Making a New City Image … or, an Eye for AI

ESRI Development Center Student of the Year Application Abstract
March 15, 2018

Brian Ho
Master in Design Engineering
Harvard Graduate School of Design (GSD) and School of Engineering and Applied Sciences (SEAS)

Advisors Robert Pietrusko (GSD) and Krzysztof Gajos (SEAS)

Abstract
Making a New City Image seeks to define a new mode of geographic analysis, based on novel techniques for street-level sensing, computer vision and deep learning.

This mode of analysis attempts to address the challenge of the plan perspective in cities: the fact that the view from above does not match the view from the ground. Can we find a way to capture the street-level perspective, texture and detail of cities in a new kind of "map," which preserves the scale and scope of the view from above? Can we leverage new technology — computer vision, deep learning, sensor systems — to unite the view from above and the view from the ground?

Problem Domain
The plan view — the view from above — is the dominant method of understanding cities, but the inherent abstraction and simplification methods[1] required to produce it result in broader challenges of urban planning, design and management. This occurs in terms of both literal representation and symbolic process. Maps and geospatial data are limited in the kind of information they can legibly express about the urban built environment, which renders them ineffective for capturing street-level perspective, texture and details. Persons who rely on the view from above must willfully overlook a certain level of information fidelity. Similarly, centralized methods of gathering and analyzing data about cities must inherently ignore — to some degree — the multitude of perspectives of people who live in them.

At the same time, the view from the ground has largely been relegated to anecdotal or ethnographic approaches: methods of analysis which are inherently qualitative and limited in scale and scope by intensive requirements for individual labor and effort. Although the view from the ground provides a great deal of information about the minutiae of urban environments, its output is currently incompatible with the overarching perspective provided by the view from above.

Within this broader problem, I am also interested in the way in which technology mediates our view — whether from above or the ground — of cities. While my own research understands that technology has always had a historical role to play in shaping human perception of the built environment, from primitive maps[2] to failed mid-century wayfinding machines[3] to today's Google Maps[4], the contemporary acceleration of technology make machine-mediated perception more present — and the need for understanding it more urgent. I argue that the tools, instruments and systems we use to navigate and understand cities have significantly altered our perception of cities. If we are to challenge the plan perspective, we will need to rethink the machinery which creates it.

Theoretical Background
Urban planning and urban interventions struggle with the problem of applying an idealized design vision to a more complex setting. In our present moment of obsession with urban technology, computation and data we face a specific evolution of the Modernist problem. Today, the diagram is an apparatus — the digital services, sensors systems, predictive algorithms, and cumulatively produces "big data" — focused on a new site for planning: the smart city.

---

[1] Scott, James C. *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale University Press, 1998.
[2] Vass, F. "The map of Nippur." *Cartography* 9.3 (1976): 168-174.
[3] "SUBWAY AID INSTALLED; Direction Machine Is Set Up in IRT's Penn Station Stop." *New York Times*, 5 Feb. 1957.
[4] Mattern, Shannon. "Mapping's Intelligent Agents." *Places Journal* (2017).

The scope of this apparatus exceeds previous measurement of the urban. There are a series of physical systems: magnetic coils embedded in the street, pole-mounted air quality sensors, sidewalk kiosks and building cameras. There are also countless devices that output useful data: cell phones, traffic signals, utility meters, public WiFi access points, elevators or automatic doors. And lastly, there are the data without hardware: the electronic traces produced by Uber, Airbnb, Waze and Yelp, or digitized versions of existing US Census or local property assessor surveys.

Critics of the smart city point out that a great distance exists between urban data and the subject they purport to represent.[5] There are also gaps in the data itself: underserved neighborhoods and populations that do not receive the high-investment infrastructure needed to implement the smart city apparatus. A smart city only measures that which it thinks to measure; for every parameter counted, many other metrics and elements of the urban go unnoticed. Models and algorithms reflect the assumptions of their makers, and their operations obscure the details and intricacy originally present on the ground.

There might yet be an alternative apparatus, capable of capturing the scale of the urban as well as making legible its constituent textures and elements. Its foundations can be found in Kevin Lynch's study of a city's image.

Lynch documented urban imageability, a quality of an environment's appearance that relates "identity, structure and meaning." In a five-year study funded by the Rockefeller Foundation and carried out in partnership with the designer Gyorgy Kepes, Lynch applied imageability as both research and cartographic process to reveal the city image of Boston, Jersey City and Los Angeles. Published as *The Image of the City*, this study relied on both "systematic field reconnaissance … by a trained observer" and "lengthy interview[s] … with a small sample of city residents," all supported by photographs, hand-drawn maps and lengthy transcripts.[6]

While Lynch's methods were highly subjective, they were grounded in a systematic effort to measure the urban environment. Lynch's identification of five elements of the city image — the path, edge, district, node and landmark — are an ambitious attempt to catalog highly perceptual information. Indeed, Lynch continued to develop them into a generalized theory of urban form.[7] The act of production of an organized but experiential dataset suggest modern equivalents. A new mode of geographic analysis, inspired by Lynch's original approach, might help blur the distinction between smart city center and its periphery.

Relevance and Stakeholders

My project has relevance for those disciplines traditionally involved in the act of understanding and mapping cities: planning, design, development, management and government. It also has relevance for newer companies engaging with the built environment through technology, making services and products that both produce and rely on machine-mediated perception of cities. I believe that better methods for generating perceptual or textural maps can provide missing information that enables qualitatively, not quantitatively, better decisions and designs.

Primary stakeholders are the users of these techniques. As a "platform" based on new kinds of urban data and methods for understanding it, my project provides stakeholders with actionable spatial intelligence. Planning and design disciplines could better understand urban context at scale (i.e. the extent and nature of architectural styles and streetscapes); management and government could assess city-wide street perception (i.e of accessibility, safety); new services and product providers could make valuable predictions (i.e. of neighborhoods of interest and potential commercial value).

Solution

My solution has three parts:

1. Instrumentation — machines that capture useful perceptual data, from devices that take 360 imagery to web interfaces that collect human opinions
2. Computation— deep learning using convolutional nerual nets applied to imagery to learn associations between the perception of the built environment and a desired set of organizational labels
3. Visualization and tool — interfaces that present the collected data and model output for productive use

[5] Mattern, S.C., 2015. *Deep Mapping the Media City*. University of Minnesota Press.
[6] Lynch, K., 1960. *The image of the city* (Vol. 11). MIT press.
[7] Lynch, K., 1984. *Good city form*. MIT press.

I have begun with a reconsideration of the work of Kevin Lynch[8] through a computational framework, a process which has combined modern day techniques with historical[9] and archival data.[10] As inputs, I have used about 2000 of Lynch's original photos of Boston, available online at Flickr[11] as part of MIT's Lynch-Kepes collection.[12] Each of these photos bears a description, but no precise geolocation information is available. I have conducted a manual geolocation process through a custom built web-based interface built in HTML, CSS and Javascript (http://www.new-city-image.io/label.html); an embedded Google Maps and Google StreetView window allow me to match an archival image to an existing view, giving a latitude and longitude as well as viewing orientation and angle.

The geolabeled Lynch images were then associated with Lynch's elements of the city image (www.new-city-image.io). This required using ESRI's ArcGIS Pro to warp a digital file of Lynch's Boston city image diagram, creating raster layer with a standard coordinate reference systems. The raster layer was traced in ArcGIS Pro to produced custom polygon vector shapefiles, one for each element of the city image: paths, edges, nodes, landmark and districts. Finally, the geolabeld images were converted into a standard CSV data file with appropriate columns for latitude and longitude; the data was XY geolocated into a point shapefile. These points were then spatially joined to the city image element layers — effectively labeling each image with its relevant category.

A convolutional neural net, written in the Tensorflow and Keras Python libraries, was used to learn the relationship between imagery and element label (https://github.com/brian-ho/IDEP) . The imagery was augmented by joining it to a plan view of the corresponding area, created from raster tile imagery from a custom Mapbox API style. While the neural net performance can still be improved, existing trials to date show accuracy rates of up to 80% for certain elements (i.e. landmarks and districts).

An updated version of this process will utilize my own collection of imagery data. To this end, I have created a bike rig for a 360 camera (a Ricoh Theta V), 3D-printed on a Formlabs Form 2 printer. This device allows me to quickly collect a large amount of data individually. I have carried out preliminary data collection and will continue to make trips to photograph Boston. Additional work will use a Raspberry Pi to access the camera's built-in WLAN server and API to automate image collection and record GPS data.

---

[8] Lynch, Kevin. *The Image of the City*. Vol. 11. MIT press, 1960.
[9] Kevin Lynch papers, MC 208. Massachusetts Institute of Technology Institute Archives and Special Collections, Cambridge, Massachusetts.
[10] https://www.flickr.com/photos/mit-libraries/sets/72157614966285159/
[11] https://www.flickr.com/photos/mit-libraries/sets/72157614966285159/
[12] https://libraries.mit.edu/archives/exhibits/kepes-lynch/index.html