Chapter 27

What Urban Cameras Reveal About the City: The Work of the Senseable City Lab

Fábio Duarte and Carlo Ratti

Abstract Cameras are part of the urban landscape and a testimony to our social interactions with city. Deployed on buildings and street lights as surveillance tools, carried by billions of people daily, or as an assistive technology in vehicles, we rely on this abundance of images to interact with the city. Making sense of such large visual datasets is the key to understanding and managing contemporary cities. In this chapter, we focus on techniques such as computer vision and machine learning to understand different aspects of the city. Here, we discuss how these visual data can help us to measure legibility of space, quantify different aspects of urban life, and design responsive environments. The chapter is based on the work of the Senseable City Lab, including the use of Google Street View images to measure green canopy in urban areas, the use of thermal images to actively measure heat leaks in buildings, and the use of computer vision and machine learning techniques to analyze urban imagery in order to understand how people move in and use public spaces.

27.1 Introduction

Cameras have become part of the urban landscape and a testimony of our social interactions with the city. They are deployed on buildings and street lights as surveillance tools, carried by billions of people daily, or as an assistive technology in vehicles with different levels of self-driving capabilities. We rely on this abundance of images to interact with the city.

In fact, 2.5 quintillion bytes of data are created each day by billions of people using the Internet. Increasingly, social media are heavily based on visual data. Among the top social media channels, several are overwhelmingly and exclusively based on images: YouTube has 1.5 billion users and Instagram has 1 billion users—as a

---

F. Duarte (✉) · C. Ratti

Senseable City Lab and Department of Urban Studies and Planning, Massachusetts Institute of Technology, Massachusetts, USA

e-mail: fduarte@mit.edu

C. Ratti

e-mail: ratti@mit.edu

© The Author(s) 2021

W. Shi et al. (eds.), *Urban Informatics*, The Urban Book Series,
https://doi.org/10.1007/978-981-15-8983-6_27

comparison, Facebook has 2.3 billion users. Such visually based social interactions are also extended to the interactions we have in our cities. In the USA, on average, a person is caught on camera 75 times per day, and over 300 times in London. Also, disruptive urban technologies such as autonomous vehicles use cameras. The challenge is to make sense of the amount of visual data generated daily in our cities in meaningful ways, beyond surveillance purposes.

In this chapter, we are not interested in the abundance of visual data available online collected by individuals and widely available on social media. The previous work used geotagged photographs available online to measure urban attractiveness (Paldino et al. 2016) or to assess the aesthetic appeal of the urban environment based on user-generated image (Saiz et al. 2018), and the visual discrepancy and heterogeneity of different cities around the world (Zhang et al. 2019). The focus of this chapter is not on the visual data produced by cameras carried by people for personal uses, but rather on the images collected by cameras specifically designed and deployed to gather visual data about the city—which we call here urban cameras.

Cameras deployed and controlled by a range of public and private organizations in urban areas are counted by the dozens of thousands in cities, from London and Beijing to New York and Rio de Janeiro. As an example, a Londoner is captured on camera more than 300 times every day; and during the same period, the UK captures over 30 million plate numbers (Kitchin 2016). Additionally, private companies, such as Google, collect and make available online hundreds of thousands of images of hundreds of cities worldwide.

Making sense of such large visual datasets is the key to understanding and managing contemporary cities. There are still many technical issues to be solved to make the use of such huge visual datasets actionable. Challenges include cloud versus local storage and processing; architecture integration, ontology building, semantic annotation, and search; and online real-time analysis and offline batch processing of large-scale video data (Shao et al. 2018; Xu et al. 2014; Zhang et al. 2015).

Besides the technical challenges, there are also ethical issues. The most prevalent among social scientists is the narrow understanding of cities when urban phenomena are equaled to available data, heading the operationalization of the urban (Luque-Ayala and Marvin 2015), mainly when “portions of the urban public space that are shadowed by the gaze of private cameras and security systems” (Firmino and Duarte 2015 p. 743) become subject to the datafication of the city, often leading to “social sorting and anticipatory governance” (Kitchin 2016 p. 4). Closed-circuit television (CCTV), deployed on public areas and aimed to assist police patrols with crime prevention, using video analytics to identify abnormal behaviors, fosters predictive policing by the profiling of subjects and places, and frequently triggers false alarms due to biases embedded in the algorithms (Vanolo 2016).

We are aware of these issues and have contributed ourselves to the literature on the risks of oversurveillance based on the abundance of data about people’s behavior in public spaces. But, in this chapter, we would like to discuss the other side of this phenomenon: how novel computational techniques can be used to make sense of the huge amount of visual data generated about cities, and how such results reveal aspects of urban life that can contribute to better understanding and design of cities.

The projects discussed in this chapter are part of the extensive work using urban cameras done by the Senseable City Lab, at the Massachusetts Institute of Technology. These works can be divided into two types: the use of visual urban data available online, and the capture of visual data by the Lab with specifically designed devices.

In the first type, we take advantage of the visual urban data available online and develop machine learning techniques to make sense of these data. The datasets used in this research are Google Street View images, which we have been using to measure a critical aspect of cities with rapid urbanization: the quantification of green canopy in urban areas using a standard method that can be deployed cheaply, and that makes possible comparisons among hundreds of cities worldwide. And, at the same time, it provides a fine-grained analysis of greenery at the street level, allowing citizens and municipalities to assess tree coverage in different neighborhoods.

In the second type, we design specific devices to collect images and deploy them ourselves. In one example, we started by using thermal cameras mounted on vehicles to measure heat leaks in buildings. Using the same devices, we developed other techniques to use thermal data to quantify and track people's movements in indoor and outdoor areas. Besides the technical advantages of the method in terms of data transmission and processing, it also addresses an important concern about the use of cameras in public spaces: Thermal cameras allow us to have accurate data about people's behavior without revealing their identities, therefore avoiding privacy concerns. Also, as part of this type of research, we address the problem of indoor navigability in large public areas. It is a well-known problem that users often have difficulty in navigating areas such as shopping malls, university campuses, and train stations, due either to their labyrinthic design or to the repetitiveness of visual cues. Here, we collected thousands of images on the MIT campus and in train stations in Paris and trained a neural network to measure the easiness to navigate these spaces, comparing the results with a survey with users.

Visual data about cities will tend to increase in the coming years, with personal photographs and videos that people use to register their daily routines in cities posted on social media, the deployment of cameras for surveillance not only for policing purposes but also for traffic management and infrastructure monitoring, and the fact that visual data will be crucial in technologies such as self-driving cars. All work dealing with visual big data needs to overcome the hurdles of manually processing this massive amount of information and generating useful empirical metrics on visual structure and perception. In this chapter, we propose to discuss how the development of novel computation methods used to analyze the abundance of visual urban data can help us to better understand urban phenomena.

27.2 Computer Vision and the City: Google Street View Images

Some of the most prolific sources of spatial data are Google Maps, Earth, and Street View. These products offer Web mapping, rendering of satellite imagery onto a 3D representation of the Earth, terrain and street maps, and 360° panoramic views of hundreds of cities around the world. GSV in particular has several advantages that allow a quantitative study of the visual features of cities, including the availability of images in hundreds of cities in more than 80 countries, the use of similar photographic equipment everywhere, all images being georeferenced, and all images are available for download. As an example of the amount of visual urban data in GSV datasets, in New York City, there are approximately 100,000 sampling points: It sums up to approximately 600,000 images, since GSV captures six photographs at each sampling point. GSV and similar services have made available an unprecedented visual database of cities around the world with comparable characteristics.

Several researchers have been using GSV to analyze cities. Khosla, An, Lim et al. (2014) have analyzed 8 million GSV images from eight cities in different countries in order to compare how accurately humans and computers can predict crime rates and economic performance. Convolutional neural networks have been used by many researchers interested in measuring how physical features of cities affect different aspects of urban life, such as chronic diseases, the presence of crosswalks, building type, and vegetation coverage (Nguyen et al. 2018; Zhang al. 2019). GSV images have also used to quantify urban perception and safety (Dubey et al. 2016; Naik et al. 2014), to detect and count pedestrians (Yin et al. 2015), to infer landmarks in cities (Lander et al. 2017), and to quantify the connection between visual features and sense of place, based on perceptual indicators (Zhang et al. 2018).

Since 2015, the MIT Senseable City Lab has been using GSV to measure green canopy in cities. Xiaojiang Li pioneered this research with the Lab, using deep convolutional neural networks to quantify the amount of green areas at the street level. In this research initiative, called *Treepedia*, the focus is on the pedestrian exposure to trees and other green areas along the streets. Streets are the most active spaces in the city, where people see and feel the urban environment in their daily lives. Street-level images have a similar view angle with to pedestrians and can be used as proxies of physical appearance of streets as perceived by humans.

Li et al. (2015) and Seiferling et al. (2017) calculated the percentage of green vegetation in streets based on large GSV datasets. The process begins by creating sample sites, usually every 100 meters along the streets, and then collecting GSV metadata, static images, and panoramas. The basic technique involves the use of computer vision and DCNN to detect green pixels in each image. Once green pixels are detected, all the remaining part is subtracted, giving a general quantification of greenery. Thus, the percentage of the total green pixels from six images taken at each site to the total pixel numbers of the six images gives the Green View Index (Li et al. 2018).

Recent development in deep learning models allows us to improve the methodology to calculate the GVI. Initiated by Bill Cai (Cai et al. 2018), another researcher with the Senseable City Lab, the goal here is to quantify what is actually vegetation in GSV images, rather than using the ratio of green pixels as proxies to street-level greenery. The process begins by labeling images in a small-scale validation dataset. In this case, five cities with different climatic conditions were selected: Cambridge (Massachusetts, USA), Johannesburg (South Africa), Oslo (Norway), São Paulo (Brazil), and Singapore. One hundred images were randomly selected for each city, and vegetation was manually labeled. The DCNN model was then trained using the pixel-labeled Cityscapes dataset. Researchers also used a gradient-weighted class activation map (Grad-CAM) to interpret the features used by the model to identify vegetation. Results show that the DCNN models outperform the original Treepedia unsupervised segmentation model significantly, decreasing the mean absolute error from 10% to 4.7%.

The Treepedia Web site counts the Green View Index for 27 cities, and we have recently released an open-source Python library that allows anyone to calculate the GVI for a city where GSV images are available.

27.3 Thermals Images of the City

The richness of urban understanding that can be derived from video cameras is well known in urban studies. In groundbreaking research in the 1970s, William Whyte (2009) employed time-lapse cameras to understand people's behavior in public spaces and used this information to inform design. The negative reactions triggered by the deployment of cameras in public areas frequently happen due to a narrow understanding of their purposes (surveillance and policing) and poor analytical techniques, often based on officers watching footage (Luque-Ayala and Marvin 2015; Firmino and Duarte 2015).

In recent years, in research initiated by Amin Amjoneshoo, the MIT Senseable City Lab has been addressing these three problems related to the deployment of cameras in urban areas. We do this by widening the spectrum of urban phenomena that we can understand using cameras, developing image processing techniques that are novel to urban studies, and employing cameras that by design do not capture people's identity features. Here, we discuss the quantification of traffic-related heat loss and people's trajectories in space using cameras mounted on street lights, and the assessment of building heat loss using cameras deployed on vehicles.

Human activities generate heat. Cooling and heating systems and transportation, to stay with examples that are part of our daily lives, generate anthropogenic heat and release it into the ambient environment. They are major sources of low-grade energy that have direct and indirect impacts on human health. Cars alone, either powered by gasoline or diesel, release 65% of the heat produced by engines into the urban environment. In order to assess vehicular heat emissions at the street level, and match

such emissions to the number of pedestrians directly exposed, we have been using thermal cameras deployed in the existing infrastructures.

Thermal cameras capture wavelengths and measure the infrared radiation emitted from objects. They have a single channel, and thermal images have lower resolution, which makes thermal data much smaller in size, in comparison with RGB visual images. Smaller data size allows faster and better data transmission and processing, being less computational intensive. Thermal data only look like images when we apply the appropriate color maps.

The previous work has used thermal cameras to identify space occupancy and count people. Qi et al. (2016) proposed the use of thermal images as a sparse representation for pedestrian detection. Gade et al. (2016) developed a system to automatically detect and quantify people in sport arenas, by counting pixel differences between two successive frames. Interestingly, they also showed that based on the movements captured by thermal cameras, they were able to differentiate the sport modality people are playing, based on the position, concentration, and trajectories of people in space.

We deployed FLIR Lepton micro thermal cameras on street lights next to MIT, in Cambridge, MA, with the goals of quantifying traffic-related heat loss and tracking pedestrian movements.

Internal combustion vehicles are one of the major sources of heat in cities. Based on the analysis of thermal images captured at this high-traffic intersection, we were able to quantify and visualize both heat intensity and traffic load. Thermal cameras showed another advantage in relation to RGB cameras: Besides the counting of vehicles and simple identification (motorcycles, cars, trucks, buses), thermal images also allowed us to measure whether the vehicle had been running for a short or long period before being scanned (Anjomshoaa et al. 2016). This analysis generated a thermal fingerprint of traffic flow at the intersection.

For the analysis of the thermal images, we propose a method based on accumulated Radon Transform, which computes the projection of images along various angles. The Radon Transform of thermal images reveals the warmer objects and at the same time preserves their locations. We used the same dataset to count pedestrians passing on the sidewalk near traffic. In order to optimize data transmission and processing, we limited the target area to a sidewalk segment next to the pedestrian crossing. It also helped us to eliminate the high thermal flux of cars, which would otherwise make detecting pedestrian thermal flux harder. With this research, we were able to study the exposure of pedestrians to various anthropogenic pollutants caused by internal combustion vehicles. Also, by detecting thermal peaks, we were able to differentiate between single individuals and groups of individuals; and by learning from many hours of image analysis and the varying amplitude of the peaks, we were able to estimate the number of people in the scene.

In the project called City Scanner, the Lab has been developing a drive-by solution in which we mount a modular sensing platform on ordinary urban vehicles—such as school buses and taxis—to scan the city. The advantage of this approach is that it does not require specially equipped vehicles, since our modular sensing platform

can be deployed virtually on any vehicle. To prove this concept, in Cambridge, MA, we deployed the sensing platform on trash trucks (Anjomshoaa et al. 2018).

Among the sensors scanning the city for a period of eight months, we had two thermal cameras capturing data from the two sides of streets. These were non-radiometric thermal cameras, in which case the thermal output is not the scene temperature, but only a display of temperature fields. Scanning the thermal signature of all street segments of the city over different seasons, we created a thermal signature of the built environment in Cambridge. With these data and continuous scanning, any anomaly in the thermal difference between neighboring buildings might trigger a detailed analysis by city officials. In the case of Cambridge, a city that has programs to help residents to improve house insulation, this constant scanning can help the public authorities to be responsive when heat leaks are detected.

27.4 Navigating Urban Spaces Using Computer Vision

The explosion of big visual data is offering new sources of data that can overcome spatial and resource constraints that are common in studies of perception and legibility of urban spaces. At the Senseable City Lab, we have been using computer vision and deep convolutional neural networks to understand how people perceive, locate themselves, and navigate spaces.

As we have explained elsewhere (Wang et al. 2019), DCNN is based on probabilistic program induction, achieved by a bank of filters whose weights are adjusted during the training phase, with the goal of obtaining the key features of the images and, more importantly, the interplay of these features.

Here, we are particularly interested in addressing the problem of indoor navigability in large public areas. It is a well-known problem that users often have difficulty in navigating areas such as shopping malls, university campuses, and train stations, due to either their labyrinthic design or to the repetitiveness of visual cues.

In order to address this challenge, we have collected hundreds of thousands of images in two space types: university campuses and train stations. We trained a deep convolutional neural network to measure the easiness to navigate these spaces, and in the case of the train stations, we compared the results with a survey of users.

We first decided to test navigability on the MIT campus—in particular in a quite bland and disorienting space: the so-called infinite corridor, the interconnected indoors corridors and atriums that links several MIT buildings. The goal was to test DCNN to recognize different locations based on spatial features. Led by Fan Zhang (Zhang, Duarte, Ma et al. 2016), the study was based on 600,000 images extracted from video footage which we took using a GoPro camera for the training dataset, and 1,697 images taken with a smartphone for the test dataset. We compared our model with two commonly used in DCNN, and regarding the location in space, we achieved 96.90% top-1 accuracy on the validation dataset—higher than the other available models. We also proposed an evaluation method to assess how distinctive an indoor place is, when compared with all other spaces in the study area, and

produced a distinctiveness map of buildings on the MIT campus, which might help to explain how people find their way (or get lost) in the infinite corridors of MIT.

Another indoor public space that might be disorienting is the train station (Wang, Liang, Duarte et al. 2019). In this research, we measured space legibility in two train stations in Paris: Gare de Lyon and Gare St. Lazare, each receiving more than 250,000 passengers daily. Legibility influences the ability of people to locate themselves and find their way—or navigate space (Herzog and Leverich 2003). We developed a device composed of a LiDAR sensor and a 360 camera. After the projection transformation, we cropped hundreds of thousands of images from panoramic images from each station to train our DCNN.

In our DCNN, we have removed the final labeling part of the neural network, because our goal was not to identify what objects are present in each image, but to understand how visual properties are used to navigate space based on visual similarities. For Gare de Lyon, we tested the model on 88,869 images and achieved 97.11% prediction accuracy of its top-1 choice, and 97.23% for Gare St. Lazare.

Although the model performed very well (more than 97% top-1 accuracy) overall, we noticed discrepancies in accuracy among different spaces in different floors and related to different uses, which could reflect different spatial legibility. Research using computer vision frequently employs surveys to test results. On one setting, in their study to compare how accurately humans and computers can predict the existence of nearby establishments, crime rates, and economic performance of urban areas, Khosla et al. (2014) used Amazon Mechanical Turk and asked participants to guess where are some establishments; on another setting, they trained the computer to recognize five visual features of the images. Their results show humans and computers with similar performance.

Thus, to prove the validity of our model, we deployed a survey on Amazon Mechanical Turk, collecting 4,015 samples. The human samples showed a similar behavior pattern and mechanism as the DCNN models. A 10-second video was shown to all participants on a Web-based survey. On the next page, we displayed one image snippet from the spatial segment shown in the video, in addition to three images (one from the same scene). From these three images, participants were asked to choose one that matched the same scene and were asked to point out three features that helped them to make the decision. We compared these results with the activation layer, which is the fully connected layer of the DCNN model. We created heatmaps of the main features used by the model and by humans to read spaces. Although in several situations both have focused on the same areas, discrepancies are also important: One example is that participants often used objects, such as TV screens or advertisement boards, to help recognize spaces and locate themselves—indicating that semantic values play an important role in spatial legibility, in addition to spatial features and visual cues. More importantly, the research showed that computer vision techniques can help us to understand space legibility even closer to how humans read space. Since the deployment of cameras is more easily reproducible than doing surveys, computer vision and DCNN are opening new avenues in the study of space legibility that can inform wayfinding and space design.

27.5 Conclusion

In this chapter, we discussed three initiatives by the Senseable City Lab, in which we proposed special devices, designed experiments, and developed machine learning methods to analyze visual urban data. Either by taking advantage of urban imagery available online or by collecting RGB and thermal images in urban areas, the goal is to demonstrate how these multiple images can help us to reveal different aspects of the city. It is only by creating novel approaches to understand the visual data generated in cities that we will be able to understand contemporary urban phenomena and inform design in innovative ways.

The abundance of images certainly raises several problems, mainly regarding individual privacy—and this topic must be taken seriously. However, we should raise other questions regarding ownership and proper use of images collected in urban areas. For example, plenty of breakthrough research has been done in the fields of urban design, computer science, and sociology, using the urban scenes available online in platforms such as Google Street View. This was done with the tacit understanding that a private company was taking pictures of public spaces and making them available for non-commercial use—including scientific research. It was almost a trade-off: We allow Google to put online images of the façades of our houses, our backyards, and our cars when parked on the streets, and, in exchange, we could use these images for the common good of deepening our understanding of cities. Recently, Google changed its rules and now forbids almost any use of Google Street View images, including for academic purposes. Thus, should we accept quietly that a private company can take millions of images of public spaces and make money out of it? And even of our private properties? The question of privacy is essential in an era of overabundance of images; but, likewise, is the question of allowing private companies to profit from common goods—and the cities are the essential common good of the modern age.

Another important aspect of the future of urban ambient sensing is that sensors will be increasingly embedded in our buildings and carried by people in different formats. In this chapter, we discussed research based on the collection of passive data from our cities: images. More and more, construction materials have sensors as their components, sensors that not only feel the environment, but also react to it. Fully transparent glass panels embedded with photovoltaic cells measure the amount of light, change the opacity to adjust to the luminosity set by the users, and, at the same time, generate energy. On the personal side, if we currently carry sensors in our cellphones, these sensors are also becoming the constituent material of our clothes, for instance. They measure the body temperature, the ambient temperature, and adjust the clothing to our optimal comfort. At the same time that glass panels or clothing are sensing and actuating at the individual level with building or user, they are also generating data that can help us to better understand the relations established between people, the built environment, and nature. Exploring new methods to understand these relations is the key to foster innovative urban design.

References

- Anjomshoaa A, Duarte F, Rennings D, Matarazzo TJ, Desouza P, Ratti C (2018) City Scanner: building and scheduling a mobile sensing platform for smart city services. *IEEE Internet of Things Journal* 5:4567–4579. <https://doi.org/10.1109/jiot.2018.2839058>
- Anjomshoaa A, Duarte F, Alvarez R, Britter R, Ratti C (2016) Quantifying the anthropogenic heat in urban areas using thermal images. 2016 International conference on computational science and computational intelligence (CSCI). <https://doi.org/10.1109/csci.2016.0128>
- Cai BY, Li X, Seiferling I, Ratti C (2018) Treepedia 2.0: applying deep learning for large-scale quantification of urban tree cover. 2018 IEEE international congress on big data (BigData Congress). <https://doi.org/10.1109/bigdatacongress.2018.00014>
- Dubey A, Naik N, Parikh D, Raskar R, Hidalgo CA (2016) Deep learning the city: quantifying urban perception at a global scale. *Computer Vision—ECCV 2016 Lecture Notes in Computer Science* 196–212. https://doi.org/10.1007/978-3-319-46448-0_12
- Firmino R, Duarte F (2015) Private video monitoring of public spaces: The construction of new invisible territories. *Urban Stud* 53:741–754. <https://doi.org/10.1177/0042098014567064>
- Gade R, Moeslund TB, Nielsen SZ, Petersen HS, Andersen HJ, Basselbjerg K, Dam HT, Jensen OB, Jørgensen A, Lahrmann H, Madsen TKO, Bala ES, Povey BØ (2016) Thermal imaging systems for real-time applications in smart cities. *Int J Comput Applications in Technology* 53:291. <https://doi.org/10.1504/ijcat.2016.076790>
- Herzog TR, Leverich OL (2003) Searching for legibility. *Environ Behav* 35:459–477. <https://doi.org/10.1177/0013916503035004001>
- Khosla A, An B, Lim JJ, Torralba A (2014) Looking beyond the visible scene. 2014 IEEE conference on computer vision and pattern recognition. <https://doi.org/10.1109/cvpr.2014.474>
- Kitchin R (2016) The ethics of smart cities and urban science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374:20160115. <https://doi.org/10.1098/rsta.2016.0115>
- Lander C, Wiehr F, Herbig N, Krüger A, Löchtfeld M (2017) Inferring landmarks for pedestrian navigation from mobile eye-tracking data and Google Street View. In: *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems, CHI EA '17*. <https://doi.org/10.1145/3027063.3053201>
- Li X, Cai BY, Ratti C (2018) Using street-level images and deep learning for urban landscape studies. *Landscape Architect Frontiers* 6:20. <https://doi.org/10.15302/j-laf-20180203>
- Li X, Zhang C, Li W, Ricard R, Meng Q, Zhang W (2015) Assessing street-level urban greenery using Google street view and a modified green view index. *Urban For Urban Greening* 14(3):675–685. <https://doi.org/10.1016/j.ufug.2015.06.006>
- Luque-Ayala A, Marvin S (2015) The maintenance of urban circulation: an operational logic of infrastructural control. *Environ Plan D: Soc Space* 34:191–208. <https://doi.org/10.1177/0263775815611422>
- Naik N, Philipoom J, Raskar R, Hidalgo C (2014) Streetscore—predicting the perceived safety of one million streetscapes. 2014 IEEE conference on computer vision and pattern recognition workshops. <https://doi.org/10.1109/cvprw.2014.121>
- Nguyen QC, Sajjadi M, McCullough M, Pham M, Nguyen TT, Yu W, Meng H-W, Wen M, Li F, Smith KR, Brunisholz K, Tasdizen T (2018) Neighbourhood looking glass: 360° automated characterisation of the built environment for neighbourhood effects research. *J Epidemiol Commun Health* 72:260–266. <https://doi.org/10.1136/jech-2017-209456>
- Paldino S, Kondor D, Bojic I, Sobolevsky S, González MC, Ratti C (2016) Uncovering urban temporal patterns from geo-tagged photography. *PLOS One*. <https://doi.org/10.1371/journal.pone.0165753>
- Qi B, John V, Liu Z, Mita S (2016) Pedestrian detection from thermal images: A sparse representation-based approach. *Infrared Phys Technol* 76:157–167. <https://doi.org/10.1016/j.infrared.2016.02.004>

- Saiz A, Salazar A, Bernard J (2018) Crowdsourcing architectural beauty: Online photo frequency predicts building aesthetic ratings. PLOS One. <https://doi.org/10.1371/journal.pone.0194369>
- Seiferling I, Naik N, Ratti C, Proulx R (2017) Green streets – Quantifying and mapping urban trees with street-level imagery and computer vision. Landscape Urban Plan 165:93–101. <https://doi.org/10.1016/j.landurbplan.2017.05.010>
- Shao Z, Cai J, Wang Z (2018) Smart monitoring cameras driven intelligent processing to big surveillance video data. IEEE Trans Big Data 4:105–116. <https://doi.org/10.1109/tbdata.2017.2715815>
- Vanolo A (2016) Is there anybody out there? the place and role of citizens in tomorrow's smart cities. Futures 82:26–36. <https://doi.org/10.1016/j.futures.2016.05.010>
- Wang Z, Liang Q, Khosla F, Zhang F, Charron L, Johnsen L, Cai B, Ratti C (2019) Quantifying legibility of indoor spaces using deep convolutional neural networks: case studies in train stations. Build Environ 160:106099. <https://doi.org/10.1016/j.buildenv.2019.04.035>
- Whyte WH, Underhill P (2009) City: rediscovering the center. University of Pennsylvania Press
- Xu Z, Mei L, Liu Y, Hu C, Chen L (2014) Semantic enhanced cloud environment for surveillance data management using video structural description. Computing 98:35–54. <https://doi.org/10.1007/s00607-014-0408-7>
- Yin L, Cheng Q, Wang Z, Shao Z (2015) 'Big data' for pedestrian volume: exploring the use of Google Street View images for pedestrian counts. Appl Geogr 63:337–345. <https://doi.org/10.1016/j.apgeog.2015.07.010>
- Zhang F, Zhou B, Ratti C, Liu Y (2019) Discovering place-informative scenes and objects using social media photos. Roy Soc Open Sci 6:181375. <https://doi.org/10.1098/rsos.181375>
- Zhang F, Zhou B, Liu L, Liu Y, Fung HH, Lin H, Ratti C (2018) Measuring human perceptions of a large-scale urban region using machine learning. Landscape Urban Plan 180:148–160. <https://doi.org/10.1016/j.landurbplan.2018.08.020>
- Zhang F, Duarte F, Ma R, Milioris D, Lin H, Ratti C (2016) Indoor space recognition using deep convolutional neural network: a case study at MIT campus. [arXiv:1610.02414](https://arxiv.org/abs/1610.02414)
- Zhang W, Xu L, Duan P, Gong W, Lu Q, Yang S (2015) A video cloud platform combining online and offline cloud computing technologies. Pers Ubiquit Comput 19:1099–1110. <https://doi.org/10.1007/s00779-015-0879-3>

Fábio Duarte is a Principal Research Scientist in the MIT Senseable City Lab and lecturer in the MIT Department of Urban Studies and Planning. Duarte is also professor at PUCPR, Brazil, and author of “Unplugging the city: the urban phenomenon and its sociotechnical controversies” (Routledge 2018)

Carlo Ratti is Director of the MIT Senseable City Lab and Professor of Practice in the MIT Department of Urban Studies and Planning. Ratti is the founder of Carlo Ratti Associati and author of “The city of tomorrow: sensors, networks, hackers, and the future of urban life” (Yale 2016).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.