Aggregating Visual Evidence from Social Media Photos to Monitor the Natural World

ABSTRACT

Social photo-sharing websites collect a huge amount of latent visual information about the world, including information about the environment and ecology. In this work, we propose to reconstruct satellite maps of environmental status across North America through millions of publicly available geo-temporal tagged images. We apply modern deep learning-based recognition techniques to identify phenomena in images, and then aggregate evidence from multiple users to estimate whether or not the phenomena were occurring in a given time and place. We then evaluate the accuracy of these estimates by comparing to actual satellite maps as ground truth. As test cases, we consider two important ecological phenomena for which high quality ground truth is available: snowfall coverage and vegetation (greenery) coverage. We find that while the automatic recognition techniques are noisy on any single particular image, we can accurately estimate the phenomena's presence when enough users have uploaded enough photos at a particular time and place. This evidence from photo-sharing websites could create new sources of data for ecologists, perhaps helping to overcome the limitations of traditional data collection techniques like manual observation (which is labor intensive) or satellites (which are not able to observe through clouds).

Keywords

ACM proceedings; LaTeX; text tagging

1. INTRODUCTION

Monitoring the meteorology and vegetation phenomenon is the cornerstone and challenge of ecology and biology research. Expensive satellite images give large scale data but struggle with cloud cover, atmospheric conditions and fine-grained localization such as flower species distribution, human interaction with nature, while citizen science provides high quality data but is also costly and is very difficult to practice over large scale areas. The enormous popularity of photo-sharing website collects images in large spatial scale, from under clouds and in close focus (compare to aerial surveillance), moreover, they are freely accessible to the public. The more than 300 million images uploaded to social media every day [**?**] potentially contain not only human activities, but also outdoor ecology and biology information intentionally and incidentally as shown in Figure 2.

The idea of reproducing satellite maps has become more and more interesting to scientists applying textual mining on **FiXme Note: citestock, ecology, election, tourists**, and recently to computer vision researches directly deriving **FiXme Note: citetemperature, cloud, mountain peak** information from visual content. In this paper, we test the feasibility of leveraging these noisy and biased images as a new approach to observe nature. We study 2 particular phenomena, snowfall and vegetation coverage as they are fundamental topic in ecology and biology study, have relatively distinct appearance to recognize, have a good chance to appear in social media, and also have satellite maps available to serve as ground truth. Our approach is illustrated in Figure 1. First, we collect a large hand-labeled data set of the existence or absence of ecology phenomena. Then, we train a classifier for each phenomenon by combining its most discriminative visual features and by using deep learning features. Finally, we collect 12 million images from entire North America over 2 years, make prediction on geo and temporal scale by aggregating this visual evidence.

This paper is built on our earlier work **FiXme Note: citewww** analyzing ecology phenomenon from image tags only. We apply a new approach understanding visual content of images, and run experiments on the exact same data set to study how vision techniques could help in social media data mining compared to using textual data alone. Also, to our best knowledge, among all the research works performing social sensing with image data, this is the first one providing continental scale quantitative performance evaluation.

2. RELATED WORK

In last few years, crowd-sourcing data from social media as a large scale and free to public data source has **FiXme Note: received lots of attention from; or (become more and more popular to)** researchers working on using textual contents to predict elections [**?**], using geo-tags to quantify tourism in nature area using geo-tag profile of social media users [**?**], **FiXme Note: talk a lot about motivation of scientific report paper since it's in nature area** to draw coastline [**?**], using geo and temporal tags to analyze people's event-based activity when large group

DOI: 10.1145/1235

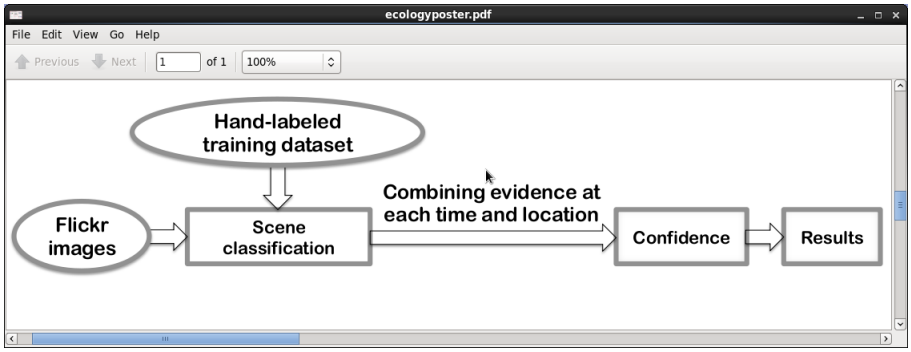

Figure 1: Overview of our approach to apply image classifiers on large scale images and make prediction by aggregating these visual evidence. FiXme Note: first classifier, then prediction

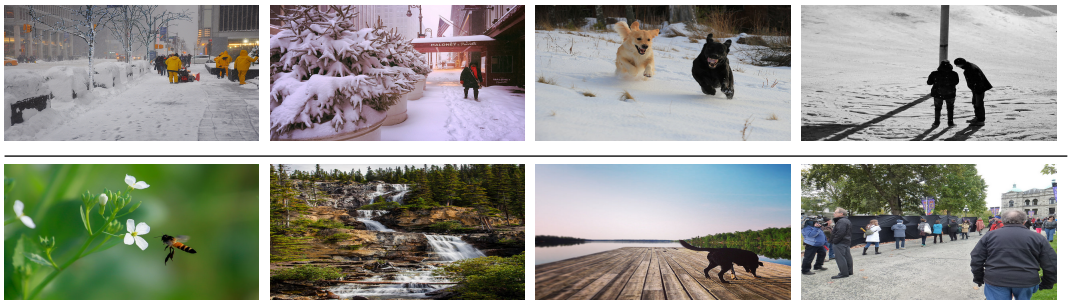

Figure 2: Flickr image examples capture snow and greenery evidence on purpose and as background.

of people gathering together during a function of time such as football match, and using both geo and textual tags to extract land use information from Panoramio [?, ?], in **FiXme Note: citewww**, Zhang *et al.* stimates **FiXme Fatal: missing letter in compiling** snowfall and vegetation coverage based on geo, temporal and textual tags of Flickr images. **FiXme Note: accuracy of geo-temporal problem, and now it's getting better.**

Since public-sharing photos provides such a huge potential in social and environmental study, it's natural to see a lot of works start analyzing image contents. Webcam providing dense temporal images is a good source to monitor the nature. A series of works explore sequences of webcam images describing outdoor scene with 40 transient attributes [?], estimating dynamic cloud maps [?, ?], exploring interactions between visual elements and the temperature **FiXme Note: or just as the title: exploring correlations between appearance and temperature** [?], and monitoring the dynamic snow phenomena at mountain areas [?, ?]. To evaluate the study of temperature, cloud, and snowfall amount, researchers can easily compare their results with satellite maps. Some works also use crowd-sourcing data from other sources, for example, Google street view provides selectively dense geo distributed images to help navigating the environment [?] and understanding urban scene and predicting urban perception [?], and Li *et al.* se the co-occurrence statistics of celebrities appears on news images to auto tag photographs of celebrity community [?] **FiXme Note: give a term like social identity?**. Unfortunately, the evaluation in these works are either not in continental scale or just via quality visualization. Performance of social activity studies, on the other hand, are even harder to evaluate. **FiXme Note: say more about our evaluation? or move this to another place?**

Flickr and Panoramio as very popular photo-sharing websites "involuntarily" support researchers identifying salient city attributes and analyzing the visual similarity among different cities in order to apply computer vision to urban planning [?] Photo-sharing websites collecting visual contents directly from people's activity and their surrounding areas which is so important, hard to collect otherwise but also very noisy. **FiXme Note: write something about so there are very few work appears and so we are working on this?**

The fact that webcam can only be placed far away from people makes it almost impossible to monitor people's activity, even not the surrounding area close to residential or **FiXme Note: crowd? I mean groups of people like downtown, not ski activity but like people going to work and back everyday also a good topic to use temporal dense images but Webcam is not good at this.** Social media, on the other hand, provides a larger freedom on location distribution. In fact, as a complementary, almost all the photos shared online are from locations people usually go to. **FiXme Note: how helpful is this to study more areas close to urban planning, market sharing, everyday living, anything related to people**

Our work take the advantage of studying ecology phenomena with **FiXme Note: easy to get, more reliable** satellite maps as ground truth and use social media data to **FiXme Note: monitor? insight?** these information from **FiXme Note: locations more related to people**. We

provide continental scale quantitative evaluation and introduce our method to tackle the problem of noisy and biased data, in order to support extended studies in other areas. **FiXme Note: just want to say more areas in natural or not only natural but also social**