within 20 meters of the exact location, whereas only 38% were within 20 meters for links. This result indicates that if people recognize an image, they tend to place the marker more accurately when the image at which they are guessing is a node rather than a link.

(a)   (b)

Figure 6. (a) Visual representation of the scores of nodes; **(b)** visual representation of the score of links

Figure 6(a) depicts the scores of all nodes. The size of the ellipses indicates the score (larger colored areas indicate higher scores). The distribution of the nodes' scores coincides closely with our observations of pedestrian traffic patterns. The nodes most familiar to people are located along Massachusetts Avenue and JFK Street, the two busiest streets. Most of the nodes along these two streets point to stores or shops on the street. Along Massachusetts Avenue, the more familiar views are mostly toward the south side of the street where shops predominate, whereas the north side is primarily occupied by institutional buildings. Likewise, among the nodes along Mt. Auburn Street, the more familiar views are all toward the south. These two parallel streets seem to act as magnets drawing attention to the area, due to the presence of their many restaurants and cafes.

Figure 6(b) depicts the scores for all links. The more familiar streets are Massachusetts Avenue and JFK Street. A pattern can be observed for the series of streets between Massachusetts Avenue and Mt. Auburn Street; the closer the street is to the center, the more familiar it is to most people. This finding again verifies the earlier observation that proximity to the center is strongly correlated with the degree of familiarity.

Our experiment required participants to guess photo locations without rotating the view (as one can normally do when using GSV). In this way, we could compare the relative familiarity of different views of the same geographic location. At most locations, the familiarity varied to some extent, and in a few cases they were dramatically different. For example, Zinnia Jewelry, the shop next to the Qdoba Mexican grill–which was the most recognized of all places–was seldom identified, although the two businesses are only 10 meters apart. Almost 70% of participants identified Qdoba's location to within 20 meters, but only 5% accurately located Zinnia within 20 meters, and nearly 80% of participants did not even guess within 800 meters, which suggested that they did not recognize this place at all. One key difference is that Qdoba is a fast-food restaurant noticed or visited by a large number of people, whereas Zinnia is a small jewelry shop that attracts relatively few people. Moreover, we speculate that the church tower at the very far end of the Qdoba image may provide a hint to the location of this street view, serving as the anchor point and spatial reference.
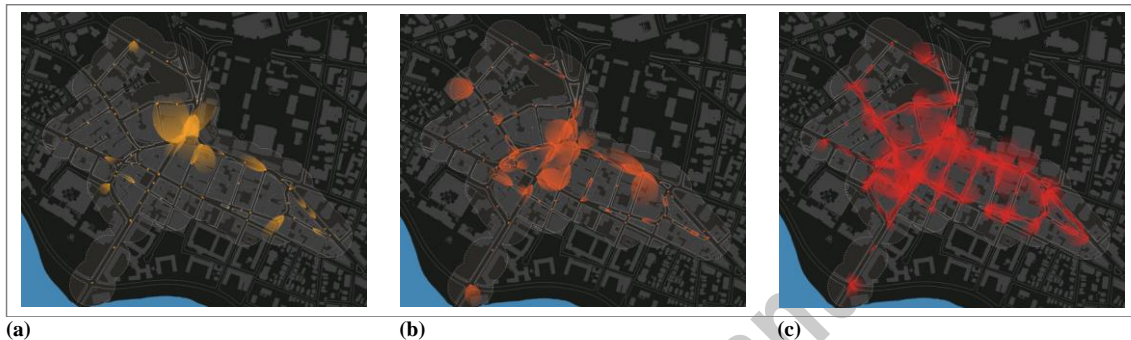

(a) (b) (c)

Figure 7. The development of people's spatial knowledge depending on the number of visits: (a) the scores of people who visited fewer than 3 times; (b) the score of people who visited 3 to 10 times; (c) the scores of people who visited more than 10 times
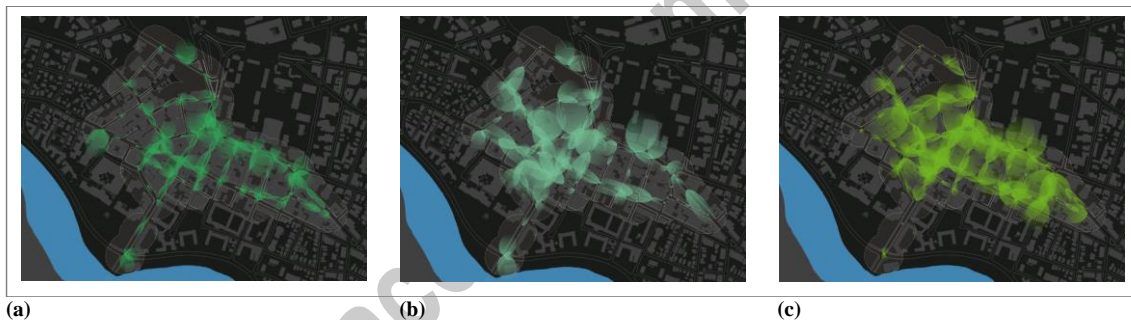

(a) (b) (c)

Figure 8. The development of people's spatial knowledge depending on the status of residency; (a) the scores of people who have lived in Cambridge; (b) the scores of people who have worked in Cambridge; (c) the scores of people who have both lived and worked in Cambridge.

Figure 7 shows how spatial knowledge differs depending on the number of visits made to the study area. When people visit once or twice, they begin to acquire spatial knowledge of the geographic area. On additional visits, they gain more detailed knowledge of the spaces located spatially close to the center, including the parts of streets that connect to the center, along with intersections one or two blocks away from the center. Finally, when people have visited the place more than 10 times, knowledge of the full spatial structure begins to appear. This observation seems to favor the anchor-point theory. Our data suggest that spatial understanding tends to grow outward from the anchor points.

Figure 8 shows how modes of interaction with the city affect spatial familiarity [28]. The familiarities of people who have only lived in Cambridge extend more toward the surrounding area, which consists mostly of residential districts, whereas the familiarities of those who have only worked in Cambridge are concentrated along the busy streets and in the center activity area and tend not to include the parallel streets between Massachusetts Avenue and Mt. Auburn Street. The familiarities among people who have both worked and lived in Cambridge are confined to the middle portion of the study area, framed by Massachusetts Avenue and Mt. Auburn Street, including the parallel streets in between. All the places in this middle portion are very well recognized by people who live and work in Cambridge. Figure 8 demonstrates that the mode of interactions can affect the distribution of familiar places. In general, people know places better if they both work and live in the area.

5. Discussion

The results indicate that places closer to the urban center are more likely to be remembered. We speculate that people's strong image of the city is formed along familiar paths or nodes, which represent the nucleus of the district. Then, their spatial knowledge expands outward, corresponding to Lynch's [1] hypothesis. In addition, inhabitants' memory of the city largely depends on their patterns of daily activities and their relationship to the city (i.e., as residents, workers, or both). People who are both living and working in the city tend to increase their spatial knowledge more fully than those who either live or work there but not both. Thus, the degree of interaction determines the level of familiarity, and modes of interaction affect the distribution of familiar places. Additionally, our findings suggest that larger or taller buildings do not necessarily evoke strong mental images, because pedestrians are more likely to look at features that are visible at eye level. The distinctive features of a building can evoke high imageability if they are readily visible from a pedestrian's perspective; however, places along a busy street may go largely unrecognized if most of the passers-by have limited or no interaction with it.

These findings are largely consistent with previous research, which has revealed that groups with different activity patterns tend to produce disparate mental images. Furthermore, the degree of local understanding of places depends heavily on the amount of time one has spent in the city [7]. There are correlations between the degree of familiarity with the city, length of residence, the location of one's job and dwelling, and personal attributes [2].
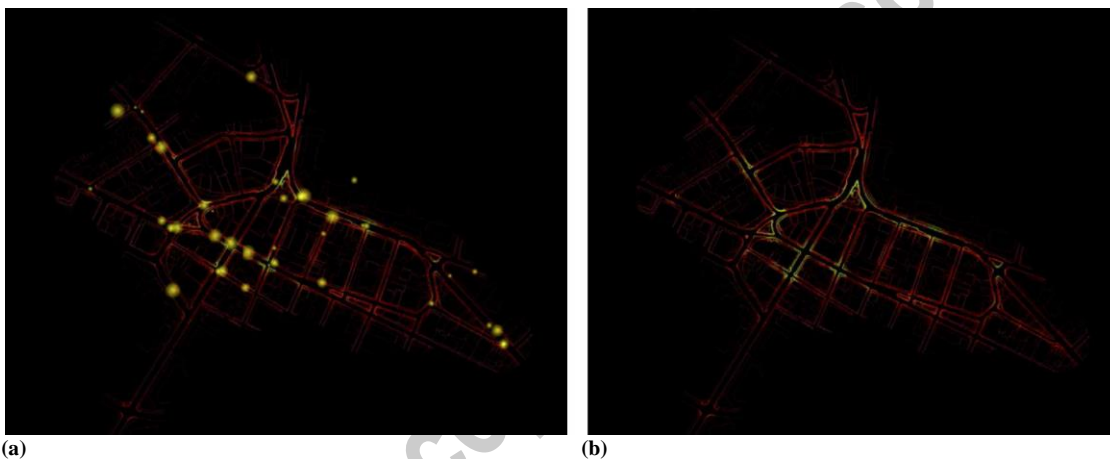

(a)   (b)

Figure 9. (a) The collective memory of the city from a viewpoint of the node; **(b)** the collective memory of the city from a viewpoint of the link.

Conversely, our results diverge from those of previous studies in some ways. First, we discovered that the irregularity of urban structure does not prevent the formation of strong mental images, as demonstrated in Figure 9, which presents the collective memory of the city based on people's responses, highlighting the most and least remembered sections of the cityscape. Lynch [1] argued that a highly legible urban structure can help to increase people's imageability. However, he also argued that such imageability derives not only from the designed and formalized structure (such as a regular grid), but also from more subjective and "fuzzy" perceptions and memories, related to a heuristic visual map of a place over time. This indicates that a distorted street is not always a contributor to visual chaos but can instead be the spatial reference that gives people a way to identify a place and orient themselves at the neighborhood level. Although our findings seem to contradict Lynch's popular interpretation (i.e., that legible structure requires geometric regularity), they confirm that a visual hierarchy of streets can result from patterns of use, even when the arrangement of streets is complex.

Second, our analysis shows that viewpoints of the same location or building from different angles can induce either extremely high or low familiarity. For example, the back view of an office building along a busy street was the sixth least remembered view among the 190 GSVs in that area, whereas the front view of the building was the eleventh most remembered view. The use of the geo-tagged GSVs for crowd-sourced data collection enables us to conduct this high-resolution analysis, which was not possible prior to this study.

Finally, we discovered that the frequency of visits to the district was directly correlated with the degree of richness of individuals' spatial knowledge. After making a few visits, people can remember a limited number of locations with which they have interacted. The more visits they make, the more spots are

added to their memory, reaching outward from the center area. This process, to some extent, aligns with Golledge's anchor-point theory [7], indicating how spatial knowledge is acquired over time.

These findings suggest that although general distribution of familiarities in a small area can be observed as highest at the center of activity and declining toward the edge of the neighborhood, differences between viewpoints of the same place can still be dramatic. Our tool enabled us to survey every corner and angle of the studied area, thereby enabling us to determine that two viewpoints of the same location or building can have dramatically different familiarities. Additionally, our methodology could easily be extended to other cities and neighborhoods, although it was applied to one particular city and a relatively compact neighborhood for this paper. Through it, we could compare results and extract the similarities or dissimilarities to explore some patterns behind seemingly different cities and neighborhoods followed as people seek to construct mental maps of cities.

6. Conclusions

The mental map of an urban environment is crucial for the daily activities of thousands of city dwellers. Until now, however, the data availability on urban perception has been limited and so has our ability to collect a large amount of fine-grained data representing how people comprehend their familiar urban environment. In this paper, we have presented a way to collect and analyze people's responses to measure their spatial familiarity.

The research method employed in this paper provides valuable and novel perspectives on the subject of the spatial familiarity, but it also has several limitations. One concern is our lack of control over participant demographics. People willing to take a survey of this type are more likely to be members of middle- or upper-income groups, who have access to the Internet and are willing to participate voluntarily in academic research. Besides, there is no means to filter irrelevant data, including the readability of images. People might carelessly click around without paying attention to the instructions, and a certain level of bias can derive from the ability to interpret a given image among individuals. Future improvements could be made by using machine learning to control and balance user demographics and by designing a better user interface to eliminate user errors.

Overall, the application of our web-based visual survey in the form of a geo-guessing game allows researchers to rapidly elicit spatial knowledge from a large number of city dwellers and conduct quantitative data analysis. We could collect a large amount of fine-grained data from numerous participants using online surveys and geo-tagged images. In contrast to conventional methods such as interviews and fact-to-face surveys [1], online surveys eliminate the time required to interact with each participant, exponentially enlarging the number of subjects who can be involved and enhancing the potential usefulness of future research. This is a piece of critical information that was not obtainable prior to this study.

6. References

1. Lynch K (1960) The image of the city. MIT Press, Cambridge, MA
2. Evans GW (1980) Environmental cognition. Psychol Bull 88:259–287.
3. Ellard C (2010) You Are Here: Why We Can Find Our Way to the Moon, but Get Lost in the Mall. Anchor, New York, USA
4. Tversky B (1993) Cognitive Maps, Cognitive Collages, and Spatial Mental Models. In: Marciana M (ed) Spatial information theory. A theoretical basis for GIS. European conference, COSIT '93. pp 14–24.
5. Appleyard D (1973) Notes on urban perception and knowledge. Image Environ Cogn Mapp Spat Behav 109–114
6. Yoshimura Y, Sobolevsky S, Ratti C, et al (2014) An analysis of visitors' behaviour in The Louvre Museum: a study using Bluetooth data. Environ Plan B Plan Des 41:1113–1131.
7. Golledge RG, Spector AN (1978) Comprehending the Urban Environment: Theory and Practice. Geogr Anal 10:403–426. https://doi.org/10.1111/j.1538-4632.1978.tb00667.x
8. Milgram S (1976) Psycholohgical maps of Paris. In: Proshansky HM, Ittelson W, Rivlin LG (eds) Environmental psychoogy: People and their physical settings. Holt, Rinehart & Winston, pp 104–124
9. Milgram S, Greenwald J, Kessler S, et al (1972) A psychological map of New York City. 60:181–194
10. Nasar JL (1998) The evaluative image of the city. Sage Publications, Thousand Oaks, CA
11. Quercia D, Pesce JP, Almeida V, Crowcroft J (2013) Psychological maps 2.0. In: Proceedings of the 22nd international conference on World Wide Web - WWW '13. pp 1065–1076
12. Salesses P, Schechtner K, Hidalgo CA (2013) The Collaborative Image of The City: Mapping the Inequality of Urban Perception. PLoS One 8. https://doi.org/10.1371/journal.pone.0068400
13. Quercia D, O'Hare NK, Cramer H (2014) Aesthetic capital: What Makes London Look Beautiful, Quiet, and

Happy? In: Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14. pp 945–955

14. Isola P, Xiao J, Parikh D, et al (2014) What makes a photograph memorable? IEEE Trans Pattern Anal Mach Intell 36:1469–1482. https://doi.org/10.1109/TPAMI.2013.200

15. Dubey A, Naik N, Parikh D, et al (2016) Deep learning the city: Quantifying urban perception at a global scale. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp 196–212. https://doi.org/10.1007/978-3-319-46448-0_12

16. Naik N, Philipoom J, Raskar R, Hidalgo C (2014) Streetscore-predicting the perceived safety of one million streetscapes. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. pp 793–799

17. Zhang F, Zhou B, Liu L, et al (2018) Measuring human perceptions of a large-scale urban region using machine learning. Landsc Urban Plan 180:148–160. https://doi.org/10.1016/j.landurbplan.2018.08.020

18. Zhou B, Liu L, Oliva A, Torralba A (2014) Recognizing city identity via attribute analysis of geo-tagged images. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp 519–534. https://doi.org/10.1007/978-3-319-10578-9_34

19. Naik N, Kominers SD, Raskar R, et al (2017) Computer vision uncovers predictors of physical urban change. Proc Natl Acad Sci 114:7571–7576. https://doi.org/10.1073/pnas.1619003114

20. Been V, Ellen IG, Gedal M, et al (2016) Preserving history or restricting development? The heterogeneous effects of historic districts on local housing markets in New York City. J Urban Econ 92:16–30.

21. Harvey C, Aultman-Hall L (2016) Measuring Urban Streetscapes for Livability: A Review of Approaches. Prof Geogr 68:149–158. https://doi.org/10.1080/00330124.2015.1065546

22. De Nadai M, Vieriu RL, Zen G, et al (2016) Are Safer Looking Neighborhoods More Lively? A Multimodal Investigation into Urban Life. In: Proceedings of the 2016 ACM on Multimedia Conference - MM '16. ACM, New York, USA

23. Gebru T, Krause J, Wang Y, et al (2017) Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. Proc Natl Acad Sci 114:13108–13113.

24. Olea PP, Mateo-Tomás P (2013) Assessing Species Habitat Using Google Street View: A Case Study of Cliff-Nesting Vultures. PLoS One 8:. https://doi.org/10.1371/journal.pone.0054582

25. Wegner JD, Branson S, Hall D, et al (2016) Cataloging Public Objects Using Aerial and Street-Level Images Urban Trees. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. pp 6014–6023

26. Golledge RG (1999) Wayfinding behavior : cognitive mapping and other spatial processes. Johns Hopkins University Press, Baltimore

27. King LJ, Golledge R. (1978) Cities, space, and behaviour: the elements of urban geography. Prentice-Hall, Englewood Cliffs, N.J.

28. Gale N, Golledge RG, Halperin WC, Coucelis H (1990) Exploring spatial familiarity. Prof Geogr 42:299–313.