American Economic Review: Papers & Proceedings 2016, 106(5): 128–132
http://dx.doi.org/10.1257/aer.p20161030

Cities Are Physical Too: Using Computer Vision to Measure the Quality and Impact of Urban Appearance[†]

By Nikhil Naik, Ramesh Raskar, and César A. Hidalgo*

In January 2016, Britain's Prime Minister David Cameron, wrote a piece in the *Sunday Times* pledging to bulldoze 100 of the United Kingdom's "bleak" postwar housing estates:

> *The riots of 2011 didn't emerge from within terraced streets or low-rise apartment buildings. As spatial analysis of the riots has shown, the rioters came overwhelmingly from these post-war estates. Almost three-quarters of those convicted lived within them. That's not a coincidence.*[1]

Cameron's ideas echo the voices of the economists, sociologists, psychologists, and urban planners, who have long pondered the relationship between the physical appearance of a city and the health, education, mobility, and criminal behavior of its citizens. Neighborhood appearance has been shown to affect rates of alcoholism, obesity, and the spread of STDs. The relationship between physical appearance and criminal activity has been, perhaps, of the greatest interest. The Broken Windows Theory (BWT) of Wilson and Kelling (1982) proposes a connection between the perception of urban disorder and criminal activity. In recent decades, the BWT literature has been characterized by a vigorous debate among scholars, who have found evidence in support (e.g., Keizer, Lindenberg, and Steg 2008) and against the theory (e.g., Sampson and Raudenbush 2004).[2]

However, the connection between the physical appearance of a city and the socioeconomic outcomes of its citizens has proved challenging to study, due to a lack of data on urban appearance. To date, urban appearance has been evaluated with low throughput tools such as field surveys (Sampson and Raudenbush 2004) or virtual audits of urban imagery (Rundle et al. 2011). These methods are time-consuming and expensive, and cover a handful of neighborhoods in a few cities at most.

I. Quantifying Urban Appearance

Imagine using street level images to survey the physical appearance of Manhattan for generating an "evaluative map." Since Manhattan has roughly 72,000 city blocks, an evaluative map with a resolution of one data point per street segment would require scoring 72,000 images. Scaling that map to New York's five boroughs would push the number of evaluations required to roughly one million. Now imagine wanting to create similar maps for tens of cities, at multiple time points, and for different evaluative measures (e.g., perceived safety, liveliness, accessibility, etc.). Such a data generation effort would require evaluating millions of images; a number that is beyond what is possible through field surveys or virtual audits.

The sheer number of data points needed to generate evaluative maps shows the need for automated surveys. To solve this problem, we propose to develop computer vision algorithms

* Naik: MIT Media Lab, 77 Massachusetts Avenue, Cambridge, MA 02139 (e-mail: naik@mit.edu); Raskar: MIT Media Lab, 77 Massachusetts Avenue, Cambridge, MA 02139 (e-mail: raskar@mit.edu); Hidalgo: MIT Media Lab, 77 Massachusetts Avenue, Cambridge, MA 02139 (e-mail: hidalgo@mit.edu). We thank Edward Glaeser, Jackelyn Hwang, Deepak Jagdish, Scott Kominers, Jade Philipoom, Robert Sampson, and Daniel Smilkov for their help and comments.

[†] Go to http://dx.doi.org/10.1257/aer.p20161030 to visit the article page for additional materials and author disclosure statement(s).

[1] David Cameron. 2016. "I've put the bulldozing of sink estates at the heart of turnaround Britain." *Sunday Times*, January 10. http://www.thesundaytimes.co.uk/sto/comment/columns/article1654318.ece.

[2] For a recent review of the literature on the connection between the physical appearance of a city and socioeconomic behavior of its residents, see Naik et al. (2015).

that quantify urban appearance using street level images. Specifically, we describe our work on an algorithm that computes the perceived safety (or "Streetscore") of streetscapes (Naik et al. 2014). We use this algorithm to create high-resolution "evaluative maps" of perceived safety for 19 US cities by scoring more than one million images.

But what are these evaluative maps useful for? First and foremost, these evaluative maps allow researchers to explore the connection between the physical appearance of a city and the socioeconomic outcomes of its citizens, at an unprecedented resolution and scale. In addition, researchers have begun using these evaluative maps to identify architectural constructs and urban planning policies that correlate with perceived safety. Been et al. (2016) find that historic district designation in New York City correlates with higher Streetscore metric of census tracts, indicating that preservation policies are protecting areas that people find more aesthetically appealing. Harvey et al. (2015) relate perceived safety to architectural constructs and show that, in New York and Boston, narrow streets with a high density of buildings are perceived as safer than wider streets with few buildings. Glaeser et al. (2015) demonstrate that the visual appearance of a neighborhood is an adequate proxy for neighborhood income.

Next, we describe our method for computing perceived safety from street level imagery in detail, followed by an analysis of the socioeconomic correlates of perceived safety using evaluative maps of 19 cities.

II. Data and Methods

We develop our algorithm for predicting perceived safety using training data from Salesses, Schechtner, and Hidalgo (2013), a crowd-sourced survey where participants repeatedly chose images from pairs in response to the question: "Which place looks safer?" These images were selected randomly from New York, Boston, Linz, and Salzburg. Here we focus on the United States and only use images for Boston and New York. This dataset contains 4,109 images and 208,738 pairwise comparisons provided by 7,872 unique participants from 91 countries. We use the pairwise comparisons to assign a score for perceived safety between 0 and 10 to each image using the Trueskill ranking algorithm

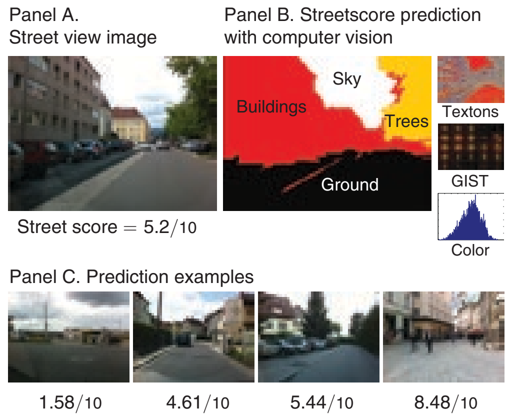

FIGURE 1. COMPUTER VISION TO PREDICT THE PERCEIVED SAFETY OF STREET VIEW IMAGES

(Naik et al. 2014). Visual inspection shows that the typical high scoring image contains houses or townhouses and streets lined with trees; while the typical low scoring image contains parking lots, empty streets, and industrial buildings (Figure 1, panel C). The images and their Trueskill scores form the dataset for training a computer vision algorithm to predict the perceived safety of new streetscapes based on image features. We call the score for perceived safety of an image, Streetscore.

To train the algorithm, we create a computational representation of images. First, we use the Geometric Layout algorithm to classify pixels as belonging to one of the four categories: "Ground," "Buildings," "Trees," or "Sky." Next, we extract three different image features separately for pixels in each of the four geometric classes: Texton histograms, CIELAB 3D color histograms, and GIST. In sum, we represent each image by a feature vector encoding its textures, colors, and shapes (Figure 1, panel B). Next, we use the aforementioned feature vectors to train a ν-Support Vector Regression with a linear kernel (ν-SVR) for predicting Streetscore. We validate the performance of the ν-SVR model using five-fold cross-validation and obtain an R^2 of 57 percent. For more details on feature computation and SVR training, we refer the reader to Naik et al. (2014).

We use the Streetscore predictor to score one million Google Street View images from 19 cities in the Northeast and Midwest of the United

Table 1—Summary Statistics ($N = 3{,}575$)

	Mean	SD	Min	Max
Panel A. Streetscore variables				
Mean Streetscore	5.628	0.485	3.399	7.798
SD of Streetscore	0.793	0.187	0	2.444
Panel B. ACS variables				
log population	3.434	0.398	0	4.166
log area	6.001	0.421	4.814	8.039
Share African American	0.410	0.375	0	1
Share college-educated adults	0.294	0.244	0	1
log median income	4.580	0.280	3.295	5.398
Gini index	0.438	0.0724	0.0330	0.701

Note: The ACS variables refer to socioeconomic indicators at the census tract level obtained from the 2006–2010 American Community Survey.

Table 2—Streetscore and Socioeconomic Characteristics

	Mean Streetscore (1)	SD of Streetscore (2)
log population	0.497***	–0.034**
	(0.031)	(0.015)
log area	–0.530***	–0.052***
	(0.018)	(0.008)
Share African American	0.204***	–0.048***
	(0.022)	(0.010)
Share college-educated adults	0.248***	–0.077***
	(0.050)	(0.023)
log median income	0.684***	–0.017
	(0.053)	(0.025)
Gini Index	0.491***	0.314***
	(0.119)	(0.055)
Observations	3,575	3,575
R^2	0.317	0.043

Notes: All results are from multivariate OLS regressions. Socioeconomic indicators at the census tract level are obtained from the 2006–2010 American Community Survey.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

States.[3] For these cities we estimate the mean and standard deviation of the Streetscores in each census tract. These 19 cities cover 3,575 census tracts according to the 2010 US census boundaries. In addition, we obtain the socioeconomic characteristics of these census tracts from the American Community Survey (ACS) using the estimates for the years 2006–2010.

III. Relating Appearance to Demographics

Table 1 provides the descriptive statistics for Streetscore measures and socioeconomic characteristics of census tracts from 19 cities. Table 2 shows the coefficients and standard errors from multivariate regressions decomposing the average Streetscore of a census tract, and its standard deviation, into socioeconomic characteristics.[4]

The first column shows that the mean Streetscore has a robust positive correlation with population and a robust negative correlation with the area of the census tract, indicating that the mean Streetscore rises with population density. This finding suggests that the architecture of densely populated places is perceived as safer than the architecture of more sparse urban areas. This is related, but not identical to the Jane Jacobs (1961) idea of "eyes on the street." Google Street View images are usually unpopulated (they are often captured early in the morning). Therefore, the observed correlation is one with the architecture of the space rather than the density of people observed in the street level images—which is very low on average.

Other statistically robust results include a strong relationship between better urban appearance and higher income of residents. Interestingly, mean Streetscore is also correlated with the Gini index, indicating that physically attractive census tracts are also more unequal in terms of their income distribution. Additionally, we observe a statistically robust positive correlation with college education, and the share of African Americans, indicating that neighborhoods with large populations of African Americans have higher perceived safety of the physical environment once the effects of other

[3] These cities are Albany, New York; Atlanta, Georgia; Arlington, Virginia; Baltimore, Maryland; Buffalo, New York; Charlotte, North Carolina; Chicago, Illinois; Cleveland, Ohio; Columbus, Ohio; Detroit, Michigan; Milwaukee, Wisconsin; Minneapolis, Minnesota; Newark, New Jersey; Philadelphia, Pennsylvania; Pittsburgh, Pennsylvania; Rochester, New York; Stamford, Connecticut; Worcester, Massachusetts; and Washington, DC.

[4] We also ran these regressions by including city-level fixed effects, which show similar trends. So we omit these results to save space.

socioeconomic characteristics are taken into account.

The second column shows the results from a multivariate regression between the standard deviation of Streetscore and socioeconomic variables. The standard deviation of Streetscore has significant but weak negative correlations with population, area, and college education. Most saliently, we find that the variation in perceived safety within a census tract rises significantly with increasing income inequality, as measured by the Gini index. This indicates that income inequality and "visual" inequality go hand in hand, and that the evaluative maps produced by Streetscore could be used to create proxies for a neighborhood or city's level of income inequality.

In sum, we find that for our dataset, the average urban appearance of a neighborhood has a strong positive correlation with median income and population density, while the variation in urban appearance within a neighborhood has strong positive correlation with income inequality.

IV. Discussion and Future Directions

In this paper we summarized our work on a computer vision technique which is able to quantify the physical appearance of streetscapes. But, this technique is not limited to cross-sectional studies of urban appearance—it can also be used to study urban change. In Naik et al. (2015), the authors measure physical urban change by calculating the difference in Streetscores for images of the same location captured in 2007 and 2014. This method enables the study of the connection between physical urban change and the socioeconomic characteristics of neighborhoods. The authors use spatial regressions to show that neighborhoods that experience physical improvements are more likely to be densely populated by highly educated people.

Beyond correlations, we could also use Streetscores, together with instrumental variables or exogenous shocks—such as construction of light-rail systems or parks—to analyze the causal effect of government spending on public goods on physical urban change. In such cases, the Streetscore algorithm could provide an accurate estimate of the physical urban change experienced by a neighborhood after the intervention.

Furthermore, there is potential for the use of street level imagery in studying urban life at a global scale with computer vision. Google alone has photographed more than 3,000 cities from 106 countries in the past decade. Traditional field studies can be used to provide training data for computer vision algorithms, like Streetscore, which will be able to extrapolate even relatively small samples of survey data over large areas. Computer vision algorithms, therefore, could become an essential tool for conducting recurrent automated surveys of the living environment at low cost and high spatial resolution.

REFERENCES

Been, Vicki, Ingrid Gould Ellen, Michael Gedal, Edward Glaeser, and Brian J. McCabe. 2016. "Preserving History or Restricting Development? The Heterogeneous Effects of Historic Districts on Local Housing Markets in New York City." *Journal of Urban Economics* 92: 16–30.

Glaeser, Edward L., Scott Duke Kominers, Michael Luca, and Nikhil Naik. 2015. "Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life." National Bureau of Economic Research Working Paper 21778.

Harvey, Chester, Lisa Aultman-Hall, Stephanie E. Hurley, and Austin Troy. 2015. "Effects of Skeletal Streetscape Design on Perceived Safety." *Landscape and Urban Planning* 142: 18–28.

Jacobs, Jane. 1961. *The Death and Life of Great American Cities.* New York: Random House.

Keizer, Kees, Siegwart Lindenberg, and Linda Steg. 2008. "The Spreading of Disorder." *Science* 322 (5908): 1681–85.

Naik, Nikhil, Scott Duke Kominers, Ramesh Raskar, Edward L. Glaeser, and César A. Hidalgo. 2015. "Do People Shape Cities, or Do Cities Shape People? The Co-evolution of Physical, Social, and Economic Change in Five Major U.S. Cities." National Bureau of Economic Research Working Paper 21620.

Naik, Nikhil, Jade Philipoom, Ramesh Raskar, and César Hidalgo. 2014. "Streetscore: Predicting the Perceived Safety of One Million Streetscapes." In *IEEE CVPR Workshops*, 793–99. Washington, DC: IEEE Computer Society.

Rundle, Andrew G., Michael D. M. Bader, Catherine A. Richards, Kathryn M. Neckerman, and

Julien O. Teitler. 2011. "Using Google Street View to Audit Neighborhood Environments." *American Journal of Preventive Medicine* 40 (1): 94–100.

Salesses, Philip, Katja Schechtner, and César A. Hidalgo. 2013. "The Collaborative Image of the City: Mapping the Inequality of Urban Perception." *PLoS ONE* 8 (7): e68400.

Sampson, Robert J., and Stephen W. Raudenbush. 2004. "Seeing Disorder: Neighborhood Stigma and the Social Construction of 'Broken Windows.'" *Social Psychology Quarterly* 67 (4): 319–42.

Wilson, James Q., and George L. Kelling. 1982. "Broken Windows: The Police and Neighborhood Safety." *Atlantic Monthly* 249 (3): 29–38.

This article has been cited by:

1. Ruifan Wang, Shuliang Ren, Jiaqi Zhang, Yao Yao, Yu Wang, Qingfeng Guan. 2021. A comparison of two deep-learning-based urban perception models: which one is better?. *Computational Urban Science* **1**:1. . [Crossref]
2. Yonglin Zhang, Shanlin Li, Rencai Dong, Hongbing Deng, Xiao Fu, Chenxing Wang, Tianshu Yu, Tianxia Jia, Jingzhu Zhao. 2021. Quantifying physical and psychological perceptions of urban scenes using deep learning. *Land Use Policy* **111**, 105762. [Crossref]
3. Mingshu Wang, Floris Vermeulen. 2021. Life between buildings from a street view image: What do big data analytics reveal about neighbourhood organisational vitality?. *Urban Studies* **58**:15, 3118-3139. [Crossref]
4. Virgilio Galdo, Yue Li, Martin Rama. 2021. Identifying urban areas by combining human judgment and machine learning: An application to India. *Journal of Urban Economics* **125**, 103229. [Crossref]
5. Liangyang Dai, Chenglong Zheng, Zekai Dong, Yao Yao, Ruifan Wang, Xiaotong Zhang, Shuliang Ren, Jiaqi Zhang, Xiaoqing Song, Qingfeng Guan. 2021. Analyzing the correlation between visual space and residents' psychology in Wuhan, China using street-view images and deep-learning technique. *City and Environment Interactions* **11**, 100069. [Crossref]
6. Thies Lindenthal, Erik B. Johnson. 2021. Machine Learning, Architectural Styles and Property Values. *The Journal of Real Estate Finance and Economics* . [Crossref]
7. Jonathan Cinnamon, Lindi Jahiu. 2021. Panoramic Street-Level Imagery in Data-Driven Urban Research: A Comprehensive Global Review of Applications, Techniques, and Practical Considerations. *ISPRS International Journal of Geo-Information* **10**:7, 471. [Crossref]
8. Jeremy Gabe, Spenser Robinson, Andrew Sanderford. 2021. Willingness to Pay for Attributes of Location Efficiency. *The Journal of Real Estate Finance and Economics* **6**. . [Crossref]
9. Haohao Ji, Linbo Qing, Longmei Han, Zhengyong Wang, Yongqiang Cheng, Yonghong Peng. 2021. A New Data-Enabled Intelligence Framework for Evaluating Urban Space Perception. *ISPRS International Journal of Geo-Information* **10**:6, 400. [Crossref]
10. Weili Guan, Zhaozheng Chen, Fuli Feng, Weifeng Liu, Liqiang Nie. 2021. Urban Perception: Sensing Cities via a Deep Interactive Multi-task Learning Framework. *ACM Transactions on Multimedia Computing, Communications, and Applications* **17**:1s, 1-20. [Crossref]
11. Arianna Salazar Miranda, Zhuangyuan Fan, Fabio Duarte, Carlo Ratti. 2021. Desirable streets: Using deviations in pedestrian trajectories to measure the value of the built environment. *Computers, Environment and Urban Systems* **86**, 101563. [Crossref]
12. Nikola Milojevic-Dupont, Felix Creutzig. 2021. Machine learning for geographically differentiated climate change mitigation in urban areas. *Sustainable Cities and Society* **64**, 102526. [Crossref]
13. Felipe Moreno-Vera. Understanding Safety Based on Urban Perception 54-64. [Crossref]
14. Felipe Moreno-Vera, Bahram Lavi, Jorge Poco. Urban Perception: Can We Understand Why a Street Is Safe? 277-288. [Crossref]
15. César A. Hidalgo, Elisa Castañer, Andres Sevtsuk. 2020. The amenity mix of urban neighborhoods. *Habitat International* **106**, 102205. [Crossref]
16. Fiona Burlig, Christopher Knittel, David Rapson, Mar Reguant, Catherine Wolfram. 2020. Machine Learning from Schools about Energy Efficiency. *Journal of the Association of Environmental and Resource Economists* **7**:6, 1181-1217. [Crossref]
17. Theo Araujo, Irina Lock, Bob van de Velde. 2020. Automated Visual Content Analysis (AVCA) in Communication Research: A Protocol for Large Scale Image Classification with Pre-Trained Computer Vision Models. *Communication Methods and Measures* **14**:4, 239-265. [Crossref]

18. Mohamed R. Ibrahim, James Haworth, Tao Cheng. 2020. Understanding cities with machine eyes: A review of deep computer vision in urban analytics. *Cities* **96**, 102481. [Crossref]
19. Thies Lindenthal, Erik Barry Johnson. 2020. Machine Learning, Architectural Styles and Property Values. *SSRN Electronic Journal* . [Crossref]
20. Christopher Yencha. 2019. Valuing walkability: New evidence from computer vision methods. *Transportation Research Part A: Policy and Practice* **130**, 689-709. [Crossref]
21. Wei Kang, Taylor Oshan, Levi J Wolf, Geoff Boeing, Vanessa Frias-Martinez, Song Gao, Ate Poorthuis, Wenfei Xu. 2019. A roundtable discussion: Defining urban data science. *Environment and Planning B: Urban Analytics and City Science* **46**:9, 1756-1768. [Crossref]
22. Mohamed R Ibrahim, James Haworth, Tao Cheng. 2019. URBAN-i: From urban scenes to mapping slums, transport modes, and pedestrians in cities using deep learning and computer vision. *Environment and Planning B: Urban Analytics and City Science* **63**, 239980831984651. [Crossref]
23. Mark A Chen, Qinxi Wu, Baozhong Yang. 2019. How Valuable Is FinTech Innovation?. *The Review of Financial Studies* **32**:5, 2062-2106. [Crossref]
24. Yongchao Xu, Qizheng Yang, Chaoran Cui, Cheng Shi, Guangle Song, Xiaohui Han, Yilong Yin. Visual Urban Perception with Deep Semantic-Aware Network 28-40. [Crossref]
25. Edward L. Glaeser, Scott Duke Kominers, Michael Luca, Nikhil Naik. 2018. BIG DATA AND BIG CITIES: THE PROMISES AND LIMITATIONS OF IMPROVED MEASURES OF URBAN LIFE. *Economic Inquiry* **56**:1, 114-137. [Crossref]
26. Kaiqun Fu, Zhiqian Chen, Chang-Tien Lu. StreetNet 269-278. [Crossref]
27. Julian TszKin Chan, Weifeng Zhong. 2018. Reading China: Predicting Policy Change with Machine Learning. *SSRN Electronic Journal* . [Crossref]
28. Susan Athey. 2017. Beyond prediction: Using big data for policy problems. *Science* **355**:6324, 483-485. [Crossref]
29. Eric A. Posner, E. Glen Weyl. 2017. Property Is Only Another Name for Monopoly. *Journal of Legal Analysis* **9**:1, 51-123. [Crossref]
30. Xiaobai Liu, Qi Chen, Lei Zhu, Yuanlu Xu, Liang Lin. Place-centric Visual Urban Perception with Deep Multi-instance Regression 19-27. [Crossref]
31. Marco De Nadai, Radu Laurentiu Vieriu, Gloria Zen, Stefan Dragicevic, Nikhil Naik, Michele Caraviello, Cesar Augusto Hidalgo, Nicu Sebe, Bruno Lepri. Are Safer Looking Neighborhoods More Lively? 1127-1135. [Crossref]
32. Eric A. Posner, E. Glen Weyl. 2016. Property Is Another Name for Monopoly Facilitating Efficient Bargaining with Partial Common Ownership of Spectrum, Corporations, and Land. *SSRN Electronic Journal* . [Crossref]
33. Abhimanyu Dubey, Nikhil Naik, Devi Parikh, Ramesh Raskar, César A. Hidalgo. Deep Learning the City: Quantifying Urban Perception at a Global Scale 196-212. [Crossref]