ISPRS Journal of Photogrammetry and Remote Sensing 135 (2018) 13–30

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

From Google Maps to a fine-grained catalog of street trees

Steve Branson [a,1], Jan Dirk Wegner [b,*,1], David Hall [a], Nico Lang [b], Konrad Schindler [b], Pietro Perona [a]

[a] Computational Vision Laboratory, California Institute of Technology, USA
[b] Photogrammetry and Remote Sensing, ETH Zürich, Switzerland

ARTICLE INFO

Article history:
Received 13 July 2017
Received in revised form 8 November 2017
Accepted 8 November 2017

Keywords:
Deep learning
Image interpretation
Urban areas
Street trees
Very high resolution

ABSTRACT

Up-to-date catalogs of the urban tree population are of importance for municipalities to monitor and improve quality of life in cities. Despite much research on automation of tree mapping, mainly relying on dedicated airborne LiDAR or hyperspectral campaigns, tree detection and species recognition is still mostly done manually in practice. We present a fully automated tree detection and species recognition pipeline that can process thousands of trees within a few hours using publicly available aerial and street view images of Google Maps™. These data provide rich information from different viewpoints and at different scales from global tree shapes to bark textures. Our work-flow is built around a supervised classification that automatically learns the most discriminative features from thousands of trees and corresponding, publicly available tree inventory data. In addition, we introduce a change tracker that recognizes changes of individual trees at city-scale, which is essential to keep an urban tree inventory up-to-date. The system takes street-level images of the same tree location at two different times and classifies the type of change (e.g., tree has been removed). Drawing on recent advances in computer vision and machine learning, we apply convolutional neural networks (CNN) for all classification tasks. We propose the following pipeline: download all available panoramas and overhead images of an area of interest, detect trees per image and combine multi-view detections in a probabilistic framework, adding prior knowledge; recognize fine-grained species of detected trees. In a later, separate module, track trees over time, detect significant changes and classify the type of change. We believe this is the first work to exploit publicly available image data for city-scale street tree detection, species recognition and change tracking, exhaustively over several square kilometers, respectively many thousands of trees. Experiments in the city of Pasadena, California, USA show that we can detect >70% of the street trees, assign correct species to >80% for 40 different species, and correctly detect and classify changes in >90% of the cases.

1. Introduction

Urban forests in the USA alone contain around 3.8 billion trees (Nowak et al., 2002). A relatively small but prominent element of the urban forest are street trees. Street trees grow along public streets and are managed by cities and counties. The most recent estimate is that there are 9.1 million trees lining the streets of California, about one street tree for every 3.4 people[2] living in an urban area, with an estimated replacement value of $2.5 billion (McPherson et al., 2016). However, the greatest value of a street tree is not its replacement value but its ecosystem services value, i.e., all economic benefits that a tree provides for a community. These benefits include: a reduction in energy use, improvement in air and water quality, increased carbon capture and storage, increased property values and an improvement in individual and community well-being (Nowak et al., 2002; McPherson et al., 2016).[3] Still, inventories are often lacking or outdated, due to the cost of surveying and monitoring the trees.

We propose an automated, image-based system to build up-to-date tree inventories at large scale, using publicly available aerial images and panoramas at street-level. The system automatically detects trees from multiple views and recognizes their species. It draws on recent advances in machine learning and computer vision, in particular deep learning for object recognition

* Corresponding author.
E-mail address: jan.wegner@geod.baug.ethz.ch (J.D. Wegner).
[1] Joint first authorship.
[2] A rough estimate for Europe is given in (Pauleit et al., 2005). The number of people per street tree strongly varies across European cities between 10 to 48 inhabitants per street tree. However, it is unclear (and in fact unlikely) if US and European census numbers rely on the same definitions.
[3] The most recent estimate of the ecosystem services value of the street trees in California is $1 billion per year or $111 per tree, respectively $29 per inhabitant.

https://doi.org/10.1016/j.isprsjprs.2017.11.008
0924-2716/

(Krizhevsky et al., 2012; Szegedy et al., 2015; Simonyan and Zisserman, 2015; Girshick, 2015; Ren et al., 2015), fine-grained object categorization (Wah et al., 2011; Angelova and Zhu, 2013; Branson et al., 2013; Deng et al., 2013; Duan et al., 2013; Krause et al., 2014; N. Zhang et al., 2014b), and analysis of publicly available imagery at large scale (Hays and Efros, 2008; Agarwal et al., 2009; Anguelov et al., 2010; Majdik et al., 2013; Russakovsky et al., 2015). The method is build around a supervised classification that uses deep convolutional neural networks (CNN) to learn tree identification and speies classification from existing inventories.

Our method is motivated by TreeMapLA,[4] which aims to build a publicly available tree inventory for the greater Los Angeles area. Its goal is to collect and combine already existing tree inventories acquired by professional arborists. In case no reasonably up-to-date data is available, which is often the case, a smartphone app is used to task users in a crowd-sourcing effort to fill in data gaps. Unfortunately only few people (so-called citizen scientists) participate. Only a small number of trees, e.g., ≈ 1000 out of more than 80,000 in Pasadena, have been mapped within the last 3 years. And often entries are incomplete (e.g., missing species, trunk diameter) or inaccurate (e.g., GPS position grossly wrong). It turns out that determining a tree's species is often the hardest and most discouraging part for citizen scientists. The average person does not know many species of tree, and even with tree identification tools, the prospect of choosing one option among tens or even hundreds is daunting.

We propose to automate tree detection and species recognition with the help of publicly available street-level panoramas and very high-resolution aerial images. The hope is that such a system, which comes at virtually no cost and enables immediate inventory generation from scratch, will allow more cities to gain access to up-to-date tree inventories. This will help to ascertain the diversity of the urban forest by identifying tree species determinants of urban forest management (e.g., if a pest arrives, an entire street could potentially lose its trees). Another benefit to a homogeneous inventory across large urban areas would be to fill in the gaps between neighboring municipalities and different agencies, allowing for more holistic, larger-scale urban forest planning and management. Each city's Tree Master Plan would no longer exist in a vacuum, but account for the fact that the urban forest, in larger metropolitan areas, spreads out across multiple cities and agencies.

Our system works as follows: It first downloads all available aerial images and street view panoramas of a specified region from a repository, in our example implementation Google Maps. A tree detector that distinguishes trees from all other scene parts and a tree species classifier are separately trained on areas where ground truth is available. Often, a limited, but reasonably recent tree inventory does exist nearby or can be generated, which has similar scene layout and the same tree species. The trained detector predicts new trees in all available images, and the detector predictions are projected from image space to true geographic positions, where all individual detections are fused. We use a probabilistic conditional random field (CRF) formulation to combine all detector scores and add further (learned) priors to make results more robust against false detections. Finally, we recognize species for all detected trees. Moreover, we introduce a change classifier that compares images of individual trees acquired at two different points in time. This allows for automated updating of tree inventories.

2. Related work

There has been steady flow of research into automated tree mapping over the last decades. A multitude of works exist and a full review is beyond the scope of this paper (e.g., see (Larsen et al., 2011; Kaartinen et al., 2012) for a detailed comparison of methods).

Tree delineation in forests is usually accomplished with airborne LiDAR data (Reitberger et al., 2009; Lähivaara et al., 2014; J. Zhang et al., 2014) or a combination of LiDAR point clouds and aerial imagery (Qin et al., 2014; Paris and Bruzzone, 2015). LiDAR point clouds have the advantage of directly delivering height information, which is beneficial to tell apart single tree crowns in dense forests. On the downside, the acquisition of dense LiDAR point clouds requires dedicated, expensive flight campaigns. Alternatively, height information can be obtained through multi-view matching of high-resolution aerial images (Hirschmugl et al., 2007) but is usually less accurate than LiDAR due to matching artefacts over forest.

Only few studies attempt segmentation of individual trees from a single aerial image. Lafarge et al. (2010) propose marked point processes (MPP) that fit circles to individual trees. This works quite well in planned plantations and forest stands with reasonably well-separated trees. However, MMPs are notoriously brittle and difficult to tune with inference methods like simulated annealing or reversible jump Markov Chain Monte Carlo, which are computationally expensive. Simpler approaches rely on template matching, hierarchies of heuristic rules, or scale-space analysis (see Larsen et al. (2011) for a comparison).

Tree detection in cities has gained attention since the early 2000s. Early methods for single tree delineation in cities were inspired by scale-space theory (initially also developed for forests by Brandtberg and Walter (1998)). A common strategy is to first segment data into homogeneous regions, respectively 3D clusters in point clouds, and then classify regions/clusters into tree or background, possibly followed by a refinement of the boundaries with predefined tree shape priors or active contours. For example, Straub (2003) segments aerial images and height models into consistent regions at multiple scales, then performs refinement with active contours. Recent work in urban environments (Lafarge and Mallet, 2012) creates 3D city models from dense aerial LiDAR point clouds, and reconstructs not only trees but also buildings and the ground surface. After an initial semantic segmentation with a breakline-preserving MRF, 3D templates consisting of a cylindrical trunk and an ellipsoidal crown are fitted to the data points. Similarly, tree trunks have been modeled as cylinders also at smaller scales but higher resolution, using LiDAR point clouds acquired either from UAVs (Jaakkola et al., 2010) or from terrestrial mobile mapping vehicles (Monnier et al., 2012).

We are aware of only one recent approach for urban tree detection that, like our method, needs neither need height information nor an infra-red channel. Yang et al. (2009) first roughly classify aerial RGB images with a CRF into tree candidate regions and background. Second, single tree templates are matched to candidate regions and, third, a hierarchical rule set greedily selects best matches while minimizing overlap of adjacent templates. This detection approach (tree species recognition is not addressed) is demonstrated on a limited data set and it remains unclear whether it will scale to entire cities with strongly varying tree shapes.

Tree species classification from remote sensing data either uses multi-spectral aerial (Leckie et al., 2005; Waser et al., 2011) or satellite images (Pu and Landry, 2012), hyperspectral data (Clark et al., 2005; Roth et al., 2015), dense (full-waveform) LiDAR point clouds (Brandtberg, 2007; Yao et al., 2012), or a combination of LiDAR and multispectral images (Heikkinen et al., 2011; Korpela et al., 2011; Heinzel and Koch, 2012). Methods that rely on full-waveform LiDAR data exploit species-specific waveforms due to specific penetration into the canopy, and thus different laser reflectance patterns, of different tree species; whereas hyperspectral data delivers species-specific spectral patterns. Most works follow

[4] https://www.opentreemap.org/latreemap/map/