Chapter 4

Street View Imaging for Automated Assessments of Urban Infrastructure and Services

Daniel Zünd and Luís M. A. Bettencourt

Abstract Many forms of ambient data in cities are starting to become available that allows tracking of short-term urban operations, such as traffic management, trash collections, inspections, or non-emergency maintenance requests. However, arguably the greatest promise of urban analytics is to set up measurable objectives and track progress toward systemic development goals connected to human development and sustainability over the longer term. The challenge for such an approach is the connection between new technological capabilities, such as sensing and machine learning and local knowledge, and operations of residents and city governments. Here, we describe an emerging project for the long-term monitoring of sustainable development in fast-growing towns in the Galapagos Islands through the convergence of these methods. We demonstrate how collaborative mapping and the capture of 360-degree street views can produce a general basis for a broad set of quantitative analytics, when such actions are coupled to mapping and deep-learning characterizations of urban environments. We map and assess the precision of urban assets via automatic object classification and characterize their abundance and spatial heterogeneity. We also discuss how these methods, as they continue to improve, can provide the means to perform an ambient census of urban assets (buildings, vehicles, services) and environmental conditions.

4.1 Introduction

Many forms of ambient data in cities are starting to allow tracking of short-term operations and services (Park et al. 2014; Townsend 2015). Uses of these technologies range from facilitating traffic management to air quality control, or the management of non-emergency requests (Park et al. 2014; O'Brien 2015). However, arguably one

D. Zünd (✉) · L. M. A. Bettencourt

Mansueto Institute for Urban Innovation and Ecology and Evolution, The University of Chicago, Chicago, USA

e-mail: dzuend@uchicago.edu

L. M. A. Bettencourt

e-mail: bettencourt@uchicago.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_4

of the greatest promises of urban analytics is to set up measurable objectives and track progress toward systemic development goals connected to human development and sustainability over the longer term (Brelsford et al. 2017). A main challenge to achieving long-term monitoring of processes in urban settings is the convergence of new technology, local knowledge, and the operations of residents and local governance. Whereas these objectives already constitute challenges for developed cities, they are even more daunting in developing country settings (Praharaj et al. 2017). In rapidly developing cities, data are often far less abundant or even non-existent. Additionally, urban environments often change at a much faster pace and in informal ways (Sarin 2016). This makes it much more difficult to track change, and specifically, to generate statistical progress in development trajectories toward sustainable development goals (Randhawa and Kuma 2015; Komninos 2015).

A good case study to research the potential of new technology in semi-informal settings, and the impact it has on managing and tracking the progress of long-term goals, are the Galapagos Islands. The archipelago, famous for its unique ecosystems, lies about 1000 km off the Pacific coast of Ecuador (the blue square in Fig. 4.1). Though most of the islands remain a natural reserve, the human presence on land and sea is growing very quickly, with four fast-growing towns concentrating most of the immigrant human population. The remote location and the unique coupled urban–natural system of these islands constitute a particularly interesting and poignant setting to study the development trajectories of urbanization (Batty et al. 2019).

Fig. 4.1 The Galapagos Islands are an archipelago in the midst of the Pacific Ocean (blue square). Their secluded location, fast-growing towns, and unique ecosystems offer a particularly interesting and poignant setting for developing models of sustainable development for coupled urban–natural systems. The manageable size of these urban areas makes it possible to study novel methods of collaborative data collection and the convergence of new technology and local knowledge. We exemplify the method on the capital of the islands, Puerto Baquerizo Moreno on San Cristóbal, depicted in the inset. Map designs are from Mapillary (2019) and OpenStreetMap (2019)

From a modeling perspective, the islands provide a unique setting due to their remote location, and the fact that all materials and goods in and out the system are registered upon arrival or departure, just as are people's migration (Bettencourt 2019), provides a good basis for assessing the impact of the island system on its external environment and vice versa.

Together with the emergence of a plan to harmonize tourism with sustainable stewardship of the local charismatic ecosystem (Rousseaud et al. 2017), the towns in the Galapagos Islands provide a unique chance to study novel approaches to urban planning, urban management of resource flows, and tracking of development toward sustainability goals (Batty et al. 2019).

We will focus in this study on the second largest town in the Galapagos, Puerto Baquerizo Moreno, which is also the regional capital and has a population of about eight thousand residents (Andrade and Ferri 2019). The town is located on the eastern part of the Archipelago, on the island of San Cristóbal, as depicted in Fig. 4.1. In terms of materials, the island is relatively independent of the other islands in the archipelago since it has its own harbor and airport that directly connect it to continental Ecuador where most people, construction materials, energy, and consumer goods originate.

Historically, the island of San Cristóbal has not been the archipelago's main tourist hotspot. However, since the airport opened in 1986, the island is increasingly attractive to a growing number of tourists—as can be seen by the number of arrivals at the airport—which shows a higher growth rate than the total growth rate of tourist arrivals across the Galapagos Islands (Izurieta 2017). The annual increase of 3.72% in tourism (about 225 thousand visitors in 2015; Izurieta 2017) creates a growing economy on the islands, but also places pressure on the urban–natural interfaces of the islands. These pressures and possible solutions remain hard to track in detail, therefore precluding a balanced path where economic opportunities may be expanded, while ecosystems in the islands are protected.

Thus, innovative approaches that track the growth and effects of urbanization on the islands are becoming paramount. Here, we exemplify how collaborative data collection and new imaging and artificial intelligence technology can support this process in the context of an emerging project for long-term sustainable development of the Galapagos Islands.

4.2 Data Collection and Object Localization

The rapid development of computer vision and object recognition has opened up efficient ways to process large image datasets (Chen et al. 2016). For urban science and policy, these capabilities have great potential to follow the trajectory of the built infrastructure and to assess the heterogeneity of urban assets and services, including the consumption of energy and materials. However, data about these issues are often lacking, outdated, or too coarse in many developing urban areas. This is even more so the case for remote locations, such as the towns in the Galapagos Islands and specifically, the town of Puerto Baquerizo Moreno. Before we started the project of

monitoring the town's built environment, very few data were available online (about a dozen images) of which only a few depicted the island's urban areas.

Monitoring the urban development, however, asks for data that capture the urban fabric as a whole and over time. In the following, we introduce a method that makes it possible to document the whole town within only a few days' work and with only minimal initial investments, thus making collaborative data collection possible. The data pipeline consists of three main steps, of which two are fully automated. The first involves capturing street-level photographs, and the second analyzes single images in order to recognize and segment objects, as depicted on the right panel of Fig. 4.2. The third step consists of identifying the same object in different images and geolocating its position in space and time.

The most time-consuming step is the collection of enough imagery to cover the whole town. The process is entirely parallelizable and can involve a group of people or vehicles. There must be enough overlap in the images so that the geolocation of objects is possible and thus becomes unambiguous. Figure 4.3 depicts an example where a store sign was recognized in six different images.

In this study, we used a 360-degree action camera able to automatically take images with a chosen temporal frequency. The camera is capable of taking images

Fig. 4.2 Street-level imagery can be captured with relatively simple tools. For this study, we collected data by attaching a 360-degree sports camera on a helmet and rode a bicycle through the town. The imagery is available through Mapillary's (2019) user interface, as depicted on the left panel. The right panel shows processed and segmented imagery. The automatic object classification identifies structures and objects out of almost three-dozen categories. However, on the island, the algorithms sometimes fail to properly identify certain objects. For example, the sidewalk on the right is classified as ground. Nevertheless, the methods provide a powerful tool to assess urban features in developing towns experiencing rapid change

Fig. 4.3 The imagery covers most of the accessible street network of Puerto Baquerizo Moreno on San Cristóbal, Galapagos. The green dots show the locations of all 360-degree imagery produced by us. When a series of images are available along a street, objects can be identified and geolocated. The inset depicts a situation in which the same store sign is recognized in six different images in the right inset panel, taken from slightly different locations, of which three are shown in the left inset panel. Map designs are from Mapillary (2019) and OpenStreetMap (2019)

that cover the whole surrounding from the current location which, with some post-processing, produced globes at each location. We attached the camera to a helmet and drove around the town with it. Since the camera also added the GPS coordinates to each image's metadata, we were able to cover about 75 km of geotagged image globes within only a couple of days. The collected imagery accounts for more than 10,000 images, of which many overlap and provide a good dataset for the next steps in the data pipeline. Each location of a 360-degree image is depicted by a trace of green dots in Fig. 4.3.

We executed steps two and three in collaboration with Mapillary (2019), a technology company dedicated to creating crowdsourced street view maps. Mapillary provides an engine that automatically processes uploaded images, including a user interface to walk from one image to the next and, thus, ultimately throughout the entire city. The left side of Fig. 4.2 depicts the interface that is accessible to the public. The images are further processed using computer vision and object recognition algorithms, of which many have been developed and optimized by the Mapillary research teams (Bulo and Kontschieder 2016; Bulo et al. 2017; Cariucci et al. 2017; Neuhold et al. 2017). The algorithms segment the images and add semantic information to different parts of the visual field.

The field of computer vision and object recognition has made significant strides in recent years by using deep-learning algorithms to perform image segmentation (Krylov et al. 2018). However, these techniques are not yet perfect and the resulting semantic information extracted from images is often only an approximation to reality. For street-level data, this is especially the case for areas that differ from the data that were used to train the object recognition classifier. Nevertheless, the algorithms are able to recognize core properties in the imagery, as depicted in the right inset panel of Fig. 4.2.

When the same object is recognized in several images, it can be geolocated uniquely in space. Figure 4.3 shows an example where a single store sign is recognized in six different images located in the right inset, three of which are shown in the left inset panel. The task of geolocating objects from different images at street level involves several major technical challenges. Besides aggregating the same object present in several images, the main challenge in processing crowdsourced street-level data is the varying qualities of the imagery, such as blurring or restricted field of view, and variability in camera positions. The latter is important, since high-quality geolocation depends on the camera position relative to the object in the field of view for accurate triangulation and location (Krylov and Dahyot 2018).

Despite these challenges, the engine was able to geolocate almost 12,000 objects in the small town of Puerto Baquerizo Moreno, including 777 trash cans, 343 store signs, 412 advertisement signs, and 224 driveways. These are the classes of objects that we use in the next section to derive the functions of certain parts of the town and to exemplify the conclusions that can be drawn from these methods, as they continue to improve.

4.3 Deriving Urban Functions from Object Statistics

The collection of data and the identification and localization of objects in space provides a basic functional mapping of an urban area. The spatial distribution of different classes of objects makes it possible to study the location and functions of different districts. For example, the density distribution of store signs in Fig. 4.4b shows the areas in Puerto Baquerizo Moreno that provides a range of specific services, typically associated with tourism (Andrade and Ferri 2019).

Figure 4.4 shows two object-class density distributions that are good indicators of residential areas: the distributions of trash cans and driveways (subfigures (a) and (c)). Trash cans in residential areas of Puerto Baquerizo Moreno are standardized vessels with a unique shape and color combination. Each household is required to have their trash cans outside of the building, close to the street for easy access for trash collectors. They additionally serve as public trash bins. The trash bins in tourist areas are different, not as prominently placed, and often obfuscated. The segmentation engine has problems identifying them as such, but this is also a clear sign of a different look and function and of an intentional effort to deal with the issue differently. The waterfront area with the most tourist services is much denser than

Fig. 4.4 Geolocated objects help to identify and locate different properties of the town. The figures depict the distribution of **a** trash cans, **b** store signs, **c** driveways, and **d** advertisement signs. The distribution of the trash cans shows the importance of local knowledge. The ones identified by the segmentation are private trash cans, whereas the public ones are not recognized and are largely in the business parts of town, close to the sea and indicated by a high volume of shop signs in **b**. The driveways in **c** indicate a lower density of houses in those areas, since they are set back from the street. The advertisement signs in **d** have a similar pattern as the store signs in **b**, but are more uniformly distributed, mainly along principal roads. Map designs are from Stamen Design (2019)

the rest of the town. The buildings are often located next to the street and not set back. This is indicated by the abundance of driveways in the residential area in the northeast and their absence in the denser locations, such as the area central of the town toward the sea. Figure 4.4c depicts this clearly.

The last indicator we want to point out in this study is the distribution of advertisement signs. Their spatial distribution is depicted in Fig. 4.4d. According to the density distributions of advertisement signs, there are three main patterns specific to places with a large accumulation of advertising signs. The first pattern is where most tourists spend their time within the town and also where most restaurants and tourist services are located, corresponding to the highest density of store signs in Fig. 4.4b.

The second area with a high density of advertisements consists of the main thoroughfares that cut through the town from east to west, each a one-way street. Within the town, these are the streets where most shops frequented by locals are located. The main street also connects further to the only other settlement on the island and is the only street that cuts through the San Cristóbal from east to west. This road constitutes the main axis in the town, together with the street that is orthogonal to it and starts at the airport on the left of the map. However, these signals are not as clear as for other indicators.

The third cluster, the one with the highest density of advertising signs according to the data, is located at the international convention center close to the center top of the image. This cluster has to be regarded with care, because many of our data collection trips started here, so that the region is oversampled in terms of imagery. The data-processing engine has some difficulties to cope with this sampling effect, separates advertisement signs that are the same, and geolocates them in very similar locations.

The above interpretations of the different density distributions in Fig. 4.4 are clearly highly reliant on local knowledge. For example, the unique form and shape of the private trash cans are not a general pattern across different urban systems, but a very local feature. There would not have been an obvious conclusion from the extracted data without knowledge of local choices, habits, and rules.

4.4 Discussion

Recent technological advancements are paving the way to novel ways of monitoring, studying, and assessing characteristics and change in urban environments that are closer to the human experience. Our present study shows how collecting street view imagery and identifying and locating associated functional objects require little initial investment. These methods are also suitable for collaborative approaches involving both image collection and interpretation of resulting spatial statistics. Thus, this type of result demonstrates that concepts of smart cities and the collection of extensive and detailed ambient urban data are no longer restricted to large investments and efforts by large corporations or universities, but are also feasible in developing towns by relatively small numbers of people.

It is desirable that local citizens take a greater part in this type of process for a number of different reasons. First, on purely technical grounds, an ongoing data collection effort helps improve the system's evidence pool in terms of coverage and accuracy of object identification statistics. Second, local knowledge is critical

for good urban planning and policy, and there have been thus far few systematic strategies that combine data and technology with people's local experiences. Third, and most important, data collections by corporations and governments rarely speak to the perspective and priorities of local communities, who, in the case of sustainable development, have a clear stake in the future of their environment and can act as the best stewards of its well-being (Burke et al. 2006). Fourth, the use of methods such as the ones discussed here provides a number of interesting educational and training opportunities that can contribute to the growth of local human capital and may have spillovers to other innovative local practices.

There are still a number of technical obstacles for turning the pilot described here into an effective system that can speak to these objectives. Object recognition in images of developing cities is far from working perfectly. This is likely due to biases in training of the artificial intelligence algorithms with imagery from more formal environments, such as cities of the Global North. As a result, the present algorithms often fail to extract all semantic information from the images in the Galapagos and thus fail to achieve high levels of accuracy in object recognition and segmentation. Nevertheless, the methods already offer powerful tools in their current state, so that we can reasonably expect that they will improve in the near future as more evidence from informal and variable environments becomes part of training corpora.

Aspects of algorithms that need improvement are likely related to increased knowledge of geographic and cultural contexts. We have seen for example that the recognition of sidewalks remains difficult as these rather irregular spaces are often classified as parts of the streets or simply as ground. Another example is the classification of beaches. In the data, we collected on the Galapagos Islands, sand beaches are often classified as snow. Simple contextual clues would certainly improve this type of classification.

Nevertheless, the methodology provides initial stages of potentially powerful artificial intelligence tools to assess the assets of cities and towns and to study the development trajectory of urban microenvironments. This will become even more powerful in the future, as the algorithms become capable of more fine-grained object classification and segmentation in a ways that can track, for example, construction processes and the materials and costs involved.

A big impact in future studies of urban areas will arise from extracting three-dimensional (3D) city models (Schläpfer et al. 2015) from the type of imagery produced and analyzed in this study. In combination with more traditional aerial and remote sensing (Qin and Fang 2014; Weng et al. 2018) and citizen engagement, high-quality 3D models of whole towns and cities are just now becoming accessible also in fast-changing settings in the developing world (see also Chap. 34). The simplicity and generalizability of data collection demonstrated here provide a way to easily and quickly track these development trajectories in ways that are closer to the experience of individuals and households living and working in these environments, and at the same time allow us to characterize material and information flows through these systems across scales.

References

- Andrade JEL, Ferri DQ (2019) The Galapagos urban context. In: Kvan T, Karakiewicz J (eds) *Urban Galapagos: transition to sustainability in complex adaptive systems*. Springer, Berlin, pp 9–22
- Batty M, Bettencourt LMA, Kirley M (2019) Understanding coupled urban natural dynamics as the key to sustainability: the example of the Galapagos. In: Kvan T, Karakiewicz J (eds) *Urban Galapagos: transition to sustainability in complex adaptive systems*. Springer, Berlin, pp 23–41
- Bettencourt LMA (2019) Critical paths to sustainability: the research challenge from island urban systems. In: Kvan T, Karakiewicz J (eds) *Urban Galapagos: transition to sustainability in complex adaptive systems*. Springer, Berlin, pp 155–160
- Brelsford C, Lobo J, Hand J, Bettencourt LM (2017) Heterogeneity and scale of sustainable development in cities. *Proc Natl Acad Sci* 114(34):8963–8968
- Bulo SR, Kotschieder P (2016) Online learning with Bayesian classification trees. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 3985–3993
- Bulo SR, Neuhold G, Kotschieder P (2017) Loss max-pooling for semantic image segmentation. In: *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, pp 7082–7091
- Burke JA, Estrin D, Hansen M, Parker AP, Ramanathan N, Reddy S, Srivastava MB (2006) Participatory sensing. Research Gate. https://www.researchgate.net/publication/228613934_Participatory_Sensing/link/544e3db20cf29473161a67f4/download
- Cariucci FM, Porzi L, Caputo B, Ricci E, Bulò SR (2017) Autodial: automatic domain alignment layers. In: *2017 IEEE international conference on computer vision (ICCV)*, pp 5077–5085
- Chen LC, Yang Y, Wang J, Xu W, Yuille AL (2016) Attention to scale: scale-aware semantic image segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 3640–3649
- Izurietta JC (2017) Behavior and trends in tourism in Galapagos between 2007 and 2015. In: *Galapagos report 2015–2016*. GNPD, GCREG, CDF and GC, 2017
- Komninos K (2015) Intelligent cities: variable geometries of spatial intelligence. In: *From intelligent to smart cities*. Routledge, London, pp 46–62
- Krylov VA, Dahyot R (2018) Object geolocation from crowdsourced street level imagery. In: *European conference on machine learning (ECML) workshops*
- Krylov V, Kenny K, Dahyot R (2018) Automatic discovery and geotagging of objects from street view imagery. *Remote Sens* 10(5):661
- Mapillary (2019) www.mapillary.com. Accessed 2019-03-05
- Neuhold G, Ollmann T, Bulo SR, Kotschieder P (2017) The Mapillary vistas dataset for semantic understanding of street scenes. In: *Proceedings of the IEEE international conference on computer vision*, pp 4990–4999
- O'Brien DT (2015) Custodians and custodianship in urban neighborhoods: a methodology using reports of public issues received by a city's 311 hotline. *Environ Behav* 47(3):304–327
- OpenStreetMap (2019) www.openstreetmap.org. Accessed 2019-03-05
- Park TH, Turner J, Musick M, Lee JH, Jacoby C, Mydlarz C, Salamon J (2014) Sensing urban soundscapes. In: *EDBT/ICDT workshops*. Citeseer, pp 375–382
- Praharaj S, Hoon Han J, Hawken S (2017) Innovative civic engagement and digital urban infrastructure: lessons from 100 smart cities mission in India. *Procedia Eng* 180:1423–1432
- Qin R, Fang W (2014) A hierarchical building detection method for very high resolution remotely sensed images combined with DSM using graph cut optimization. *Photogramm Eng Remote Sens* 80(9):873–883
- Randhawa A, Kuma A (2015) Exploring sustainability of smart development initiatives in India. *Int J Sustain Built Environ* 6(2):701–710

- Rousseaud A, Cruz E, Naula E, Ramos A, Granda M, Calvopiña M, León P, Sanchez D, Zapata F, Guerrero J, Babriel L, Falconi E (2017) Plan Galapagos: an instrument for the holistic sustainable development of the province. In: Galapagos report 2015–2016. GNPD, GCREG, CDF and GC
- Sarin G (2016) Developing smart cities using internet of things: an empirical study. In: 3rd international conference on computing for sustainable global development (INDIACom), pp 315–320
- Schläpfer M, Lee J, Bettencourt LMA (2015) Urban skylines: building heights and shapes as measures of city size. arXiv preprint [arXiv:1512.00946](https://arxiv.org/abs/1512.00946)
- Stamen Design (2019) maps.stamen.com. Accessed 2019-03-04
- Townsend A (2015) Cities of data: examining the new urban science. *Public Cult* 27(2):201–212
- Weng Q, Quattrochi D, Gamba PE (eds) (2018) Urban remote sensing. CRC Press, Boca Raton, FL

Daniel Zünd is a Postdoctoral Fellow at the Mansueto Institute for Urban Innovation and a Postdoctoral researcher in Ecology and Evolution at the University of Chicago. He holds a Ph.D. in Architecture and Urban Planning, as well as a Master's degree in Computer Science.

Luís M. A. Bettencourt is the Inaugural Director of the Mansueto Institute for Urban Innovation and Professor of Ecology and Evolution at the University of Chicago, as well as an External Professor of Complex Systems at the Santa Fe Institute.

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.