C-IMAGE: city cognitive mapping through geo-tagged photos

Liu Liu, Bolei Zhou, Jinhua Zhao & Brent D. Ryan

GeoJournal
Spatially Integrated Social Sciences and Humanities

ISSN 0343-2521

GeoJournal
DOI 10.1007/s10708-016-9739-6

Springer

GeoJournal
DOI 10.1007/s10708-016-9739-6

C-IMAGE: city cognitive mapping through geo-tagged photos

Liu Liu · Bolei Zhou · Jinhua Zhao · Brent D. Ryan

Abstract Traditional research categorizes people's perceptions towards city into Kevin Lynch's five elements: node, path, edge, district, and landmark. However, enabled by the proliferation of crowd sourced data this paper utilizes geo-tagged photos to detect, measure, and analyze people's perceptions. This paper introduces a project called C-IMAGE, which analyzes the interactions between city and human perception through the massive amount of photos taken in 26 different cities: one based on the metadata and the other based on image content. Important discoveries through them include that (1) C-IMAGE can partially confirm Kevin Lynch's city image efficiently; (2) There are mainly four prototypes among the tested 26 cities, based on the 7 urban perceptions based C-IMAGE; (3) C-IMAGE shows the gap between subjective perceptions and objective environment while compared to traditional urban indicators.

Keywords City image · Cognitive mapping · Urban computing · Geo-tagged photos

L. Liu (✉)
China Academy of Urban Planning and Design, 168 Linhong Rd, Shanghai, People's Republic of China
e-mail: liuilu@caupd.com

B. Zhou
Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: bzhou@csail.mit.edu

J. Zhao
Department of Urban Studies and Planning, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA
e-mail: jinhua@mit.edu

B. D. Ryan
Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: bdr@mit.edu

Introduction

What does a city look like? For centuries, scientists, architects, planners, sociologists have been trying to answer this question. As a urban designer, Kevin Lynch innovatively offered his answer through mapping city images (Lynch 1960). Based on social surveys and interviews, he collected a large number of perceptions from the public and mapped the image of the city by compiling all of them. Decades have passed, his answer still remains insightful and applicable to current cases. Inspired by the proliferation of crowd sourced photos from the web, this research suggested a new way to tackle the question through cognitive mapping, which could align with Lynch's work for city perception.

The technological trigger comes from the explosive growth of the Internet. It has facilitated an unprecedented social network so that people are connected. Everyone in a city is somehow transformed to a "sensor" through his mobile client, receiving and sharing information all the time. Meanwhile, it becomes tangible to collect large amount of shared information such as photos and tweets from the Internet within short amount of time. From the perspective of urban researchers, learning knowledge from the information shared by those human mobile sensors could lead to the deeper understanding on how people cognize the city and the urban environment.

The project proposed in this paper is named "C-IMAGE". The capital letter of "C" stands for a comprehensive meaning of city, cognition, and computing techniques. Borrowing from Lynch's concept of "city image", the "IMAGE" in this project name is the fundamental research object.

Given the rapidly developed data collection and processing technology, C-IMAGE introduced in this paper provides a new method to extract city perception and its potential usage for urban issues. Generally speaking, it is trying to tackle the following questions: (1) How to extract the holistic cognition and collect the scattered perceptions from the public through modern techniques? (2) How to apply such extracted knowledge to urban problems about detecting shifts in city changes, planning strategies, and urban land use? (3) What are the implications from the findings, such as comparability to city image or judging planning suitability, after applying this new method?

Literature review

Two main approaches have laid the foundation of this research: the cognitive mapping approach and the computation approach. The first approach, as an essential expression of individual's internal entities (Marr and Vision 1982), is an ideal form for C-IMAGE. For the computation approach, it covers modern technology about the acquisition and analysis of big data, which provides the technical support for C-IMAGE.

Cognitive mapping approach

The term "cognitive mapping" does not originate from urban study. This label comes from the simple idea to find an alternative to the stimulus–response model of humans in rats and men experiments (Tolman 1948). The phrase of "cognitive maps" is first mentioned in describing the nervous system and later is mainly mentioned in research from environmental psychology.

Over half a century ago, Lynch (1960) illuminated how people perceive their environment and summarized five basic elements that are operable and referable in urban design practice. To capture perceptions from the public, he invented a series of methods, and the core of them is asking people to draw mental maps of the city (LeGates and Stout 2011, p 499). Compiling all the sketches, he used visual method to distinguish the degree of recognition for each part of the city. Figure 1 shows two typical Lynch's city images of peninsular Boston based on sketches and interviews. For example, if more than 75 % of the respondents draw a line corresponding to Massachusetts Avenue on the map of Boston, he will use a fairly thick line for the position of Massachusetts Avenue of the final map to indicate a 75 % recognition from people and thus reflect the significance of this street in their perception. Finally, by counting different percentages, his team created comprehensive mental maps of Boston, Jersey City, and Los Angeles. Based on the definition of cognitive mapping, Kevin Lynch successfully transplanted its concept to the research of city image.

According to these cognitive maps, he concluded the well-known five elements—paths, edges, districts, nodes, and landmarks—that enhance a city's identity. They have been widely used as principal rules for improving city legibility.

During the following decades, many related studies to some extent confirmed the stability of those five elements (Appleyard 1970; de Jonge 1962; Francescato and Mebane 1973; Harrison and Howard 1972). For example, Appleyard conducted research by requiring local residents to draw a map of the whole city between the steel mill and San Felix in Ciudad Guyana. Through analyzing those "inhabitant maps", he pointed out that such subjective maps provide rich information about urban perception correlated with the visible, functional, and social character of the city.

Other research slightly questioned the "five-elements" schema of urban perception (Gulick 1963; Klein 1967; Rapoport 1977). These disagreements came from the diversity of soci-cultural and physical context of different areas and populations. For

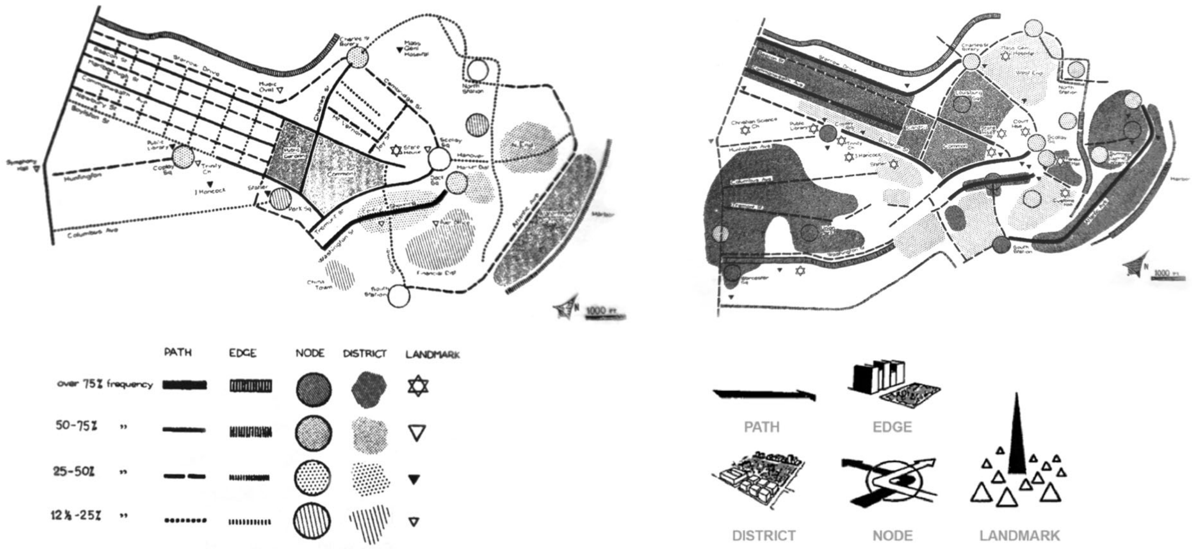


Fig. 1 The image of Boston from sketches (*top left*), the image of Boston from interviews (*top right*), and the *symbols* of the five elements (*bottom left*) (the *top figure* page 21, and the *bottom figure* page 19 Lynch 1960)

example, Gulick proposed that urban imageability was "a product of the perception of visual form and of the conception of social significance" based on the mental map from 35 Tripolitan students in Lebanon. He argued that to understand the visual experience of laymen required knowledge beyond treating the elements purely as visual concepts. It is because that people's defining criteria are social and behavioral (Gulick 1963).

Computation approach

Like the application of cognitive mapping, computation approach is not an independent research field derived from urban study. It is a cross-platform field of a variety of urban issues. Many applications in this newly born approach have inspired this research. For example, a team from Microsoft research center detects "flawed urban planning", which means a long-term traffic overload in specific areas,[1] using the GPS trajectories of taxicabs traveling in urban areas (Zheng et al. 2011). The project U-Air proposed a new methodology of comprehensive utilization of existing sensors, infrastructure for a detailed prediction of urban air quality (Zheng et al. 2013). An European media team reports a novel deployment of UBI-hotspots in a city center to establish an ecosystem infrastructure for conducting diverse urban computing research and business in authentic urban setting (Ojala et al. 2010). These attempts all have shed light on using sensor data for urban study research.

The project named "City Pulse" developed by a team from the Media Lab at MIT is using thousands of geo-tagged images to measure the perception of safety, class and uniqueness in the cities of Boston and New York in the United States, and Linz and Salzburg in Austria. The research finds that the range of perceptions elicited by the images of the US cities is larger than that of the other two European cities, which was interpreted by the team that cities of Boston and New York are more contrasting, or unequal, than those of Linz and Salzburg. Besides they also validate a significant correlation between the perceptions of safety and class and the number of homicides in a NYC zip code, after controlling other dependent variables (Salesses et al. 2013). The dataset of their research is built on 4136 geo-tagged photos cropped from Google Street View and the evaluation of those

[1] In their paper, the author partitions a city into some disjoint regions using major roads. And they project the taxi trajectories of each day into these regions and detect the salient region pairs having traffic beyond its capacity. When the overloaded region pairs are repeatedly detected across many days, they define such a situation as a "flawed planning". Here the word "planning" they used is somehow weird to the normal one in urban planning area.

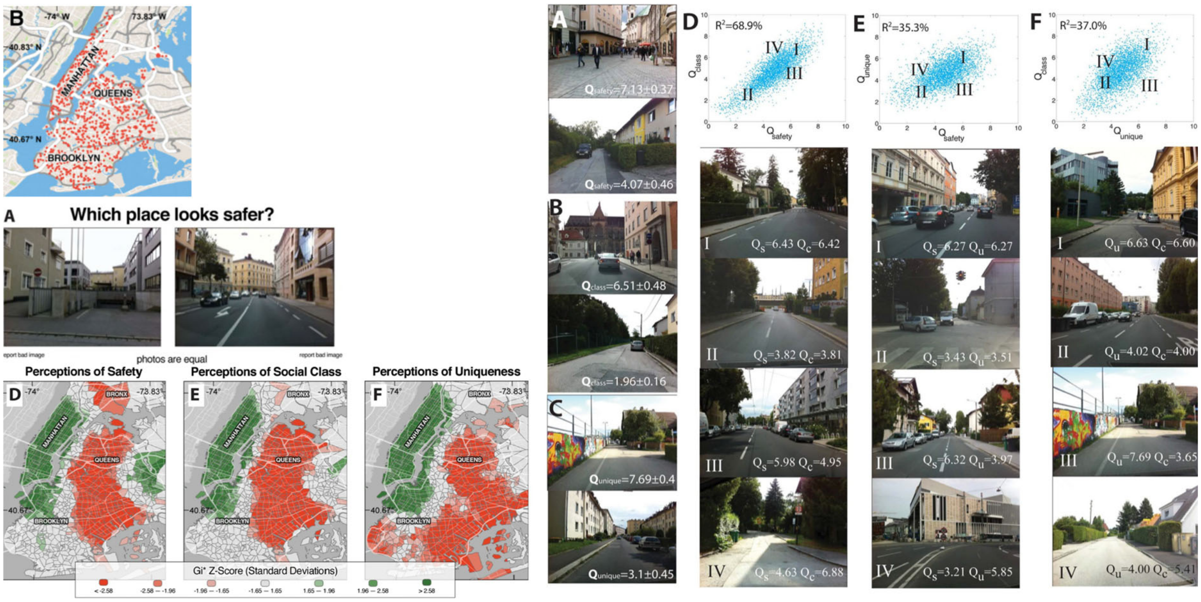


Fig. 2 The locations of street views collected from Google Street Views in New York City (*top left*), a typical questions for online voting (*middle left*), map of NYC showing statistically significant clusters of high and low Q-scores (the Q-score is calculated as a weighted score through an algorithm defined by the researchers based on the result of photo votes. They scale the score to a 0–10 range, which means the score of 10 represents highest degree of safety/class/uniqueness and the score of 0 stands for the lowest degree) for the perception of safety/class/uniqueness (*bottom left*), and identifying places associated with different urban perceptions (*right*) (Salesses et al. 2013)

images are achieved through an on-line survey from a website where they uploaded all the photos for the project. For each vote, there is a pair of photos put together along with a question such as "Which place looks safer?" on the screen (Fig. 2), and all the respondents need to do is to click the picture that they believe is of higher security. The statistical analysis and modeling are based on a total number of 208,738 votes.

Another team from CMU demonstrates that geographically representative image elements can be discovered automatically from Google Street View imagery in a discriminative manner and these elements are visually interpretable and perceptually geo-informative, which can support a variety of computational geography tasks (Doersch et al. 2012).

Both approaches activate the concept of this research theoretically and technically. The city image approach redraws the attention of urban designers back to the interaction between forms and cognition, while the computational approach arms researchers with the scope of big data so that the knowledge boundary is extended.

Method statement

Before starting this research, we need to define the term "city cognitive mapping": First, the geographic research scope of city cognitive map is refined to the physical boundaries of cities.[2] City is also the basic unit of how image datasets are all collected and stored. Second, conforming to the essence of cognitive map, the information prepared for the maps should stay subjective, and the process of cognitive map is a translation of subjective information to a readable format. Third, collecting the knowledge for city cognitive map is a "bottom-up" choice, which means the data source should come from the public so that the results are reflection of the perception from the public.

Research objects

Typically texts and photos are two forms of information carrying information that are subjective and pervasive. There are a great variety of platforms

[2] The city boundary data is downloaded from GADM, http://www.gadm.org.

providing geo-located texts and photos all over the world, but usually texts convey senders' opinions in a more direct way. However, texts cannot avoid the language barrier, especially when we hire computer to read them. In contrast, photos are more appropriate for this study, since visual contents are universally and objectively recognizable.

Two image sources are available for the study of city cognitive mapping. One is from Google Street View and the other is from social media platforms such as Flickr and Panoramio. The difference between them lies in photo takers. For streetviews, cameras on Google Street View cars automatically capture photos. Yet when it comes to photos, they are all taken and shared manually. Choosing Google Street View is somehow contradicting the underlying principle of "choose subjective information", because those images from Google Street View are photos taken by machines instead of human beings. Although this technology brings back a large quantity of elegant images even from unpopulated places, the photos from this source are not representing anyone's motivation and the prerequisition of taking a Google Street View photo is the existence of a street and the city must be involved Google's coverage area.

Nevertheless, it would be different to collect photos from social media, because the data from this source is subjective information, which exactly reflects individual preferences. Many such platforms as Flickr and Panoramio have already built up a huge database of photo collections covering most areas in the world. Almost every large city in this study has a substantially large number of geo-tagged photos.

Data Collection Here we take collecting data from the city of San Francisco as an example. First, we build a network with 500 $\times$ 500 m cells in ArcGIS, shown as in Fig. 3. Then we calculate each pair of the diagonal vertices' latitude and longitude, with which to compose the request ready for using Panoramio API.

Once a request has been sent out, a JSON[3] file is returned and parsed by our script in order to abstract the metadata of all the photos within that bounding box. Tables 1 and 2 completely display a typical record of Panoramio. Apart from the coordinates, some other downloaded features are or can be highlighted in the following application, such as the upload date and the owner id.

The downloaded metadata is kept in the C-IMAGE database, which is powered by postgres. For both current and future research purpose, in the C-IMAGE project we have collected photos from 26 cities with a total number of around 2.3 million, which means there are nearly 90,000 photos for each city on average. To diversify the dataset, this paper selects 26 cities from Europe, Asia, and North America, which are listed in Table 3.

Applications

There are two main applications of this research: (1) comparing the metadata with Kevin Lynch's mental maps, and (2) disclosing city image through computationally recognizing images' content.

C-IMAGE based on metadata

In the book "The Image of the City", Kevin Lynch provided two city images of downtown Boston[4]: one from respondents' sketches and the other from interviews. Here comes the question: Is it possible to use the online geo-tagged photos to achieve the cognitive maps similar to those old ones? We choose Boston, which is a moderately large city, from his book as the case for the comparison. Noticing that Panoramio only provides 16,041 photos for that area, we extend our photo collection to Flickr, from which around seven times photos have been downloaded. Table 4 lists the metadata of these two datasets.

As shown in Fig. 4, the patterns of photos from Panoramio and Flickr are structurally similar but different in density. Following this C-IMAGE created by modern technology, Lynch's "city image", also combined with the underlying map of downtown Boston in Fig. 5 diagrams the five elements (paths, edges, nodes, landmarks, and districts), adding the frequency of recognition by degree of darkness.

[3] JSON, which is an open standard format that uses human-readable text to transit data objects consisting of attribute-value pairs. (Wikipedia, http://en.wikipedia.org/wiki/JSON).

[4] For those two maps, lease refer to Fig. 1.

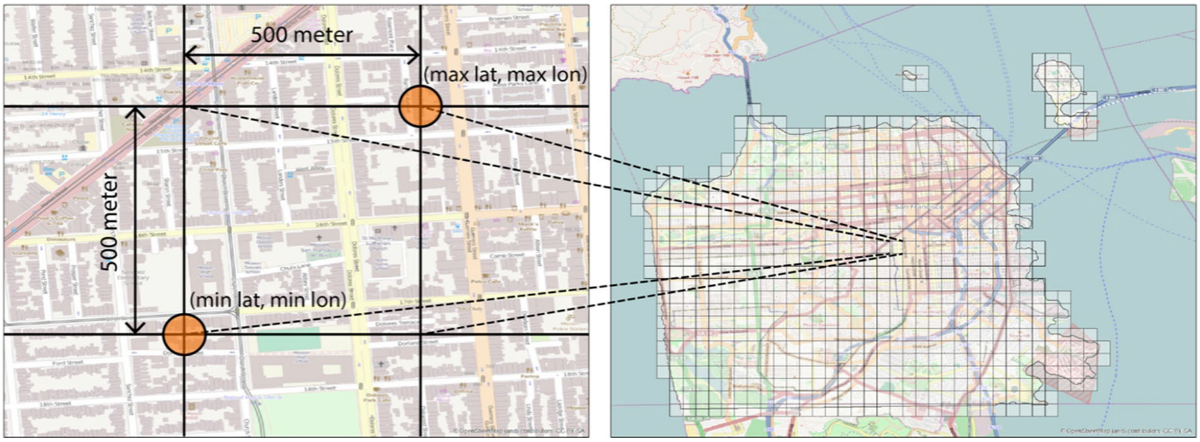


Fig. 3 The geographic coordinates as parameters used for defining the cells

Table 1 A sample record of a geo-tagged photo from Panoramio (the height and width refer to the original size)

Fields	Upload_date	Owner_name	Owner_id	Photo_id	Longitude	Latitude	Height	Width
Sample record	2013-12-01	KWO Tsoumenis	2549084	100000822	2.289131	48.862171	2484	3312

Table 2 A sample record of a geo-tagged photo from Panoramio (here the url is the medium size of the photo)

Fields	Photo_title	Owner_url	Photo_url	Photo_file_url
Sample record	FRA Paris eiffel tour from trocadero {in the blue hour} by KWOT	http://www.panoramio.com/user/2549084	http://www.panoramio.com/photo/100000822	http://static.panoramio.com/photos/original/100000822.jpg

C-IMAGE and the traditional five elements

(1) Clues about the path As defined by Lynch, the path is "the channel along which the observer customarily, occasionally, or potentially moves" (Lynch 1960, p 41). A path usually refers to linear public space, and in most cases is a partial or entire section of a street or walkway. In fact, the path is the most common element in Lynch's maps. According to the recognition frequency,[5] the paths of the highest identity in downtown Boston are Commonwealth Avenue, Massachusetts Avenue, Washington Street, Beacon Street, and Storrow Drive. Similarly, the mixed plotting of photos in C-IMAGE also clusters around these highly recognized roads in Fig. 6.

(2) Clues about the edge As the linear existence to break spatial continuity and confine boundaries, the edge is not often seen in the urban context. Typical example of an edge could be a river, a railroad, or even a wall, because an edge can be "seams, lines along which two regions are related and joined together" (Lynch 1960, p 41). In this case, only the waterfront of Boston has been marked as an edge. Since the C-IMAGE extends the research boundary to part of the ocean and the river, it is also clear to see there is a set of dots scattered on the water. These dots somehow form a wide belt. Interestingly, a great number of dots are located on the Charles River. These on-river photos may come from photographers taking pictures on a watercraft. Alternatively, another explanation is the owners of the photos chose wrong locations while manually uploading (Fig. 7).

[5] Please refer to the legend of Fig. 5.

Table 3 Data about the downloaded cities

City	Photo analyzed	Area (km^2)	Population	Uploaders	Date of download	Country and region
Amsterdam	67,853	219	0.780 (2011)	8452	2013/11/10	Netherland, Europe
Bangkok	100,808	1569	6.355 (2000)	8983	2014/02/04	Thailand, Asia
Barcelona	114,867	101.9	1.621 (2012)	12,281	2013/11/10	Spain, Europe
Beijing	64,631	1231.3	18.251 (2013)	9096	2013/10/04	China, Asia
Berlin	148,119	891.8	3.502 (2012)	15,047	2013/11/01	Germany, Asia
Boston	26,288	232.1	0.636 (2012)	4031	2013/10/03	US, North America
Brussels	45,354	161.4	1.119 (2011)	5642	2013/11/10	Belgium, Europe
Hong Kong	152,147	1104	7.155 (2012)	11,702	2013/10/05	China, Asia
Kuala Lumpur	45,927	243	1.589 (2010)	5300	2013/11/20	Malaysia, Asia
London	209,264	1572	8.174 (2011)	23,402	2013/10/31	UK, Europe
Madrid	125,055	605.8	3.234 (2012)	10,586	2013/11/18	Spain, Europe
Milan	44,545	181.8	1.316 (2010)	5320	2014/02/15	Italy, Europe
Moscow	291,371	2510	11.5 (2010)	18,347	2014/02/15	Russia, Europe
New Delhi	38,664	42.7	0.302 (2011)	6341	2013/11/18	India, Asia
New York	159,393	1213	8.337 (2012)	17,129	2013/11/12	US, North America
Paris	154,437	105.4	2.211 (2008)	16,274	2014/02/15	France, Europe
Prague	74,984	496	1.257 (2011)	8825	2013/11/01	Czech Republic, Europe
Rome	97,578	1285	2.753 (2010)	10,726	2013/11/10	Italy, Europe
San Francisco	64,592	600.6	0.826 (2012)	10,546	2013/11/12	US, North America
Seoul	89,007	605.2	9.82 (2005)	5758	2014/02/15	Korea, Asia
Shanghai	35,722	1499	15.247 (2011)	6063	2013/10/02	China, Asia
Singapore	66,364	710	5.312 (2012)	6585	2013/11/10	Singapore, Asia
Tokyo	242,468	2188	13.23 (2013)	11,920	2013/10/30	Japan, Asia
Toronto	58,125	630	2.503 (2006)	6141	2013/11/20	Canada, North America
Vienna	89,380	414.6	1.731 (2012)	7713	2014/02/15	Austria, Europe
Zurich	31,940	87,88	366.765 (2009)	3054	2013/02/04	Swiss, Europe

Area data GADM, *population source* UNdata

Table 4 Information about the downloaded data from Panoramio and Flickr

Data source	Photo analyzed	Date of downloaded	Uploaders	Maximum uploading/taking time	Minimum uploading/taking time
Panoramio	16,041	2013/10/03	2754	2013/10/03	2005/10/13
Flickr	113,693	2014/04/08	7438	2014/04/07	Unclear[a]

For Panoramio, the last two columns stand for the latest and earliest uploading time of all the photos, and for Flickr, the last two columns stand for the latest and earliest taking time of all the photos.

[a] Because of the column of "datetaken" is quite bias due to incorrect time setting on cameras, the minimum time of all the dataset is unclear

(3) Clues about the node According to Lynch's definition, the node is "a spot entitled with unique means, function, or any other uniqueness that enhance its sense of existence" (Lynch 1960, p 41). Typical nodes in Lynch's map are often connected with paths or landmarks. While providing connections for paths and landmarks, in return, the node also benefits from the various flows attracted by the other elements. As

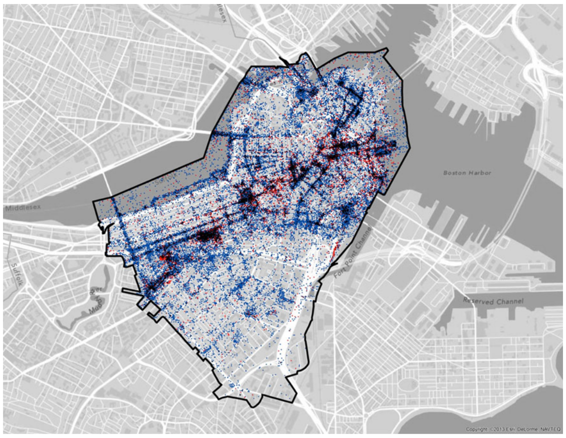


Fig. 4 The combination of the photo distribution of Panoramio (16,041 *red dots*) and Flickr (113,693 *blue dots*) data (*left*). (Color figure online)

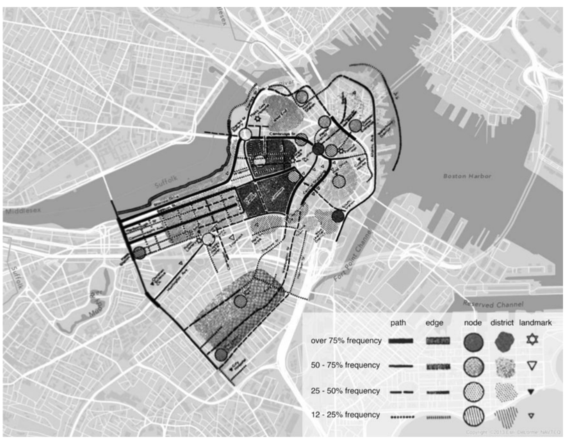


Fig. 5 Lynch's "city image" map based on interview and combined with the map of downtown Boston [this map is actually modified in Adobe Photoshop to combine with current map of Boston, the upper layer of the map is actually the same as the one shown in Fig. 1 (the *middle one*)] (*right*) (for a clearer look of the two maps, please refer to the "Appendix 1")

Fig. 8 shows, Boston's Copley Square is a five-pointed crossing hit by the Huntington Ave at a strange angle. However, "it enjoys its fame as a highly popularized place surrounded by a series of unique buildings such as the Public Library, Trinity Church, the Copley Plaza Hotel, the sight of the John Hancock Building" (Lynch 1960, p 58). All kinds of activities embedded in those eminently contrasting buildings enhance diversity and reduce unclarity caused by the non-perpendicularly angled intersection. In the map of C-IMAGE, the square is so popular that it is almost fully occupied by black dots.

(4) Clues about the landmark Different from nodes, the landmark usually refers to a construction that holds interior space. It is well defined as "external presence with physical shapes such as buildings, sculptures, or mountains" (Lynch 1960, p 48). For downtown Boston, typical landmarks are highlighted, such as the State House, the Faneuil Hall, and the Public Library. All of them look different from their adjacent buildings in terms of volume, outlook, and place: The gold dome of the State House, the tower upon the Faneuil Hall, and the monumental Catalan Vaults from the McKim building (Public Library) have all been symbolized in city dwellers' and visitors' memory. Based on the information provided by C-IMAGE, the locations of landmarks marked by "city image" all have a cluster of dots. However, these clusters cannot prove the existence of the buildings. In other words, the clusters on the C-IMAGE maps can to some extent provide clues for landmarks, but the validation of that needs a close look of the contents of those assembling images.

(5) Clues about the district As Kevin Lynch mentions, the district is to some extent "the special ones" among all the five, because of its potentially large scale. Or to say, it can hide itself with other elements. One perfect match between Figs. 4, 5 is the Public Garden who has over 75 % frequency of recognition from interviews and an obvious rectangle with dense dots, as shown in Fig. 9. Unfortunately, such a close match is not common throughout the comparison of districts. As a district also scored over 75 % recognition, it is difficult to identify Beacon Hill from C-IMAGE. A district may include other elements, and structures such as the Charles Street, the Louisburg Square, and the State House appear clearly within Beacon Hill area in Fig. 10. However, such occurrence of single element will increase the difficulty in visually recognizing the district. Although this C-IMAGE contains the information from 140,000 geo-tagged photos from Panoramio and Flickr, it is still a point-based map and the technical difficulty of transforming a dot-based map to a polygon-based one prevents from discovering the district in such maps.

(6) Four prototypes Different from the five elements defined by the "city image" maps, there are primarily four different prototypes in summarizing the scattered points in this C-IMAGE: lines, clusters, dense areas, and sparse areas as displayed in Fig. 11.

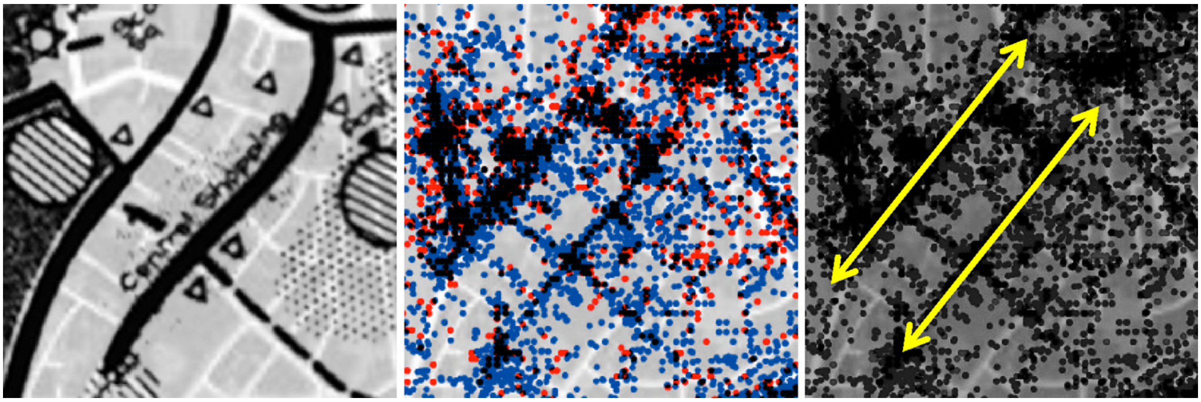


Fig. 6 Tremont Street and Washington Street as paths from Lynch's map (*left*) and C-IMAGE of that area (*middle*), with two linear clustered dots (*right*)

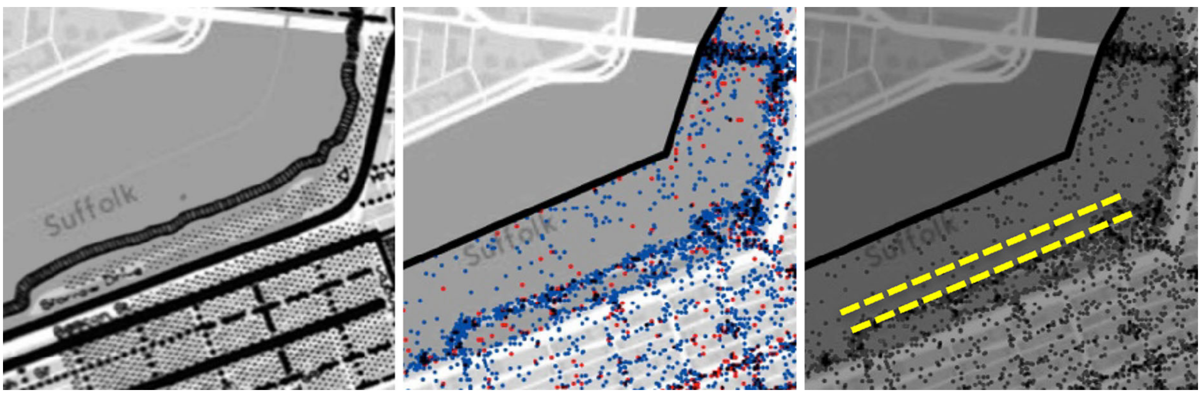


Fig. 7 The water front of Boston along Charles River as an Edge from Lynch's Map (*left*) and C-IMAGE of the area (*middle*), with a wall-like space (*right*)

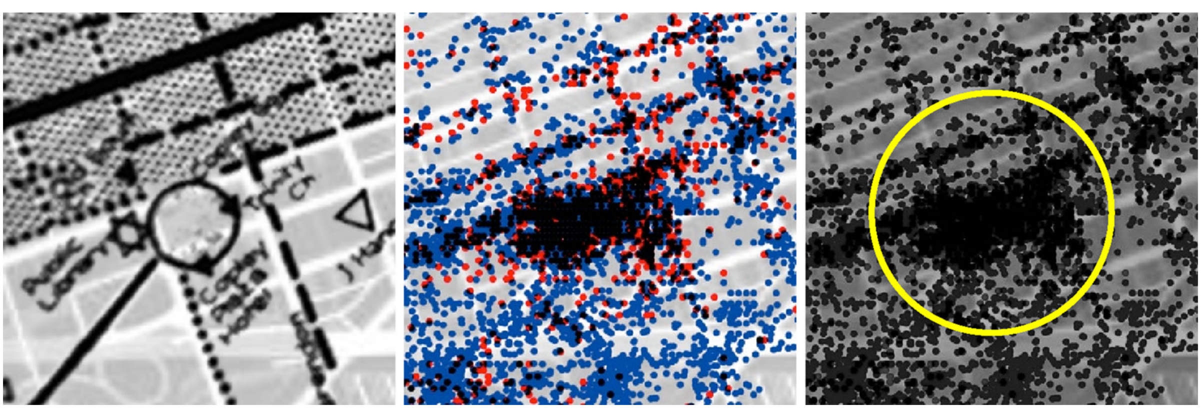


Fig. 8 Copley Square as a node from Lynch's map (*left*) and C-IMAGE of that area (*middle*), together with a clustered area (*right*)

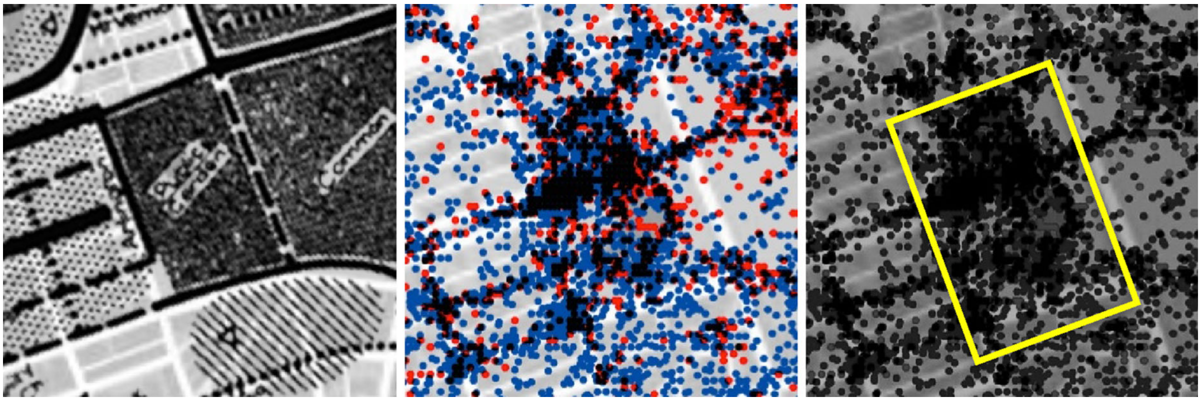


Fig. 9 The Public Garden as a district from Lynch's map (*left*) and C-IMAGE (*middle*), together with a clustered area (*right*)

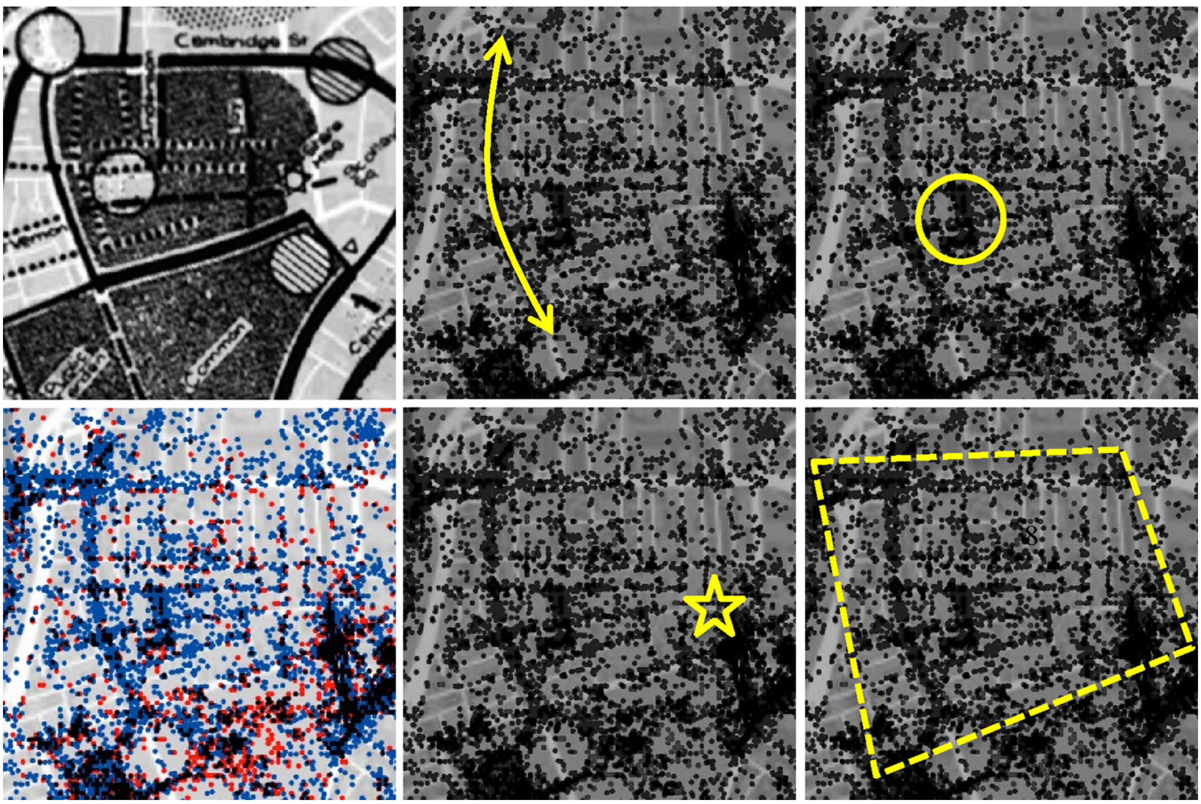


Fig. 10 Beacon Hill as a district from Lynch's map (*top left*), and C-IMAGE of that area (*bottom left*), together with the linear cluster indicating the Charles Street as a path (*top middle*), the Louisburg Square as a node (*top right*), the State House as a Landmark (*bottom middle*), and a map showing the ambiguous boundary as a district (*bottom right*)

Comparison with the city problem map

As discussed above, the geo-tagged photos can somehow reflect the intensiveness of social activities like what Lynch's cognitive map did. Confirming his theory through C-IMAGE takes much less time and human resources: any single researcher can create such a map in simply several hours based on the process in section. Although C-IMAGE is not equal to Lynch's map, these findings can still answer the second research question via providing an application of the knowledge from today's large-scale data to

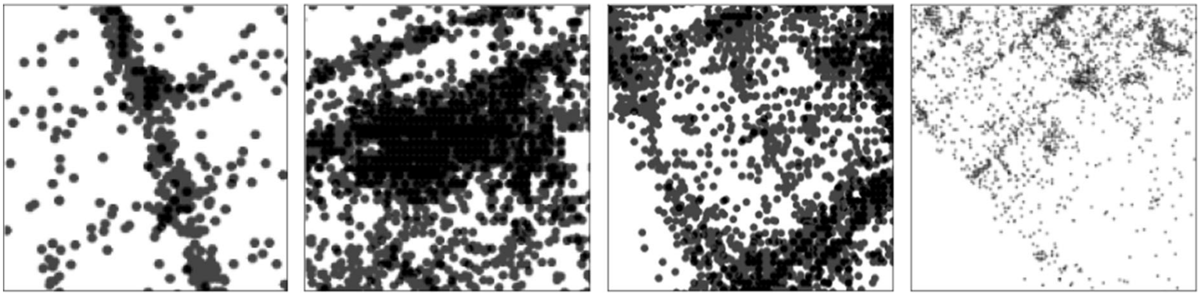


Fig. 11 The four different patterns extracted from C-IMAGE: *lines*, *clusters*, *dense areas*, and *sparse areas* (from *left* to *right* respectively)

confirm the city image theory by computational technology.

Apart from the well known "city image" of Boston peninsular, Lynch has also summarized, in his book, a map that describes the problems in Boston from the public perspective in Fig. 12. The findings from above section have confirmed some of the visual elements proposed by Lynch, and they have established a relationship between C-IMAGE and the conventional city image. Given that Lynch's map was conducted in 1960s while C-IMAGE was based on photos that are mostly taken in recent 10 years, matching these two maps provides a way to computationally understand shift in city form since 1960. According to previous section, the problem areas in Lynch's map are not easily identifiable, so lines, clusters, dense areas should not be found around those areas from C-IMAGE. Otherwise, if one of these three patterns shows up in a problematic place, we call such case a mismatch, which is usually connected with a shift in urban form. With the problem map as a checklist, it becomes possible to examine whether these negative spaces have been improved today. Table 5 displays all the mismatches and underlying urban changes.

C-IMAGE based on image content

To further explore this new method based on geo-tagged photo, we attempt to include the image content information into our analysis through the technology

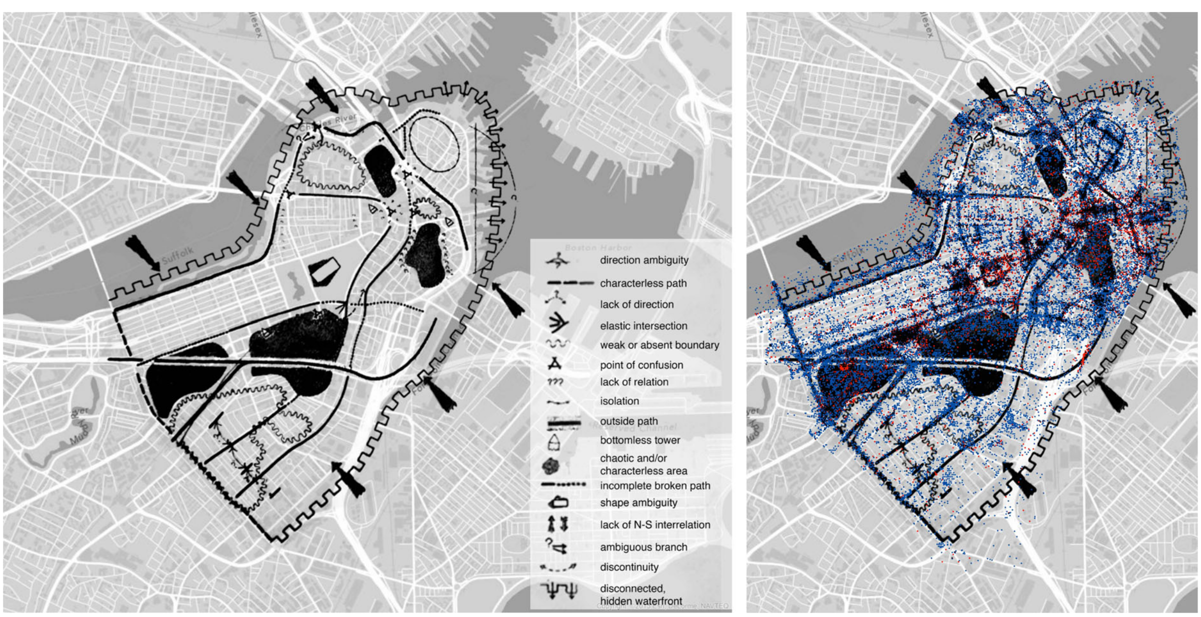


Fig. 12 Map of problems in the image of Boston (the problem map of Boston: page 24 Lynch 1960) (*left*), and map showing the patches from the Panoramio and Flickr photos' distribution overlaid with the problem map of Boston image (*right*)

Table 5 The comparison between C-IMAGE and Kevin Lynch's Boston problem map

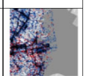
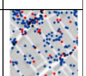
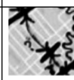
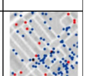
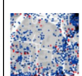
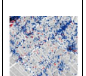
Type	Problem number according to Lynch's map	Prototypes in C-IMAGE	Related projects if possible[a]	Magnified segment in problem map	Magnified segment in C-IMAGE
Direction ambiguity	1 waterfront area (partially improved)	Dense Area (north) Lines (south)	the big dig, Christopher Columbus Waterfront Park (1976)		
Lack of direction	2 Columbus Ave and Tremont St (no change)	No lines			
	3 Tremont St and Washington St (no change)	No lines			
Weak or absent boundary	4 West End (no change)	Sparse Area	The tear down of West End and the construction of Charles River Park (1958–1965)		
	5 South End (no change)	Sparse Area	Neighborhood housing programs are begun including the South End neighborhood housing initiative (1985–1988)		

Table 5 continued

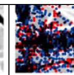
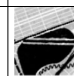
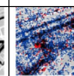
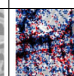
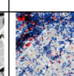
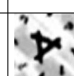
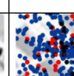
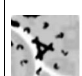
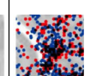
	6 Faneuil Hall (improved)	a Cluster	Renovation Faneuil Hall Marketplace[b] (1976)		
Outside path	7 Boston and Albany railway (significantly improved)	Clusters and dense Areas (all color)	Prudential center (1960–1964), underground of Tpke E		
	8 John F.Fitzgerald Expressway (significantly improved)	Four perpendicular lines and dense area	the Big Dig (1982–2007), the Rose Kennedy Greenway in the Big Dig (2008)		
	9 Huntington Ave (improved)	A line (blue)	Southwest Corridor Park[c] (1970s)		
Point of confusion	10 Haymarket station (improved)	A cluster (blue)	Government center garage (1950s?)		
	11 Government Center (improved)	A cluster (dark)	Government Center construction from Scollay Square (1958–1965), including JFK federal building (1966), City Hall (1968)		

Table 5 continued

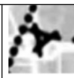
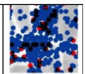
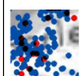
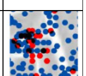
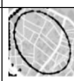
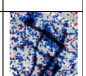
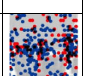
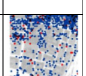
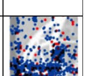
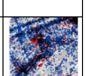
	12 GMH station (partially improved)	A cluster and two Lines (blue)	Renovation of MGH station (2007), pedestrian system around Charles Circle		
	13 science park station (partially improved)	A cluster (blue)	Construction of science park station (1955)		
	14 park square (no change)	A cluster (dark)			
Isolation	15 North End (significantly improved)	Two lines towards inland	the big dig (1982–2007[d])		
Incomplete broken path	16 Boylston Street (partially improved)	A line (blue)			
	17 Washington Street (no change)	Sparse Area in the south			
	18 Storrow Drive (no change)	No lines			
Chaotic or Characteraless area	19 prudential (old rail station) (significantly improved)	Clusters and Dense Areas	Prudential center (1960–1964) Tent City (1984)		

Table 5 continued

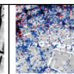
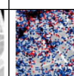
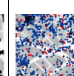
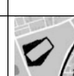
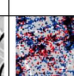
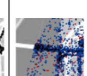
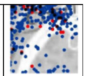
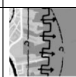
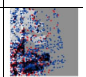
	20 park square area (partially improved)	+Clusters and Dense Areas in the north, Sparse Area in the south (mostly blue)	New John Hancock Building (1973), development of Copley Place (1984)		
	21 east to downtown crossing (improved)	+Dense Area (blue and orange)	downtown office building boom (1970s–1980s) Post Office Square (1992, 1997 name changed to Leventhal Park)		
	22 east to West End (no change)	–Dense Area (blue)	The tear down of West End and the construction of Charles River Park (1958–1965)		
Shape ambiguity	23 Boston Common (no change)	Dense Area inside (orange), lines in the east			
Lack of N–S interrelation	24 north and south boundary (improved)	Lines in the north (dark), no Lines in the south	Callahan Tunnel opens (1961)		

Table 5 continued

Ambiguous branch	25 end of Charles Street (no change)	Sparse Area			
Disconnected	26 hidden waterfront (partially improved)	Clusters in the south, but no Line	Downtown waterfront urban renewal plan (1964) Rowes Wharf (1987) Christopher Columbus Waterfront Park (2003)		

The location map of these problems can be found in "Appendix 1"

[a] Some of the changes, if not specified, come from the Boston Chronology in the book "Mapping Boston" (Krieger et al. 2001, pp 245–246)

[b] For more information about the marketplace: http://en.wikipedia.org/wiki/Faneuil_Hall

[c] For more information about the park: http://en.wikipedia.org/wiki/Southwest_Corridor_Park

[d] The entire timeline for the central artery/tunnel project (we call the big dig) can be found from the official website of MassDOT https://www.massdot.state.ma.us/highway/TheBigDig.aspx

named "scenes understanding", which is specifically explained in another paper of ours (Zhou et al. 2014a). Through such a tool, the content of each photo can be related to 1 of the 102 attributes, and following paragraph provides a summary of this computational technology.

To train the scene attribute classifier, the project resorts to SUN attribute database (Patterson and Hays 2012), which consists of 102 scene attributes labeled on 14,340 images from 717 categories from the SUN database (Xiao et al. 2010). These scene attributes, such as "natural", "eating", and "open area", are well tailored to represent the content of visual scenes. Then we use a deep convolutional network pre-trained on Places (Zhou et al. 2014b) to extract features from images in the SUN attribute database, since deep learning features area shown to outperform other features in many large-scale visual recognition tasks. A 4096 dimensional vector then represents every image from the output of the pre-trained network's final hidden layer. These deep learning features are then used to train an SVM classifier for each of the 102 scene attributes using libSVM with the default settings (Chang and Lin 2011). In Fig. 13, the comparison of our approach to the methods of using single feature GIST, HoG, Self-Similarity, Geometric Color Histogram, and to the combined normalized kernel method (Zhou et al. 2014a; Patterson and Hays 2012) indicates that our approach outperforms the current state-of-the-art attribute classifier with better accuracy and scalability.

After applying the algorithm discussed above, eventually for each photo a vector of 102 numbers is granted, and each single number indicates the probability of whether the photo belongs to that scene category. In each photo, the largest one among the 102 numbers indicates the most likely scene the photo should be recognized.

102 attributes to 7 urban perceptions

Since the 102 attributes are borrowed from vision computing field, this categorization needs refinement because of three problems: (1) some of the attributes such as "trees" and "vegetation" are visually similar to each other, and ten attributes from the 102 scenes can be merged to "green space"; (2) other attributes such as "carpet" are irrelevant to urban issues; (3) attributes such as "man-made" are too broad to use and they cover almost 80 % of the images among all the photo libraries. We cut down the 102 attributes to 7 urban perceptions by either merging similar attributes or deleting irrelevant ones.[6] Due to the diversity of the photo library, these perceptions may come from different dimensions as follows:

(1) Green perception It is one of the most frequent perceptions showing up in the entire image database. Although the 102 attributes can somehow distinguish more detailed information by categorizing green space photos into shrubbery, grass, or trees. For the purpose of plotting the general image for the city, merging

[6] The complete list of the 102 attributes and their relation to the 7 urban perceptions is provided in the "Appendix 2".

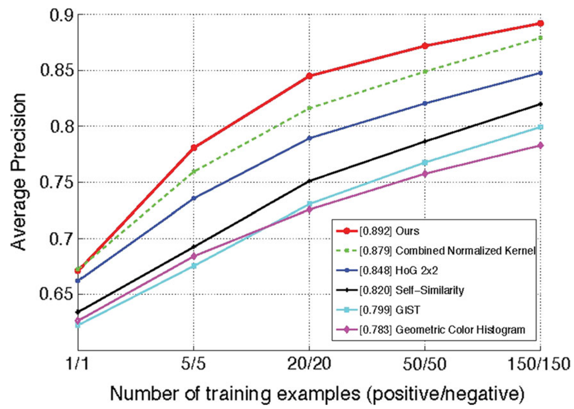


Fig. 13 Average precision (AP) averaged over all the scene attributes for different features. Our deep learning feature classifier is more accurate than the other individual feature classifiers and the combined kernel classifier (Zhou et al. 2014a)

them together is a better way of comprehending the core information. Since creating a library of landscape photos is the primary goal of Panoramio, the green perception occupies a dominant proportion in most cities. Figure 14 take the city of London as an example, showing both the distribution of green perception and sample photos of this category.

(2) Water perception As they are shown in Fig. 15, water perception refers to the photos that contain a large proportion of water in their contents. It is relatively reasonable that inland cities have fewer photos in this category due to a lack of ocean view.

(3) Transportation perception Transportation perception, as shown in Fig. 16, means those photos contain any vehicles. Although the transportation is one of the key functions for almost every city, it is less likely to be captured by photo takers because usually a road full of vehicles is a less attractive landscape.

(4) High-rises perception Theoretically speaking, this category was named as a perception of "vertical buildings" because the technology in this paper is not capable of calibrating the real height of the building only from a photo. However, most part of the photos is related with high-rises buildings with over 10 floors. Besides, even if the buildings in the photos just have four or five floors, the impressions conveyed by the photos are high enough. So since this research is trying to detect the city image on a mental level rather than talking about the definition of different building types, naming this category as the high-rises perception is appropriate in referring to psychological feelings (Fig. 17).

(5) Architecture perception Merging from attributes such as "praying", "shingles", and "marble", usually this category means some historical or traditional buildings in the city. The word "Architecture" may not exclude other types of buildings particularly those high-rises objects which are not categorized into this type, but it is fairly close to describing those within its category as shown in Fig. 18.

(6) Socializing perception Socializing perception means photos that are related with a wide range of activities varying from a small group to larger ones such as a large parade, a long queue, or a crowd of tourists. The typical photos of this category and their geographic distribution are given in Fig. 19.

(7) Athletic perception This category is a subset of socializing perception, but since it takes a relatively large proportion of that perception, so during the preliminary stage, it is separated from socializing it. Figure 20 offers the distribution of this perception in London, and the samples photos of this category.

Typology from the seven-perception C-IMAGE

Figure 21 displays a comprehensive map of London using 7 colors to plot all the perceptions. The map is processed in R via the package of "ggmap"[7] and CouldMade map.[8] Iterating the plotting script for the 21 cities, we get all the C-IMAGEs based on the 7 perceptions.[9] Rendered in 7 colors, those dots are granted with an alpha value to be transparent to keep each other visible, while staying different sizes as a reflection to their diverse probability score.

[7] Package of "ggmap": http://cran.r-project.org/web/packages/ggmap/ggmap.pdf (page 6 describe how to use the "get_cloudmap" to create the C-IMAGE in this section).

[8] Unfortunately, CloudMade has shut down its service for Map Tile, Geocoding, Routing and Vector Stream since May 1, 2014.

[9] All the 21 cities' urban perception C-IMAGEs are included in the "Appendix 2".

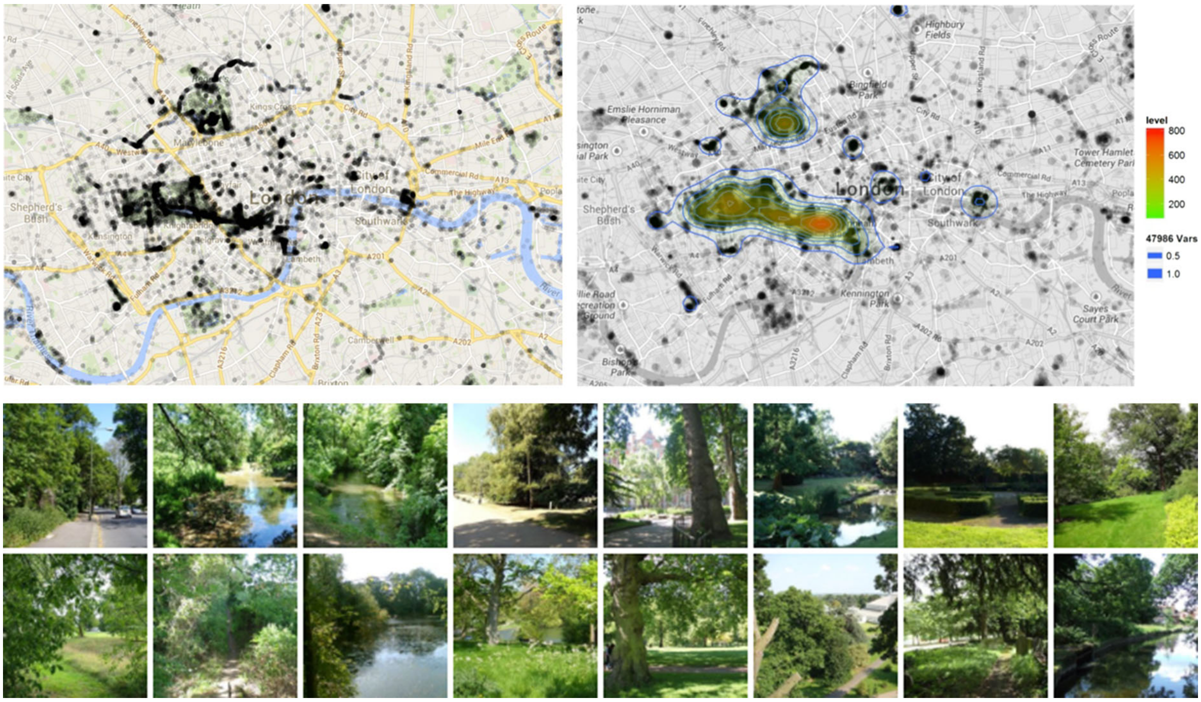


Fig. 14 The distribution of the "green perception" photos on map of London (*left*), the distribution with kernel density map (*right*), and the sample photos of "green perception" in London (*bottom*). (Color figure online)

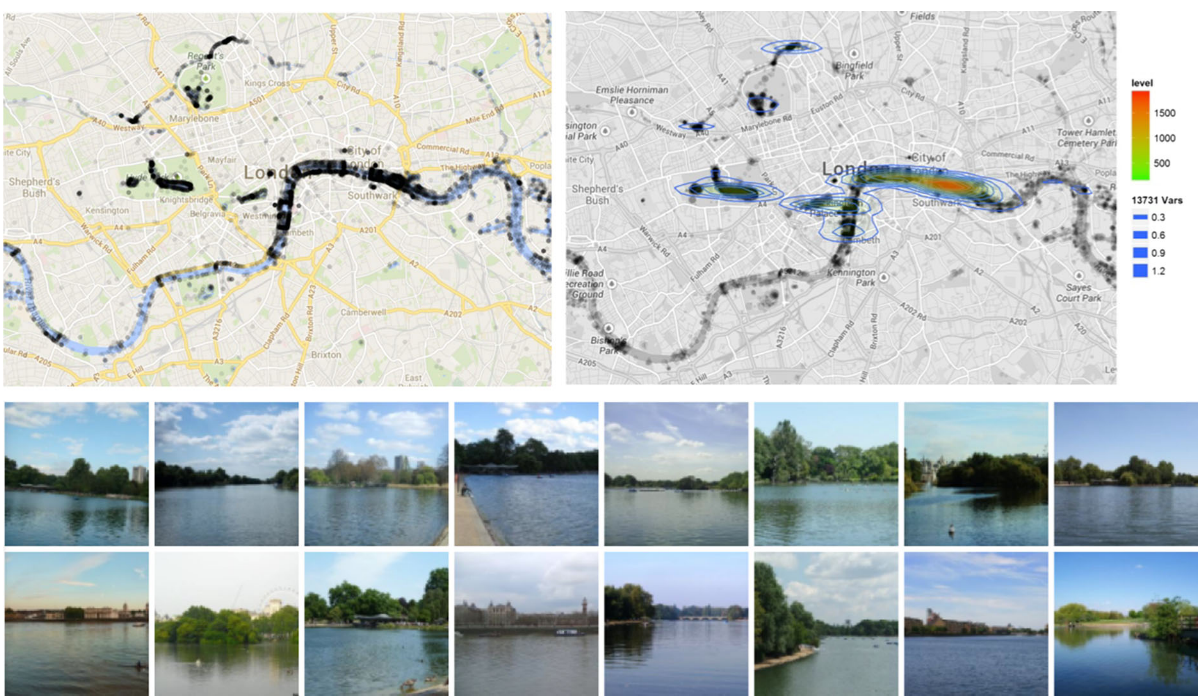


Fig. 15 The distribution of the "water perception" photos on map of London (*left*), the distribution with kernel density map (*right*), and the sample photos of "water perception" in London (*bottom*). (Color figure online)

Fig. 16 The distribution of the "transportation perception" photos on map of London (*left*), the distribution with kernel density map (*right*), and the sample photos of "transportation perception" in London (*bottom*). (Color figure online)

Fig. 17 The distribution of the "high-rises perception" photos on map of London (*left*), the distribution with kernel density map (*right*), and the sample photos of "high-rises perception" in London. (Color figure online)

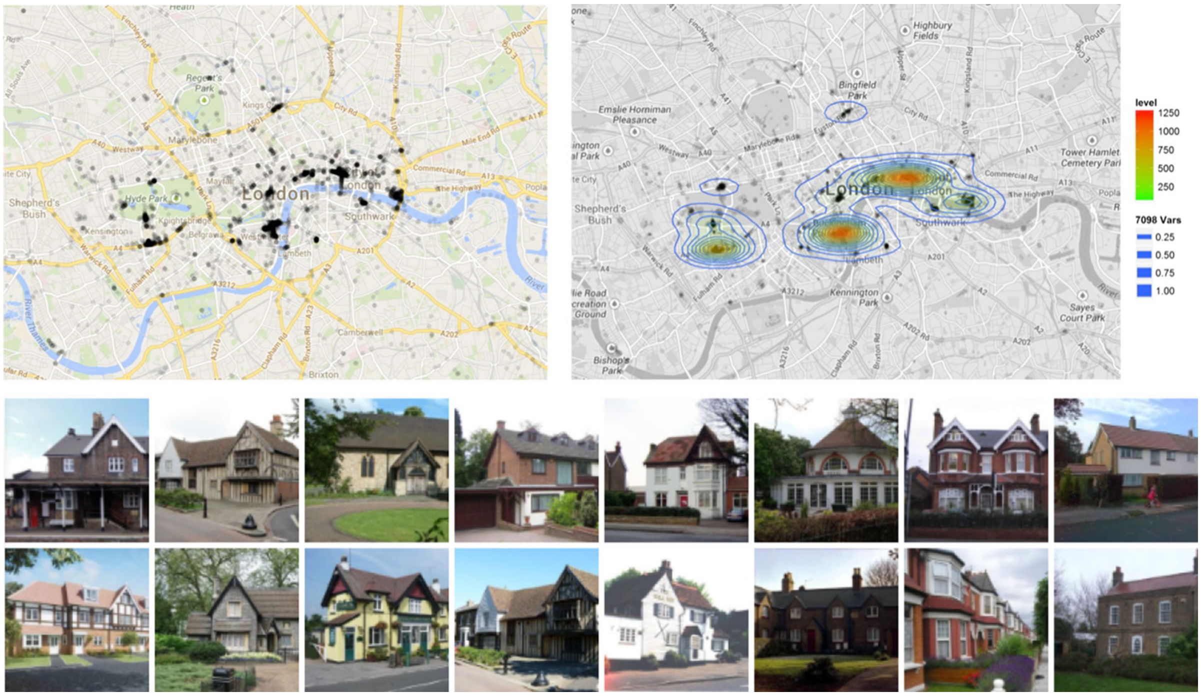


Fig. 18 The distribution of the "architecture perception" photos on map of London (*left*), the distribution with kernel density map (*right*), and the sample photos of "architecture perception" in London (*bottom*). (Color figure online)

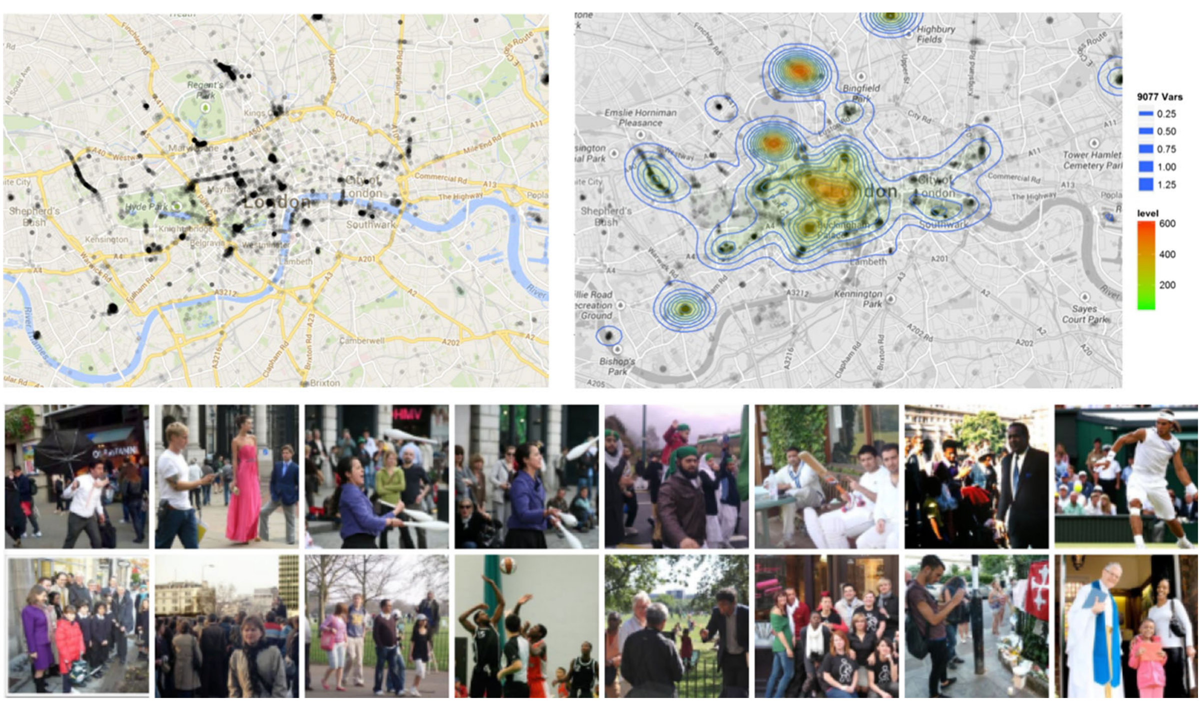


Fig. 19 The distribution of the "socializing perception" photos on map of London (*left*), and the distribution with kernel density map (*right*), and the sample photos of "socializing perception" in London (*bottom*). (Color figure online)

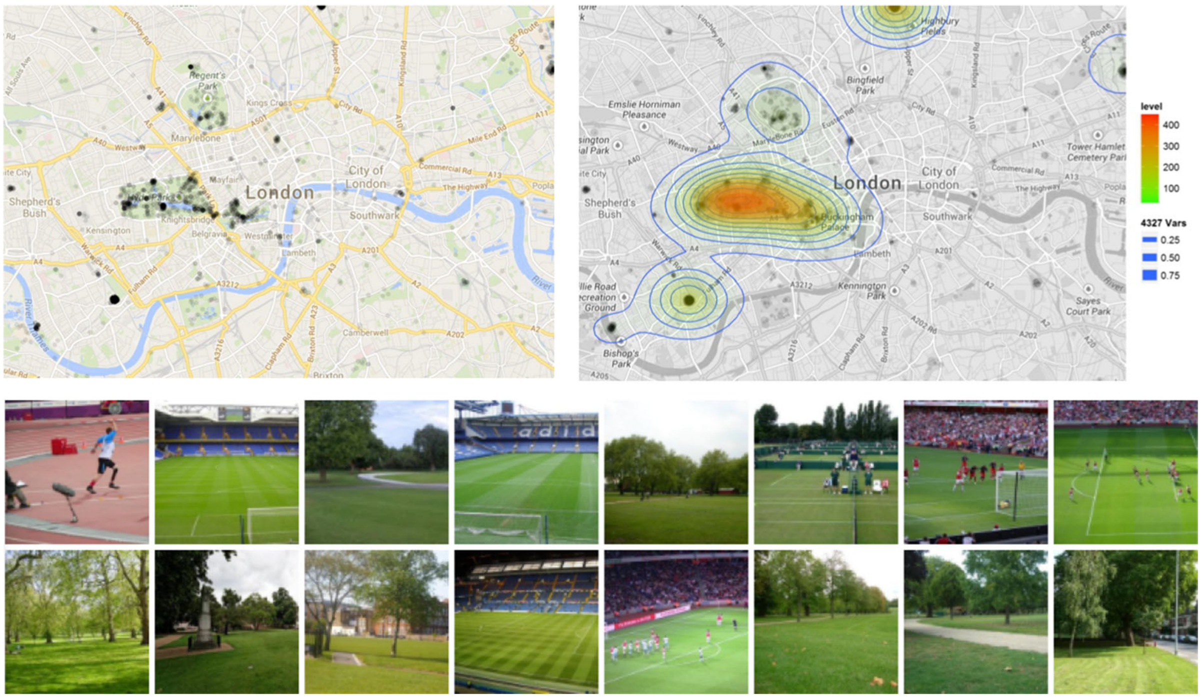
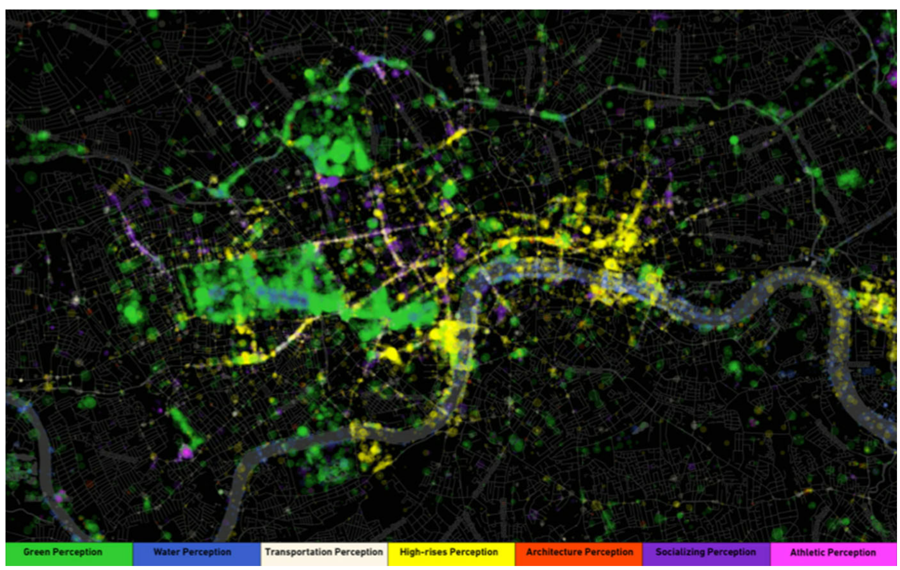


Fig. 20 The distribution of the "athletic perception" photos on map of London (*left*), the distribution with kernel density map (*right*), and the sample photos of "athletic perception" in London (*bottom*). (Color figure online)

Fig. 21 A typical C-IMAGE of London, UK based on the seven perceptions. (Color figure online)



After synthesizing all the 21 cities' seven-color maps, there are basically four types of structures on perceptional level:

(1) Green perception dominated city (Singapore) From the perspective of overall look of this type, the C-IMAGEs of these cities are full of green dots, as the two maps in Fig. 22. As the green dots means the locations of photos with contents associated to green space, it means someone must have seen a green vision, taken a picture of it and uploaded to share

Fig. 22 The green space dominated cities: Singapore (*left*) with its image (*top right*), and Prague (*middle*) with its image (*bottom right*) (for the description of legend, please refer to Fig. 24. For the description of legend, please refer to Fig. 24. The *top right figure* http://ggw.fransoringardend.netdna-cdn.com/wp-content/uploads/2011/03/sg1.jpg, and the *bottom right figure* http://1.bp.blogspot.com/-6oVXRUbkykw/Tke2NzXi56I/AAAAAAAAAFI/wZTc-WqKsdI/s1600/PragueCastleFromAbove.jpg). (Color figure online)

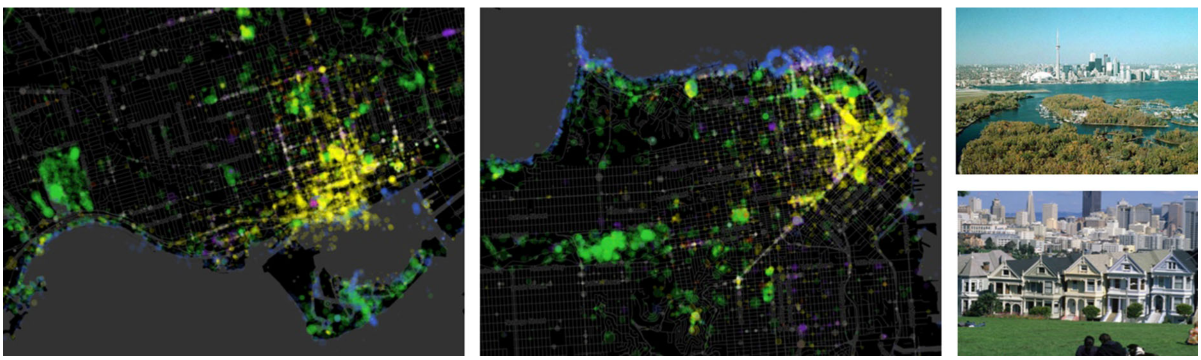


Fig. 23 The green and high-rises separated cities: Toronto (*left*) with its image (*top right*), and San Francisco (*middle*) with its image (*bottom right*) (the *top right figure* http://inanutshell.ca/wp-content/uploads/2012/08/toronto-islands.jpg, and the *bottom right figure* http://www.westinsf.com/var/cdev_base/storage/images/media/images/masthead-images/homepage-1220x667/downtown_san_francisco_panoramic_view/1031-3-eng-US/downtown_san_francisco_panoramic_view_home_masthead.jpg). (Color figure online)

publicly. Therefore, a clustered group of green dots on the map does not have a direct connection with the green space in the real world. It means what planners care more: green space that is perceivable or even liked by the public. What if a city had a large green space in its downtown that was full of crimes? Probably no one would like to pay a visitor there with a camera. Whatever reasons there are, green spaces with limited public accessibility face a failure of creating sufficient green perceptions.

Particularly in the case of Singapore, close to the equator, its unique geographic location raises a rich ground for all kinds of greens with adequate lighting and precipitation. Gifted by such an ideal geographic location, Singapore has still spent great efforts in building up a Garden-like ambience by promoting the greenway movement for an island-wide network of green corridors since the late 1980's. Nowadays, Singapore has achieved a perceivable green environment with a luxuriance of greenery along its tree-lined roads and hierarchy of parks ranging from the large regional open parklands to intimate pocket parks within the city and in residential neighborhoods (Tan 2006).

(2) Green and high-rises separated city (Toronto) The second type of city stands for another living style with a typical pattern of separated green perceptions and high-rises showing in Fig. 23. Many cities of this group come from the North America, and areas with colorful dots are more concentrated than those from other cities. The place between green perception and

Fig. 24 The photos with high-rises perceptions in Paris, in which the buildings are not very high in reality but are of a "high impression" (the *left figure* http://www.panoramio.com/photo/75777925, the *middle figure* http://www.panoramio.com/photo/49571711, and the *right figure* http://www.panoramio.com/photo/68482412)

high-rises perception looks like a vacuum area. Most of these large-scale blank areas are communities, because compared to other spaces community is the least likely place to attract attention. Such a phenomenon is fairly common among North America cities, where single-family housing is the most popular prototype for residence. In most cases of this type of cities,[10] it is reasonable to see that the high-rises perceptions stand for office buildings as working places, while the green perceptions are public parks, offering the recreational function.

(3) Green and high-rises perceptions mixed cities (Paris) Strictly speaking, the high-rises perceptions do not mirror the real height of the building in Fig. 25, because each photo has its own measurement that is hard to be uniformed, the high-rises spots are where photos with content brings an impression of tall buildings. Showing on the map, cities in this category have relatively well-balanced look of both green and yellow areas, because their urban structures tend to have most amenities evenly distributed and embeded within homogeneous urban fabricates. Most of the high-rises perceptions are not really high but quite dense, like what Fig. 24 shows, it can be generalized that cities in this category plant green perceptions into their concrete forests, together with convenient public access.

These evenly distributed green perceptions within high-rises perceptions are relatively common in cities from Europe, which is probably a result of planning efforts. In Europe, the European Environment Agency (EEA) recommends that "people should have access to green space within 15 min walking distance" (Barbosa et al. 2007). A Europe-wide assessment of access to green space reported that all citizens in Brussels, Copenhagen, Glasgow, Gothenburg, Madrid, Milan and Paris live within 15 min walks of urban green space, as well as the residents of many smaller cities (Stanner and Bordeaux 1996).

(4) High-rises perception dominated cities (Manhattan) If without the central park, the overall look of the C-IMAGE of Manhattan is full of yellow dots indicating high-rises perceptions almost everywhere. When it comes to the fourth category in Fig. 26, cities of this category usually display an impression of overwhelmingly crowded buildings. As the pioneer of regulating skyscrapers, the landmark 1916 law reached a compromise solution to the reconcile the conflicts between economic and public interests for the real estate development in the commercial district of the island: it permits developers to construct a higher tower if they preserve more amount of open space (Weiss 1992).

Tokyo versus Shanghai

A particular example jumps out when juxtaposing the C-IMAGEs of Tokyo and Shanghai as in Fig. 27. Tokyo looks much greener than Shanghai, yet both as one of the densest cities in the world, they are similar in many aspects especially in the construction of green space. Based on a study of Asian Green City Index in 2011 (Economist Intelligence Unit et al. 2011, pp 103,

[10] The downtown of New York City is different, which will be discussed later.

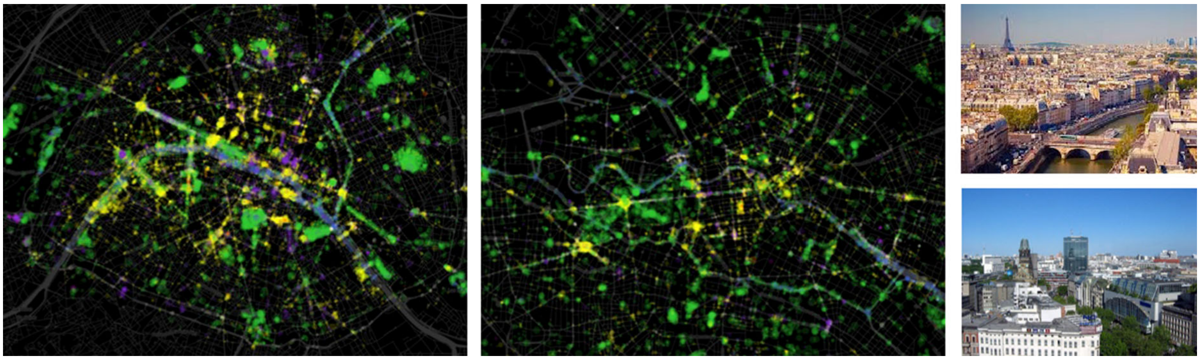


Fig. 25 The green and high-rises mixed cities: Paris (*left*) with its image (*top right*), and Berlin (*middle*) with its image (*bottom right*) (the *top right figure* http://www.worldretailcongresslatam.com/media/18623/paris.jpg, and the *bottom right figure* http://www.berlin-germany-hotels.com/images/Kudamm.JPG). (Color figure online)

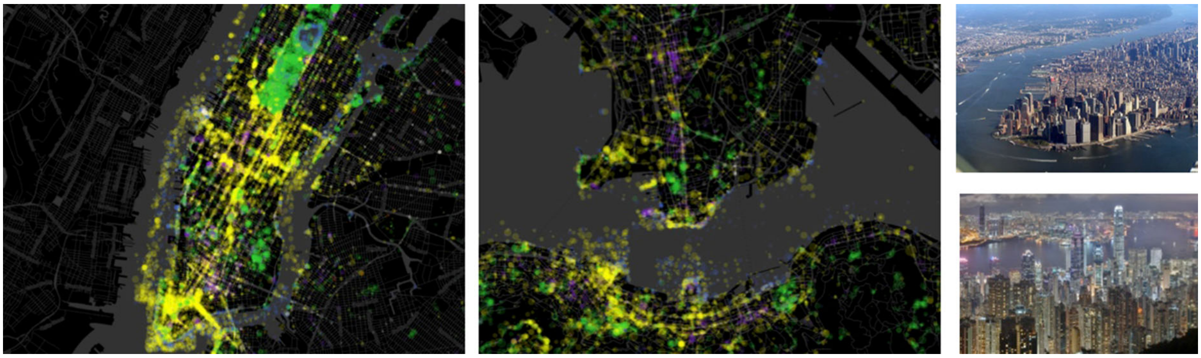


Fig. 26 The high-rises dominated cities: Manhattan of New York City (*left*) with its image (*top right*), and Hong Kong (*middle*) with its image (*bottom right*) (the *top right figure* http://bungobox.com/images/uploads/Manhattan.jpg, and the *bottom right figure* http://upload.wikimedia.org/wikipedia/commons/1/18/Hong_Kong_Night_Skyline.jpg). (Color figure online)

Fig. 27 The C-IMAGE of Tokyo (*left*) and Shanghai (*right*). (Color figure online)

115), Table 6 lists a series of urban indicators showing their similarity.

Given those numbers in Table 6, as the largest financial center of China, Shanghai still has a long way to catch up with its neighbor city Tokyo in terms of energy efficiency. However, they both have to face the intense conflict between the giant population and the shortage of land resources. As a widely recognized

Table 6 the quantitative indicators of Tokyo and Shanghai (Economist Intelligence Unit et al. 2011, pp 103, 115)

Urban indicators	Tokyo	Shanghai
Total population (million)	13.0	19.2
Administrative area (km^2)	2187.7	6340.5
GDP per person (current prices) (US$)	70,759.6	11,463.7
Population density (persons/km^2)	5946.9	3030.2
Temperature (24-h average, annual) (°C)	15.0	16.0
CO_2 emissions per person (tonnes/person)	4.8 (2008)	9.7 (2008)
Energy consumption per US$ GDP (MJ/US$)	1.2 (2008)	14.8 (2008)
Green spaces per person (m^2/person)	10.6 (2005)	18.1 (2008)
Superior public transport network, covering trams, light rail, subway and BRT (km/km^2)	0.14 (2010)	0.07 (2010)
Share of waste collected and adequately disposed (%)	100.0 (2005)	82.3 (2009)
Waste generated per person (kg/person/year)	375.1 (2008)	369.5 (2009)
Water consumption per person (l/person/day)	320.2 (2008)	411.1 (2008)
Water system leakages (l/person/day)	3.1 (2008)	10.2 (2008)
Population with access to sanitation (%)	99.4 (2008)	72.5 (2009)
Share of wastewater treated (%)	100.0 (2009)	78.4 (2008)
Daily nitrogen dioxide levels (μg/m^3)	39.5 (2007)	53.0 (2009)
Daily sulfur dioxide levels (μg/m^3)	5.7 (2007)	35.0 (2009)
Daily suspended particulate matter levels (μg/m^3)	33.1 (2007)	81.0 (2009)

Fig. 28 The satellite images of Tokyo (*left*) and Shanghai (*right*). (Color figure online)

urban indicator to evaluate the quality of green space of a city (Newton 2001), the "green spaces per capita" for Tokyo is only 10.6 and 18.1 m^2/person for Shanghai, which are below the average score of 38.6 m^2/person among the Asian cities in the 2011 research. Their shortage of green space is even clear to see from the satellite maps (Fig. 28). This vision-based comparison also has quantitative support. In adjusting its previous Liveability Index, EIU published a report about city rankings in 2012, which includes a new method of evaluating cities' quality of green space through detecting and calculating green areas from satellite images. The scores from this study of Tokyo and Shanghai are the same, as low as 3.3 (E. I. Unit 2012, p 12).

The performance of green perception in Tokyo becomes more confusing if taking Singapore into account. Tokyo has an amazingly large proportion of "green perception" even compared to Singapore as listed in Fig. 29. Different from Singapore, Tokyo does not obtain a tropical climate to support all year round green coverage, but it has to deal with an intensive conflict between population and limited land resources. It seems that the only chance for Tokyo is to encourage its residents to upload green photos for their city. Is Tokyo cheating, and how?

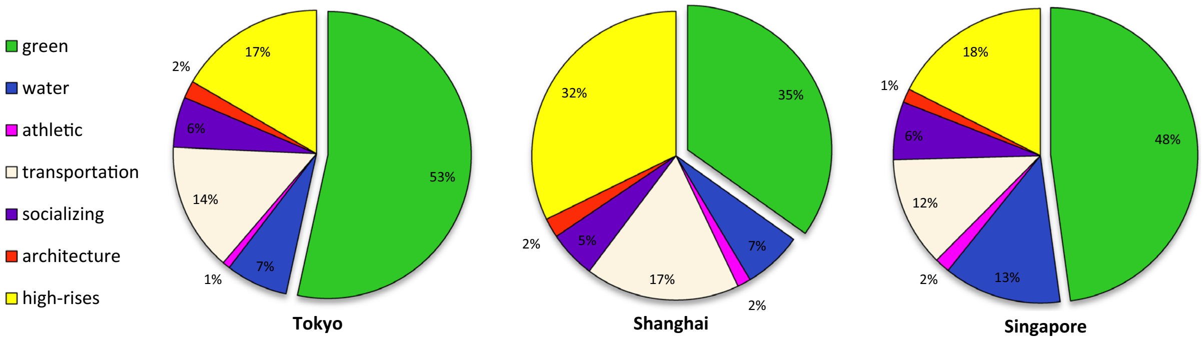


Fig. 29 The proportion of the seven perceptions in Tokyo (*left*), Shanghai (*middle*), and Singapore (*right*). (Color figure online)

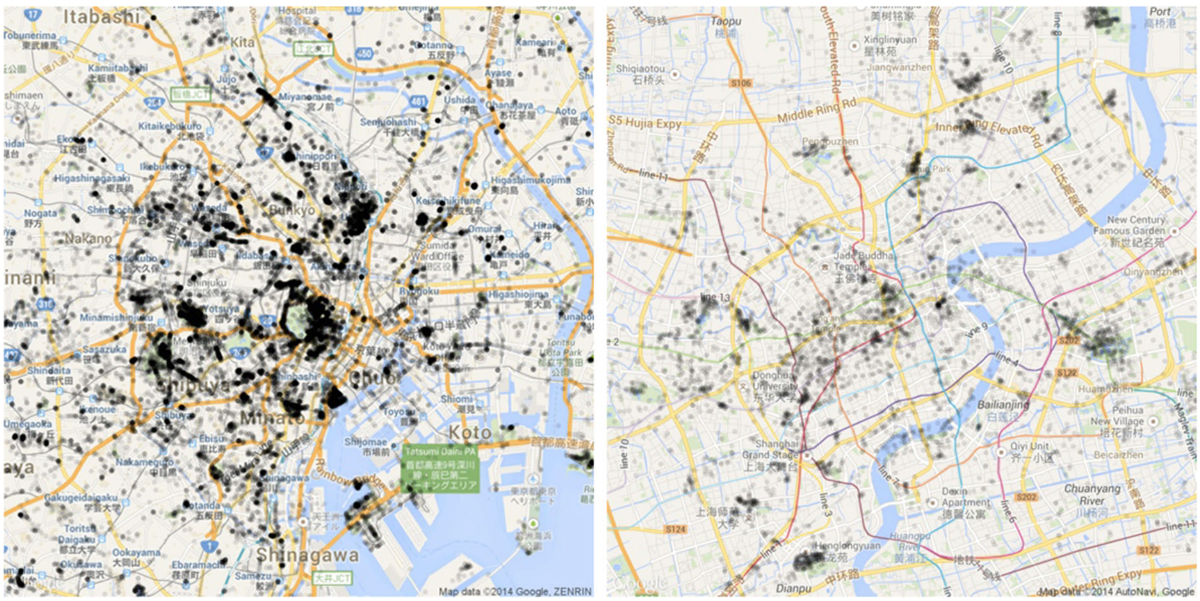


Fig. 30 The distribution of green perception in Tokyo (*left*) and Shanghai (*right*). (Color figure online)

There can be a lot of explanations for this difference that needs further study. For example, as listed in Fig. 30, the concentrated areas of green perceptions in Tokyo and Shanghai are opposite—photos in Tokyo are centralized while those in Shanghai are closer to the periphery. Parks such as Akasaka Palace, Aoyama Park, Roppongi Hills, and the Imperial Palace are all located near the downtown with convenient accessibility to the public. On the contrary, despite Shanghai is also providing large bulks of green space as Tokyo, a lot of them are located in the periphery of the city with a constrained public accessibility. Other studies between these two cities have also mentioned this difference. (Xu 2005), in his compare study between Tokyo and Shanghai's urban green space, also owes the difference to an unbalanced structural layout of urban parks in Shanghai and the loss of small-scale green spaces.

Interestingly, at the beginning of the twentieth century, Tokyo and Shanghai both published their metropolitan master plans, including the planning for the construction of urban green space. The 2000 green plan in Tokyo does not touch any specific issues such as the allocation of urban land resources, but plays as guidance for a set of following regulations and policies. From the disaster-prevention circle to neighborhood green spaces, from roadside trees (Jim 2004) to water system, from green trust to thematic youth center, and from ecological preservation to green building technology, the plan has set up a framework that covers almost all the aspects that can be related to green space construction.

On the contrary, Shanghai's green plan in 2003 sets up a set of numbers as development goals, emphasizing the increase in total volume of green space. For example, the plan defines green bands by different width according to the hierarchy of the roads or rivers. Moreover, for various stages, it also prepares the explicit increase number of green area for different districts. After the approval of this plan, which clearly states all of its construction targets, there is no more any other green-related planning at the municipal level. The implementation of it 10 years later is not that smooth. According to Shanghai's 2013 statistical yearbook, the average park area per capita is only 13.29 m^2/person, but since the adjustment in the definition of "green space", the pure volume of green space have surged to 124,204 ha. Following this adjustment, barely calculating the green space per person can directly produce a high value of green spaces per capita up to 52 m^2/person, which is almost two times of its goal of 2003!

Setting up goals by number brings efficiency, but chasing numbers also accompanies with a severe ignorance of quality. There has been a long-term debate about which green space should include and which should not. In the green plan of Shanghai, however, it is hard to find any suggestions about specific utilization of those green space the government has planned.

Conclusion

This paper innovatively proposes a methodology in recognizing the public perception of a city—C-IMAGE—and provides two applications for a better understanding of the public cognition of urban environment. Section "Introduction" throws a light on the chance to develop a new method for city cognition study based on nowadays crowd-sourcing technology. Section "Literature review" briefs the study of urban cognition mapping in both traditional urban study area and today's urban computing stream. Section "Method statement" demonstrates the general concept and method to achieve the data for C-IMAGE. Section "Applications" provides two usages of such a new dataset, together with two findings. Based on the metadata, the first application addresses the similarity and difference between C-IMAGE and the traditional mental maps. The second one composes city maps according to photos' contents.

Key contributions

(1) Efficient method This paper proposes a duplicable method of detecting public perception of the city via geo-tagged photos in a quite efficient way. Such a research tool achieves theoretical support as a successor of the traditional cognitive mapping tool, such as the city image from Kevin Lynch. It also achieves practical acknowledgement as a trailblazer to integrate vision computation with urban subject.

(2) Cognition as elements In this application, the results from C-IMAGE can partially verify and match the five elements from the conventional mental maps. Although coming from different techniques, they do share similarities in discovering Paths, Nodes, and Edges. With the technology of image understanding, C-IMAGE is also proved to be capable of detecting Landmarks. But it is difficult to detect "districts" as those defined in Lynch's map.

(3) Perceptions as indicators This paper is also among the first ones who illustrate cities' visual contents based on image content level. Both qualitatively and quantitatively, it treats perceptions as a measurement of cities' performance in comparison with that of using traditional urban indicators.

Further study

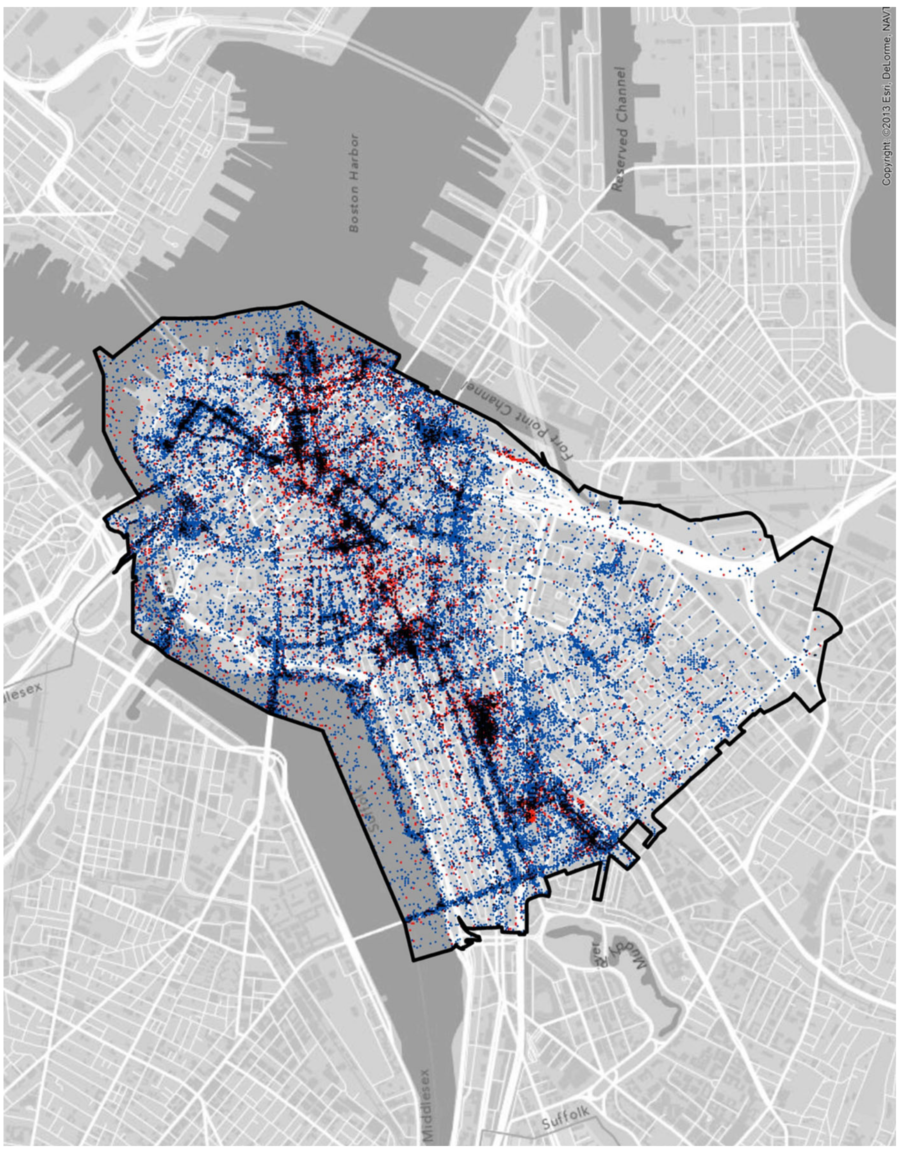
The C-IMAGE project sheds a light on the study of city cognition based on today's crowd-sourcing technology. But it is just the start with a lot of areas needs further explorations.

Above all, both applications in this paper need to be extended. The identification of the dots' clusters in the metadata based C-IMAGE is still through a manual process, which is less convincing and effective. For example, to what degree of density a pattern can be recognized as a node is very yet to be determined, and probably an algorithm of cluster is necessary to substitute the manual clustering. As to the second application, there is much more information to explore from the C-IMAGE maps composed by the 7 urban perceptions. Cases such as the comparison between Tokyo and Shanghai that shows the gap between perceptions and realities also deserve further study.

For the technical part, the vision computation is far from to be complete, so the accuracy and quality of image analysis is yet not to be satisfied. With the development of this technology, it is very likely to have a deeper categorization of the image content so that researchers may disclose more interesting findings.

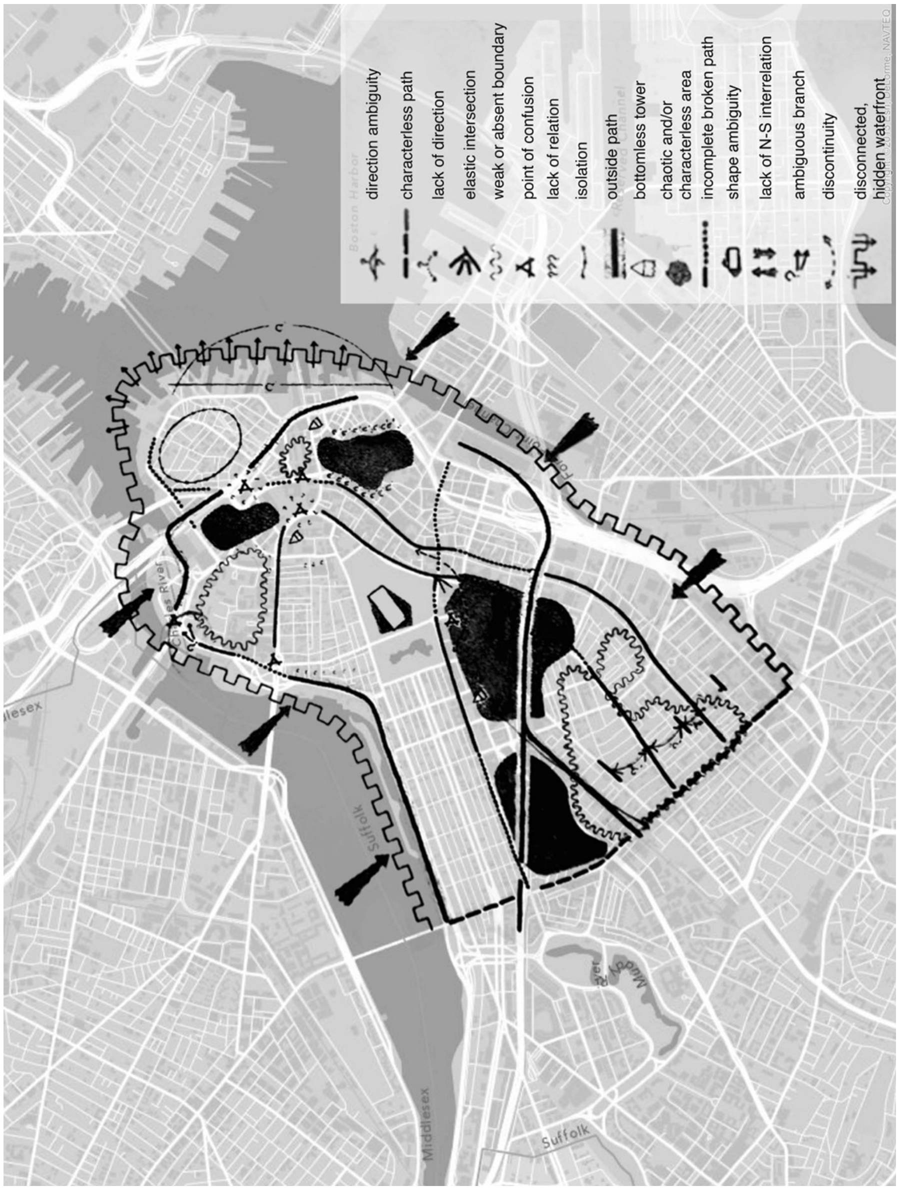
Appendix 1

Larger map of of Fig. 4.

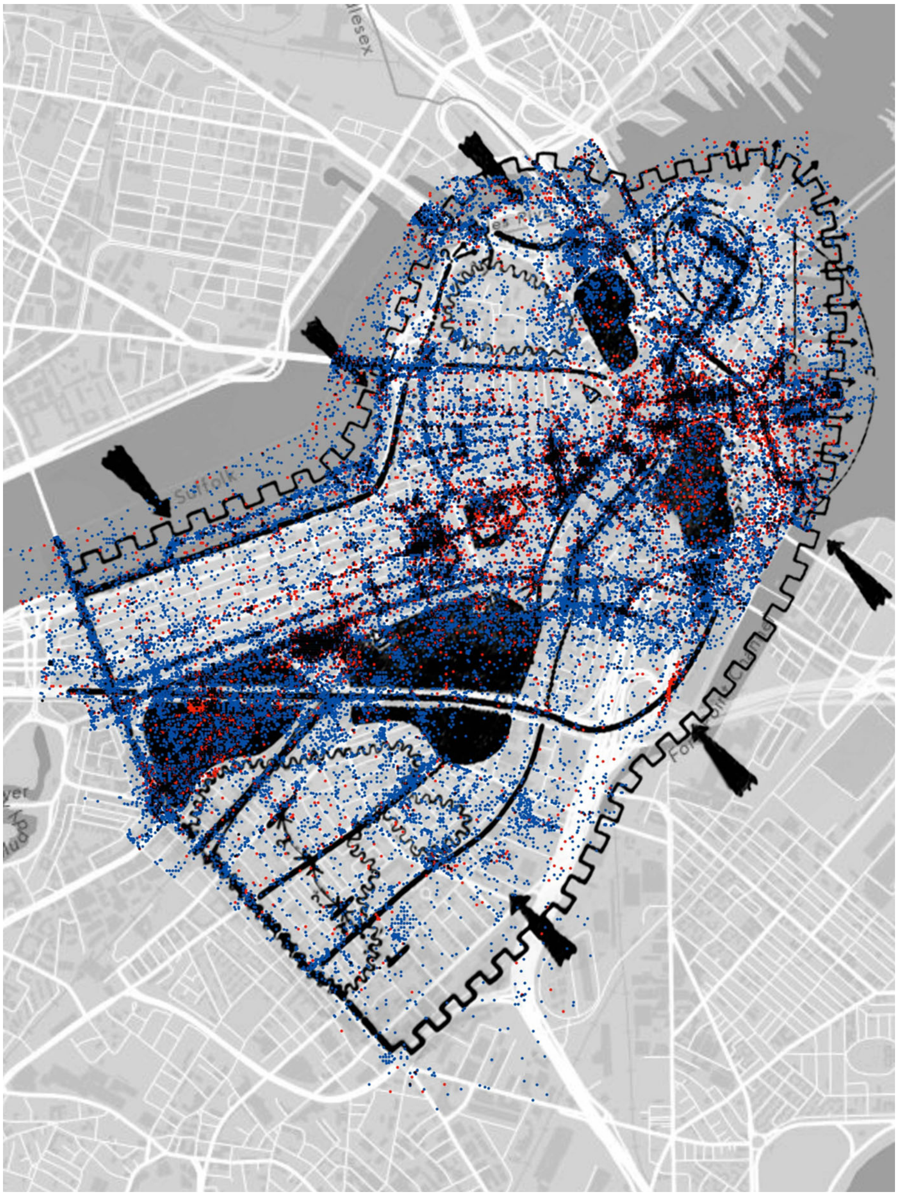


Larger map of Fig. 5.

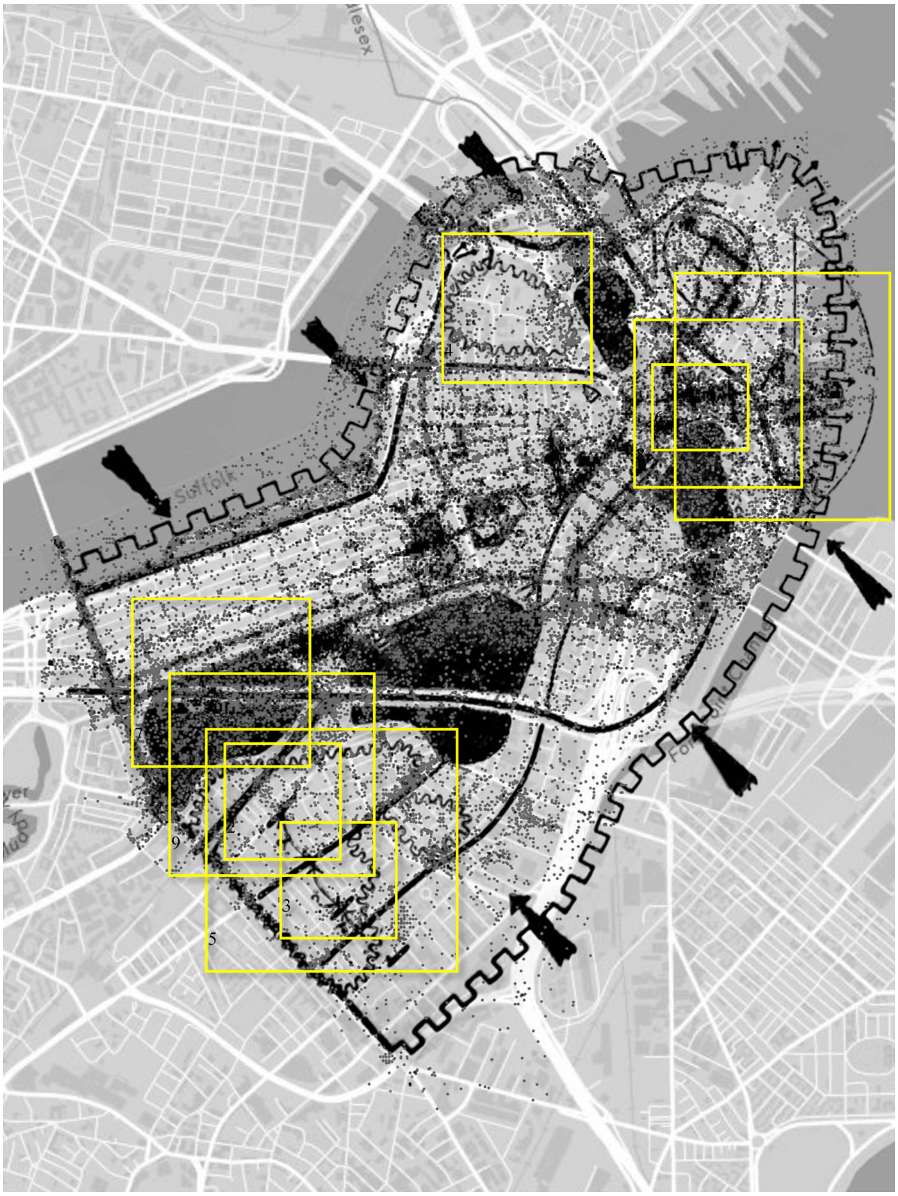
Larger map of Fig. 12 (left).



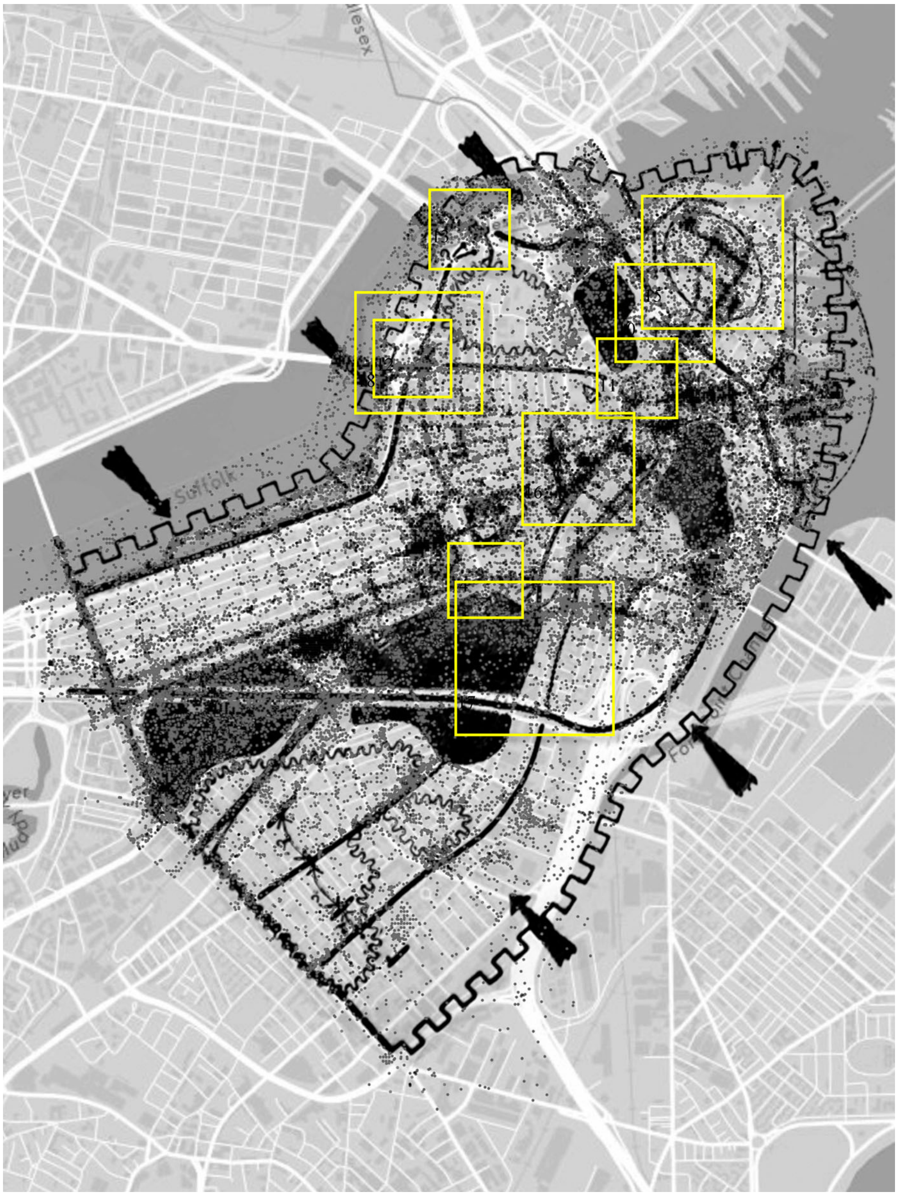
Larger map of Fig. 12 (right).



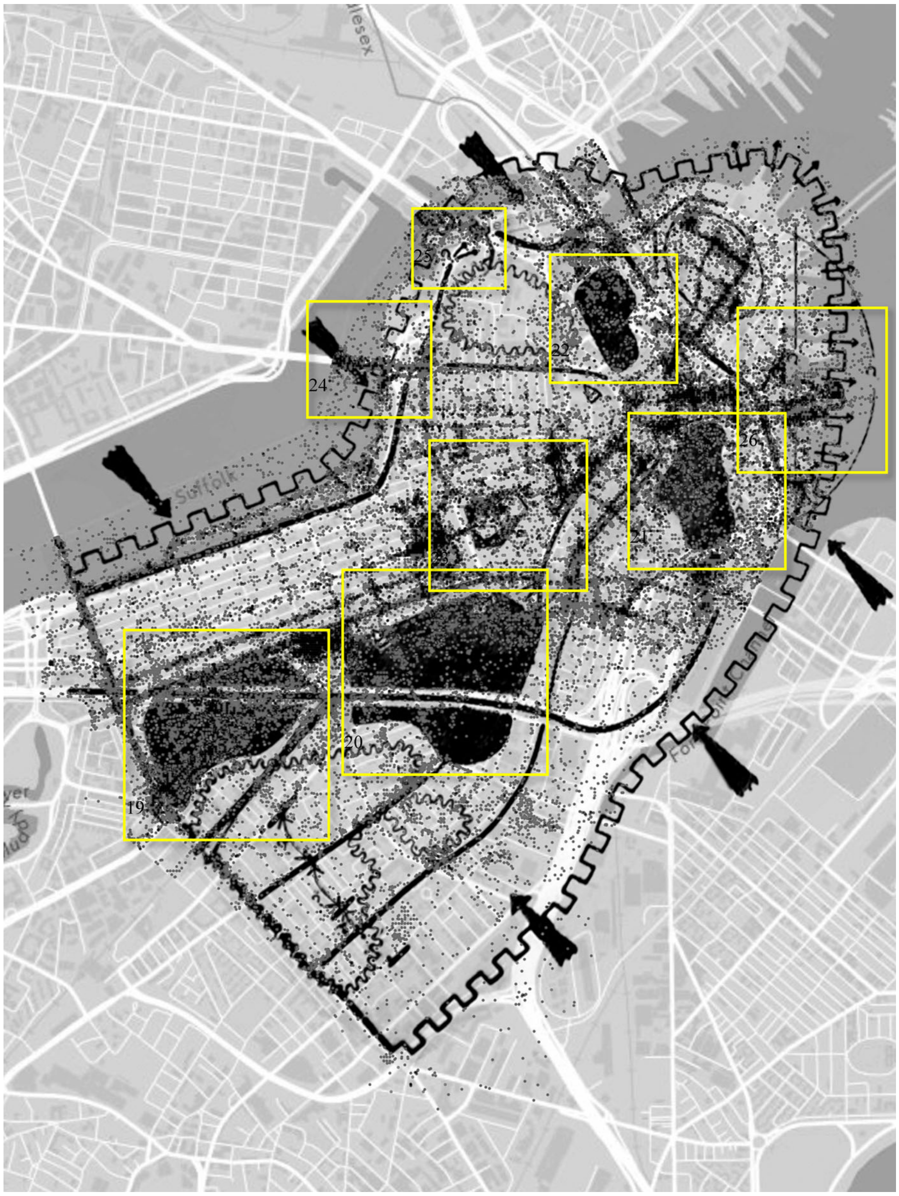
Reference map for Table 5 (problem 1 to 9).



Reference map for Table 5 (problem 9 to 18).

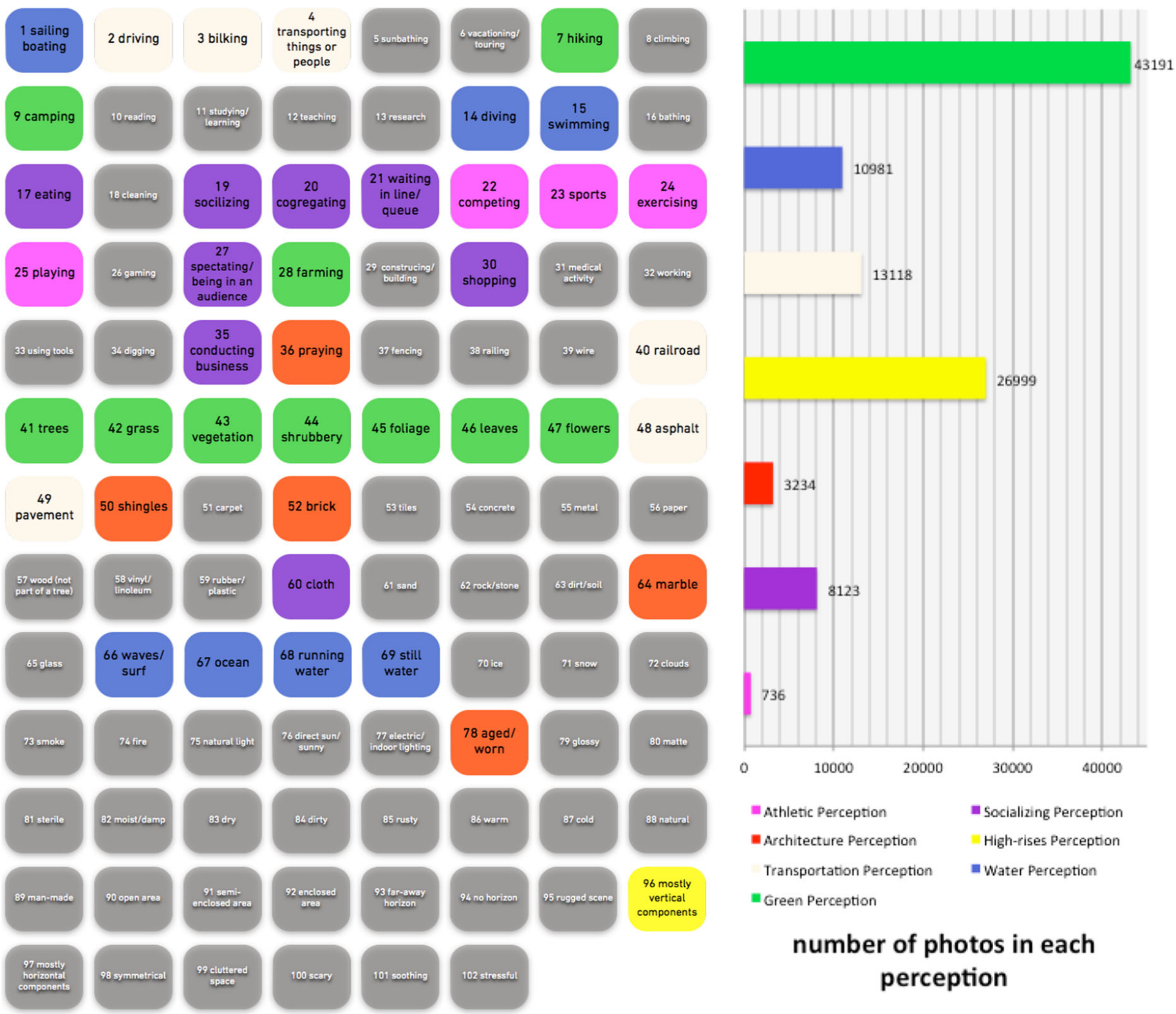


Reference map for Table 5 (problem 19 to 26).



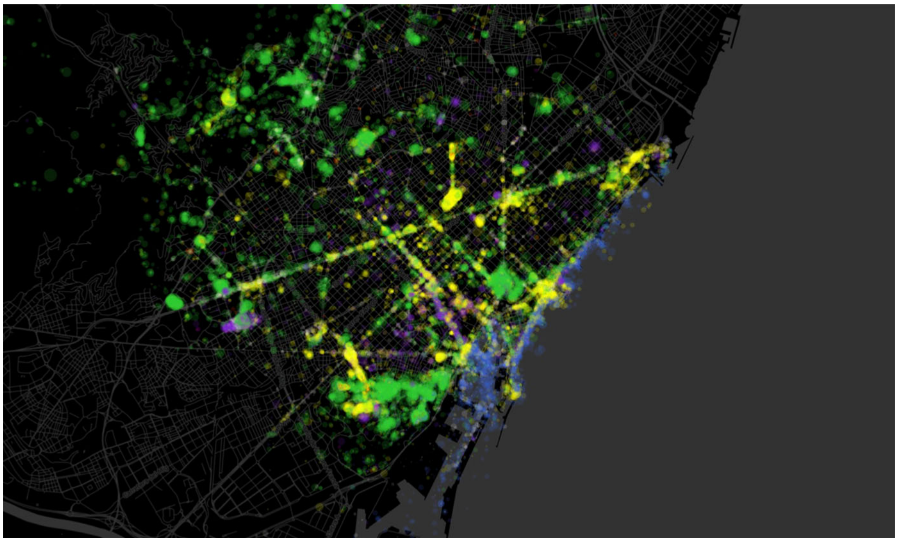
Appendix 2

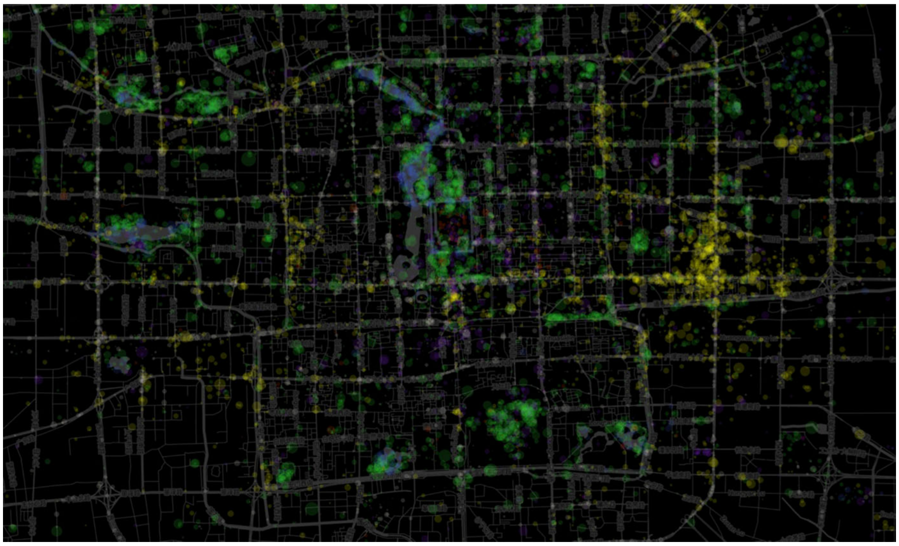
The 7 urban perceptions (colored rectangles) derived from the 102 attributes (left) and the histogram of all the photos of each perception in London[11] (right).

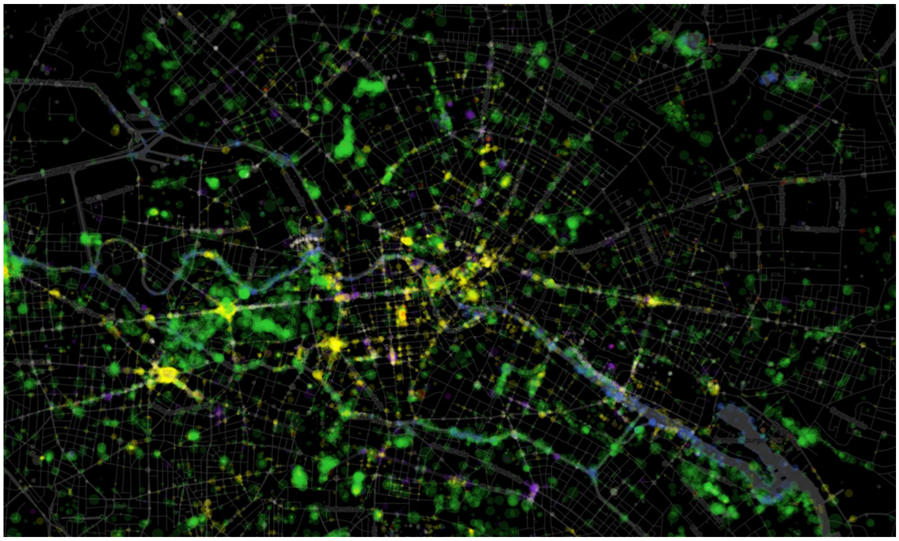


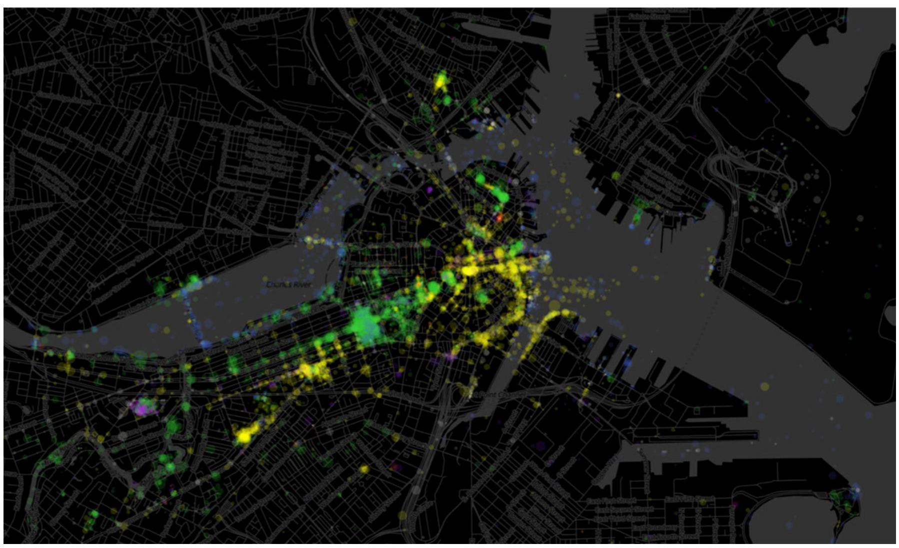
[11] The complete set of all the histograms of the 26 cities can be found in "Appendix 2".

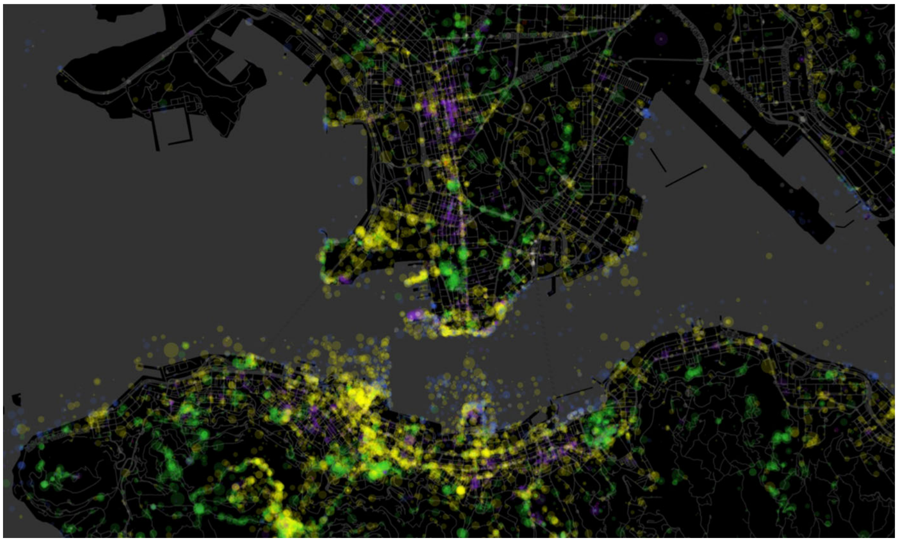
The infographics based on the seven perceptions of 21 cities.

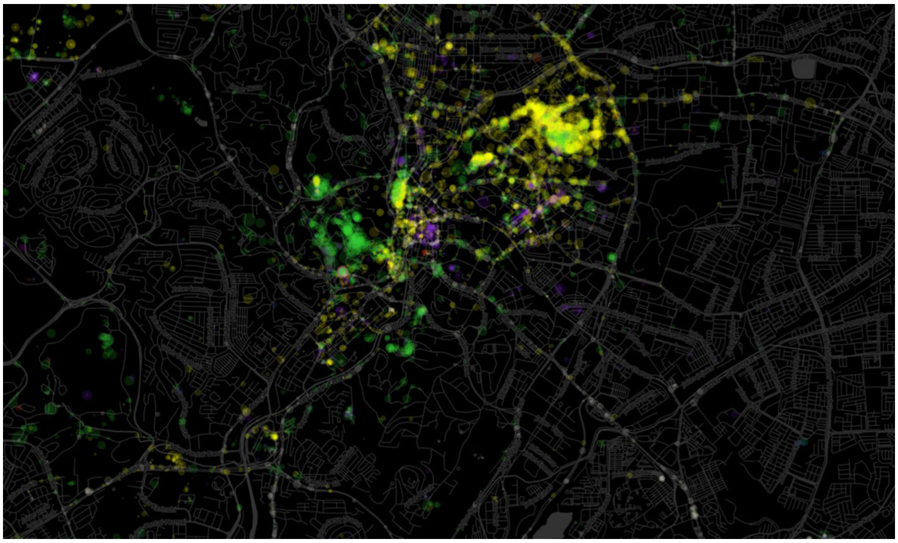


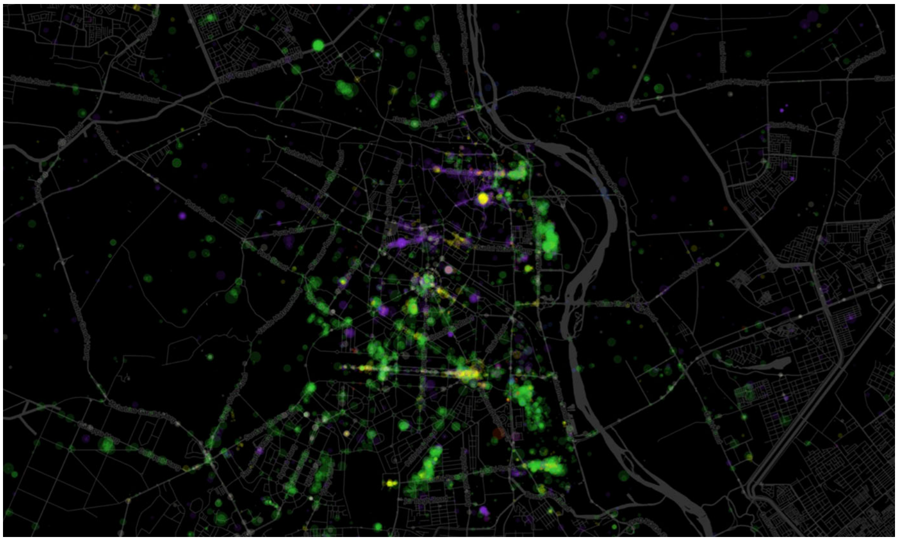


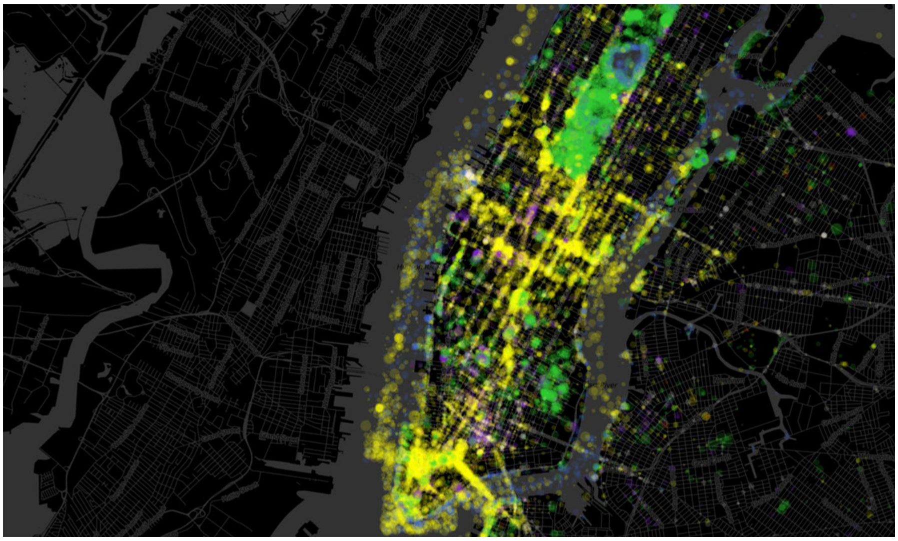


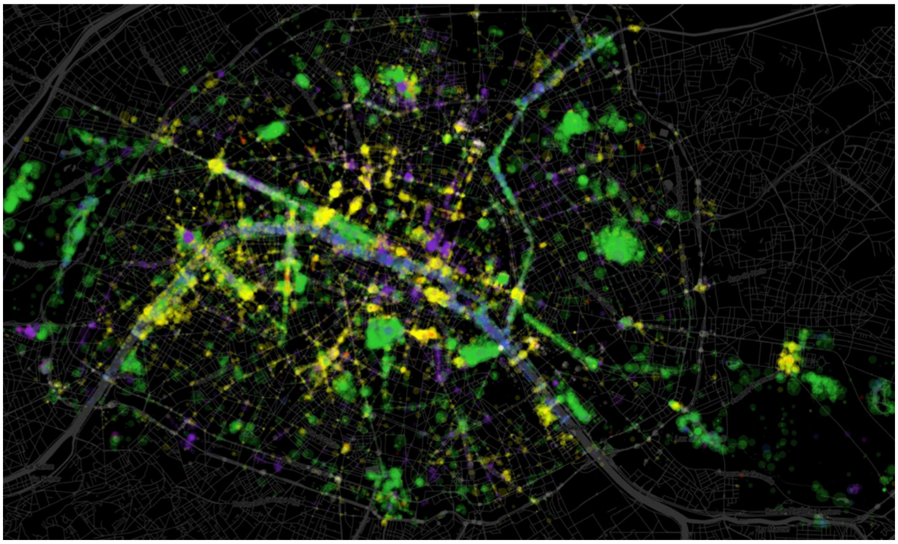


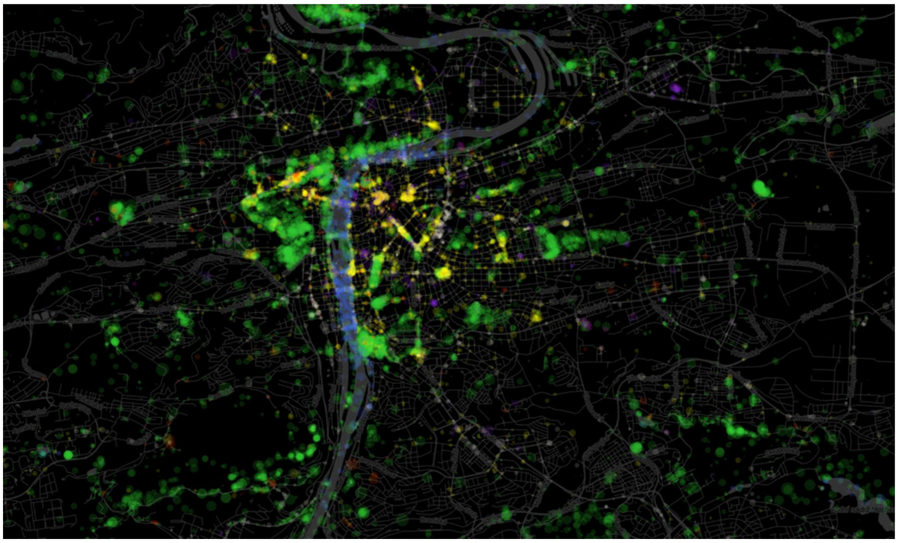


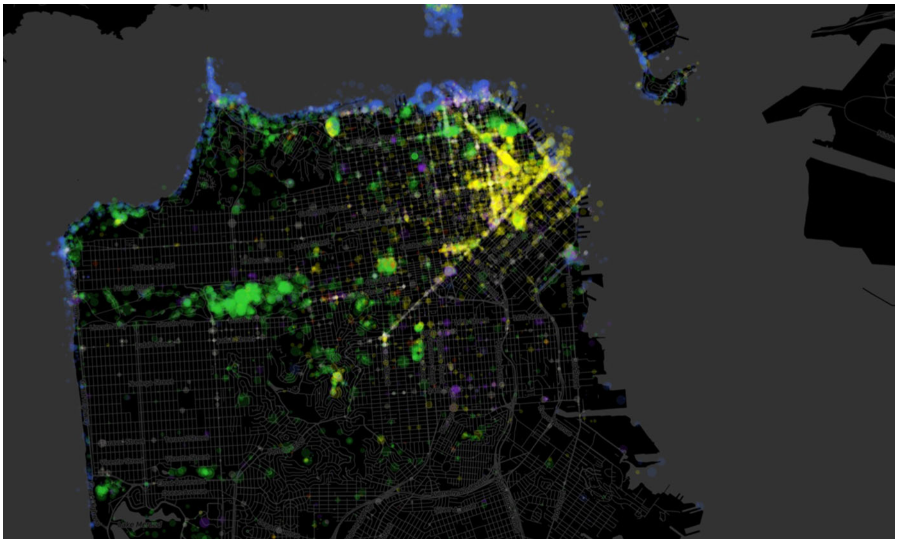


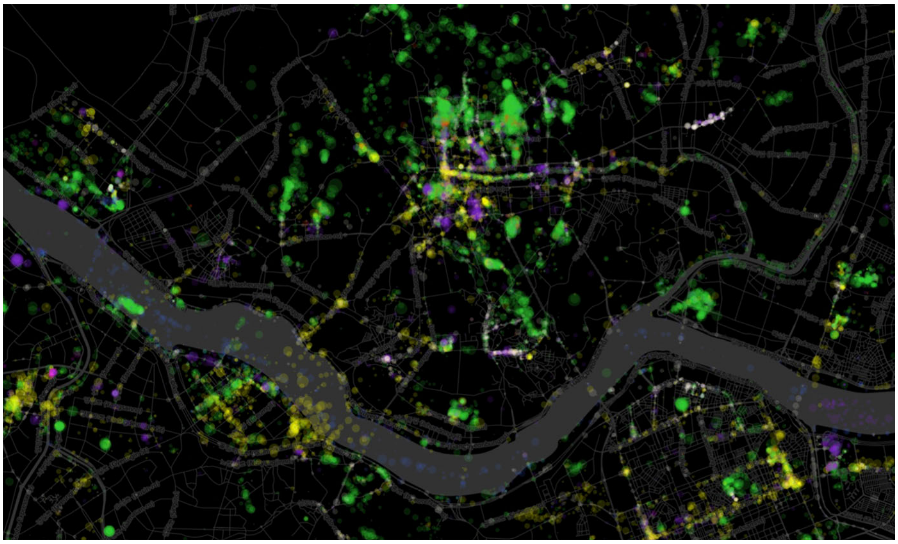


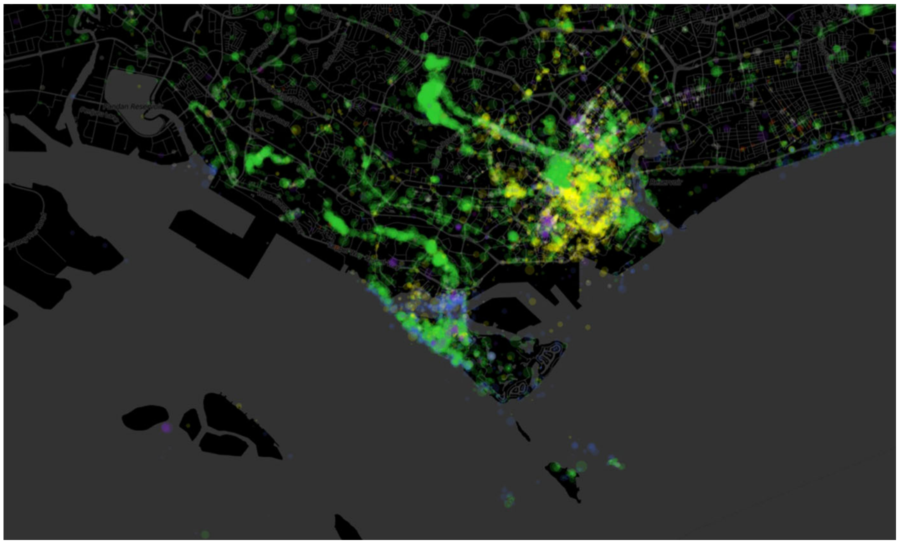


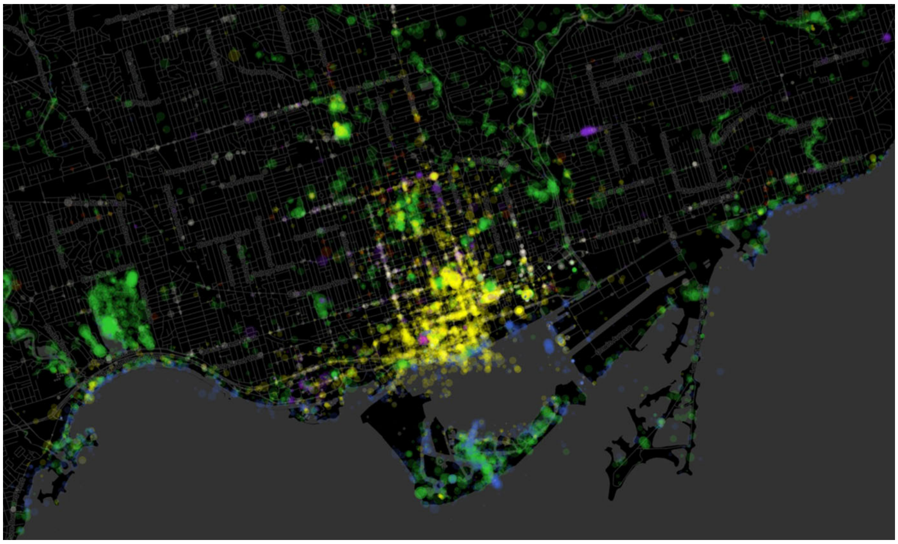


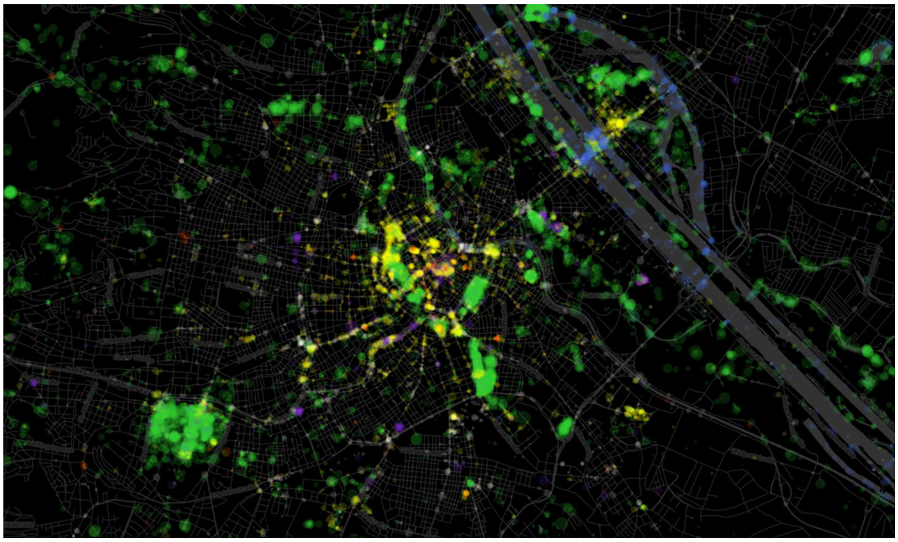


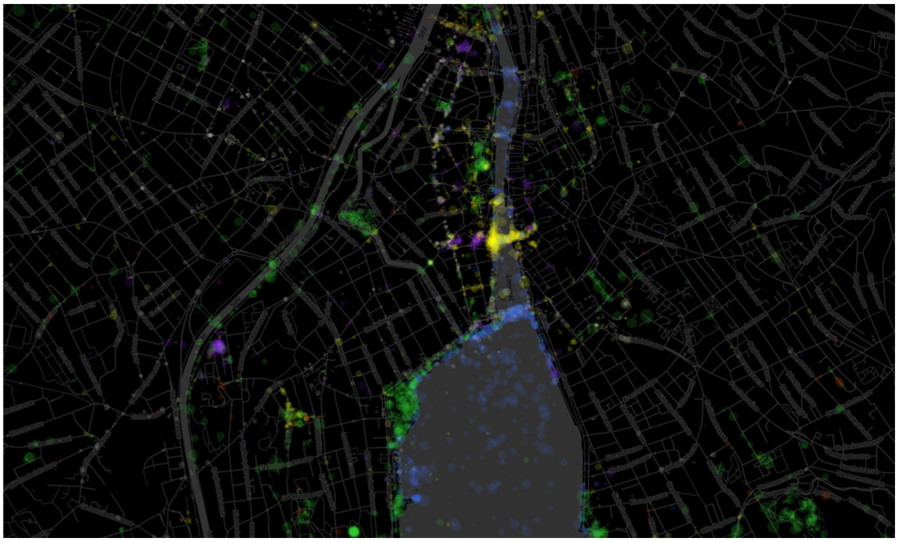












References

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

Barbosa, O., Tratalos, J. A., Armsworth, P. R., Davies, R. G., Fuller, R. A., Johnson, P., et al. (2007). Who benefits from access to green space? A case study from Sheffield, UK. *Landscape and Urban Planning, 83*(2–3), 187–195. doi:10.1016/j.landurbplan.2007.04.004.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*(3), 27:1–27:27. doi:10.1145/1961189.1961199.

de Jonge, D. (1962). Images of urban areas their structure and psychological foundations. *Journal of the American Institute of Planners, 28*(4), 266–276. doi:10.1080/01944366208979452.

Doersch, C., Singh, S., Gupta, A., Sivic, J., & Efros, A. A. (2012). What makes Paris look like Paris? *ACM Transactions on Graphics, 31*(4), 101:1–101:9. doi:10.1145/2185520.2185597.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., & Darrell, T. (2013). DeCAF: A deep convolutional activation feature for generic visual recognition. http://arxiv.org/abs/1310.1531

Francescato, D., & Mebane, W. (1973). How citizens view two great cities: Milan and Rome. In R. M. Downs & D. Stea (Eds.), *Images and environment: Cognitive mapping and spatial behaviour* (pp. 131–147). Chicago: Aldine.

Gulick, J. (1963). Images of an Arab City. *Journal of the American Institute of Planners, 29*(3), 179–198. doi:10.1080/01944366308978063.

Harrison, J. D., & Howard, W. A. (1972). The role of meaning in the urban image. *Environment and Behavior, 4*(4), 389–411.

Jim, C. Y. (2004). Green-space preservation and allocation for sustainable greening of compact cities. *Cities, 21*(4), 311–320. doi:10.1016/j.cities.2004.04.004.

Klein, H.-J. (1967). The delimitation of the town centre in the image of its citizens. *Urban Core and Inner City,* 286–306.

Krieger, A., Cobb, D. A., Turner, A., & Bosse, D. C. (Eds.). (2001). *Mapping Boston*. Cambridge: MIT press.

LeGates, R. T., & Stout, F. (2011). *The city reader*. London: Routledge.

Lynch, K. (1960). *The image of the city* (Vol. 11). Cambridge: MIT Press.

Marr, D., & Vision, A. (1982). A computational investigation into the human representation and processing of visual information. *WH San Francisco: Freeman and Company*. http://www.contrib.andrew.cmu.edu/~kk3n/80-300/marr2.pdf

Newton, P. (2001). *Urban indicators for managing cities. Urban indicators for managing cities: Cities data book*. Manila: Asian Development Bank.

Ojala, T., Valkama, V., Kukka, H., Heikkinen, T., Lindén, T., & Jurmu, M., et al., (2010). UBI-hotspots: Sustainable Ecosystem Infrastructure for Real World Urban Computing Research and Business. In *Proceedings of the international conference on management of emergent digital ecosystems* (pp. 196–202). New York, NY, USA: ACM. 10.1145/1936254.1936288

Patterson, G., & Hays, J. (2012). SUN attribute database: Discovering, annotating, and recognizing scene attributes. In *2012 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 2751–2758). doi:10.1109/CVPR.2012.6247998

Rapoport, A. (1977). *Human aspects of urban form: Towards a man—environment approach to urban form and design/Amos Rapoport* (p. 1977). Oxford: Pergamon Press.

Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013). The collaborative image of the city: Mapping the inequality of urban perception. *PLoS One, 8*(7), 1–12. doi:10.1371/journal.pone.0068400.

Stanner, D., & Bordeaux, P. (1996). The urban environment. *The State of the Environrnent (The Dobris Assessment)*.

Tan, K. W. (2006). A greenway network for Singapore. *Landscape and Urban Planning, 76*(1–4), 45–66. doi:10.1016/j.landurbplan.2004.09.040.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55*(4), 189–208. doi:10.1037/h0061626.

Unit, E. I. (2012). Best cities ranking and report: A special report from the Economist Intelligence Unit. *The Economist*.

Unit, E. I., Britain, G., & Aktiengesellschaft, S. (2011). *Asian Green City Index: Assessing the environmental performance of Asia's major cities: a research project*. Berlin: Siemens.

Weiss, M. A. (1992). Skyscraper zoning: New York's pioneering role. *Journal of the American Planning Association, 58*(2), 201–212. doi:10.1080/01944369208975794.

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. In *2010 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 3485–3492). doi:10.1109/CVPR.2010.5539970

Xu, H. (2005). The analysis and comparison study on Tokyo urban open space and our country's city. *Urban Planning Overseas, 20*(06), 27–30.

Zheng, Y., Liu, F., & Hsieh, H. P. (2013). U-Air: When urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1436–1444). ACM.

Zheng, Y., Liu, Y., Yuan, J., & Xie, X. (2011). Urban computing with taxicabs. In *Proceedings of the 13th international conference on ubiquitous computing* (pp. 89–98). New York, NY, USA: ACM. doi:10.1145/2030112.2030126

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2014a). Learning deep features for scene recognition using places database. In *Advances in neural information processing systems* (pp. 487–495).

Zhou, B., Liu, L., Oliva, A., & Torralba, A. (2014b). Recognizing city identity via attribute analysis of geo-tagged images. In *European conference on computer vision* (pp. 519–534). Springer International Publishing.