Applied Geography 63 (2015) 337–345

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

'Big data' for pedestrian volume: Exploring the use of Google Street View images for pedestrian counts

Li Yin [a,*], Qimin Cheng [b], Zhenxin Wang [c], Zhenfeng Shao [d]

[a] Department of Urban and Regional Planning, University at Buffalo, The State University of New York, Buffalo, NY 14214, USA
[b] Department of Electronics and Information Engineering, Huazhong University of Science & Technology, Wuhan 430074, China
[c] Center for Human-Engaged Computing, Kochi University of Technology, 185 Miyanokuchi, Tosayamada-Cho, Kami-Shi, Kochi 782-8502, Japan
[d] The State Key Laboratory of Information Engineering on Surveying Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

ARTICLE INFO

Article history:
Received 9 March 2015
Received in revised form
19 July 2015
Accepted 19 July 2015
Available online xxx

Keywords:
'Big data'
Google Street View
Pedestrian count
Walkability

ABSTRACT

New sources of data such as 'big data' and computational analytics have stimulated innovative pedestrian oriented research. Current studies, however, are still limited and subjective with regard to the use of Google Street View and other online sources for environment audits or pedestrian counts because of the manual information extraction and compilation, especially for large areas. This study aims to provide future research an alternative method to conduct large scale data collection more consistently and objectively on pedestrian counts and possibly for environment audits and stimulate discussion of the use of 'big data' and recent computational advances for planning and design. We explore and report information needed to automatically download and assemble Google Street View images, as well as other image parameters for a wide range of analysis and visualization, and explore extracting pedestrian count data based on these images using machine vision and learning technology. The reliability tests results based on pedestrian information collected from over 200 street segments in Buffalo, NY, Washington, D.C., and Boston, MA respectively suggested that the image detection method used in this study are capable of determining the presence of pedestrian with a reasonable level of accuracy. The limitation and potential improvement of the proposed method is also discussed.

1. Introduction

"[B]ig data and computational sciences are changing how we can analyze and understand individual and collective human behavior" (Ruppert, 2013; p269). Planners and social scientists have been using survey, interview, and field work to collect empirical data for years (Ruppert, 2013). Recent studies suggested to attend to new forms of empirical data and conceptions (Lury and Wakeford, 2012; Ruppert, 2013). 'Big data' and computational analytics create important opportunities for interdisciplinary approach to study new phenomena or to study old questions with new data and insights (Arribas-Bel, 2014; Ruppert, 2013). They have a potential to help study social phenomena in ways never before imagined and possible (Arribas-Bel, 2014; Watts, 2007).

New data sources and technologies have simulated research on walkability (Lee & Talen, 2014). Responding to the growing demand for walkable and transit-oriented development in recent years, pedestrian activity has been used to study what it is about the built environment that gets people active such as walk and bike in their neighborhood (Ewing & Clemente, 2013). Pedestrian count is a quantitative measure of pedestrian volume to help evaluate pedestrian activity, walkability and how it correlates with land use, and other built environment characteristics (Ewing & Clemente, 2013; Hajrasouliha & Yin, 2014). The count data can also be used as baseline data to help inform planning and funding decisions. Even though some methods have been developed to estimate pedestrian volumes (Ercolano, Olson, & Spring, 1997; Landis, 1996), they are usually not designed for actual pedestrian counts and many of them do not have a fine-grained geographic scale (Scheneider et al., 2009). Collection of detailed information about non-motorized activity is insufficient and inefficient in many transportation and built environment studies, especially at a large scale.

A large amount of data on many aspects of human behavior are available on various websites nowadays (Arribas-Bel, 2014). Google Street View provides panoramic views along public streets in the U.S. and many countries around the world. It has recently been used

* Corresponding author.
E-mail address: liyin@buffalo.edu (L. Yin).

http://dx.doi.org/10.1016/j.apgeog.2015.07.010

to help audit the built environments (Badland, Opit, Witten, Kearns, & Mavoa, 2010; Ben-Joseph et al., 2013; Clarke, Ailshire, Melendez, Bader, & Morenoff, 2010; Lee & Talen, 2014; Rundle, Bader, Richards, Neckerman, & Teitler, 2011) and for pedestrian counts (Ewing & Clemente, 2013). Current studies, however, are still limited and subjective with regard to the use of Google Street View for environment audits or pedestrian counts because of the manual information extraction and compilation, especially for large areas.

Building on recent studies with manual collection of pedestrian counts (Purciel et al., 2009; Ewing & Clemente, 2013), our study contributes in three ways. First, we propose an alternative method to collect pedestrian volume information more consistently and objectively and at a larger scale using automatic information extraction on Google Street View images. It can possibly be used for environment audits by automatically detecting benches, trees, building shapes, etc. Pedestrian volumes are found to be correlated with land use and development density, street network, and a sense of safety including crime and traffic (Chu, 2005; Hajrasouliha & Yin, 2014; Ozer & Kubat, 2007). Previous studies called for better data on pedestrian volumes and more effective methodologies for counting and modeling pedestrian volumes with existing data (Lee &Talen, 2014; Scheneider et al., 2009). Our method can help identify places with different pedestrian needs using readily available Google Street View data to prioritize investment in order to improve pedestrian environment with more safety, comfort and convenience for building walkable environment.

The second and third contributions lie in how this study can stimulate and push forward the interdisciplinary discussion of the use of online 'big data' and recent computational advances for planning and design. Google and other websites make a large amount of information available. However, many companies and websites, such as Google, target mainly for the third party web-based development. APIs were provided for third party websites to display Google maps and street views. Google Street View does not explicitly provide direct and enough information on their street view images, such as image parameters and how the images were stored and assembled for other purposes. Going through Google Street View and Google map websites and API codes, we report and describe related Google Street View information needed to automatically download and assemble images, as well as other image parameters for a wide range of analysis and visualization. The automatically downloaded images can be transformed into images for different research purposes, such as neighborhood audits. Finally, we present how 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.

2. Pedestrian activity, 'big data', and Google Street View images

Building walkable and healthy communities is a heightened and widespread interest in recent years among researchers and practitioners. Many studies are in an effort to improve the pedestrian environment and pedestrian has become the subject of increasing attention among planners, engineers and public health officials (Clifton et al., 2007; Ewing & Bartholomew, 2013). The built environment, and streets in particular as one important element, should be designed not only to enhance mobility choices but also to reinforce walkability, livability, and sustainability (Ewing & Clemente, 2013; Yin, 2014). How people use their built environment and how the built environment characteristics influence physical and psychological health for people of all ages has increasingly being studied in recent years (Frank, Sallis, Conway, Chapman, & Saelens, 2006; Forsyth, Schmitz, Hearst, & Oakes, 2008; Yin et al., 2013; Kim & Susilo, 2013). Pedestrian count has been used as an important measure for these studies.

Pedestrian count data has been traditionally collected along sampled streets through field work, self-reported survey, or automated counting. Automated counting technology for pedestrian is less developed even though it has been used for motor vehicles for many years. Most pedestrian counts are done manually. Like in-person audits for walkability, the significant limitation of current pedestrian count method is mainly on cost, time, data accuracy, and subjectivity (Badland et al., 2010; Ben-Joseph et al., 2013; Purciel et al., 2009; Rundle et al., 2011). The observation time, number of observers, and training sessions required can be substantial (Lee & Talen, 2014). There are data errors due to human mistakes in counting and data entry. Collection of such data is also less unfeasible with large and spatially dispersed samples (Purciel et al., 2009; Rundle et al., 2011). As suggested in Ewing and Clemente (2013), sample sizes are usually small when the process is manually done. In addition, Field work based on observations and self-reported surveys are more subjective than automatic counts using video-taping or sensors. The counting methods used vary in different studies (Rundle et al., 2011; Ewing & Clemente, 2013). Finally, pedestrian counts cannot be acquired easily as a secondary data source; in other words, they are usually collected and used by the same group without being verified or made available to the public. Using existing data can potentially help to increase accuracy and improve efficiency, as well as increase reuse of the data (Lee & Talen, 2014).

With the recent rapid development of internet and cloud computing, we are entering the era of 'big data' with the 'Internet of Things' and People (O'Leary, 2013). 'Internet of Things' is configured to include inputs from humans and many different things linked to the internet (O'Leary, 2013). 'Big data' was described as data gathered from different online sources by Goodchild (2013). Another definition of 'big data' is about the effort to make the rapidly expanding amount of digital information analyzable and "the actual use of that data as a means to improve productivity, generate and facilitate innovation and improve decision making" (O'Leary, 2013, p54).

Google Street View is a component of Google Map and Google Earth that serves millions of people daily with images captured in many cities in over 20 countries across four continents (Anguelov et al., 2010). It allows users to see panoramic images from points along public streets to replicate an eye-level experience and to virtually walk down the street (Ewing & Clemente, 2013; Ben-Joseph et al., 2013). Google Street View has provided an unprecedented source of visual information about our streets, for instance, pedestrians, trees, and building features. It has become a source of 'big data' and it is readily available to anyone with access to internet.

Google Street View "has rarely been utilized in published research" until recently (Ewing & Clemente, 2013, p85). A number of studies related to urban planning and public health that used Google Street View have been published since 2010 and they all found that Google Street View offered a reliable alternative for neighborhood audits associated with walking and cycling (Badland et al., 2010; Ewing & Clemente, 2013). Web-based tools such as Google Street View offer a more resource-efficient substitute for on-site audits to save time and cost by allowing for preliminary audits to be performed accurately from remote locations, and increasing the effectiveness of subsequent on-site visits (Badland et al., 2010; Ben-Joseph et al., 2013). However, current studies manually processed information from Google Street View. This is partly because Google Street View targets primarily third party web-based development with little and limited information available for other purposes; partly because recent tools developed in other disciplines to process such information have not been well applied by planners and social scientists.