Landscape and Urban Planning 165 (2017) 93–101

Contents lists available at ScienceDirect

Landscape and Urban Planning

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

Research Paper

Green streets – Quantifying and mapping urban trees with street-level imagery and computer vision

Ian Seiferling[a,b,*], Nikhil Naik[c], Carlo Ratti[a], Raphäel Proulx[b]

[a] Senseable City Laboratory, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Room 10-485, 77 Massachusetts Avenue, Cambridge, MA, 02139, United States
[b] Canada Research Chair in Ecological Integrity, Centre de Recherche sur les Interactions Bassins Versants-Ecosystemes Aquatiques, Universite du Quebec a Trois-Rivieres, 3351 Boulevard des Forges, Trois-Rivieres, Quebec, G9A 5H7, Canada
[c] MIT Media Lab, 75 Amherst St, Cambridge, MA 02139, United States

ARTICLE INFO

Keywords:
Urban trees
Computer vision
Streetscapes
Tree cover
Greenspace

ABSTRACT

Traditional tools to map the distribution of urban green space have been hindered by either high cost and labour inputs or poor spatial resolution given the complex spatial structure of urban landscapes. What's more, those tools do not observe the urban landscape from a perspective in which citizens experience a city. We test a novel application of computer vision to quantify urban tree cover at the street-level. We do so by utilizing the open-source image data of city streetscapes that is now abundant (Google Street View). We show that a multi-step computer vision algorithm segments and quantifies the percent of tree cover in streetscape images to a high degree of precision. By then modelling the relationship between neighbouring images along city street segments, we are able to extend this image representation and estimate the amount of perceived tree cover in city streetscapes to a relatively high level of accuracy for an entire city. Though not a replacement for high resolution remote sensing (e.g., aerial LiDAR) or intensive field surveys, the method provides a new multi-feature metric of urban tree cover that quantifies tree presence and distribution from the same viewpoint in which citizens experience and see the urban landscape.

1. Introduction

With the growing consensus that nature and multi-functional ecosystems are intrinsic to sustainable cities, decision makers, designers and the broader public alike are looking to trees as urban keystone flora that provide natural infrastructure and services – to reduce air pollution, support biodiversity, mitigate heat island effects, increase land value, improve aesthetics and even improve human health (Kardan et al., 2015; Lothian, 1999; Lovasi et al., 2008; McPherson et al., 1997; Nowak, Hirabayashi, Bodine, & Greenfield, 2014; Thayer and Atwood, 1978). Urban tree effects may even extend to cultural and psychological behaviours with, for example, a high abundance of street trees being linked to urban scenes that were perceived to be safe (Naik, Philipoom, Raskar, & Hidalgo, 2014). The fact remains however that urban trees come with costs and are currently threatened by climate change, pests and diseases. Conflicting land uses and cost-benefit tradeoffs cause contention at many levels of society. Such contentions can be alleviated through a better understanding of the role of trees in the complex and cluttered landscapes that are cities. To this end, tools to quantify and monitor presence, abundance and health of urban trees are needed. Governments, particularly cash-strapped ones, are evermore looking for low-cost ways to establish baseline data, manage and engage the public on urban trees.

Traditionally, urban tree cover has been quantified using coarse-scale methods developed for naturally forested landscapes and exposure to "nature" as an urban quality indicator has been quantified by measuring the total land area covered by greenspace (*i.e.,* city park area) in cities (Fuller and Gaston, 2009; Richardson, Pearce, Mitchell, & Kingham, 2013; Schroeder, 1986). In either case, these methods primarily rely on long-range remotely-sensed image processing to classify landcover (*i.e.,* satellite imagery such as LANDSAT, ortho-aerial photographs or, more recently, LiDAR) (Homer et al., 2007) or data derived from field surveys (Kardan et al., 2015). Substantial drawbacks exist within each case, many of which present particular challenges in an urban context. For example, traditional remote-sensing techniques for vegetation cover have, most often, been based on moderate-resolution imagery (*e.g.,* 30 m in the case of openly available data) which has limited utility at the scale of cities. Recent

* Corresponding author at: Senseable City Laboratory, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Room 10-485, 77 Massachusetts Avenue, Cambridge, MA, 02139, United States.
E-mail address: ianseifs@mit.edu (I. Seiferling).

http://dx.doi.org/10.1016/j.landurbplan.2017.05.010
Received 6 July 2016; Received in revised form 10 May 2017; Accepted 13 May 2017

efforts exploiting high resolution active sensing like LiDAR are proving well-suited for urbanscapes (MacFaden, O'Neil-Dunne, Royar, Lu, & Rundle, 2012), however they can be hindered by specialized proprietary software, high data-acquisition costs and significant labour inputs. On the other hand, field-based surveys lack the automation and the scale of big data sets (*i.e.,* low-throughput), are prone to sampling errors (Dickinson, Zuckerberg, & Bonter, 2010) and require enormous organizational efforts. These methodological impediments also make it difficult to achieve periodic resampling to asses changes in tree cover and health over time.

Chiefly through machine learning models, computer vision scientists are teaching computers to see the world at astounding rates of success. However, few disciplines outside of the strict artificial intelligence fields (e.g., robotics, driverless cars, software) have utilized these advancements. One of the few examples bridging ecology and computer vision technologies is the mobile app, *Leafsnap*, which identifies plant species using automatic visual recognition (Kumar et al., 2012). If a computer can learn to detect and quantify features of an environmental scene from digital photographs (*i.e.,* scene understanding), it stands that those algorithms can be used to objectively quantify real-world features and their spatial distribution within a landscape for a multitude of applications. For instance, Naik et al. have developed computer vision algorithms that process street-level imagery to quantify urban appearance (Naik et al., 2014), urban change (Naik, Kominers, Raskar, Glaeser, & Hidalgo, 2015), or even socio-economic indicators (Glaeser, Kominers, Luca, & Naik, 2015). Opportunely, we now also have access to entire cities in the form of geo-tagged, street-level images.

Using Google Street View images that represent a ground-based perspective of city streets − *streetscapes* − and which cover a city-wide extent, we develop and test a new method of rapid quantification and mapping of urban vegetation, specifically trees. The method applies a trained predictor to segment the amount of tree cover in a given image of a city streetscape using multiple image features. We aim to demonstrate that we can quantify the presence and perceived cover of street-side trees with high spatial resolution at the city-scale by: *i)* sampling a series of sequential neighbouring image scenes of the streetscape; *ii)* predicting the amount of tree cover present in them and; *iii)* modelling the relationship between the tree cover of these neighbouring view-points. To estimate the accuracy and utility of this approach we compare our method to contemporary remote-sensing techniques used to estimate urban tree canopy cover (*i.e.,* object-based image analysis (OBIA) of high-resolution LIDAR data and multispectral imagery).

The goal of this study is to present a novel method of measuring trees in a city at extremely high-throughput; one that may not replace existing techniques, but offers clear benefits such as being relevant to the human perspective (i.e., *the perceived tree cover*), cheap, independent of proprietary software and easily scaleable across cities.

2. Methods

2.1. Study areas and image datasets

We collected data on urban tree cover by using 456,175 geo-tagged images from the two cities of New York (336,998 images) and Boston (119,177 images) in the United States. However, for the vast majority of the results presented, we focus on New York because the best-suited tree canopy cover maps and street tree survey data we could acquire were of New York. Images were sourced from the Google Street View (GSV) application program interface (API) (Google Inc., 2014), were acquired in 2014 and represent a ground-level, side-view perspective of the city streetscape (Fig. 1C). All image collection points along city roads were downloaded for a target city and this resulted in a GSV image roughly every 15 m along a given roadway; these image samples are hereafter referred to as *GSV sampling points*. However, due to the protocol of the GSV system the 15 m interval could deviate by approximately +/− 5 m. Given this, we define a neighbour sample points as two GSV sampling points on the same road segment and a minimum of 10 m and maximum of 20 m apart. Some GSV sampling points, road segments or areas of the city did not have data for various reasons (*e.g.,* corrupt or missing data, no-coverage area). Notwithstanding those instances, the sampling regime covered the full extent of the cities' official boundaries, though for the case of New York it did not cover Staten Island (Fig. 1A & B).

Each digital photograph (Red-Green-Blue color channel jpeg image) was acquired from the GSV API at a resolution of 400 by 300 pixels, at a 90° horizontal field of view, 90° heading (east) and a 10° pitch. The level of pitch was chosen in order to optimize the capture of the streetscape (*i.e.,* decrease the amount of foreground composed only of roadway). Fixing the image heading to 90° east for every sampling point allowed us to compare how the road-to-image orientation would affect the metrics and, ultimately, the ability to estimate tree cover. As such, all sampling points were grouped into one of four categories based on their road orientation, given 22.5° intervals around 360°: 1) N-S: GSV sampling points lying on roads that are oriented in a north-south direction (± 22.5° from 0° or 180°); 2) E-W: GSV sampling points lying on roads oriented in an east-west direction (± 22.5° from 90° or 270°); 3) NW-SE: GSV sampling points lying on roads oriented in a diagonal northwest- southeast direction (± 22.5° from 135° or 315°); 4) NE-SW: GSV sampling points lying on roads oriented in a diagonal northeast-southwest direction (± 22.5° from 45° or 225°).

In order to estimate the real-world surface area covered by each GSV image, we modelled the 2-dimensional (horizontal and vertical) surficial field of view (FOV) represented in an image at each sampling point; *i.e.,* the camera's horizontal field of view (90°) and depth of field projected onto the earth's surface. We computed this FOV polygon for each GSV sampling point which was then projected on the horizontal surface plane to associate a surface area with the sampling point (Fig. 1d). The length of the polygon (i.e., length of the right-angle bisector) represents the approximate image depth of field. However, in reality the depth of field varies with the presence, size and proximity of objects occluding the horizon. We assume that a given length should, on average, be representative of an urban streetscape. Therefore, we varied this depth of field parameter and created four levels: 15 m, 25 m, 35 m and 45 m from the GSV sampling point. In addition to the road-to-camera orientation groups, we run our analysis at each of these depth of field levels in order to determine which provides the best spatial context for predicting real-world tree cover.

2.2. Tree detection using computer vision

We estimated the total area covered by trees in each image by applying a multi-step image segmentation method developed by Hoiem et al. (Hoiem, Efros, & Hebert 2005). On a per-image basis, the objective of the method is to model geometric classes that depend on the orientation of a physical object with relation to the scene and with respect to the camera. Specifically, each image pixel is classified into one of a few geometric classes: *i)* the ground plane; *ii)* surfaces that stick up from the ground (vertical surfaces); *iii)* part of the sky plane. Further, vertical surfaces are subdivided into planar surfaces facing left, right or towards the camera and either porous (e.g. trees and their leafy vegetation) or solid (e.g. a person or lamp post) non-planar surfaces. Although this recognition approach differs from those that instead model semantic classes (*e.g.,* car, house, person, vegetation), it has proven exceptionally powerful and efficient in cluttered outdoor scenes like urban streetscapes and, most relevant to our application here, in distinguishing human built structures from natural ones like trees.

The algorithm operates by first grouping image pixels into *superpixels*, which are groups of pixels assumed to share a single label (e.g., ground or sky) and respect coarse-level segment boundaries (e.g., edges) (Felzenszwalb & Huttenlocher, 2004). The algorithm then groups regions of the image into homogenous segments using a