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

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.

© 2013 Elsevier B.V. All rights reserved.

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.

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:

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

Here, $1/r$ is the scaling factor and $\ln|\nabla J|/r$ is the offset, which we treat as unknown in our prediction models below.

4. Our approach: Supervised learning

We learn our prediction models by using supervised learning. Given a labeled set of observations–the training set–we learn a predictive regression model. In our case, the observations are the images, or, more precisely, the image contrast measured from one or more image regions, and the labels are the values of the coefficient of extinction associated with the images as measured with a transmissometer. These values are considered the ground truth. We evaluate the learned models by predicting the coefficient of extinction for a separate set of images–the test set–based on image contrast and then comparing the predicted values with the ground truth values.

More formally, we want to learn a function $y = f(x)$ where the input is the image contrast and the output is the coefficient of extinction. We only consider linear models. In the univariate case, which corresponds to Eq. (7), the function $f(x)$ has the form

$$f(x) = \alpha_1 \cdot x + \alpha_0 \tag{8}$$

where x is the log of the image contrast, α_1 is the scaling factor, and α_0 is the offset.

The problem is now reduced to learning the unknown coefficients α_1 and α_0 using the labeled training set. We formulate this as a least-squares optimization problem. Given l observations x_i and corresponding labels y_i where $i = 1, \ldots, l$, we find the values of the coefficients that minimize the sum of the squared differences between the predicted and ground truth labels:

$$\alpha_1^*, \alpha_0^* = \underset{\alpha_1, \alpha_0}{\arg\min} \frac{1}{l} \sum_{i=1}^{l} (y_i - f(x_i))^2. \tag{9}$$

5. Dataset and evaluation

We evaluate our prediction models by using a set of images and ground truth extinction readings from the Arizona Department of Environmental Quality, which manages the PhoenixVis.net visibility web cameras website (Phoenixvis).[3] This website contains live images from six visibility cameras of scenic urban and rural vistas in the Phoenix, Arizona region. Our dataset consists of the following acquired over 2006:

- Digital images of South Mountain (SOMT) captured every 15 min.
- The extinction coefficient b_{ext} measured every hour using a transmissometer. This serves as the ground truth.

The images measure 1536×1152 pixels in size, are in the RGB color space, and (unfortunately) have been JPEG compressed at an unknown quality level. We initially consider only images taken at the top of each hour, since this is when the transmissometer readings are made, and during daylight hours, approximately 10 am to 4 pm. This results in a dataset of 1963 images from the SOMT camera which is "labeled" with ground truth extinction b_{ext} values. Note that these images are taken over the entire year and are not specific to a particular season, or time of month or week. In our work on semi-supervised learning in Section 7 where we incorporate unlabeled observations, we consider all the images–i.e., images captured every 15 min–but on a smaller dataset. Sample images corresponding to good and poor visibility conditions are shown in Fig. 1.

We evaluate the accuracy of our models using the coefficient of determination R^2 between the predicted and ground truth values. Let $y_{i'}$ and y_i be the predicted and true extinction coefficients for image i then

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{i'} - y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{10}$$

where n is the number of images in the evaluation set and $\overline{y}$ is the mean of the true values. R^2 is unitless and ranges from zero to one with higher values indicating a more accurate model.

In order to provide an intuitive feel for the predictions, we also compute the mean absolute error (MAE) between the predicted and true values:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_{i'} - y_i|. \tag{11}$$

MAE has the same units as B_{ext}, Mm^{-1}, and has a minimum value of zero with higher values indicating a less accurate model.

6. Incorporating multiple image regions

The hypothesis and motivation behind incorporating multiple image regions are that no single image region will be effective for estimating light extinction under different visibility regimes. 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.

We experiment with a number of image partitionings. First, we partition the images into square blocks of different sizes including: a 24×18 grid of 64×64-pixel blocks (B64); a 12×9 grid of 128×128-pixel blocks (B128); an 8×6 grid of 192×192-pixel blocks (B192); and a 6×4 grid of 256×256-pixel blocks (B256). We also consider regions corresponding to rows of these blocks; i.e. regions composed of the grouping of all blocks in a row. This results in three additional partitionings: a 24×1 grid of 64×1536-pixel regions (R64); a 12×1 grid of 128×1536-pixel regions (R128); and an 8×1 grid of 192×1536-pixel regions (R192). The motivation here is that each such row represents a large image region which is situated at approximately the same distance. These extended horizontal regions might thus be more effective than smaller blocks for estimating visibility. Table 1 summarizes the different partitionings and also indicates the dimensionality of the feature vectors that result since we extract one contrast value per region using Eq. (6). The terms in parenthesis above are the IDs assigned to the different partitionings.

In Sections 6.1 and 6.2 below, we explore and compare two different methods for incorporating multiple regions from the above partitionings into the visibility estimation.

6.1. Method: Regression trees

Transmission is an inverse exponential function of both distance and the extinction coefficient. For areas with a low transmission value, there will be a minimal change in contrast for small shifts in the extinction coefficient. However, these same areas can become very informative as the visibility improves. Regression trees generate a decision-making process by which multiple linear fits can be considered within the same model. As the conditions of the image change, different image regions may be used to generate the output. By using a regression tree, the system can be trained to observe closer landmarks if distant points become overly attenuated.

We extend our earlier work by allowing for the use of all regions in the image and optimizing their contributions to the overall estimation. This problem is handled by partitioning the image into regions (see above)

---

[3] http://phoenixvis.net.

(a) b_{ext} = 25 Mm^{-1}

(b) b_{ext} = 174 Mm^{-1}
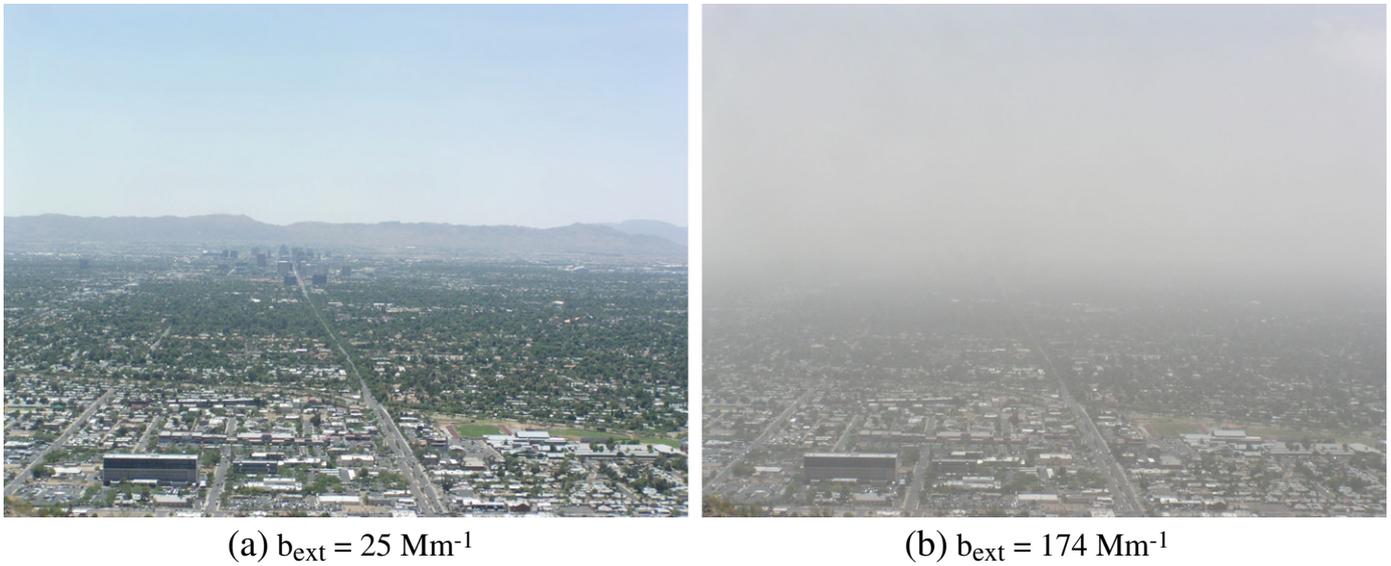
Fig. 1. We investigate methods to estimate light extinction b_{ext} using visibility cameras. Shown above are images corresponding to good and poor conditions from a visibility camera in Phoenix, Arizona. Ground truth readings from a transmissometer appear in the captions.

and using an M5' regression tree (Quinlan, 1992; Wang & Witten, 1997) to estimate visibility. That is, multiple features, one from each region, are provided to the regression tree which, during the training phase, learns which features (regions) to use under different visibility regimes. This approach results in a discontinuous but piece-wise linear predictive model.

Regression trees are a type of decision tree learning method. They are very similar to classification trees, but the primary goal is to output a real number instead of a class. The performance at each iteration of tree growth is measured by the information gain. This provides a quantitative value for the quality of each data split, and is calculated as the standard deviation of the examples assigned to the resulting leaf nodes. Intuitively, this means that we select splits that group similar regimes together. Regression trees select inequalities that split data in a way that maximizes this information gain. Each leaf node is then represented by a numerical value or a simple linear model to produce the output. An example tree using two features is shown in Fig. 2.

M5' regression generates a decision tree using standard deviation to calculate the information gain. The feature with the largest information gain as measured in reduction of standard deviation, represents the best way to split the current examples into two distinct groups. Standard deviation reduction (SDR) is calculated as

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i) \qquad (12)$$

where $sd(T)$ is the standard deviation of the parent node, $sd(T_i)$ is the standard deviation of each child node, $|T|$ is the number of data points in the parent node, and $|T_i|$ the number of data points in the child

nodes. This splits the training data in a way that minimizes the standard deviation of each subset proportional to the size of those sets. Because variables contain continuous values, we perform the fit based on inequalities (e.g. $X_2 > 4.5$). The function stops splitting when the SDR becomes sufficiently low or the number of examples in the set becomes too small to continue splitting. The algorithm then begins pruning the tree for reduced error. This is necessary because the building process relies solely on SDR and does not test fits as it progresses. If the summed error of child nodes is greater than a parent node, the parent node will become a leaf node instead. Each leaf node consists of a set of data points which can be characterized by a single value or a linear model. We use the M5PrimeLab implementation of M5' regression trees (M5PrimeLab).

We evaluate the regression trees using fivefold cross validation. The 6897 labeled images are partitioned randomly into five approximately equal sized sets. A regression tree is then learned using four of these partitions–the training set–and then evaluated on the fifth partition–the test set. This is then repeated four times using different training/test

Table 1

The image partitionings considered and the dimensionality (Dim) of the resulting feature vectors (one contrast value per region). See the end of section 6 for more details on the partitionings, which are identified by ID.

ID	Dim	Description
B64	432	Blocks 64 × 64 px
B128	108	Blocks 128 × 128 px
B192	48	Blocks 192 × 192 px
B256	24	Blocks 256 × 256 px
R64	24	Rows 64-px tall
R128	12	Rows 128-px tall
R192	8	Rows 192-px tall
All	224	All of the above, except B64

Fig. 2. A mock-up example of a regression tree for a 2-feature visibility problem. Here, B_{ext} is expressed in terms of $X1$ and $X2$. The number of training points that are used to learn each of the constituent parts of the model is indicated by the count N.

sets and the final results are averaged. This is performed for each of the image (spatial) partitionings described in Section 5.

Table 2 shows the results for the different image partitionings. It is immediately apparent that using regression trees to incorporate multiple regions significantly improves upon our previous approach (Graves & Newsam, 2011), which used only a single region and resulted in an R^2 value of 0.646 and a MAE value of 12.5 Mm^{-1}. We discuss these results further below.

6.2. Method: Multivariate linear regression

In addition to the regression tree approach above, we also consider multivariate linear regression as a way to incorporate multiple image regions into the visibility estimation. Given n image regions, we have n instances of Eq. (7). Assuming a constant value of the light extinction b_{ext} for the scene (i.e., the atmosphere is approximately uniform), we can sum these instances to get

$$nb_{ext} = \sum_{j=1}^{n} \frac{ln\, C_j}{r_j} - \frac{ln|\nabla J_j|}{r_j} \tag{13}$$

where C_j, r_j, and J_j are the measured contrast, distance, and true contrast of region j respectively. We can rewrite this as

$$b_{ext} = \sum_{j=1}^{n} \frac{ln\, C_j}{nr_j} - K \tag{14}$$

where the individual offsets have been grouped into a single offset K. We can consider this as a multivariate linear regression model in which the (logs of the) contrasts of the regions C_j are the variables, the $1/nr_j$ are the scaling factors, and K is the offset. Using the notation of Section 4, this model can be written as

$$f(\mathbf{x}) = [\alpha_n...\alpha_1] \cdot \mathbf{x} + \alpha_0. \tag{15}$$

Given a training set, we can then use least-squares to estimate the coefficients of the model α_j for $j = 1,...,n$.

We evaluate this model using the same fivefold cross validation and image (spatial) partitionings as for the regression tree model. The results for the multivariate linear regression model are summarized in Table 2. This model is shown to outperform the regression tree model.

6.3. Discussion

We make the following observations based on the results in Table 2. First, not surprisingly, models which incorporate multiple image regions are more effective than those with only a single region. The best results for the regression tree and multivariate linear regression models are $R^2 = 0.780$ (MAE = 8.90 Mm^{-1}) and $R^2 = 0.845$ (MAE = 7.95 Mm^{-1}) respectively, while the best result using

a single region in our previous work was $R^2 = 0.646$ (MAE = 12.5 Mm^{-1}). This represents improvements of 20.7% and 30.8% in terms of R^2. This improvement is likely in large part due to the fact that scene regions at different distances are effective for estimating light extinction under different visibility regimes. 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.

We also observe that models incorporating a larger number of smaller regions are better than those incorporating a smaller number of larger regions. This is true for both the regression tree and multivariate linear regression models. For the most part, more small blocks performed better than fewer large blocks and more small rows performed better than fewer large rows. Using the smaller blocks also performed better than the larger rows. This finding is further evidence that it is useful to have regions at varying distances so as to better model different visibility regimes.

The multivariate linear regression model is shown to outperform the regression tree model for all image partitionings. While it is difficult to make any general conclusions based on this result since these are two very different approaches, it is likely that the regression trees are over-fitting the training data. Many of the leaf nodes of the trees contain very few samples–some as few as four–which makes it difficult to fit even a simple linear univariate model. It is possible that more aggressive pruning of the tree could improve this. This is potentially a topic for further investigation.

7. Incorporating unlabeled observations

In this section, we propose a framework to incorporate unlabeled observations using semi-supervised learning to improve the prediction models. The motivation here is that ground truth data is typically more difficult to obtain than observed data. In remote sensing in the geographic sciences, this is due to the effort involved in collecting ground level field data and results in ground truth data which is sparsely sampled and/or distributed spatially. In image understanding in computer vision, this is due to the time required to manually annotate images and results in relatively few labeled image examples compared to the total number of images that are typically available. In our case, our observations–the images–are sparsely labeled temporally since the transmissometer only provides readings every hour whereas the images are captured every 15 min.

The goal of semi-supervised learning is to use the unlabeled observations to improve the learned model. This of course requires that some assumptions are made about the relationship between the observations. In remote sensing, it is assumed that spatially nearby observations will be similar based on the first law of geography: all things are related but nearby things are more related that distant things. In our problem, we assume that visibility does not change rapidly and that observations (images) obtained a short time apart will correspond to similar atmospheric conditions in terms of the coefficient of extinction. The technical challenge in semi-supervised learning is how to exploit this assumption of similarity to smooth the model so that observations that are close in time are not assigned drastically different labels. An alternate view is that we would like to in some sense "propagate" the labels from the labeled observations to the unlabeled observations to learn a better model.

7.1. Method: Laplacian regularized linear least squares

Our approach to semi-supervised learning is inspired and based on the paper by (Xie et al., 2010). We use a weighting scheme to make sure our predictions do not vary significantly for related observations. In our case, these weights are based on the time between observations. In the following, we describe a general mechanism to perform this

Table 2
Results of the regression tree and multivariate linear regression predictive models for different image partitionings. See Table 1 for details on the partitionings. MAE is in units of Mm^{-1}.

Partitioning	Regression trees		Multi. linear reg.	
	R^2	MAE	R^2	MAE
All	0.780	8.90	0.845	7.95
B64	0.776	8.88	0.830	8.21
B128	0.762	9.22	0.813	8.94
B192	0.735	9.82	0.830	8.21
R64	0.733	9.72	0.748	10.5
R128	0.722	10.3	0.729	11.1
R192	0.698	10.5	0.719	11.1

weighting in a linear regression framework. We then investigate several weighting schemes.

To state it formally, given l labeled and u unlabeled observations x_i with temporal index i, we want to minimize $(f(x_i) - f(x_j))^2$, the difference in our predictions for these observations, for i and j values that are close in time. That is, we do not want to assign drastically different coefficients of extinction to images that are taken a short time apart since it is unlikely that the atmospheric conditions changed drastically during this period. This is of course in addition to the prediction error on the labeled training data $(y_i - f(x_i))^2$. In the general case, we use w_{ij} to weight $(f(x_i) - f(x_j))^2$. This results in the following optimization problem in which mma determines the relative weighting of the smoothing component:

$$f' = arg\min_{\alpha} \frac{1}{l} \sum_{i=1}^{l} (y_i - f(x_i))^2 + \gamma \sum_{i,j}^{l+u} w_{ij}\left(f(x_i) - f(x_j)\right)^2 \quad (16)$$

where

$$f(x) = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_n x_n. \quad (17)$$

Note how the objective function simultaneously minimizes the prediction error on the labeled training set and the prediction between closeby observations whether they are labeled or not. If we construct a graph matrix W that captures the (temporal) distance between observations, the solution to this optimization problem can be solved using the graph Laplacian. We thus refer to this method as Laplacian regularized linear least-squares (LRLLS) regression.

Let $D_{ii} = \sum_{j=1}^{l+u} w_{ij}$. The graph Laplacian can then represented as $L = D - W$. Our optimization problem can now be rewritten as

$$f' = arg\min_{\alpha} \frac{1}{l} J(\overline{Y} - X\alpha)^2 + \gamma (X\alpha)^T L X\alpha \quad (18)$$

where α are our regression coefficients, X is the stacked observations, Y is the stacked labels, and J is a matrix used to separate labeled and unlabeled observations. It is an $(u + l) \times (u + l)$ matrix with form

$$J_{ii} = \begin{cases} 1 & \text{if } x_i \text{ is labeled} \\ 0 & \text{if } x_i \text{ is unlabeled} \end{cases}.$$

The optimization problem now becomes

$$f' = arg\min_{\alpha} \frac{1}{l} J(\overline{Y} - X\alpha)^2 + \gamma (X\alpha)^T L X\alpha . \quad (19)$$

The optimal α for a fixed γ is found by taking the partial derivative and solving with respect to α.

$$\frac{\partial}{\partial \alpha} \frac{1}{l} J(\overline{Y} - X\alpha)^T (\overline{Y} - X\alpha) + \gamma (X\alpha)^T L X\alpha$$
$$= 0 \; \frac{J}{l}\left(-X^T\right)(\overline{Y} - X\alpha) + \gamma X^T L X\alpha$$
$$= 0 - JX^T\overline{Y} + JX^T X\alpha + \gamma X^T L X\alpha \quad (20)$$
$$= 0 \left(JX^T X + \alpha X^T L X\right)\alpha = JX^T\overline{Y}\alpha$$
$$= \left(JX^T X + \gamma X^T L X\right)^{-1} JX^T\overline{Y}.$$

Eq. (20) is used to compute the coefficients α for the linear model.

We consider two cases for the temporal connections between nearby observations: adjacent and Gaussian weights. For adjacent weights, $w_{ij} = 1$ if i connected to j as shown in W_a. Adjacent here means that two observations were taken 15 min apart. Alternatively, these weights can be determined using a Gaussian scaling. Examples of both W_a and W_G are shown below, where the Gaussian connectivity matrix uses

$\sigma = 20$ min. The motivation behind the Gaussian weighting is to model the smoothly changing atmospheric visibility.

$$W_a = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$W_G = \begin{bmatrix} 0 & 0.47 & 0.22 & 0.11 & 0.05 \\ 0.47 & 0 & 0.47 & 0.22 & 0.11 \\ 0.22 & 0.47 & 0 & 0.47 & 0.22 \\ 0.11 & 0.22 & 0.47 & 0 & 0.47 \\ 0.05 & 0.11 & 0.22 & 0.47 & 0 \end{bmatrix}$$

Note that the W_a and W_G above are reduced versions of the graph matrices corresponding to a dataset of only five observations. The actual matrices are of the size of the number of observations used in the learning stage and have values of zero for pairs of observations that are not adjacent in the case of W_a, and zeros for pairs of observations that are more than one hour apart in the case of W_G.

To evaluate the proposed LRLLS approach, we construct a dataset that contains both labeled and unlabeled training samples, a validation set for determining optimal values of model parameters, and two test sets. Labeled here means the image was acquired at the top of the hour at approximately the same time as the transmissometer reading. Unlabeled means images acquired at other times during the hour. Table 3 provides details on this dataset. Note that since the main purpose of this section it to investigate the benefits of the proposed semi-supervised learning framework and not to improve on the results in the previous sections of the paper, we use a smaller set of images to simplify the analysis.

We use a bivariate linear regression model in which the (log) of the contrast from two 128×128-pixel blocks is the input to the model and the output is the predicted coefficient of extinction. The two blocks were manually chosen to correspond to a distant ground region and a sky region just above the horizon. Note that the goal of this section is to evaluate the benefit of incorporating unlabeled observations independent of whether the selected image regions are necessarily optimal.

The validation set is used to determine the values of two model parameters, γ, which determines the relative weighting of the smoothing component of Eq. (16), and the width of the Gaussian weighting matrix W_G. Simple grid search was used to find the optimal values of these two parameters.

Table 4 compares the results of the proposed LRLLS approach to that of standard linear least squares (LLS). In the standard LLS approach, the training data consists of only labeled observations. This includes both the labeled training data (TR) as well as the validation data (V) for a total of 51 labeled observations (see Table 3). In the proposed LRLLS approach, the training data consists of just 21 labeled observations as well as 63 unlabeled observations.

7.2. Discussion

While the above results are based on a relatively small evaluation dataset, they indicate that the proposed LRLLS approach results in a better prediction model. LRLLS results in improved R^2 and MAE for both test

Table 3
Dataset used to evaluate the proposed LRLLS approach.

	# Labeled images	# Unlabeled images
Training set (TR)	21	63
Validation set (V)	30	0
Test set 1 (T1)	119	0
Test set 2 (T2)	119	0

Table 4
Comparison of the proposed LRLLS approach to incorporate unlabeled observations and standard LLS. MAE is in units of Mm^{-1}.

Training set	Validation set	Model	Connectivity	Test set	R^2	MAE
TR + V	–	LLS	–	T1	0.6803	9.7042
TR	V	LRLLS	Gaussian	T1	0.6978	9.5680
TR	V	LRLLS	Adjacency	T1	**0.6981**	**9.4886**
TR + V	–	LLS	–	T2	0.7144	11.3082
TR	V	LRLLS	Gaussian	T2	**0.7368**	**10.6980**
TR	V	LRLLS	Adjacency	T2	0.7346	10.7410

Best results shown in bold.

sets. This is significant since unlabeled observations are typically easier to obtain than labeled observations for learning prediction models.

There is no clear winner however between the two different weighting schemes, strict adjacency and Gaussian smoothed. While they both improve upon the standard LLS approach, their relative performance depends on the test set. Investigating this further, as well as a linearly smoothed weighting scheme, is the subject of future work.

8. Conclusion

We investigated image processing and pattern recognition techniques for estimating atmospheric visibility using off-the-shelf cameras. We proposed a linear regression framework in which image contrast is related to the coefficient of light extinction through atmospheric transmission. We demonstrated that prediction models which incorporate multiple image regions, either through regression trees or multivariate linear regression, outperform models which consider only a single image region. We also described a semi-supervised learning framework which incorporates unlabeled observations to improve the prediction model.

There are many directions to take this work. One of the long-term goals of this work is to learn prediction models for visibility camera systems for which we do not have access to ground truth data from a transmissometer or similar instrument. Such models would, of course, be limited to providing relative and not absolute estimates of visibility. They could, however, be deployed in the growing number of cameras already present in our ecosystem such as web, surveillance, and traffic cameras.

A second goal is to link the observations more directly to atmospheric pollution. We are currently in the process of acquiring images and corresponding particulate pollution measurements as a first step towards this goal. We expect future work on that problem to be directly informed by the findings in this paper.

References

Baumer, D., Versick, S., Vogel, B., 2008. Determination of the visibility using a digital panorama camera. Atmospheric Environment 42, 2593–2602.

Betts, J., 1971. The instrumental assessment of visual range. Proc. IEEE 59, 1370–1371.

Caimi, F., Kocak, D., Justak, J., 2004. Remote visibility measurement technique using object plane data from digital image sensors. Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, pp. 3288–3291.

Fattal, R., 2008. Single image dehazing. ACM SIGGRAPH, pp. 72:1–72:9.

Graves, N., Newsam, S., 2011. Using visibility cameras to estimate atmospheric light extinction. IEEE Workshop on Applications of Computer Vision (WACV), pp. 577–584.

Graves, N., Newsam, S., 2012. Visibility cameras: where and how to look. Proceedings of the 1st ACM International Workshop on Multimedia Analysis for Ecological Data, pp. 7–12.

He, K., Sun, J., Tang, X., 2009. Single image haze removal using dark channel prior. IEEE Conference on Computer Vision and, Pattern Recognition, pp. 2341–2353.

Kim, K., Kim, Y., 2005. Perceived visibility measurement using the HSI color different method. J. Korean Phys. Soc. 46, 1243–1250.

Krishnakumar, V., Venkatakrishnan, P., 1997. Determination of the atmospheric point spread function by a parameter search. Astron. Astrophys. Suppl. Ser. 126, 177–181.

Lee, P., Hoffer, T., Schorran, D., Ellis, E., Moyer, J., 1982. Laser transmissometer—a description. Sci. Total. Environ. 23, 321–335.

Luo, C.H., Wen, C.Y., Yuan, C.S., Liaw, J.J., Lo, C.C., Chiu, S.H., 2005. Investigation of urban atmospheric visibility by high-frequency extraction: model development and field test. Atmos. Environ. 39, 2545–2552.

M5PrimeLab, b. M5primelab: M5' regression tree and model tree toolbox for matlab/octave http://www.cs.rtu.lv/jekabsons/ (last accessed Jan. 20 2012).

Molenar, J.V., Cismoski, D.S., Schreiner, F., Malm, W.C., 2004. Analysis of digital images from Grand Canyon, Great Smoky Mountains, and Fort Collins, Colorado. Regional and Global Perspectives on Haze: Causes, Consequences and Controversies Visibility Specialty Conference.

Namer, E., Schechner, Y.Y., 2005. Advanced visibility improvement based on polarization filtered images. Proc. SPIE 5888: Polarization Science and Remote Sensing II, pp. 36–45.

Namer, E., Shwartz, S., Schechner, Y.Y., 2009. Skyless polarimetric calibration and visibility enhancement. Opt. Express 17, 472–493.

Narasimhan, S., Nayar, S., 2000. Chromatic framework for vision in bad weather. IEEE Conference on Computer Vision and, Pattern Recognition, pp. 598–605.

Narasimhan, S., Nayar, S., 2001. Removing weather effects from monochrome images. IEEE Conference on Computer Vision and, Pattern Recognition, pp. 186–193.

Narasimhan, S., Nayar, S., 2002. Vision and the atmosphere. Int. J. Comput. Vis. 48, 233–254.

Narasimhan, S., Nayar, S., 2003a. Contrast restoration of weather degraded images. IEEE Trans. Pattern Anal. Mach. Intell. 25, 713–724.

Narasimhan, S., Nayar, S., 2003b. Interactive (de)weathering of an image using physical models. ICCV Workshop on Color and Photometric Methods in Computer Vision.

Narasimhan, S., Nayar, S., 2003c. Shedding light on the weather. IEEE Conference on Computer Vision and, Pattern Recognition, pp. 665–672.

Phoenixvis, s. Phoenix visibility web camera website managed by the Arizona Department of Environmental Quality (ADEQ) http://www.phoenixvis.net/ (last accessed Jan. 20 2012).

Quinlan, J., 1992. Learning with continuous classes. Australian Joint Conference on, Artificial Intelligence, pp. 343–348.

Raina, D.S., Parks, N.J., Li, W.W., Gray, R.W., Dattner, S.L., 2004. Innovative monitoring of visibility using digital imaging technology in an arid urban environment. Regional and Global Perspectives on Haze: Causes, Consequences and Controversies Visibility Specialty Conference.

Schechner, Y., Narasimhan, S., Nayar, S., 2001. Instant dehazing of images using polarization. IEEE Conference on Computer Vision and, Pattern Recognition, pp. 325–332.

Schechner, Y., Narasimhan, S., Nayar, S., 2003. Polarization-based vision through haze. Appl. Optics. Spec. Issue 42, 511–525.

Seinfeld, J., Pandis, S., 2006. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. Wiley.

Shwartz, S., Namer, E., Schechner, Y.Y., 2006. Blind haze separation. IEEE Conference on Computer Vision and, Pattern Recognition, pp. 1984–1991.

Tan, R., 2008. Visibility in bad weather from a single image. IEEE Conference on Computer Vision and, Pattern Recognition, pp. 1–8.

Wang, Y., Witten, I., 1997. Induction of model trees for predicting continuous classes. European Conference on, Machine Learning, pp. 128–137.

Xie, L., Carreira-Perpinan, M.A., Newsam, S., 2010. Semi-supervised regression with temporal image sequences. IEEE International Conference on Image Processing, pp. 2637–2640.