this paper makes the first attempt to address the two-class weather classification problem from a single outdoor image.

it's an easy task for human, but it's challenging for computer. naive schemes based on image brightness or color statistics are failed in this problem.

the first contribution of this paper is proposing various weather cues to form the weather feature. These everyday weather cues (sky, shadow, reflection, contrast, haze) are what human are still using for weather observing. It's a 621-D feature vector.
since not all of these cues are necessarily present in a given outdoor image, so there is the existence vector, where each of the scalar score in [0, 1] indicates the confidence in the corresponding weather cue being present in the given image. since image contrast is present in both sunny and cloudy photos, therefore it is not included.

THE first one, Sky. the sky is the most important cue for weather labeling. firstly, the sky region is detected in a pixel-wise manner. this paper collects 20 thousand sky and non-sky patches, and extract a 131-D feature which contains the SIFT descriptor and mean HSV color. then random forest classifier is learned on the two patch classes.
THE second one, Shadow. Hard shadow boundaries present another useful cue as they are often found in outdoor photos shot in sunny days. Unlike detecting sky in an image, shadow detection is still a challenging problem. they apply the shadow detection tool to rank the resulting shadow boundary confidence scores and take the 10th highest score.
THE third one, Reflection. Strong sunlight reflected from shiny objects is another powerful cue. they apply image matting at the detected white pixels, then compute the alpha matte distributions to construct this feature.
THE fourth one, Contrast. they collect all saturation percentile ratios to build this feature.
THE last one, Haze. Cloudy weather may come with haze. haze priors have been well studied in computer vision, the dark channel prior is effective to form the haze feature.

Traditional classifiers such as SVM cannot achieve good performance on our weather feature because they assume all components to present simultaneously in all images, which unfortunately may not be correct in many tasks. For example, in our setting, outdoor images do not always contain the sky region. So the learning strategy is to partition training images into disjoint clusters of homogeneous voters. Given a test image, voters closer to it are given more weights for correctly finding the weather label. The partitioned sets correspond to different weather cue patterns, such as “sky + shadow”, “sky + haze”.

In the testing phase, given a weather feature x with existence vector e, we rely on training data whose existence vectors are similar to e. Following this idea, the classifier is based on weighted voting scheme. sign is the function outputting 1 for non-negative input. s is a similarity function. and h i hat is the trained homogeneous voter.

The weather dataset contains 14 thousand labelled images. the label is either sunny or cloudy, half and half. the dataset has 3 sources.

This work is the first attempt in dealing with weather labeling given single images, apparently, this problem is far from solved. we could add more useful weather cues in this framework. And this approach is limited in labeling two weather types, we could generalize to labeling more weather types. In addition, This approach rely on handcrafted features, may adopt some deep learning based approaches to improve it.