Exploring outdoor appearance changes with transient scene attributes

Pierre-Yves Laffont, James Hays
Brown University*

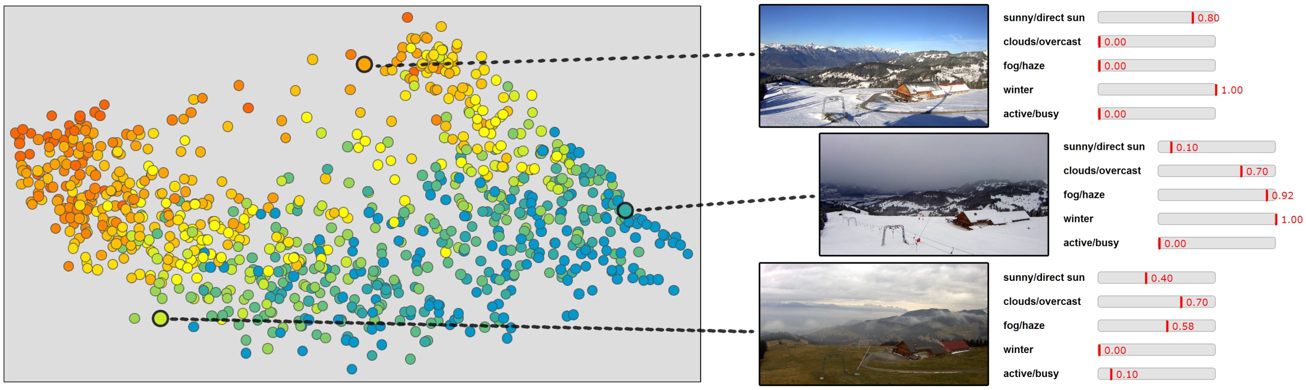

Figure 1: *Each image in a photo collection is represented as a point in **attribute space**, where each dimension corresponds to a scene property which can vary with time, weather, or lighting conditions. Left: projection of all images on the dominant plane of attribute space; each image is represented as a dot, color-coded according to its value of the "sunniness" attribute. Right: values of a few **transient attributes** for three photographs. The scene appearance and its attributes vary widely between the three images, despite the fixed viewpoint.*

The appearance of outdoor scenes changes dramatically with lighting and weather conditions, time of day, and season. We relate visual changes to ***scene attributes***, which are human-nameable concepts used for high-level description of scenes. They carry semantic meaning and are more flexible than a categorical representation of scenes. While the discriminative scene attributes proposed in [Patterson and Hays 2012] distinguish scenes from each other, we focus on ***transient attributes*** which describe changes in appearance within each scene under real-world conditions.

Using online webcams to gather many photographs of outdoor scenes, crowdsourcing to collect human annotations, and machine learning to train classifiers, we:

- discover which attributes are likely to vary among images of an outdoor scene,
- learn to recognize significant attributes in outdoor pictures,
- propose a user interface to browse collections of photographs, based on transient attributes.

Our Approach

Since we are interested in appearance changes of outdoor scenes, we focus on images captured by static webcams over several months. We gather images from 35 webcams, and extract 60-120 high quality frames which are representative of the appearance variations of each scene.

Discovering transient scene attributes. We conduct a crowdsourcing experiment on Mechanical Turk to find out which attributes can describe the changes of appearance of one scene. We first collect a list of adjectives and nouns frequently recurring in descriptions of outdoor scenes. We define a candidate list of 92 scene attributes: while some appeared in prior work, such as *spatial envelope* properties ("natural", "enclosed area"), we add attributes related to lighting ("daylight", "sunrise"), weather ("snow", "warm") and season.

In each crowdsourcing task, we show workers images of a single outdoor scene. We propose some attributes in our list, along with their definition, and ask which ones appear in "all / some / none" of the images. Results show that most of the scene attributes described in prior work do not vary much across images of one scene. However, properties related to weather, lighting, or emotions induced by the viewing, can vary drastically across images of one scene. Examples of such ***transient attributes*** are "daylight", "sunrise/sunset", "sunny", "fog/haze", "winter".

Recognizing scene attributes. We adopt an approach based on machine learning to recognize the presence of transient attributes in an image. We conduct a second crowsourcing experiment to collect annotations for 19 attributes and images of 10 webcams. In each task, workers are asked to rank photographs of a single scene according to how much each image exhibits a particular attribute; possible answers are "totally / a little / not at all". Up to 7 annotations are gathered for each image/attribute pair to establish consensus.

We create an individual SVM classifier for each attribute using labeled data collected with our crowdsourcing experiment. We consider that an attribute is present in an image when $\alpha > 0.7$ and absent when $\alpha < 0.3$, with $\alpha = 0.5 + \frac{P-N}{2C}$ where P is the number of "totally" annotations, N the number of "not at all" annotations, and C the number of annotations different from "unsure" for this image. Following [Patterson and Hays 2012], we train each SVM classifier using an average kernel generated from gist, HOG 2x2, self-similarity, and geometric context color histogram features.

We train our classifiers using random splits of our images from 10 different webcams, and evaluate performance on a testing set consisting of 20% of our annotated images. We obtain a high average precision (AP) score of 0.92, averaged over all attributes.

We further evaluate the performance of our classifiers by using all images of 8 webcams as our training set, and using all images of 2 separate webcams as our testing set. We obtain an AP score of 0.76, which shows that our classifiers generalize moderately well to scenes that were not seen in the training data. Performance is likely to increase with more labeled training data.

Browsing photo collection with transient attributes. We show an application of transient attributes for browsing a photo collection. In Figure 1 (left), images are represented as colored dots and laid out according to their attribute values. Clicking on a specific dot displays the corresponding image and attribute values (right).

References

PATTERSON, G., AND HAYS, J. 2012. Sun attribute database: Discovering, annotating, and recognizing scene attributes. In *CVPR*.

*e-mail: {pylaffont, hays}@cs.brown.edu