Ecological Informatics 23 (2014) 62–68

Contents lists available at ScienceDirect

Ecological Informatics

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

Camera-based visibility estimation: Incorporating multiple regions and unlabeled observations

Nathan Graves, Shawn Newsam *

Electrical Engineering & Computer Science, University of California, Merced, CA 95343 United States

ARTICLE INFO

Article history:
Received 31 January 2013
Received in revised form 13 August 2013
Accepted 19 August 2013
Available online 31 August 2013

Keywords:
Image processing
Atmospheric visibility
Multivariate linear regression
Regression trees
Semi-supervised learning

ABSTRACT

This paper investigates image processing and pattern recognition techniques to estimate atmospheric visibility based on the visual content of images from off-the-shelf cameras. We propose a prediction model that first relates image contrast measured through standard image processing techniques to atmospheric transmission. This is then related to the most common measure of atmospheric visibility, the coefficient of light extinction. The regression model is learned using a training set of images and corresponding light extinction values as measured using a transmissometer.

The major contributions of this paper are twofold. First, we propose two predictive models that incorporate multiple scene regions into the estimation: regression trees and multivariate linear regression. Incorporating multiple regions is important since regions at different distances are effective for estimating light extinction under different visibility regimes. The second major contribution is a semi-supervised learning framework, which incorporates unlabeled training samples to improve the learned models. Leveraging unlabeled data for learning is important since in many applications, it is easier to obtain observations than to label them. We evaluate our models using a dataset of images and ground truth light extinction values from a visibility camera system in Phoenix, Arizona.

1. Introduction

Atmospheric visibility can be a useful indicator of atmospheric pollution resulting from suspended particulates especially in drier climates. This coupled with the rapidly growing number of cameras in our ecosystem motivates image-based visibility estimation as an appealing complement to traditional means of monitoring air pollution since specialized equipment for measuring pollution is comparatively expensive. So-called visibility camera systems are already seeing widespread deployment. For example, the Interagency Monitoring of Protected Visual Environments (IMPROVE)[1] program has installed and maintains cameras in over two dozen national parks in the United States. In addition, regional air quality agencies[2] have deployed visibility camera systems in over 30 cities. More broadly, though, there are potentially tens of thousands of web, surveillance, traffic, and other cameras, which could be used to monitor atmospheric visibility and thus air pollution.

The work in this paper represents a step towards using multimedia data, in particular images from off-the-shelf cameras, to perform quantitative estimation of atmospheric visibility. We investigate image processing and pattern recognition techniques to derive prediction models of light extinction based on image content. Light extinction captures the joint effects of light scattering and absorption that result from particulates in the atmosphere.

Our major contributions are twofold. First, we demonstrate that models which incorporate scene regions located at different distances from the camera are more effective than models which incorporate only a single region. This result is due to the fact that far regions are not useful when visibility is relatively poor since they are not observable at all, and close regions are not useful when visibility is relatively good since there is not enough intervening atmosphere to reduce visual acuity by a measurable amount. Our second major contribution is a semi-supervised learning framework which incorporates unlabeled training samples to improve the learned models. Leveraging unlabeled data for learning is important since, in many applications, it is easier to obtain observations than to label them.

The rest of the paper is organized as follow. First, Section 2 discusses related work. The problem is formally defined in Section 3. Section 4 describes the general framework of our approach and Section 5 describes the evaluation dataset and methodology. Sections 6 and 7 describe the proposed methods for incorporating multiple image regions and incorporating unlabeled observations, including the experimental results. Section 8 concludes the paper.

2. Related work

There is a sizable body of work on the related problem of improving the fidelity of images taken under hazy or otherwise atmospherically

* Corresponding author. Tel.: +1 209 228 4167.
E-mail address: snewsam@ucmerced.edu (S. Newsam).

[1] http://vista.cira.colostate.edu/improve.

[2] http://www.hazecam.net, http://www.mwhazecam.net, http://www.wyvisnet.com.

1574-9541/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecoinf.2013.08.005

degraded conditions. This includes work by Narasimhan and Nayar on using physics-based models to improve a single image (Narasimhan & Nayar, 2003b, 2003c) and using multiple images of the same scene but under different conditions (Narasimhan & Nayar, 2001, 2002, 2003a); work by Schechner and colleagues on using polarization to improve one or more images (Namer & Schechner, 2005; Namer et al., 2009; Schechner et al., 2001, 2003; Shwartz et al., 2006); and work by (He et al., 2009) on using a dark channel prior to dehaze a single image. The objective of this paper, however, is to derive quantitative estimates of atmospheric visibility and so these works are not directly applicable.

There is a much smaller body of work on using images to measure atmospheric visibility. (Caimi et al., 2004) review the theoretical foundations of visibility estimation using image features such as contrast, and describe a Digital Camera Visibility Sensor system, but they do not apply their technique to real data. (Kim & Kim, 2005) investigate the correlation between hue, saturation, and intensity, and visual range in traditional slide photographs. They conclude that atmospheric haze does not significantly affect the hue of the sky but strongly affects the saturation of the sky, but they do not use the image features to estimate visibility. (Baumer et al., 2008) use an image gradient based approach to estimate visual range using digital cameras but their technique requires the detection of a large number of targets, some only a few pixels in size. This detection step is sensitive to parameter settings and is not robust to camera movement. Also, for ranges over 10 km, they only compare their estimates to human observations, which have limited accuracy. (Luo et al., 2005) use Fourier analysis as well as the image gradient to estimate visibility but they also only compare their estimates to human observations. (Raina et al., 2004) do compare their estimates to measurements taken using a transmissometer-like device but their approach requires the manual extraction of visual targets. The work by (Molenar et al., 2004) is closest to the proposed technique in that it is fully automated and the results are compared to transmissometer readings. However, their technique uses a single distant and thus small mountain peak to estimate contrast and thus is very sensitive to camera movement and is unlikely to perform well under varying visibility regimes.

In contrast to the works above, our approach is fully automated, does not rely on the detection and segmentation of small targets, is robust to modest camera movement, and performs favorably when compared to ground truth measurements acquired using specialized equipment.

In our previous work (Graves & Newsam, 2011), we compared different methods for computing image contrast as the basis for estimating visibility. We considered Sobel filters in the spatial domain, low-, band-, and high-pass filters in the frequency domain, and an image haze model based on the so-called dark channel prior (He et al., 2009). We concluded that Sobel filters worked best. This paper extends that work in two fundamental ways: 1) we consider multiple image regions using regression trees as well as multivariate linear regression (this was introduced in our earlier workshop paper (Graves & Newsam, 2012)); and 2) we investigate semi-supervised learning to incorporate unlabeled observations.

3. The problem

Our goal is to estimate visibility from a static image. Reduced visibility by the intervening atmosphere is mainly due to three factors: 1) light radiating from the scene is absorbed before it reaches an observer; 2) light radiating from the scene is scattered out of the visual pathway of an observer; and 3) ambient light is scattered into the visual pathway of an observer. The combined effect of the absorption and scattering is referred to as the total *light extinction*. The higher the light extinction, the poorer the visibility.

Light extinction is typically measured using a transmissometer (Betts, 1971; Lee et al., 1982). This device consists of a light source (transmitter) and light detector (receiver), generally separated by a distance of several kilometers, and assesses visibility impairment by measuring the amount of light lost over this known distance. The transmitter emits a uniform light beam of known constant intensity. The receiver separates this light from ambient light, computes the amount of light lost, and reports the extinction coefficient b_{ext}, which is commonly measured in units of inverse megameters ($1\ \mathrm{Mm}^{-1} = 1.0 \times 10^{-6}\ m^{-1}$).

Our goal is to measure b_{ext} using a camera instead of a transmissometer. We do this by noting that b_{ext} is inversely related to *atmospheric transmission t* through the exponential equation (Seinfeld & Pandis, 2006)

$$t = exp^{-b_{ext} r} \tag{1}$$

where r is the distance of the scene. Further, atmospheric transmission t can be related to the observed image I through (Fattal, 2008; He et al., 2009; Narasimhan & Nayar, 2000, 2002; Tan, 2008)

$$I = Jt + A(1-t) \tag{2}$$

where J is the scene radiance and A is the ambient (atmospheric) light. The first term on the right side of this equation is inversely related to the amount of light radiating from the scene that is scattered out of the visual pathway and thus increases with improved transmission. The second term is the amount of ambient light typically from the sun that is scattered into the visual pathway and thus decreases with improved transmission. In the extremes, the perceived image can either be just the scene radiance, i.e., no atmospheric interference, or just the scattered ambient light.

Intuitively, reduced visibility results in an image with less detail especially in the distance. This reduced acuity is caused by two factors: the objects and their backgrounds become more similar due to increased attenuation and scattering; and the atmosphere acts as a low-pass filter (Krishnakumar & Venkatakrishnan, 1997), suppressing the higher-frequency image components or details. We use the term local contrast to refer to image acuity and define it as the magnitude of difference in image intensity over a short spatial distance $C = |\nabla I|$ where the gradient is with respect to the two-dimensional image space. The same spatial difference can be computed on the right side of Eq. (2) to get

$$|\nabla I| = |\nabla(Jt + A(1-t))| \tag{3}$$

$$= |\nabla Jt| \tag{4}$$

$$= t|\nabla J|. \tag{5}$$

Line 4 results from the assumption that the ambient light A is locally constant and line 5 results from the positivity of transmission t and the assumption that it is locally constant as well. The quantity $|\nabla J|$ is the "true" contrast of the scene when imaged under perfect transmission; i.e. when there is no intervening atmosphere to reduce visibility. This equation shows that transmission has the intuitive interpretation as the ratio of the observed contrast to the true contrast.

We use Sobel filters to estimate the image gradient at each pixel. To compensate for slight camera movement and other sources of image noise, we compute image contrast C as the average of the gradient magnitude over an image region Ω:

$$C = \frac{1}{|\Omega|}\sum_{\Omega}|\nabla I|. \tag{6}$$

Finally, putting it all together, we can relate the quantity we are trying to estimate, the coefficient of extinction b_{ext}, to what we measure from the image, contrast C (or, more precisely, the log of the contrast) through the linear relation:

$$b_{ext} = \frac{\ln C}{r} - \frac{\ln|\nabla J|}{r}. \tag{7}$$