

How we can use Google Street View data to help build healthier communities.

The last decade has seen an explosion of 'Big Data' sources about society which have opened up new avenues of research for social scientists. In new research, [Li Yin](#) looks at how Google Street View can be used to make more accurate pedestrian counts in urban areas compared to more traditional methods. She writes that this information can in turn be used by policymakers to inform their decision-making about building healthier communities.



The explosion of available data about human behavior from digital sources in recent years has provided an incredible resource that might help social scientists to study phenomena and test hypotheses in ways that were previously impossible. One such data source is Google Street View, a component of Google Maps and Google Earth that serves millions of people daily with street images captured for cities in over 20 countries across four continents. Street view images taken along public streets allow users to replicate an eye-level experience and to virtually walk down a street. The unprecedented source of visual information about streets, for instance, pedestrians, trees, and building features on these images is readily available to anyone with access to internet and has become a source of 'big data'. How can planners and social scientists use this data for design and planning purposes?

In new research, my co-authors and I show how Google Street View images can be used to automatically compile pedestrian volume data more consistently and objectively and at a larger scale. Our findings suggest that this method is capable of determining the presence of pedestrians with a reasonable level of accuracy, data which can be used to help build healthy communities.

Planning policies and how communities are designed influence the health of residents and their quality of life. Studies have shown that increased physical activity through walking and other exercises leads to better health. The past decade has seen a dramatic increase in pedestrian oriented design and planning to facilitate physical activity and active living. Pedestrian count is a quantitative measure of pedestrian volume to help evaluate pedestrian activity and walkability. In addition to helping study how pedestrian activity correlates with land use and other built environment characteristics, it has also been used as baseline data to help inform planning and funding decisions. Pedestrian count data has been traditionally collected manually along sampled streets through field work and self-reported survey. Therefore, the significant limitation of current pedestrian count method is mainly on cost, time, data accuracy, subjectivity, and sample size. Data errors due to human mistakes in counting and data entry are also a real concern. Collection of this kind of data is also less feasible when samples are large and spatially dispersed.

Can we build better databases on pedestrian counts more effectively and for a larger area using 'big data' and recent computational advances? Massive new sources of data are increasingly produced and made available via web, mobile device, and other technologies at an unprecedented rate, such as opinions and images shared via Twitter and other social media platforms. These data can help to design experiments to test social theories for a deeper understanding of the social world and move away from coarse level estimates from sampled population.

Most of the information made available by Google and other websites, however is targeted toward third party web-based development, e.g. for third party websites to display Google maps and street views. There is limited information explicitly and directly disclosed for other purposes, e.g. image parameters and how the images were stored and assembled for a wide range of analysis and visualization purposes such as neighborhood audits. In our research, we borrowed a tool developed in another discipline, in particular, from the most recent development in machine learning to help detect and extract pedestrian volume for design and planning of walkable environments.

Our research used the city of Buffalo in the State of New York as a case study. [Figure 1](#) illustrates how we downloaded Google Street View images for one street block and transformed them into images for pedestrian detection and counting. As shown in the upper part of [Figure 1](#), one street block can have a number of shooting locations along the street, as marked by “x” in the figure. There are about 130,000 locations with panoramas from Google Street View in a city like Buffalo. The middle section of [Figure 1](#) shows an example of downloaded panoramic images for one of the shooting locations (marked by the bolded x) on the example street. Image one (marked both in the images in the upper right corner and in the second row) covers the lower side of the street and image two covers the upper side of the street.

Figure 1 – Retrieving Google Street View images

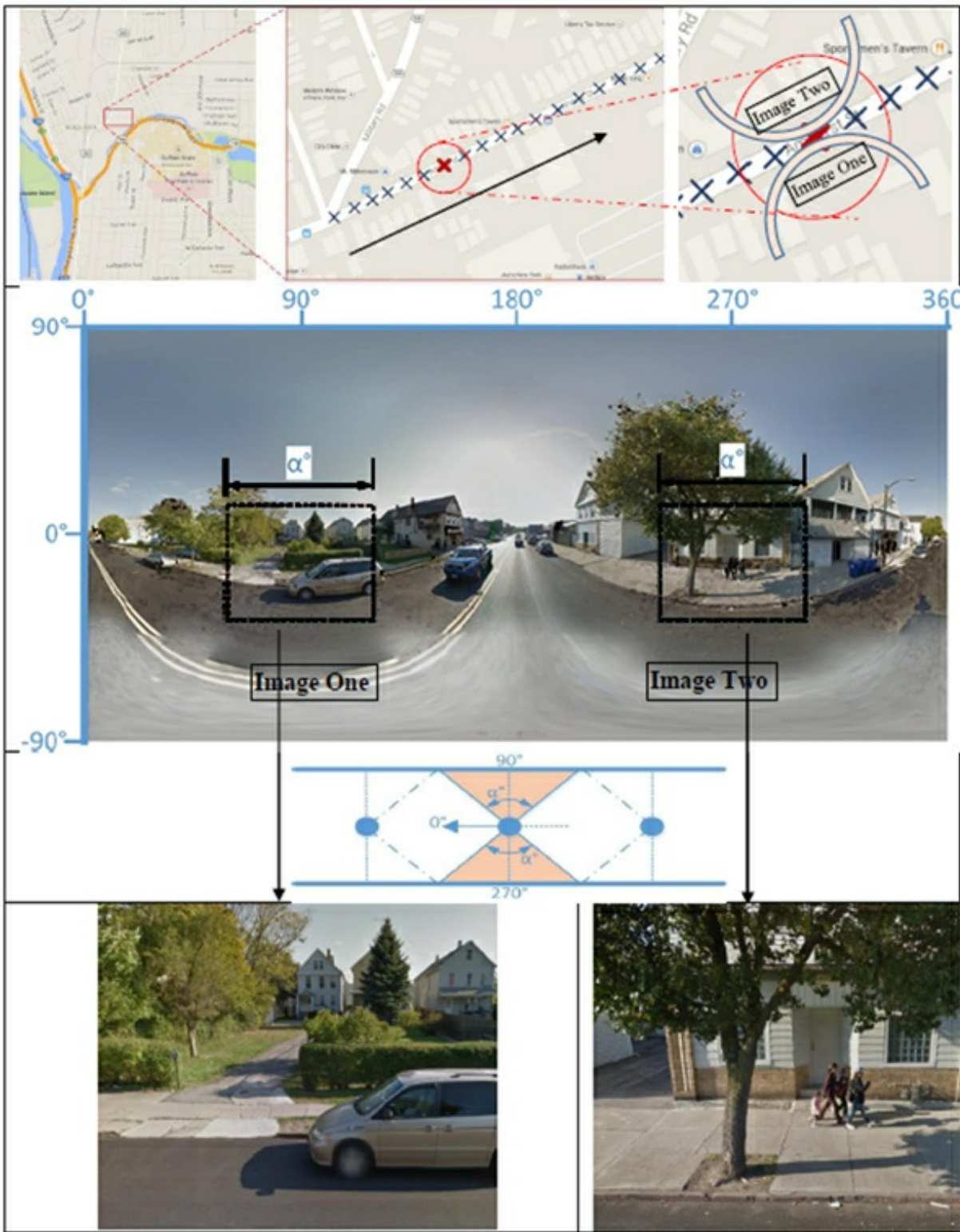


Figure 2 illustrates how images downloaded and processed were used for automatic pedestrian detection. The upper left corner is a street side view image used as the input for pedestrian detection. The upper right shows the results from pre-processing through image segmentation. At the end of this primary stage, background was removed, and some of the environmental interference was reduced to get a more effective area for the fine detection stage. This helps to reduce the false detection rate and unnecessary complex calculations at the fine detection stage to increase accuracy and processing speed. The image on the bottom shows the green rectangles as outlined pedestrian areas for pedestrian counts. One rectangle represents one pedestrian. One pedestrian was not identified. The missing error could happen on any street and may be offset when the algorithm is used for a large scale pedestrian count across a city.

Figure 2 – Results from ACF Pedestrian Detection



With the large quantities and diverse datasets in both computing and spatial context comes the challenge of analyzing 'big data'. Our work shows that collaborative research can help social scientists to meet the emerging big data challenges and investigate complex socioeconomic dynamics by bringing in big data technologies and advanced computational analytics transcending traditional disciplinary boundaries. Our efforts can help to stimulate and push forward the interdisciplinary discussion of the use of online 'big data' and recent computational advances for planning and design.

This article is based on the paper, 'Big data' for pedestrian volume: Exploring the use of Google Street View images for pedestrian counts', in Applied Geography.

Featured [image](#) credit: [Michelle Milla](#) (Flickr, [CC-BY-2.0](#))

[Please read our comments policy before commenting.](#)

Note: This article gives the views of the author, and not the position of USAPP – American Politics and Policy, nor the London School of Economics.

Shortened URL for this post: <http://bit.ly/1QgBZ1W>

About the author

Li Yin – *University of Buffalo, State University of New York*

Li Yin is an Associate Professor in the Department of Urban and Regional Planning at the University of Buffalo, State University of New York. Her research focuses on practical applications of spatial models, joining amenity and location theory with applied GIS and simulation methods to explore the complexity and dynamic processes of urban systems for environmental planning, urban design, and sustainable development.



- CC BY-NC 3.0 2015 LSE USAPP