Image and Vision Computing 52 (2016) 154–166

Contents lists available at ScienceDirect

Image and Vision Computing

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

Cloudmaps from static ground-view video ☆

Nathan Jacobs [a,*], Scott Workman [a], Richard Souvenir [b]

[a] Department of Computer Science, University of Kentucky, Lexington, KY40506, United States
[b] Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC28223, United States

ARTICLE INFO

Article history:
Received 12 August 2015
Received in revised form 15 March 2016
Accepted 27 May 2016
Available online 22 June 2016

Keywords:
Image formation
Time-lapse
Clouds
Lighting estimation
Solar forecasting
Scene factorization

ABSTRACT

Cloud shadows dramatically affect the appearance of outdoor scenes. We describe three approaches that use video of cloud shadows to estimate a cloudmap, a spatio-temporal function that represents the clouds passing over the scene. Two of the methods make assumptions about the camera and/or scene geometry. The third method uses techniques from manifold learning and does not require such assumptions. None of the methods require directly viewing the clouds, but instead use the pattern of intensity changes caused by the cloud shadows. An accurate estimate of the cloudmap has potential applications in solar power estimation and forecasting, surveillance, and graphics. We present a quantitative evaluation of our methods on synthetic scenes and show qualitative results on real scenes. We also demonstrate the use of a cloudmap for foreground object detection and video editing.

1. Introduction

Clouds are a significant factor in determining the available solar energy and, consequently, have a significant impact on processes ranging from plant growth [1], solar power generation [2] and climate change [3]. For such applications, the most common method for assessing available solar energy relies on point source samples using specially designed solar radiation sensors. This can lead to inaccurate estimates, especially if the sensor is located far from the object of study. Vision-based methods have the potential to extend the coverage of such sensors, thereby increasing the accuracy of solar radiation estimates. However, recent vision-based methods for outdoor scene understanding either explicitly eliminate images captured on cloudy days [4–6] or only estimate a single scalar cloudiness parameter per frame [7–9].

The most prominent approaches for estimating cloud cover rely on sky cameras [10–14], which are imaging systems setup to capture a full view of the sky from a single location. In such systems the sun is often in the field of view, which causes artifacts. Therefore, these cameras often incorporate moving physical barriers to block the sun. The main drawback of these approaches is the cost associated with purchasing, maintaining, and deploying the equipment. Additionally, limitations inherent to a single viewpoint are unavoidable. A single viewpoint reduces the value of a sky camera when the clouds are near the ground or when they are vertically thick. In both cases, it is difficult to estimate the amount of cloud occlusion anywhere except directly toward the sun, which, in most setups, is blocked.

While sky cameras are rare, surveillance cameras and webcams are ubiquitous. These types of cameras are inexpensive to purchase, and often view very little sky but a wide area on the ground. We propose to use such cameras to estimate a *cloudmap*, a time-varying 3D function, defined in world coordinates, that describes the clouds passing over a scene (see Fig. 1). Our methods take advantage of shadows cast by moving clouds to simultaneously estimate a cloudmap and a mapping between image pixels and cloudmap coordinates. We estimate the amount of sunlight attenuation at each pixel in each frame and combine these individual pixel estimates into a globally coherent model.

Given a video of an outdoor scene, we first extract a scalar time series for each pixel describing sunlight attenuation, i.e., cloudiness. For each pair of pixels, we estimate the temporal delay between the corresponding time series, and then denoise the time series using the assumption that the time series of other pixels, directly in-line with the cloud motion direction, will be similar. We then estimate the scene model and cloud motion direction using one of three methods. Finally, the denoised cloudiness time series, temporal delays,

☆ This paper has been recommended for acceptance by Enrique Dunn.

* Corresponding author.
E-mail addresses: jacobs@cs.uky.edu (N. Jacobs), scott@cs.uky.edu (S. Workman), souvenir@uncc.edu (R. Souvenir).

http://dx.doi.org/10.1016/j.imavis.2016.05.013
0262-8856/

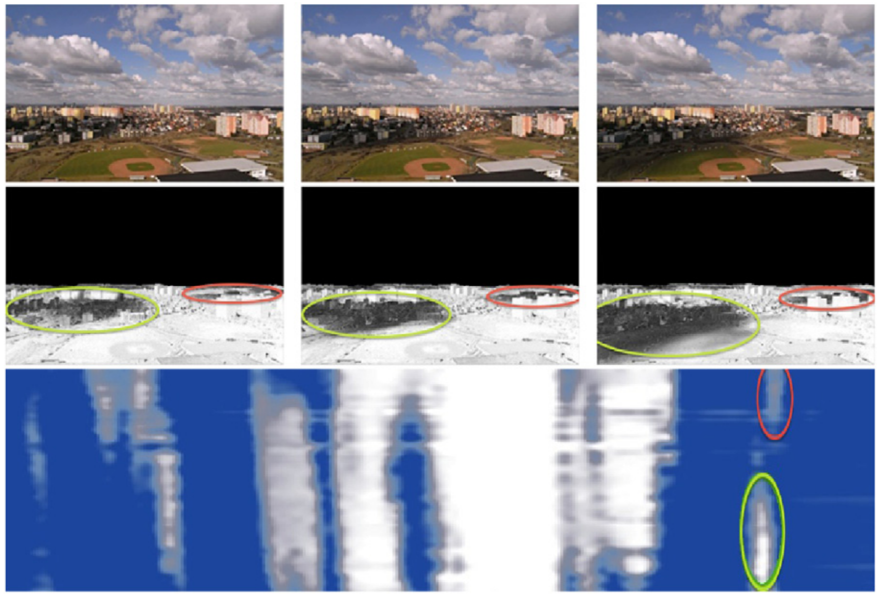

Fig. 1. Given a video from an outdoor static camera, we estimate the sunlight attenuation due to clouds (middle) by analyzing the time series of intensity changes in pixels from the ground. We introduce several methods, which do not require directly viewing the clouds in the sky, for combining these time series into a cloudmap (bottom), a geo-temporal function that describes the shape and thickness of set of clouds that passed over the scene. This cloudmap summarizes ~25 min of video and the clouds corresponding to the visible shadows in the image are outlined.

scene model, and cloud motion direction are used to estimate a cloudmap which is a time-varying 3D function. If required for an application, we render the cloudmap in 2D by making some assumptions about how the cloudmap changes.

This paper makes several contributions: three techniques for estimating cloudmaps from outdoor video, techniques for denoising and rendering cloudmaps, and a method for estimating the cloud motion direction from a video with known geometry. We present quantitative and qualitative results on a variety of scenes, as well as show several proof of concept applications.

2. Related work

Images have been used to estimate cloud cover for a variety of purposes, including image retrieval [15], graphics [16], weather estimation [8] and solar power forecasting [2, 17]. This paper extends our previous work [18], which was, to our knowledge, the first work to attempt to use ground shadows to estimate a cloud layer. Here, we describe work on several related problems.

2.1. Outdoor video understanding

Recent work has shown the benefits of explicitly reasoning about the underlying causes of outdoor appearance variations. For example, color changes due to sun motion are strong cues to scene shape and albedo [4, 5, 6, 19] and transient clouds and their shadows are strong cues to scene geometry [20, 21] and camera calibration [22]. Outdoor videos have also been used to estimate dynamic scene properties, including low-dimensional cloudiness models [7–9]. Our work is the first to construct highly detailed cloudiness models from video of an outdoor scene.

2.2. Shadows in outdoor scenes

Methods for detecting and handling shadows appear in several settings. Most similar to our work is in the surveillance domain [23, 24] where it is important to reduce the number of false-positive detections. These methods focus on modeling individual pixel variations and do not explicitly model the motion of the clouds in the scene. More sophisticated methods, such as Ref. [25], can detect shadows in individual frames, but are computationally intensive and are less robust because they do not take advantage of the available temporal information. Our approach takes advantage of the image time series to obtain accurate per-pixel, per-frame attenuation estimates and explicitly fits a model that accounts for cloud motion.

2.3. Estimating outdoor illumination

Estimating a cloudmap is a special case of the more general problem of estimating illumination conditions. Recent work in several areas has attempted to solve this problem from a single image of an outdoor scene. This is a challenging line of research, with many inherent ambiguities.

Lalonde et al. [26] estimate illumination conditions from a single outdoor image and obtain an estimate of the sun direction, but only consider three cloudiness states: clear, partly cloudy, and overcast. Li et al. [27] address the problem of estimating the degree to which each sky pixel in an image is occluded by clouds by directly observing the sky. They use a Gaussian mixture model to represent a set of simple color features combined with a Markov Random field for enforcing spatial coherence. Peng et al. [16] address the same problem, but incorporate a physically based sky appearance model [28]. Veikherman et al. [14] propose a tomographic approach to building a 3D model of clouds from imagery captured by a network of sky cameras. This approach is complementary to ours because it relies on a different imaging geometry.

Murdock et al. [8] propose a method for estimating cloud cover from a collection of ground-based cameras. The authors use many images and corresponding pixel intensities from existing satellite images to train camera-specific regression models that predict the cloudiness given a single image from the webcam. The scalar cloudiness estimates from simultaneously captured images from many webcams are interpolated to estimate a synthetic satellite image. A key difference in our approach is our generative model, which is necessary because satellite imagery of sufficient resolution is not