ARTICLE IN PRESS

Ecosystem Services xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Ecosystem Services

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

Using image recognition to automate assessment of cultural ecosystem services from social media photographs

Daniel R. Richards [a,*], Bige Tunçer [a,b]

[a] ETH Zurich, Future Cities Laboratory, Singapore-ETH Centre, Singapore
[b] Architecture and Sustainable Design, Singapore University of Technology and Design, Singapore

ARTICLE INFO

Article history:
Received 8 March 2017
Received in revised form 2 September 2017
Accepted 6 September 2017
Available online xxxx

Keywords:
Machine learning
Recreational ecosystem services
Singapore
Urban ecology
Recreation

ABSTRACT

Quantifying and mapping cultural ecosystem services is complex because of their intangibility. Data from social media, such as geo-tagged photographs, have been proposed for mapping cultural use or appreciation of ecosystems. However, manual content analysis and classification of large numbers of photographs is time consuming. This study develops a novel method for automating content analysis of social media photographs for ecosystem services assessment. The approach applies an online machine learning algorithm – Google Cloud Vision – to analyse over 20,000 photographs from Singapore, and uses hierarchical clustering to group these photographs. The accuracy of the classification was assessed by comparison with manual classification. Over 20% of photographs were taken of nature, being of animals or plants. The distribution of nature photographs was concentrated around particular natural attractions, and nature photographs were more likely to occur in parks and areas of high vegetation cover. The approach developed for clustering photographs was accurate and saved approximately 170 h of manual work. The method provides an indicator of cultural ecosystem services that can be applied rapidly over large areas. Automated assessment and mapping of cultural ecosystem services could be used to inform urban planning.

1. Introduction

Ecosystem services are the benefits that nature provides to people, including provisioning, regulating and cultural services (Millenium Ecosystem Assessment, 2003). To integrate ecosystem services within environmental decision making, we require information on a broad range of these services, so as to identify trade-offs among different objectives (Fish, 2011). Cultural ecosystem services are the non-material benefits that nature can provide, including recreational, spiritual and heritage values (Hernández-Morcillo et al., 2013). Quantifying cultural ecosystem services has traditionally been a time-consuming process, involving interviews (Plieninger et al., 2013), focus groups (Norton et al., 2012), or social surveys (Pleasant et al., 2014). More recently, data extracted from social media, such as geo-tagged photographs, have been used as indicators of recreational and aesthetic cultural ecosystem service value (Casalegno et al., 2013; Gliozzo et al., 2016; Keeler et al., 2015; Nahuelhual et al., 2013; Richards and Friess, 2015; Tenerelli et al., 2016; van Zanten et al., 2016; Wood et al., 2013). Data from social media offer great potential for quantifying ecosystem services rapidly and over large areas, though for media such as photographs this still involves human supervision, making it a time-consuming process (Richards and Friess, 2015). This study presents an approach for automating the content analysis of social media photographs to facilitate rapid and large-scale ecosystem service mapping.

Previous studies that have used social media data in ecosystem service mapping have mapped the density of geo-tagged photographs as a proxy for public interest in an area (Casalegno et al., 2013; Keeler et al., 2015; van Zanten et al., 2016; Wood et al., 2013). Such studies have demonstrated that social-media data can be valuable in helping to map cultural ecosystem services over large areas, but there is a wealth of additional information in social media that is currently under-used. To extract more from the resource provided by social media, it is necessary to analyse the content of photographs, firstly to ensure that the photographs are relevant to the natural environment, and secondly to understand what aspects of the environment are of most interest to people in a particular area (Richards and Friess, 2015).

The density of photographs in a location corresponds closely to its popularity with visitors (Wood et al., 2013), but does not necessarily relate to public interest in the environment. The presence of a photograph does not tell us why people visited a location; for

* Corresponding author.
E-mail address: richards@arch.ethz.ch (D.R. Richards).

http://dx.doi.org/10.1016/j.ecoser.2017.09.004
2212-0416/

example, there may be a high density of photographs because of high public interest in a natural feature, or due to interest in a nearby popular restaurant. Photograph density as an index of cultural ecosystem service quality may be particularly confounded in urban areas, because recreational green spaces are found in close proximity to built infrastructure with high recreational potential, such as malls, cinemas, and restaurants. The risk of overestimating public interest in ecosystems can be reduced by discounting photographs that fall over non-natural areas (Gliozzo et al., 2016). However, this approach requires a complete knowledge of where those areas are, which may be difficult to obtain in heterogeneous urban-rural landscapes. In urban areas, many human-nature interactions occur in very small spaces, including community or private gardens (Marco et al., 2010; Seik, 2000). Such spaces may be hidden within larger areas that are assumed to be urban, and excluding all photographs from these areas could undervalue the importance of nature. The risk of overestimating public interest in the environment can also be limited by applying filters to the textual content (''tags'') associated with each image (van Zanten et al., 2016). However, photograph tagging relies on classification of the images by users, some of whom are less diligent in recording a broad range of tags, or may not record any tags at all.

There is a wealth of information held within photographs of the environment that can be analysed alongside social surveys and interviews to infer how, and why, people interact with nature (Dorwart et al., 2009). The density of nature photographs in a place can be considered as an indicator of public interest in nature in that area, which to some extent relates to recreational and cultural heritage ecosystem services (Richards and Friess, 2015). However, interpreting the information held within photographs can be a challenge for research into cultural uses of the environment, because the choice of what to take a photograph of is subjective on behalf of the photographer, and the motivation for taking the photograph may be unclear without additional contextual information (Scott and Canter, 1996). Despite the caveats in analysing the content of photographs, images from social media websites have previously been analysed by manual classification, to assess the relative importance of different cultural and recreational uses (Richards and Friess, 2015; Thiagarajah et al., 2015). Manual analysis of photographs can be consistent between assessors, and can be relatively rapid at small spatial scales; indeed, a sufficient analysis of one natural site can be completed in around 30 min (Richards and Friess, 2015). However, manual classification of photograph content does not scale up easily, as the time investment required to compare a large number of sites would be substantial. To allow rapid assessment of cultural ecosystem services over large areas, we require automated content analysis of photographs from social media.

The capability of image recognition software has improved rapidly over the past few years, spurred by the availability of large image datasets and high-power cloud computing (Agrawal et al., 2015; Kwak and An, 2016). Generally-applicable image recognition algorithms are now accessible online, such as Google's Cloud Vision, and Microsoft's Computer Vision (Google Cloud Vision, 2017; Microsoft Computer Vision, 2017). Google Cloud Vision, launched in 2016, provides an Application Programming Interface (API) for image recognition, allowing users to analyse individual images for their content, which is described in terms of keywords (Google Cloud Vision, 2017). Generalised image recognition APIs are now being applied to analyse the content of large number of images for research purposes, for example to categorise the content and gender balance of images in the global news (Kwak and An, 2016).

Automated content analysis of photographs from social media could help to ensure that only relevant photographs are included in indices of cultural ecosystem service quality, and may also provide more information, by distinguishing between different cultural ecosystem services, or aspects of the environment that are of interest (Di Minin et al., 2015; Richards and Friess, 2015). This study demonstrates a novel method for analysing photographs taken from social media with the Google Cloud Vision image recognition algorithm, and applies this approach across the city-state of Singapore. The objectives of the study were (1) to quantify the occurrence of photographs relating to nature, (2) to analyse the drivers of spatial variation in nature photographs, and (3) to assess the accuracy of the automated classification by comparison with a manual classification conducted by a human.

2. Method

2.1. Cultural ecosystem services in Singapore

Singapore is an island city-state of 5.6 million people, located in Southeast Asia (Department of Statistics, 2016). The population is relatively wealthy and there is a high rate of mobile phone ownership (IDA, 2014), making it a suitable country for analyses of social media data (Richards and Friess, 2015). While largely urban, around 56% of the land area is covered in managed and spontaneous vegetation, including public parks, gardens and nature reserves (Yee et al., 2011). Recreation is an important use of green space in Singapore, with a number of designated parks and gardens as well as networks of footpaths (Henderson, 2013; Tan, 2006).

2.2. Extraction of images and image recognition

Flickr is a photograph-sharing website with over 70 million users and 200 million geo-tagged photographs (Wood et al., 2013), which has become a commonly-used source of social-media photographs for assessing cultural ecosystem services (Richards and Friess, 2015; van Zanten et al., 2016; Wood et al., 2013). To download the photograph population across mainland Singapore, Sentosa Island, and Ubin Island, we laid out a regular grid of 165 locations and extracted all geo-tagged photographs from the social media website Flickr. The sample locations were arranged on a 2 km by 2 km grid. A random subset of 25000 photographs from Flickr (approximately 20% of the total) was then analysed further using image recognition. A random subset of all photographs was used, rather than stratifying the random sample by photographer. As such there is likely to be some bias in the data, as photographers that uploaded large numbers of photographs to Flickr were more likely to be represented. Each image was sent to the Google Cloud Vision API, which used a machine learning algorithm to assign keywords to the images (Google Cloud Vision, 2017). The Google Cloud Vision API was accessed through the RoogleVision package for the R statistical programming language (Teschner, 2016). A maximum of five keywords were returned for each image (Teschner, 2016).

2.3. Hierarchical clustering of images into general categories

The Google Cloud Vision image recognition algorithm returned up to five keywords that were associated with each photograph. A hierarchical clustering algorithm was applied to group the photographs according to their keywords (Oteros-rozas et al., 2017). A distance matrix was generated by comparing the proportion of the keywords assigned to one photograph that did not match the keywords assigned to another photograph. Hierarchical clustering was then applied using Ward's distance, as implemented in the hclust function for the R statistical programming language (R Core Team, 2015). The appropriate number of clusters for analysis was assessed by plotting the average difference between the