Tracking Natural Events through Social Media and Computer Vision

Jingya Wang¹
wang203@indiana.edu
Mohammed Korayem¹²
mkorayem@indiana.edu
Saúl A. Blanco¹
sblanco@indiana.edu
David J. Crandall¹
djcran@indiana.edu

¹ School of Informatics and Computing
Indiana University
Bloomington, Indiana
² CareerBuilder, LLC
Atlanta, Georgia

1 Introduction

Recent dramatic improvements in object and scene recognition raise the potential for new applications to help organize and extract latent knowledge from large-scale photo collections. Consider one particular example. Monitoring the state of the natural world over time and geographic space is crucial for a variety of scientific fields. Satellites can observe at a large scale but only for phenomena that can be seen from far above, like movements of weather systems but not migration of birds or flowering of flora. Satellite observations are also affected by clouds and other atmospheric conditions [11]. A potentially rich alternative data source is to mine consumer photos from geo-tagged, time-stamped, large-scale public social media collections for evidence of natural events, because photos often contain evidence (either on purpose or accidentally) of the state of the natural world.

A few recent papers have explored how “big visual data” [12] could create novel data sources to complement more traditional data collection techniques. For example, Zhou et al. [15] and Lee et al. [8] estimate demographic, geographic, and other properties of physical places based on social image data. Wang et al. [13] try to recognize snowfall in images to monitor snow storms, but use hand-crafted features and rely heavily on text tag features. Other work has used data from webcams to monitor changing weather and other natural properties [3, 9].

In this extended abstract, we summarize our work that tests the feasibility of using noisy image collections to observe nature using modern deep learning-based computer vision to recognize visual content automatically [14]. As a case study, we investigate two particular phenomena: continental-scale snowfall and vegetation coverage. We first collect millions of geo-tagged, timestamped, public photos from Flickr, and daily snow and weekly vegetation satellite maps for North America. By cross-referencing the photo geo-tags and timestamps with the maps, we automatically label each image with whether or not it was taken in a place with actual snow or green vegetation. We then train state-of-the-art Convolutional Neural Networks and Support Vector Machines to recognize these phenomena in individual images. Of course, these classifiers are imperfect, in part because social image data is noisy with inaccurate timestamps and geo-tags, and the satellite data is also incomplete. We thus train an additional classifier that aggregates evidence from multiple images taken at a given time and place, yielding more accurate observations. We evaluate at a large scale, training and testing on millions of Flickr images and quantitatively evaluating the performance at hundreds of thousands of places and times. Finally, we present a tool to visualize the combination of satellite and social photo-derived observations. The tool is general and can be applied to a wide range of phenomena with minimal additional effort.

This extended abstract is a summary of our recent paper at ACM Multimedia [14], that we feel would also be of interest to the audience at the Web-scale Vision and Social Media workshop.

2 Data

We collected images geo-tagged in North America and time-stamped between 2007 and 2015 using Flickr’s public API (similar to [2]). We removed photos with inaccurate geo-tags time-stamps yielding 77.6 million images. Throughout our experiments, we used the 2007–2010 data for training and reserved 2011–2015 as a separate test set.

For the ground truth for training and testing, we used NASA’s Terra satellite [1, 4, 11], which gives daily snow and bi-weekly vegetation cover maps gridded into $0.05\degree \times 0.05\degree$ latitude-longitude bins (roughly 5km $\times$ 5km at the middle latitudes). Unfortunately, this data is neither complete nor fully accurate, primarily because many satellites cannot make accurate observations through clouds. For each day and each bin (which we call a “day-geobin”), the satellite data records the percentage of the bin that was visible, the percentage of the visible area that was covered by snow or greenery, and confidence scores. To identify day-geobins with reliable ground truth, we excluded low-confidence bins, computed a probability as a function of the snow (or greenery) and visibility percentages, and labeled those below 0.15 as non-snow (or greenery) day-geobins, and over 0.85 as snow (or greenery) day-geobins. The remaining day-geobins were ignored.

3 Method

We investigate two specific types of conditions: (1) whether there was snow on the ground, and (2) whether there was green vegetation. Both of these properties change over time and over geospatial location on Earth. To do this we require two key steps: deciding whether or not there is evidence of snow or greenery in an individual image, and then integrating this (very noisy) evidence across multiple images to estimate the actual real-world natural state at that time and place.

**Image classification.** In training, we consult the satellite data to find all labeled day-geobins and label all these images as positive or negative exemplars, respectively. It is very noisy but it permits cheap, scalable training with little human effort.

We consider two types of features: text tags and visual content. For text tags, we built a vocabulary consisting of the 1,000 most frequent tags in the training set and represented each image as a 1000-d binary vector indicating presence or absence of each tag. We then trained a linear Support Vector Machine [6] to predict whether or not the tags have evidence of the event. For visual features, we learned a model using Convolutional Neural Networks (CNNs), which are the state-of-the-art in image classification [7]. We used the AlexNet network architecture and the Caffe open-source software framework [5], and followed the popular procedure of initializing CNN weights based on a network trained on ImageNet, and then fine-tuning using our training set [10].

**Aggregating evidence.** We combine classification results from multiple images taken at the same time and place, taking into account the image classifier’s confidence. In particular, for each day-geobin, we build a histogram of quantized confidence scores, recording how many of the photos were classified as snow and non-snow (or green/non-green) at 20 quantized confidence levels. While this improves results compared to considering single images, it suffers from the problem that users with many photos have a disproportionate influence. We thus build a histogram over users instead of photos, so that each of the 20 histogram bins counts how many users took at least one photo at that confidence level. We then trained an SVM to estimate environmental state from these histograms.

4 Experiments and evaluations

We trained classifiers using data from North America for the years 2007 to 2010. To make results more easily interpretable and to prevent problems with unbalanced classes, we randomly sampled from the larger class to yield a roughly equal number of positive and negative exemplars for each event. For snow, there were 626,522 such photos taken by 49,462 distinct users in 87,586 distinct day-geobins; for vegetation, there were 645,694 photos by 35,510 users in 84,921 day-geobins. We tested using data from

2011–2015, again balancing the classes, for a total of 577,186 test images for snow and 769,992 for vegetation.

**Individual image classifier.** We first tested accuracy on the individual image classification problem. The tag features achieve 63.0% accuracy for snow and 67.5% for vegetation, compared to random baselines of 50.0%. Visual features, in contrast, performed at 69.2% accuracy for snow and 80.5% for vegetation. A visualization of some sample visual classification results are shown in Figure 1.

**Day-geobin classifier.** Our accuracy on this task for snow was about 60.8% for textual features alone, 69.3% for visual features, and 71.7% for the combination of concatenating visual features and textual features, compared to 50.0% random baseline; for vegetation, accuracies were 71.3% for tags, 79.4% for visual features, and 81.9% for the combination.

We have observed that most incorrectly detected day-geobins occur in places with very few observed photos contributed by few users, where the classifier is basing its entire decision on very little evidence. Figures 2(a) plot ROC curves for snow as a function of the number of distinct users in each day-geobin; vegetation curves are not shown due to space constraints, but the trend is similar. Increasing the number of distinct users improves accuracy dramatically, up to nearly 95% for 10 users. In many applications, it may be more important for scientists to retrieve places and times when specific users contributed photos, rather than broad regions, and the number of good users will be relatively small. Figures 2(b) shows precision-recall curves that adopt this retrieval view. At 80% recall, precision nears 90% even for day-geobins with single users, and reaches 99% for 20 users.

**Browsing and visualization tool.** We have developed a web-based tool\(^1\) that allows users to explore and compare satellite and social media data. Users can click on any geobin of interest to see photos taken at that time and place, organized by distinct user, and the visualization also shows the classification results estimated for each image.

5 Conclusions

We presented a technique and visualization tool for combining automatic image analysis of public Flickr photos with satellite maps for tracking natural events. We considered snow and vegetation as test cases, but the automatic classification techniques and visualization tools are general enough to be applied to a wider range of events. We hope our work inspires further interest in using social photo collections and computer vision as a novel source for environmental data.

6 Acknowledgments

This work was supported in part by the National Science Foundation through CAREER grant IIS-1253549 and the IU Data-to-Insight Center, and used compute facilities donated by NVidia. We thank Dennis Chen and Alex Seewald for assisting with initial data collection and system configuration.

References


\(^1\) *http://vision.soic.indiana.edu/snowexplorer/*

Figure 1: Classification results on random images from times and places where satellites reported snow (top), no snow (second row). Images are ordered according to the classifier’s confidence, from highly certain of absence (left), to uncertainty either way (middle), to highly certain of presence (right). Faces obscured for privacy.

Figure 2: Performance on estimating snow presence for about 98,000 North American day-geobins from 2011–2015, in terms of (a) ROC and (b) Precision-Recall, as a function of number of distinct users per bin.