Title
Image-based meteorologic visibility estimation

Author
Graves, Nathan

Publication Date
2011-12-11

UNIVERSITY OF CALIFORNIA
MERCED

Image-Based Meteorologic Visibility Estimation

A Thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Science

in

Electrical Engineering and Computer Science

by

Nathan Graves

Committee in Charge:

Prof. Shawn Newsam[1], Chair

Prof. Arnold Kim[2]

Prof. David Noelle[1,3]

December 2011

---

[1] Electrical Engineering and Computer Science Graduate Group, University of California, Merced
[2] Applied Mathematics Graduate Group, University of California, Merced
[3] Cognitive & Information Sciences Graduate Group, University of California, Merced

The Thesis of
Nathan Graves is approved:

Prof. Arnold Kim[2]

Prof. David Noelle[1,3]

Prof. Shawn Newsam[1], Committee Chairperson

December 2011

Image-Based Meteorologic Visibility Estimation

Abstract

Image-Based Meteorologic Visibility Estimation

Nathan Graves

Awareness and understanding of atmospheric visibility has strong implications for our daily lives. In addition to being critical for navigation, it acts as an indicator for air quality and pollution. The devices traditionally used to measure visibility, transmissometers and nephelometers, are expensive and often require field maintenance and calibration. Visibility camera systems are increasingly being deployed to measure atmospheric visibility; however, their use has so far been limited to qualitative analysis. The primary focus of this study is to develop image analysis techniques to derive quantitative measurements of visibility from such camera systems. We take advantage of the Beer-Lambert law, which defines the exponential relation by which light is attenuated when traveling through a medium. This is used to define a standard visibility model, which then allows us to frame the problem as a simple log-linear relation. We investigate several numerical models to estimate visibility, including single and multivariate linear least square regression, Laplacian-regularized linear least squares regression, and approximation with the M5' linear regression tree algorithm. We demonstrate the effectiveness of these algorithms on images and ground truth visibility measurements from the PhoenixVis.net visibility camera system. The features chosen by the multivariate feature selection process provide insight into the benefit of edge contrast and color saturation as indicators for poor visibility. In addition, we investigate Lambertian lighting and a dark channel prior as cues for salient regions for visibility estimation.

Acknowledgements

First and foremost, I would like to give my utmost gratitude and respect to my thesis supervisor, Professor Shawn Newsam, for his ongoing insight, encouragement, and contributions to my work. He has provided many valuable opportunities, from undergraduate research to graduate school, that I would not have pursued without his motivation.

In addition, I would like to show my appreciation for the rest of my advisory committee: Professor Arnold Kim and Professor David Noelle. Thank you for taking the time to evaluate my work and allowing me to contribute to UC Merced's growing body of research.

I would like to thank all of the graduate students that were a part of the Computer Vision Lab for the many meetings and discussions that have contributed to my success at UC Merced. I'd like to thank Ling Xie in particular, whose original work in visibility estimation provided a foundation for my research.

I am grateful for the ongoing work done by the Arizona Department of Environmental Quality to monitor visibility and pollution in the environment. The ground truth data provided by their Phoenix Visibility Web Cameras made my work at UC Merced possible.

Lastly, I'd like to thank my family for their ongoing support for the duration of my stay at UC Merced, as well as my loving girlfriend Brandi for her encouragement, understanding, and patience throughout the development of my thesis.

Contents

List of Figures

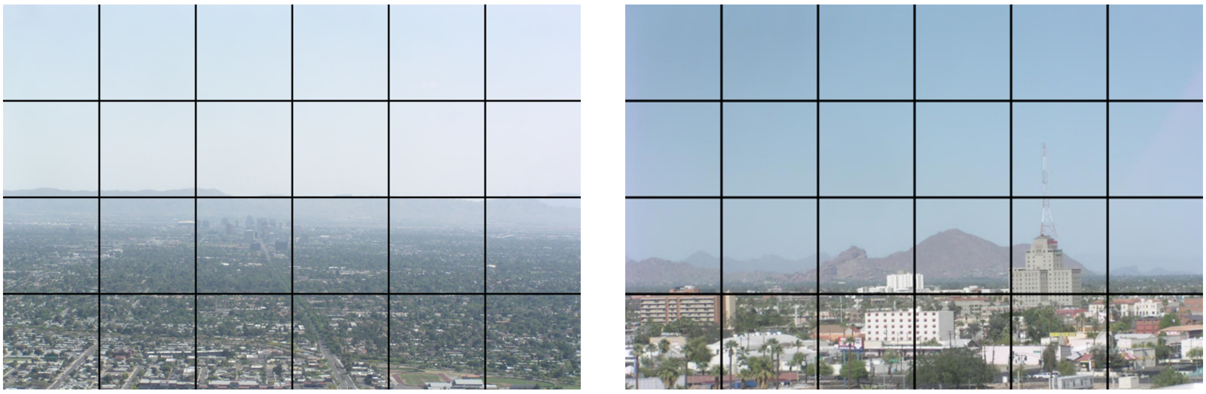
List of Tables

Chapter 1

Introduction

Quantitative measures of atmospheric visibility are increasingly being used for purposes other than navigation. For example, measures of visibility are being used as indirect estimates of air pollution especially where direct measurements are not available. They are being used to estimate solar irradiance which is important for determining where to situate solar energy farms and for forecasting the near term energy output of existing farms. And, visibility measurements are central to the United States' Environmental Protection Agency's (EPA) goal for improving visual air quality in the Class I Federal areas which include 156 national parks and wilderness areas. In 1977, Congress amended the Clean Air Act with legislation to prevent future and remedy existing impairment of visibility in Class I areas and in 1999, the EPA issued the Regional Haze Rule which mandates that state and federal agencies work together to actually improve the visibility.

Expanding visibility monitoring is key to the EPA's mandates and the agencies charged with monitoring typically use a combination of three techniques. First, they utilize specialized equipment such as transmissometers, which measure light extinction, and nephelometers, which measure light scattering. Second, they use Mie scattering the-



(a) SOMT: $b_{ext} = 41$ Mm^{-1}

(b) CAME: $b_{ext} = 41$ Mm^{-1}

(c) SOMT: $b_{ext} = 190$ Mm^{-1}

(d) CAME: $b_{ext} = 214$ Mm^{-1}

Figure 1.1: We investigate methods for estimating the coefficient of extinction b_{ext} using visibility cameras. Shown above are images corresponding to good and poor conditions taken from two such cameras, SOMT and CAME. Ground truth readings from a transmissometer appear in the captions.

ory to calculate visibility based on measurements of airborne particulates. Finally, and relevant to this work, they deploy networks of visibility cameras. For example, the Interagency Monitoring of Protected Visual Environments (IMPROVE) program has installed and maintains cameras in over two dozen national parks. In addition, regional air quality agencies have deployed visibility camera systems in over 30 cities.

This study focuses on image analysis techniques for deriving quantitative measurements of visibility from such systems. *Visibility cameras are currently used for qualitative purposes only* such as providing visual examples of good and bad days. We feel, however, there is significant opportunity to use these images to derive quantitative measures of visibility perhaps not as accurately as specialized equipment but at much lower cost and possibly even by piggy-backing onto existing web-connected cameras.

We describe methods for estimating the coefficient of light extinction which is a standard measure of atmospheric visibility. We investigate several numerical models, including single and multivariate linear fits and linear regression trees using the M5' algorithm. One extension of our linear model incorporates additional unlabeled data with a temporal Laplacian weighting. These models use contrast as well as a recently proposed dark channel prior [1] as quantitative visual features from the images.

This thesis is organized as follows: Section 2 shows related work on dehazing and visibility estimation. Section 3 provides background information on atmospheric visibility and obtaining ground truth data. Section 4 provides justification and results with our linear model. Section 5 expands on our linear model to include unlabeled data. Section 6 employs an M5' regression tree framework to fit our model. Section 7 details features that may be used to detect salient regions for visibility applications. Section 8 provides implementation details and results for our image registration procedure, while we conclude and discuss future work in Section 9.

Chapter 2

Related Work

2.1 Dehazing

There is a sizable body of work on the related problem of improving the fidelity of images taken under hazy or otherwise atmospherically degraded conditions. These algorithms estimate the signal loss at each pixel and then attempt to restore the original color. This includes work by Narasimhan and Nayar on using physics-based models to improve a single image [2, 3] and using multiple images of the same scene but under different atmospheric conditions [4–6]; work by Schechner and colleagues on using polarization to improve one or more images [7–11]; and work by He et al. on using a dark channel prior to dehaze a single image [1]. While the primary focus of this thesis is to derive quantitative estimates of atmospheric visibility, dehazing work can potentially be used to inform the problem because both models are dependent on signal loss due to haze. We later demonstrate that the dark channel prior can be used as an indicator of visibility.

2.2 Atmospheric Visibility Estimation

There is a much smaller body of work on using images to measure atmospheric visibility. Caimi et al. [12] 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 and Kim [13] 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. [14] 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. [15] 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. [16] 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. [17] 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.

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. We also perform a more thorough investigation into different image features and learning models than any of the works above.

Chapter 3

Background

This section discusses why visibility is reduced by the atmosphere and describes a standard model for the formation of a hazy image that relates atmospheric transmission to the observed image. It then relates transmission to light extinction, the quantity being estimated. Finally it introduces specialized instrumentation for measuring the extinction of light through the atmosphere–transmissometers–and measuring the scattering of light by the atmosphere–nephelometers. These instruments provide the ground truth data for our experiments.

3.1 Atmospheric Transmission

Atmospheric light extinction is inversely related to transmission through the following exponential equation [18]

$$t(x) = \exp^{-\beta_{ext} r(x)} \tag{3.1}$$

where β_{ext} is the extinction coefficient, $r(x)$ is the length of the visual pathway, and x is each pixel in our image. The transmission is a scalar value that ranges from 0 to 1, and represents the fraction of a signal that is preserved as it passes through a medium. This equation, often referred to as the Beer-Lambert law, assumes the atmosphere is homogeneous along the pathway. We further assume a homogeneous atmosphere throughout a scene. Inverse megameter (Mm^{-1}) is the typical unit of measurement for the extinction coefficient.

3.2 Why is Visibility Reduced?

Reduced visibility by the intervening atmosphere is mainly due to three first-order processes: 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. Absorption and scattering are due to gases and aerosols (particles) suspended in the atmosphere. The combined effect of the absorption and scattering is referred to as the total light extinction. Previous work suggests that most of the extinction in the atmosphere is due to scattering alone [19] and so in this work we anticipate that the effects of absorption will be negligible.

3.3 Atmospheric Attenuation Model

Atmospheric transmission describes the fraction of light that reaches an observer after radiating from a scene. It is a positive scalar quantity ranging from 0 to 1 where larger values indicate improved visibility. Transmission is commonly related to image formation through [1, 5, 20–22]

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{3.2}$$

where x is a two dimensional spatial variable, $I(x)$ is the observed image, $J(x)$ is the scene radiance, A is the ambient (atmospheric) light, and $t(x)$ is the atmospheric transmission. The first term on the right side of eq. (3.2) 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. As $t(x) \to 0$, we expect that the perceived image will be equivalent to the scattered ambient light A. Alternatively, we expect to see the original scene radiance $J(x)$ as $t(x) \to 1$.

3.4 Obtaining Ground Truth Data

Many devices are available that can provide valuable ground truth data for this problem. The most important devices, transmissometers and nephelometers, have a direct correlation with visibility as defined in eq. (3.2). Luminance meters can be used to obtain additional information about the illumination of the scene, allowing for the normalization of images with varying lighting conditions.

3.4.1 Transmissometer

A transmissometer [23–25] measures light extinction. It 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 β_{ext}.

3.4.2 Nephelometer

A nephelometer [23, 26] measures light scattering. It is a compact instrument which measures the amount of light scattered by gases and aerosols in a sampled air volume. It also consists of a transmitter and receiver but configured at an angle so the receiver only receives scattered light. A nephelometer calculates the scattering coefficient β_{sp} which when added to the absorption coefficient β_{abs} gives the total extinction coefficient $\beta_{ext} = \beta_{sp} + \beta_{abs}$. However, as mentioned above, extinction in the Earth's atmosphere is mostly due to scattering and so we consider β_{ext} and β_{sp} as equivalent.

3.4.3 Luminance Meter

A luminance meter measures the amount of illumination received by the sensor. This is necessary for some techniques that rely heavily on the amount of ambient lighting, such as the detection of Lambertian surfaces [27]. This data could also be used to normalize images from cameras that rely on automatic exposure settings. While a luminance meter was not available for our ground truth data, an alternative method proposed for use with the dark channel prior [1] was used to estimate the atmospheric light.

Chapter 4

Estimating Atmospheric Visibility

The goal of this work is to estimate light extinction β_{ext} given an image $I(x)$. We do this by first estimating transmission $t(x)$ from $I(x)$ using eq. (3.2) and then using eq. (3.1) to compute β_{ext}. We investigate two methods for estimating transmission: 1) based on local image contrast as computed in either the spatial or frequency domain; and 2) using a dark channel prior. These are then separated into single variable and multivariate linear models for evaluation.

4.1 Local Image Contrast

Intuitively, reduced visibility results in an image with less detail especially in the distance. This reduced acuity results from two sources: the objects and their backgrounds become more similar due to increased attenuation and scattering; and the atmosphere acts as a low-pass filter [28], 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: $C_l(x) = |\nabla_x I(x)|$. The same spatial difference can be computed on the right side of eq. (3.2) to get

$$
\begin{aligned}
|\nabla_x I(x)| &= |\nabla_x \left(J(x)t(x) + A(1 - t(x))\right)| && (4.1) \\
&= |\nabla_x J(x)t(x)| && (4.2) \\
&= t(x)|\nabla_x J(x)| \,. && (4.3)
\end{aligned}
$$

Eq. (4.2) results from the assumption that the ambient light A is locally constant and eq. (4.3) results from the positivity of transmission $t(x)$ and the assumption that it too is locally constant. The quantity $|\nabla_x J(x)|$ is the "true" contrast of the scene when imaged under perfect transmission. This equation shows transmission has the intuitive interpretation as the ratio of the observed contrast to the true contrast.

4.1.1 Contrast in the spatial domain

It is natural to consider $|\nabla_x I(x)|$ as the magnitude of the image gradient as computed in the spatial domain. This is calculated using Sobel filters. Each image is convolved with a set of 3×3 kernels that emphasize horizontal and vertical edges. These can then be combined to estimate the gradient magnitude at each pixel. To compensate for slight camera movement and other sources of image noise, we compute local contrast in the spatial domain C_{lsd} as the average of the gradient magnitude over an image region Ω:

$$
C_{lsd} = \frac{1}{|\Omega|} \sum_{x \in \Omega} |\nabla_x I(x)| \,. \qquad (4.4)
$$

Transmission $t(x)$ is assumed constant over this region.

4.1.2 Energy in the frequency domain

A standard way to measure visual acuity is through frequency analysis in the Fourier domain. To understand how this is interpreted in images, it is important to understand the output of the Fourier transform. The one-dimensional Fourier transform (1D-DFT) takes a discretely indexed 1D time signal and outputs a discretely indexed signal which gives information on the frequency content of the time signal, including the magnitude and phase information. In the two-dimensional case, the input is a discretely indexed 2D signal, such as a digital image, and the output is a 2D discretely indexed signal known as the 2D-DFT which gives information on the frequency content of the input. The distance and angle of each value from the center in the 2D-DFT provide information on the frequency and the direction respectively on the content of the image.

For our experiments, we detect the strength or amount of energy in the higher-frequency regions of the Fourier space by summing the Fourier energy spectral density. This sum provides a single numerical value for the strength of edges in a specified frequency range. Given lower and upper frequencies w_l and w_u, we compute the local contrast in the frequency domain for an image region C_{lfd} as the sum of the square of the magnitude of the two-dimensional discrete Fourier transform (2D-DFT) $F(u,v)$ in band-pass regions defined by concentric circles centered at the zero-zero or DC frequency:

$$C_{lfd} = \sum_{w_l < \sqrt{u^2+v^2} \leq w_u} |F(u,v)|^2. \tag{4.5}$$

The cutoff frequencies w_l and w_u can range between 0 and the Nyquist frequency w_{Ny} and determine whether the energy is computed in a low-pass, band-pass, or high-pass region. The DC frequency is never included since it is the average value of an image region and thus not indicative of acuity. Example outputs for the 2D-DFT are shown in Figure 4.1.

4.1.3 Dark Channel Prior

We also estimate transmission using a dark channel prior based on the work by He et al. [1] on single image dehazing. The intuition is that one can reasonably expect any natural image to have a dark region which has very low intensity values in at least one of the color channels when imaged under perfect transmission. This can result from bright colors, shadows, or dark objects in the image. According to our image model, all pixels are incremented by $A(1 - t(x))$. As $t(x)$ decreases, the minimum pixel value is increased in relation to the ambient light. Therefore, the difference between the observed intensity and the expected low intensity for these image regions–the prior–is indicative of the loss of transmission in relation to atmospheric scattering. He et al. use estimated transmission based on a dark channel prior to perform image correction (dehazing). We use it here to estimate light extinction.

The derivation is as follows [1]. Starting with the haze image formation model, we determine the minimum intensity value in color channel c for an image region Ω:

$$\min_{x \in \Omega} \left(I^c(x) \right) = \min_{x \in \Omega} \left(J^c(x) t(x) + A^c (1 - t(x)) \right) \ . \tag{4.6}$$

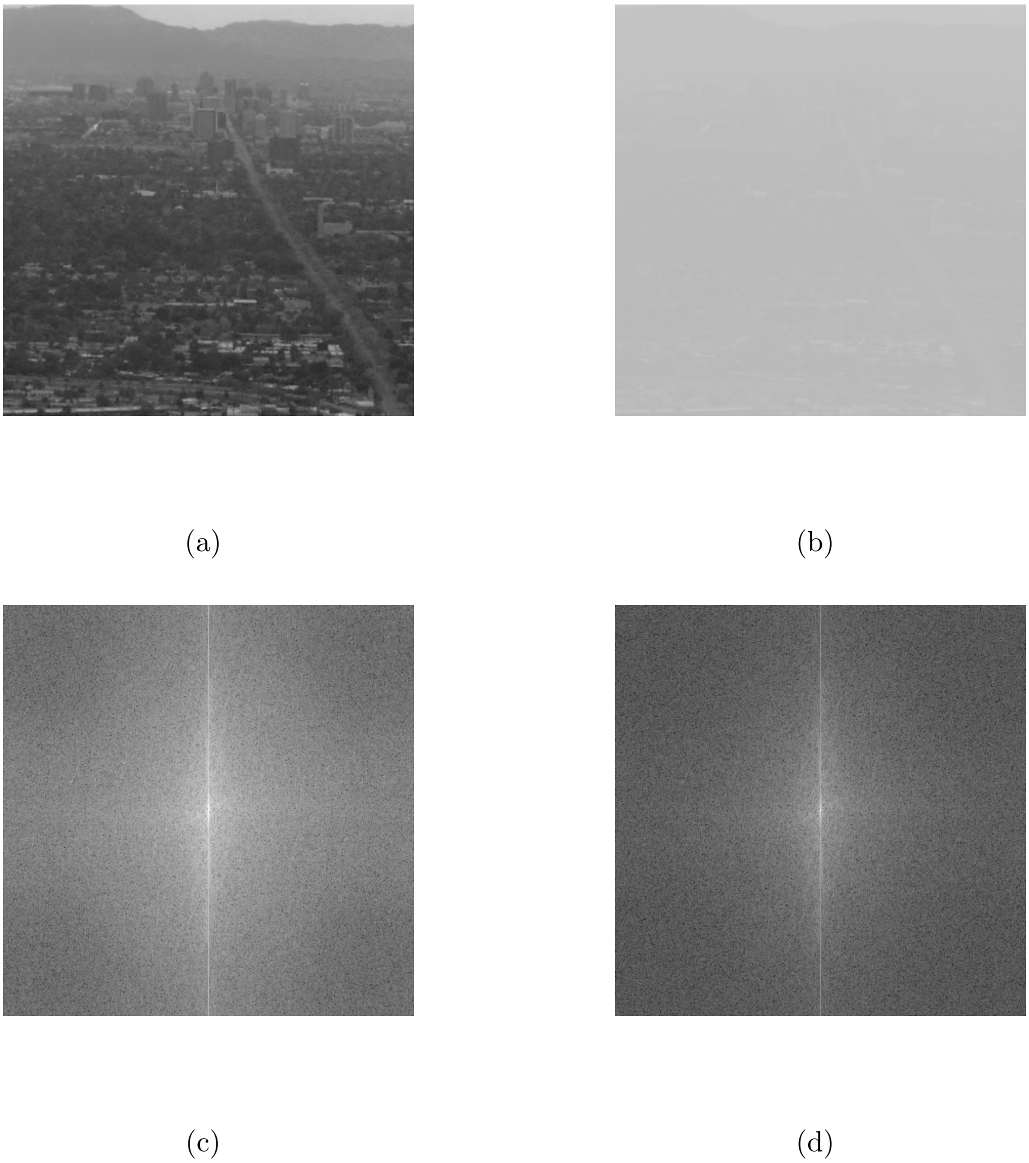


Figure 4.1: Images (a) and (b) show the same scene at different levels of visibility, where images (c) and (d) are their respective Fourier transforms. While the same shape is maintained, the two Fourier transforms have notably different magnitudes at most frequencies. In particular, the spectrum in (c) has larger values further from the center indicating that image (a) has more high-frequency content or edge components.

Assuming that the transmission and ambient light are constant in the region, this is equivalent to

$$\min_{x\in\Omega}\left(\frac{I^c(x)}{A^c}\right) = t(x)\min_{x\in\Omega}\left(\frac{J^c(x)}{A^c}\right) + (1-t(x))\ . \tag{4.7}$$

Now the minimum is computed with respect to each color channel

$$\begin{aligned}\min_{c}\left(\min_{x\in\Omega}\left(\frac{I^c(x)}{A^c}\right)\right) = t(x)\min_{c}\left(\min_{x\in\Omega}\left(\frac{J^c(x)}{A^c}\right)\right) \\ +(1-t(x))\ .\end{aligned}$$

Looking more closely at the right hand side of this equation, we realize that based on the dark channel prior, which again assumes there is some region with zero or near-zero haze-free intensity in one of the color channels, that

$$\min_{c}\left(\min_{x\in\Omega}\left(\frac{J^c(x)}{A^c}\right)\right) = 0\ . \tag{4.8}$$

since A^c is positive. We thus get

$$t(x) = 1 - \min_{c}\left(\min_{x\in\Omega}\left(\frac{I^c(x)}{A^c}\right)\right)\ . \tag{4.9}$$

With a naive approach, ambient lighting could be calculated as the brightest pixel in the image. However, this estimate could then change drastically based on a single pixel. In order to be robust to any outlier pixels, we estimate the ambient light A^c as the maximum first percentile of pixel intensities in a region just above the horizon. Likewise, we compute the minimum ambient-light normalized image intensity for a region $\min_{x\in\Omega}\left(\frac{I^c(x)}{A^c}\right)$ as the minimum first percentile rather than finding the absolute minimum. These values are more representative of how we perceive the image and provide reliable results.

4.2 Linear Least Squares Regression

Again, the primary objective is to estimate light extinction β_{ext} given an image $I(x)$. Taking the log of both sides of eq. (3.1) gives a linear relationship between extinction and transmission

$$\ln t(x) = \beta_{ext} r(x) \ . \tag{4.10}$$

In this case of transmission based on local contrast computed in either the spatial domain C_{lsd} or frequency domain C_{lfd} this becomes

$$\ln C_l = \ln |\nabla_x J(x)| + \beta_{ext} r(x) \tag{4.11}$$

where $|\nabla_x J(x)|$ is the true contrast of the scene. Rearranging we get

$$\beta_{ext} = \frac{\ln C_l}{r(x)} - \frac{\ln |\nabla_x J(x)|}{r(x)} \tag{4.12}$$

and use linear least squares regression (LLSR) to learn the scaling $\frac{1}{r(x)}$ and offset $-\frac{\ln |\nabla_x J(x)|}{r(x)}$ parameters from a labelled training set.

In the case of transmission based on the dark channel prior, eq. (4.10) becomes

$$\beta_{ext} = \frac{\ln t(x)}{r(x)} \ . \tag{4.13}$$

However, we found that pure scaling results in poor performance so we include an offset to accommodate for errors in the model–perhaps there is no dark pixel in the image region–and/or errors in the observations such as unreliable estimates of the ambient light. We again use LLSR to learn the scaling parameter and the offset.

4.3 Data Access and Dataset Generation

We evaluate our method using an extensive set of images and ground truth extinction readings from the the Arizona Department of Environmental Quality which manages the PhoenixVis.net visibility web cameras website [23]. 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 minutes.
- Digital images of Camelback Mountain (CAME) captured every 15 minutes.
- The extinction coefficient β_{ext} measured every hour using a transmissometer.
- The scattering coefficient β_{sp} measured every hour using a nephelometer.

The SOMT camera is located on a mountain north of Phoenix and faces south. Figures 1.1(a) and 1.1(c) contain examples of good and bad visibility for the SOMT camera. The CAME camera is located on a tall structure in downtown Phoenix and faces north east. Figures 1.1(b) and 1.1(d) contain examples of good and bad visibility for the CAME camera. The transmissometer and nephelometer are located in downtown Phoenix and are approximately within the field of view of both cameras.

All images are in the RGB colorspace and have been JPEG compressed at an unknown quality level. The SOMT images measure 1536×1024 pixels. The CAME images are a mix of 1536×1024 and 1230×820 pixels so we transform all CAME images to a common size of 1230×820 using bilinear interpolation. Each image is partitioned using a 6×4

grid and a prediction model is trained and evaluated for each block separately. Figure 1.1 shows the grid layout for the two scenes.

We only consider images taken at the top of each hour, since this is when the transmissometer and nephelometer readings are made, and during daylight hours, approximately 10 am to 4 pm. The ground truth data include a warning flag for any visibility readings that are unreliable due to mechanical or environmental conditions. These images are not included in our training or test sets. This results in a labeled dataset of 8,598 images from the SOMT camera and 7,676 images from the CAME camera.

4.4 Experiments

We evaluate our method based on how well the learned model is able to predict the (known) extinction coefficient β_{ext} corresponding to an image $I(x)$ using only the image features. We perform five-fold cross-validation to observe how well our method generalizes. The labeled images are randomly partitioned into five equal-sized sets. The model is learned using four of the sets and used to predict the extinction coefficient for the images in the fifth held-out set. We evaluate the accuracy of our model using the coefficient of determination R^2 between the predicted and ground truth values. Let ${\beta^i_{ext}}'$ and β^i_{ext} be the predicted and true extinction coefficients for image i respectively, then

$$R^2 = 1 - \frac{\sum_{i=1}^{n}({\beta^i_{ext}}' - \beta^i_{ext})^2}{\sum_{i=1}^{n}(\beta^i_{ext} - \overline{\beta_{ext}})^2} \tag{4.14}$$

where n is the number of images in the evaluation set and $\overline{\beta_{ext}}$ is the mean of the true values. R^2 has a maximum value of 1 with higher values indicating a more accurate model.

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |{\beta_{ext}^{i}}' - \beta_{ext}^{i}| \tag{4.15}$$

The values of R^2 and MAE reported below are averages over the five training/test splits.

4.4.1 Single Variable Analysis

The evaluation is performed on each of the 24 image blocks separately. The visual distance $r(x)$ and transmission $t(x)$ in eqs. (4.12) and (4.13) are assumed to be constant over a block. The image region Ω used to compute local contrast in the spatial domain in eq. (6.1) and the transmission based on the dark channel prior in eq. (4.9) is taken as an image block. Contrast in the frequency domain is computed by applying the 2D-DFT to an image block.

We perform a series of experiments to: determine which image feature is most effective for predicting the coefficient of extinction; whether the predictions are more correlated with the transmissometer or nephelometer readings; the effect of scene geometry; and the optimal lower and upper cutoff frequencies for the Fourier analysis.

The results are summarized in Table 4.1. For each combination of *image feature*–local contrast in the spatial (C_{lsd}) or frequency (C_{lfd}) domain, or dark channel prior; *ground*

Table 4.1: Summary of results for each combination of *image feature*–local contrast in the spatial (C_{lsd}) or frequency (C_{lfd}) domain, or dark channel prior; *ground truth reading*–transmissometer or nephelometer; and *scene*–SOMT or CAME. R^2 and MAE (Mm^{-1}) values are given for the image block that results in the best model as ranked by R^2 for the training set (higher is better.)

	SOMT				CAME			
	Transmissometer		Nephelometer		Transmissometer		Nephelometer	
	R^2	MAE	R^2	MAE	R^2	MAE	R^2	MAE
C_{lsd}	**0.646**	**12.5**	0.407	10.4	0.103	21.6	0.199	12.0
C_{lfd}	0.549	14.8	0.372	10.2	0.103	21.5	0.169	12.3
Dark Channel	0.237	20.3	0.309	10.8	0.081	21.3	0.061	13.0

truth readings–transmissometer or nephelometer; and *scene*–SOMT or CAME–it lists the R^2 and MAE values for the image block that results in the best model as ranked by R^2. The 6×4 image blocks are numbered 1 through 24 in raster-scan order (see Figure 1.1). The values reported for C_{lfd} are the best over a range of lower and upper frequency bounds. We now discuss these results.

Image Features Local contrast consistently outperforms the dark channel prior across scenes and ground truth labeling. Further, contrast computed in the spatial domain using the image gradient generally outperforms contrast computed in the frequency domain using the Fourier energy spectral density. These two contrast features are of course related and will be discussed further below.

Ground Truth Readings For the SOMT scene, local contrast is a better predictor of the transmissometer than the nephelometer readings while the reverse is true for the dark channel prior. While our model assumes that the effect of absorption is negligible, as is commonly done in atmospheric modelling, the ground truth transmissometer and nephelometer values in our dataset are different indicating there is a non-zero absorption component β_{abs}; i.e., $\beta_{ext} \neq \beta_{sp} + \beta_{abs}$. Table 4.2 gives the statistics of the ground truth readings. Performing a linear least squares fit between the 8,598 transmissometer and nephelometer readings associated with scene SOMT gives $\beta_{ext} = 1.18\beta_{sp}+24.6$ with an R^2 value of 0.571 and an MAE of 15.1 Mm^{-1}. This indicates that there is a nonlinear relation between these two measurements which cannot be accounted for in the scaling and offset parameters of our linear model of β_{ext} and $\ln C_l$ (eq. 4.12) or $\ln t(x)$ (eq. 4.13). It also shows interestingly that *once calibrated, the image features can provide a better estimate of light extinction (as measured using the transmissometer) than the nephelometer*. This shows that absorption plays a large role in light extinction for the observed image and may also have implications for detecting pollution.

Returning to Table 4.1, we see that things are reversed for the the CAME scene: local contrast is a better predictor of the nephelometer than the transmissometer readings while the dark channel prior is a better predictor of the transmissometer than the nephelometer readings.

Scene Comparison

The image based approach to estimating light extinction is significantly more effective for SOMT than the CAME. This is true for all feature and ground truth combinations. This might be due in part to the different image resolutions, particularly for the local

Table 4.2: The statistics for the ground truth transmissometer and nephelometer readings. All values are in Mm^{-1} except for R^2.

	SOMT (8,598 pts)		CAME (7,676 pts)	
	Trans.	Neph.	Trans.	Neph.
min	9	0	9	0
max	218	153	218	155
mean	57.8	28.2	57.7	28.1
median	50	24	50	24
std dev	31.2	20.1	31.0	19.8
R^2	0.571		0.567	

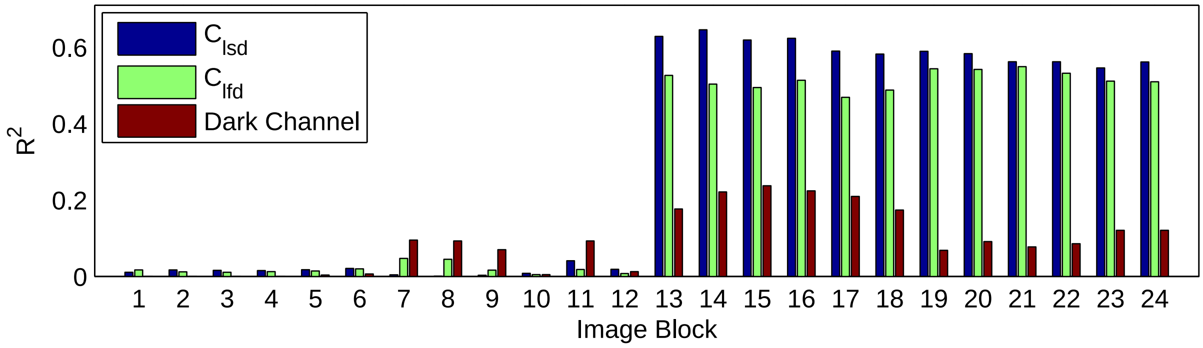


Figure 4.2: R^2 for different image blocks for SOMT and transmissometer readings.

contrast approaches, but is more likely due to different scene geometry as discussed in the next section.

Image Regions

Figures 4.2-4.5 plot the R^2 values for each of the three image features over all 24 image blocks. Figure 4.2 contains the results for scene SOMT and the transmissometer readings; Figure 4.3 for scene SOMT and the nephelometer readings; Figure 4.4 for scene CAME and the transmissometer readings; and Figure 4.5 for scene CAME and the nephelometer readings.

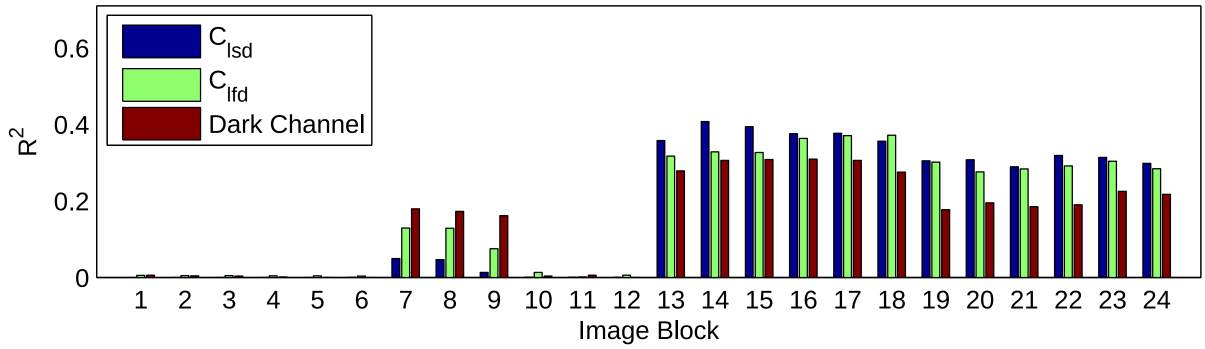


Figure 4.3: R^2 for different image blocks for SOMT and nephelometer readings.

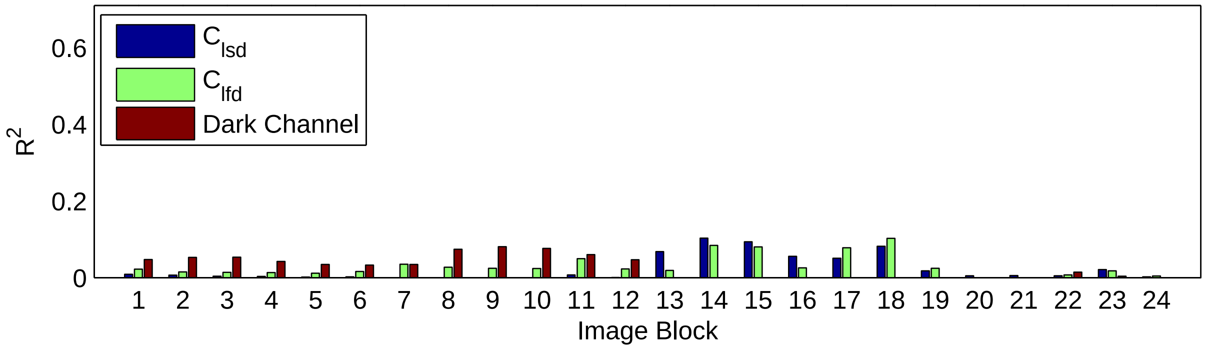


Figure 4.4: R^2 for different image blocks for CAME and transmissometer readings.

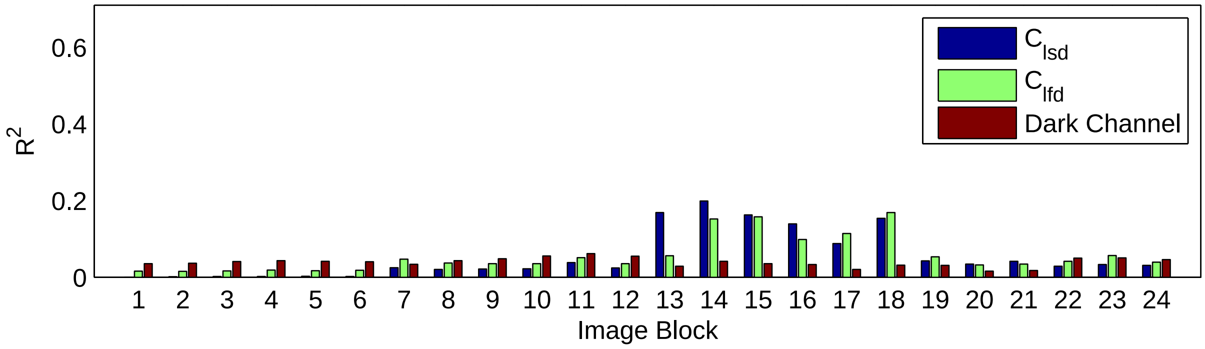


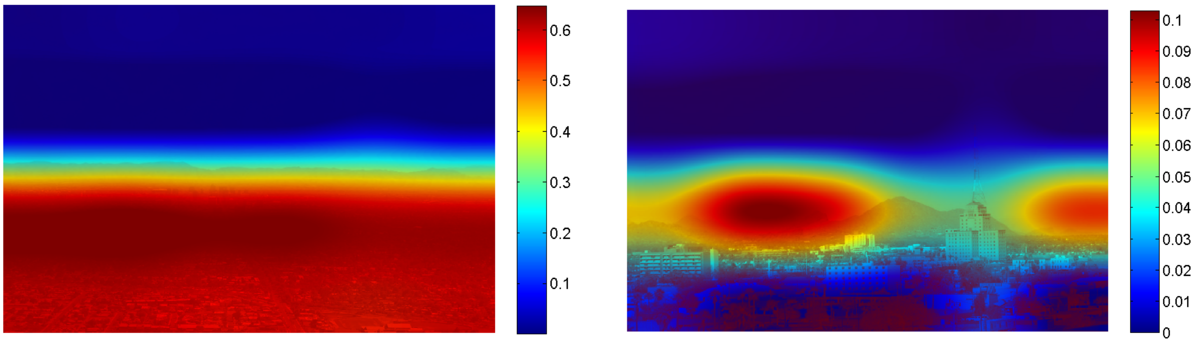
Figure 4.5: R^2 for different image blocks for CAME and nephelometer readings.

These region level results provide insight into why SOMT is the more effective scene. The R^2 values for all features and ground truth readings are relatively large for all the blocks below or containing the horizon in SOMT. However, the bottom row of blocks (19-24) in CAME, which represent the closest parts of the foreground, all perform poorly. This foreground region is much closer than any of the SOMT regions and thus is too close to the camera to estimate light extinction. There is simply not enough atmosphere to cause sufficient variation in the image features. This is evident in figures 1.1(c) and 1.1(d) in which the foreground regions of the two scenes are affected very differently by a similar increase in light extinction. Further, the lower vantage point of the CAME camera results in a perspective with very little distant scenery in terms of image area. The image features are now extracted from blocks containing sky regions which results in worse performance than SOMT.

The effect of different scene geometry is visually depicted in in figures 4.6(a) and 4.6(b) using smoothed colormaps of R^2 overlaid on SOMT and CAME images. These results correspond to predicting the transmissometer readings using local contrast in the spatial domain. The distant regions are the most effective in both scenes; the ideal regions for CAME, however, are much smaller. Due to the poor performance of the CAME dataset, all remaining sections will be completed using the SOMT camera exclusively.

Frequency Bands

The LLSR fit of local contrast in the frequency domain C_{lfd} is performed for lower w_l and upper w_u cutoff frequencies ranging from 0 to 1 in increments of 0.05 where 1



(a) SOMT

(b) CAME

Figure 4.6: Smoothed colormaps of R^2 overlaid on scene images indicating the effect of scene geometry. The distant regions are the most effective in both scenes; the ideal regions for CAME, however, are much smaller.

corresponds to the Nyquist frequency. The R^2 values reported in table 4.1 and figures 4.2-4.5 represent the optimal cutoffs. Figure 4.7 shows how the optimal frequencies vary by image block for SOMT using the transmissometer readings. In particular, the values for non-sky blocks (7-9 and 13-24) decrease as the scene distance increases. This is expected because even in relatively good conditions, the atmosphere still acts as a low-pass filter whose attenuation increases with distance and so the higher frequency image signal components for distant scenes do not vary enough to be informative. We also see that w_u never equals the maximum frequency in non-sky regions even when they are close-by. This may be in part due to the low-pass filtering of the atmosphere but is more likely due to JPEG compression which discards the higher frequency signal components.

The two measures of local contrast are of course related since convolution with the Sobel kernels in the spatial domain corresponds to applying a related filter in frequency domain. The fact that contrast in the spatial domain provides the best result overall

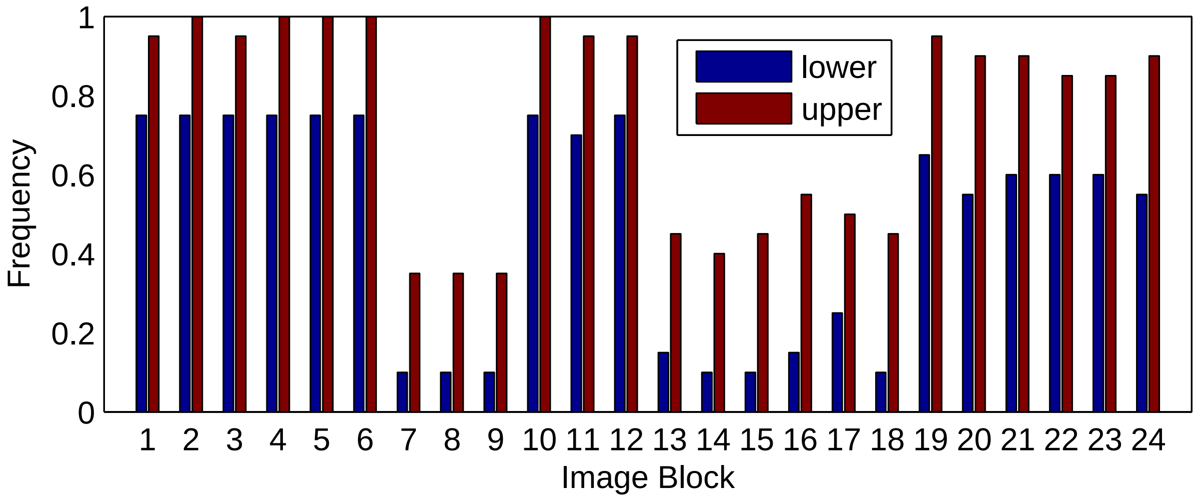


Figure 4.7: Per block values of the optimal lower w_l and upper w_u frequencies for local contrast in the frequency frequency domain. This is for the SOMT scene and transmissometer readings. A value of 1 corresponds to the Nyquist frequency.

indicates that the optimal filter for estimating atmospheric light extinction does not have a band-pass frequency response but is more complex. The frequency response of the Sobel kernels might provide a good initial estimate of this filter. This filter would of course need to be tuned to scene distance just like our simple band-pass filters.

Visibility Index Based on Deciview Analysis

In 2003, the Arizona Department of Environmental Quality defined a five valued visibility index [29] to track regional visibility conditions over a multi-year period. This index is based on deciviews which are linear with respect to perceived visual changes analogous to how decibels are for sound [26]. Deciview readings DV are derived from transmissometer estimates of light extinction:

$$DV = 10 \ln(\beta_{ext}/10 \text{ Mm}^{-1}) \ . \tag{4.16}$$

The visibility index is then determined by binning DV into five ranges corresponding to “excellent”, “good”, “fair”, “poor”, and “very poor”.

We applied the same deciview conversion and binning to the predictions of our best model, local contrast features in the spatial domain for block 14 of SOMT. Figure 4.9 compares the predicted index values with those computed using the ground truth transmissometer readings. Note the similarity in the distributions. This fit corresponds to an R^2 value of 0.980 if the values are mapped to 1,...,5. Figure 4.8 shows a scatter plot for the results of deciview prediction with a linear regression model.

4.4.2 Multivariate Analysis and Feature Selection

Due to the success of the linear model for a single block, the same concept was then expanded to a multidimensional feature vector that combined various blocks. In addition to the local spatial contrast, the mean and standard deviation were calculated for each channel in the HSV color model. Because haze causes an overall decrease in saturation, we expect additional information about the hue and saturation to be useful. Due to the high number of blocks and features, feature selection must be performed in order to create a functional feature vector. Based on our knowledge of transmission, we would expect that contrast and saturation would be the strongest indicators of visibility. We instead found that local variation in saturation provided more information than saturation alone. An overview of the selected features is provided in table 4.3. The local contrast C_l includes both $|\Delta I|$ and V_{std}, the standard deviation of the value layer of the HSV image.

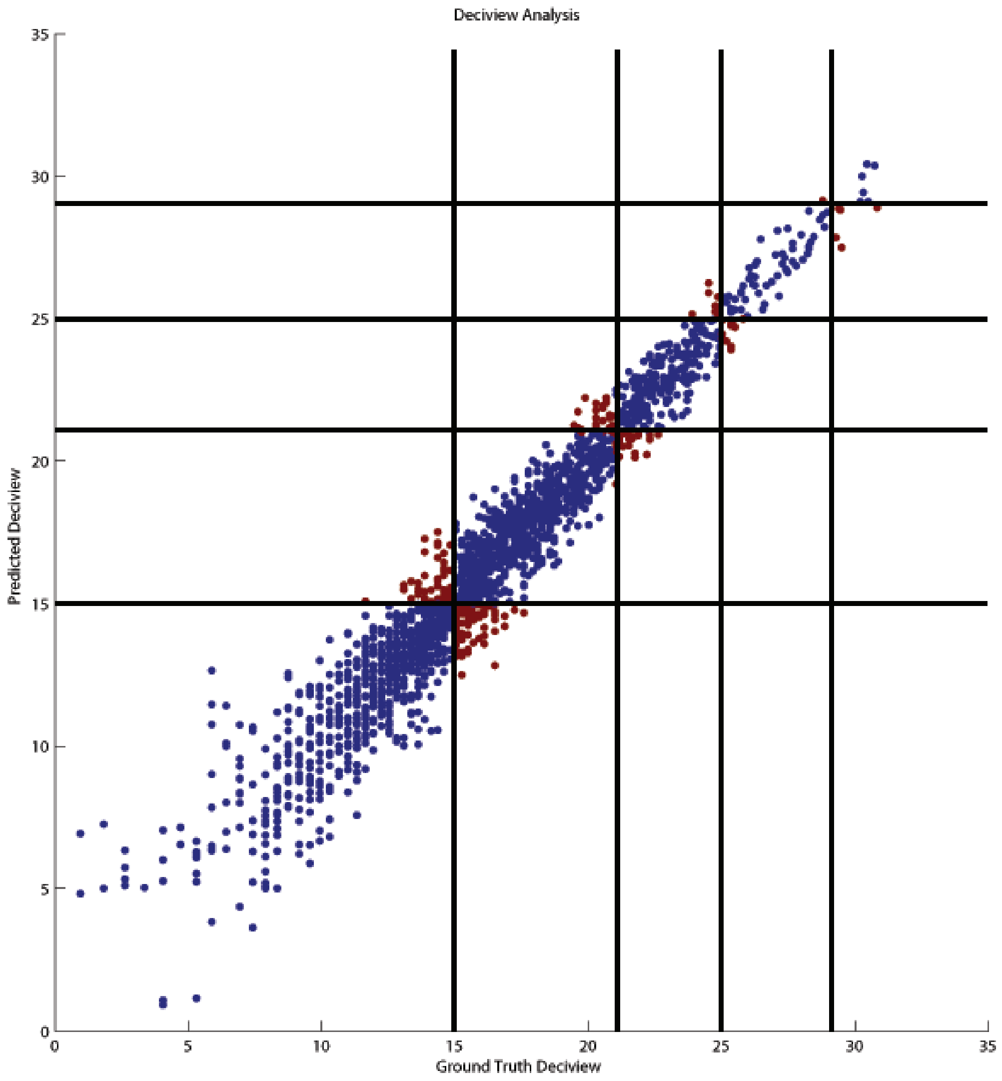


Figure 4.8: Predicted deciview values are plotted against the ground truth data with the five subcategories separated by vertical lines. Note that all misclassifications (red dots) occur very close to the boundary lines. The displayed results correspond with a 89.2% accuracy rate.

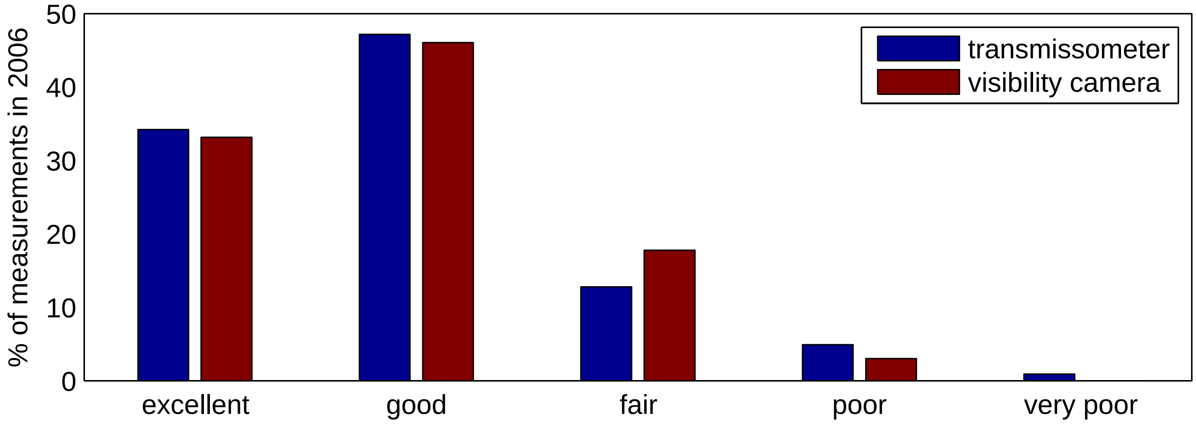


Figure 4.9: Comparison of visibility index measured using a transmissometer and a visibility camera.

Block and feature selection was performed with a greedy search algorithm. While we could reuse the top blocks from Section 4.4.1, many of these blocks can be highly correlated with one another. In this case, improvements are minimal because the two blocks are often modeling the same general feature from the image. Instead, we iteratively expand our feature vector by searching through our feature space. At each iteration, we add an additional feature that provides the lowest error on our training set when combined with our previous feature vector. This results in a complementary set that provides the best model for our images across all combinations of blocks and features. Graphs are shown in Figures 4.10 and 4.11. Table 4.4 is of particular interest because it shows which features are chosen at every iteration. While the local contrast is most important, we can see that the saturation and hue are used immediately afterwards to achieve the largest increase in the R^2 value.

These results demonstrate that our linear model is expandable to a high-dimensional feature search that is able to generalize to our test sets. This provided some insight onto

Table 4.3: Summary of features selected over a 100-iteration feature search.

Feature	Iter. Chosen
Local Contrast	32
Hue	14
Hue Contrast	10
Saturation	6
Saturation Contrast	36
Pixel Intensity	2

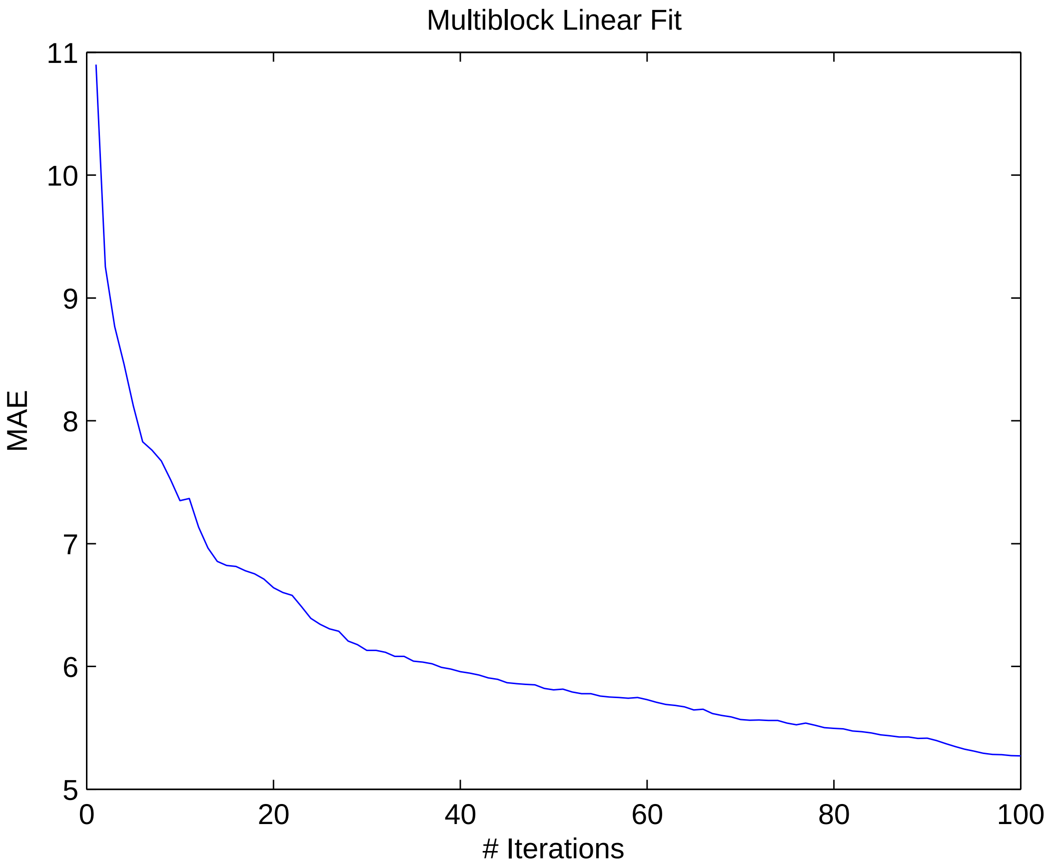


Figure 4.10: Results showing the decrease in the mean absolute error (MAE) over 100 iterations of the multiblock search.

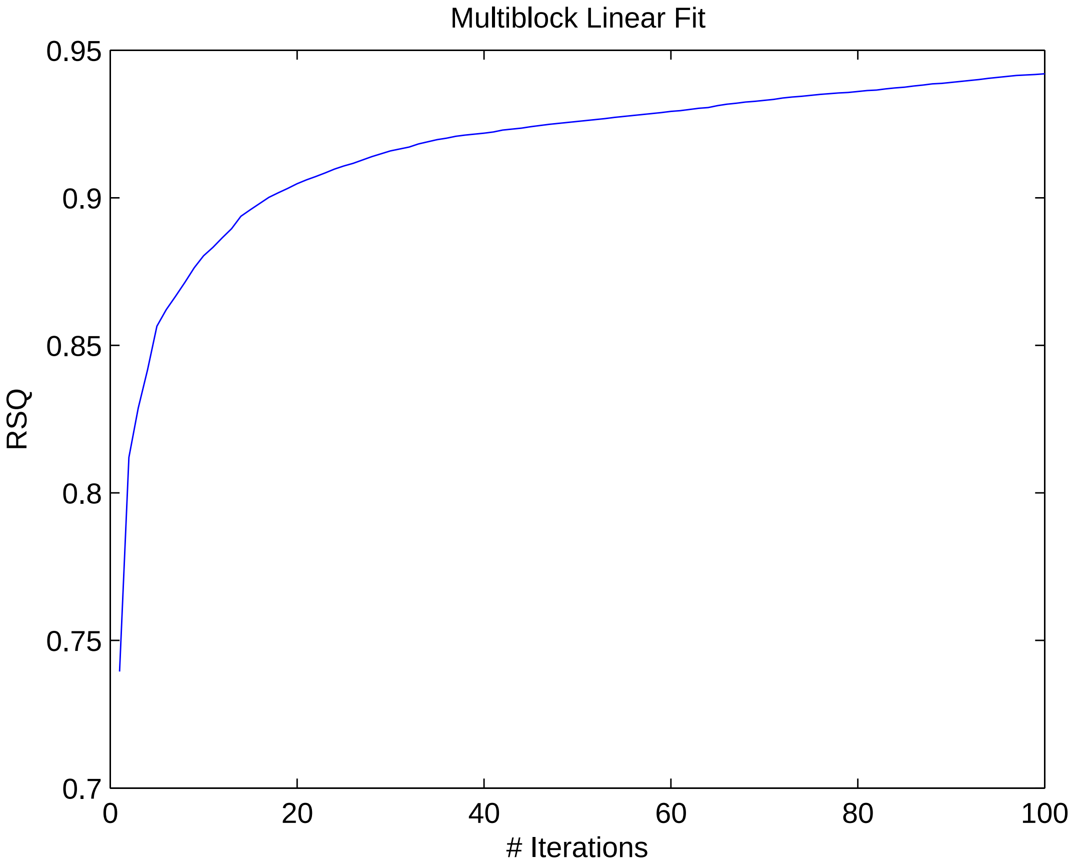


Figure 4.11: Results showing the increase in the R^2 value over 100 iterations of the multiblock search.

Table 4.4: Details of features selected over a 25-iteration feature search. Blocks 1-672 correspond to the sky region. Blocks 673-1728 correspond to the regions from the horizon to the foreground in raster scan order.

Iteration	Block #	Feature	R^2	MAE
1	920	C_{lsd}	0.7395	10.9002
2	817	S_{mean}	0.8120	9.2539
3	1035	S_{std}	0.8287	8.7681
4	789	V_{std}	0.8417	8.4624
5	853	H_{mean}	0.8565	8.1198
6	1352	H_{mean}	0.8621	7.8284
7	893	S_{std}	0.8666	7.7596
8	855	C_{lsd}	0.8713	7.6737
9	1082	V_{std}	0.8763	7.5188
10	384	V_{mean}	0.8832	7.3667
11	804	H_{mean}	0.8803	7.3492
12	1111	H_{mean}	0.8864	7.1334
13	1686	S_{std}	0.8895	6.9641
14	1393	S_{std}	0.8937	6.8544
15	1006	S_{std}	0.8959	6.8226
16	115	S_{mean}	0.8980	6.8141
17	910	C_{lsd}	0.9001	6.7783
18	239	H_{mean}	0.9017	6.7544
19	1179	V_{std}	0.9031	6.711
20	959	S_{std}	0.9048	6.6413
21	1119	S_{std}	0.9061	6.6011
20	1299	C_{lsd}	0.9072	6.5776
21	1557	S_{std}	0.9084	6.4873
22	1208	H_{std}	0.9097	6.3923
23	767	H_{std}	0.9108	6.3421
24	1386	S_{std}	0.9117	6.3057
25	906	S_{mean}	0.9128	6.2862

the types of features selected by our algorithm, most notably the usage of contrast within the saturation layer.

Chapter 5

Incorporating Unlabeled Training Data

A common problem with training data is that the observations and ground truth measurements are not captured at the same frequency. This is the case with the Phoenxvis.net dataset where the images are acquired every 15 minutes but the transmissometer and nephelometer ground truth measurements are acquired only every hour. This will result in $\frac{3}{4}$ of our images being unlabeled and thus unused when training standard regression models.

This section provides details for incorporating these unlabeled images into our linear framework described in Chapter 4. By constraining the structure of the affinity matrix over a set of sequential data, we can solve for temporal weights that minimize the error of the system. We compare different connectivity schemes for the graph Laplacian in order to incorporate unlabeled data and exceed the performance of a standard linear fit which was only on labeled data.

5.1 Introduction

The intuition behind this section is that images taken within a short span of time will have similar image features and β_{ext} values. Even if some of these images do not have ground truth data, weighting can be established between neighboring images that allows them to be considered when the linear fit is being modeled. We take advantage of temporally adjacent images using a Laplacian regularized linear least squares (LRLLS) approach. The primary goal is to minimize the prediction error by constraining the structure of the affinity matrix and optimizing its components. This will leverage the features of temporally adjacent images to improve the model, and is expected to increase the accuracy of our visibility estimation system due to the relative similarity of images taken within a small timeframe.

By incorporating the graph Laplacian, we can leverage these unlabeled images when we develop our regression model. The affinity matrix determines the strength of the connection between various images, and can be used as a temporal prior by adding bias to images in close temporal proximity with a sequential dataset. By constraining the structure and parameterizing these relationships, we can optimize the use of unlabeled data for our system. Using this approach, the unlabeled images can still be analyzed and influence the usage of nearby labeled data. Two schemes, adjacency and Gaussian weighting, are assessed for their ability to incorporate unlabeled data into an accurate model. Furthermore, the sensitivity of the Gaussian model with respect to the width parameter σ will be analyzed. Unlike our previous work, an intermediate validation stage is now being used to assign model parameters.

5.2 Incorporating Unlabeled Observations

This work is inspired and based on the paper by Xie et al. [30]. Given that we have 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. This is 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:

$$f' = arg\,\underset{\alpha}{min} \frac{1}{l} \sum_{i=1}^{l} (y_i - f(x_i))^2 + \gamma \sum_{i,j}^{l+u} w_{ij} (f(x_i) - f(x_j))^2 \tag{5.1}$$

where

$$f(x) = \alpha_0 + \alpha_1 x_1 + ... + \alpha_n x_n \tag{5.2}$$

Note how the objective function simultaneously minimizes the prediction error on the labeled training set and the prediction between closeby points whether they are labeled or not. If we construct the graph matrix W, the solution to this optimization problem can be solved using the graph Laplacian.

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

$$f' = arg\,\underset{\alpha}{min}\frac{1}{l}J(\bar{Y} - X\alpha)^2 + \gamma(X\alpha)^T LX\alpha \tag{5.3}$$

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

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

The optimization problem now becomes

$$f' = arg\,\underset{\alpha}{min}\frac{1}{l}J(\bar{Y} - X\alpha)^2 + \gamma(X\alpha)^T LX\alpha \tag{5.4}$$

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

$$\tfrac{\partial}{\partial\alpha}\tfrac{1}{l}J(\bar{Y} - X\alpha)^T(\bar{Y} - X\alpha) + \gamma(X\alpha)^T LX\alpha = 0$$

$$\tfrac{J}{l}(-X^T)(\bar{Y} - X\alpha) + \gamma X^T LX\alpha = 0$$

$$-JX^T\bar{Y} + JX^T X\alpha + \gamma X^T LX\alpha = 0$$

$$(JX^T X + \alpha X^T LX)\alpha = JX^T\bar{Y}$$

$$\alpha = (JX^T X + \gamma X^T LX)^{-1} JX^T\bar{Y} \tag{5.5}$$

Eq. (5.5) is the used to compute the coefficients α for the linear model. With an appropriate validation set and γ value, we expect that this will result in an improved visibility estimation.

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. 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 $\sigma = 20$ minutes.

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

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

5.2.1 Connectivity Metrics

The use of unlabeled images relies on the weightings that relates temporally nearby data. Different schemes are available to weight these values, and results will vary depending on the nature of the environment being observed. The two approaches being tested for this project are adjacency weighting and Gaussian weighting. The adjacency weighting only considers observations that are adjacent. The Gaussian weighting will provide an exponentially decreasing weight dependent on the σ-value used.

5.3 Dataset

We evaluate our method using an extensive set of images and ground truth extinction readings from the the Arizona Department of Environmental Quality which manages the PhoenixVis.net visibility web cameras website [23]. Our dataset consists of the following acquired over 2006:

- Digital images of South Mountain (SOMT) captured every 15 minutes.
- The extinction coefficient β_{ext} measured every hour using a transmissometer. This serves as our ground truth data for training, validation, and testing.

All images are in the RGB colorspace and have been JPEG compressed at an unknown quality level. The SOMT images measure 1536×1024 pixels. Each image is partitioned into 128×128 pixel blocks and a prediction model is trained for a given combination of blocks.

These images were subdivided into a training set, a validation set, and two test sets as shown in table 5.1. Training set values were chosen across the span of three randomly chosen days. The validation and test sets were randomly selected from remaining labeled data.

Table 5.1: Summary of dataset composition used in experiments. Mean and standard deviation values refer to the β_{ext} values from the transmissometer and do not include any unlabeled images.

	Labeled Images	Unlabeled Images	Mean	StdDev
Training (TR)	21	63	42.0476	17.2496
Validation (V)	30	0	36.5	10.7599
Test Set 1 (T1)	119	0	51.0336	26.2507
Test Set 2 (T2)	119	0	56.9916	31.1711

Figure 5.1: Two blocks in an SOMT image that were manually chosen for the feature extraction process.

5.4 Experiments

If we assume that $r(x)$ is a constant over a chosen image block, visibility estimation with our Laplacian regularized linear least squares approach uses the same theoretical foundation as Section 4. Two blocks, shown in Figure 5.1, were chosen to derive features from the images.

We evaluate our method based on how well the learned model is able to predict the (known) extinction coefficient β_{ext} corresponding to an image $I(x)$ using only the image features. Two different types of tests are used to monitor the effectiveness of this algorithm.

The first test learns a linear model using standard least squares regression using only the labeled images in TR+V as the training data. The model is then evaluated using a held-out test set.

The second test uses the proposed LRLLS to learn a model. Now, both the labeled and unlabeled images in TR are used to learn the model and the values of σ and γ are optimized using a validation set (V). The model is evaluated again on a test set.

We evaluate the accuracy of the models using the coefficient of determination R^2 between the predicted and ground truth values in the test set. Let ${b^i_{ext}}'$ and b^i_{ext} be the predicted and true extinction coefficients for image i then

$$R^2 = 1 - \frac{\sum_{i=1}^{n}({b^i_{ext}}' - b^i_{ext})^2}{\sum_{i=1}^{n}(b^i_{ext} - \overline{\beta_{ext}})^2} \tag{5.6}$$

where n is the number of images in the evaluation set and $\overline{\beta_{ext}}$ is the mean of the true values. R^2 has a maximum value of 1 with higher values indicating a more accurate model.

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|{b^i_{ext}}' - b^i_{ext}|. \tag{5.7}$$

Table 5.2: Summary of results for each combination of validation and test sets. R^2 and MAE (Mm^{-1}) values are reported from the trained model.

Training Set	Validation Set	Model	Connectivity	σ	Test Set	R^2	MAE
TR+V	-	LLS	-	-	T1	0.6803	9.7042
TR	V	LRLLS	Gaussian	20	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	20	T2	**0.7368**	**10.6980**
TR	V	LRLLS	Adjacency	-	T2	0.7346	10.7410

5.5 Results

The results are summarized in table 5.2. The R^2 and MAE values are given for each combination of the validation sets, test sets, and connectivity schemes, as well as the simple linear models described above. These results reflect the best model after performing a linear fit and adapting with the validation set as needed. Values for γ were most effective in the range from -0.2 to -0.4.

Training Set The training set is based on both labeled and unlabeled data. The size and temporal relevance of the unlabeled data will determine how effective this improvement is for the linear fit. In addition, a very large training set may overshadow any beneficial effects of the validation set. This may suggest that this algorithm will offer the most improvement when dealing with a limited amount of data.

Validation Set The validation set can have a strong effect on the success of the model. In the standard linear model, adding the validation data to the training set will have a limited effect on the outcome. However, the validation approach assumes that this set has a similar distribution and feature set to the test images. A non-representative validation set will significantly reduce the quality of the fit. Likewise, we would expect to see better results if the statistics for our validation set were more closely aligned with the test sets.

Test Set Both test sets have an increased R^2 and reduced MAE when incorporating the unlabeled data. This is an encouraging confirmation that our method is able to provide a better fit than the standard linear approach.

Gaussian vs. Adjacency Connectivity The σ value chosen determines how temporally relevant the unlabeled images are in relation to labeled images from the training data. This is specific to each problem type because it is based on the variation in ground truth data for our particular data set. If we are viewing a region that is prone to sudden drastic changes in visibility, it may be more important to constrain the usage of unlabeled data. The fact that $\sigma = 20$ minutes for our training data, where images are typically 15 minutes apart, shows that an emphasis is being put on temporally close images. An example Gaussian weighting matrix with this σ value is shown in Section 5.2.

As shown above, the adjacency connectivity provides similar performance to the Gaussian distribution. The low σ decreases the width of the Gaussian and nearly mirrors the effects of a strictly adjacent connectivity scheme. An alternative approach could be to use linear weighting for some scenes, as it may be the case that exponentially decreasing weights do not accurately represent our data when extrapolated. Ultimately, the best weighting method is dependent on the temporal properties of the dataset.

5.6 Discussion

This section introduced a method for incorporating unlabeled data when estimating the coefficient of light extinction using visibility cameras. This method appears to result in a reasonable improvement in prediction models over a standard linear model. Comparing Gaussian and adjacency connectivity metrics provided informative insight into the temporal relation of images for our particular dataset. Alternative connectivity matrices can be used to better represent the temporal relationship between images for a given scene.

While this work should generalize to higher dimensionality features, it becomes increasingly important to have a relevant γ and validation set to produce accurate results. Furthermore, room for improvement is dependent on the ratio of labeled:unlabeled images in the dataset.

Chapter 6

M5' Regression Trees

Transmission is an inverse exponential function of both the depth 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 area 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 features and 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.

This chapter seeks to expand on previous 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 segmenting the image into block regions and using an M5' regression tree to estimate visibility. Unlike the linear multiblock model from Chapter 4, this models a discontinuous piece-wise linear function based on our features. We expect that this will have an advantage over a linear fit by allowing for non-linear changes in the environment, such as the effects that a cloudy day might have on visibility.

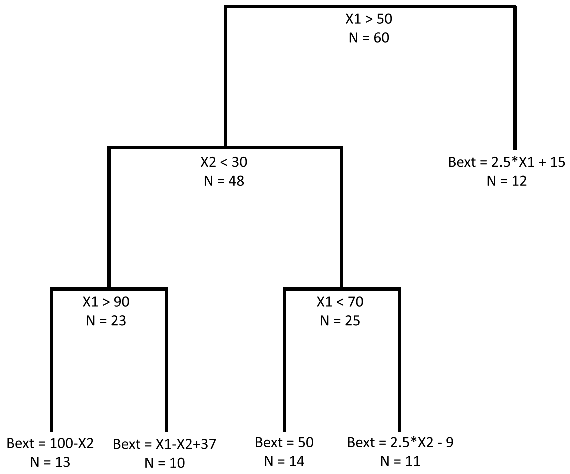


Figure 6.1: A mock-up example of a regression tree for a 2-feature visibility problem.

6.1 Regression Trees

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 with the entropy of the resulting leaf nodes. Intuitively, this means that we try to select splits that group similar objects 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 figure 6.1. When working with a large number of blocks and features, the decision-making tree becomes very complex and can be difficult to interpret by hand.

6.2 Methods

The following sections detail the use of image features, linear regression, and incorporating M5' regression trees. Based on the functionality and methodology of the M5' tree, I then tested the use of multiple scales in an attempt to improve the resulting models.

6.2.1 M5' Regression

To implement M5' regression trees I used a Matlab toolbox called M5PrimeLab [31]. M5' regression generates a decision tree similar to ID3. ID3 involves using a measure of entropy in order to calculate the information gain. The feature with the largest information gain represents the best way to split the current examples into two distinct groups. Rather than entropy, M5' looks at the standard deviation of each set. This is important because we are evaluating the system over continuous values rather than trying to solve a classification problem. Standard deviation reduction (SDR) is calculated as:

$$SDR = sd(T) - \sum_{i} \frac{|T_i|}{|T|} \times sd(T_i) \tag{6.1}$$

where $sd(T)$ is the standard deviation of our parent node, $sd(T_i)$ is the standard deviation of each leaf 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 up 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 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 small set of data points which can be characterized by a single value or linear model. Because it is still performing linear regression over localized regions, we use the same theoretical foundation provided in Section 4.

6.2.2 Multi-scale Features

Multi-scale features were incorporated to leverage the method used by the M5' regression trees. Although we use equal-size regions for our image segmentation, this does not mean that all regions are equally informative. The other option is to consider that a combination of regions as a whole may be more informative than any individual block. However, this information may never be utilized because the separate blocks will not create a suitably large standard deviation reduction. To counter this, I experimented with combining multiple blocks in an effort to improve fit results over the data. The first method of scaling was performed over larger block sizes. Rather than use the 24x18 segmentation grid, this was exchanged for a combination of larger blocks (e.g. 12x9 or 8x6 grids). Next, I experimented with summing entire rows into a single feature. The primary motivation here was that a row would represent a large image region with an approximately constant distance. Furthermore, this may take advantage of the sky as a whole by detecting clouds. I expanded on this by also summing multiple rows into the

ID	Dim	Description
B64	432	Blocks 64x64 px
B128	108	Blocks 128x128 px
B192	48	Blocks 192 x 192 px
B256	24	Blocks 256x256 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, excluding B64

Table 6.1: Overview of the various features used for the fitting the linear and M5' models.

same feature. For example, the 24x18 grid would be condensed into a 24x1, 12x1, or 8x1 feature. Example figures displaying the segmented features are shown in the Appendix.

6.3 Dataset

The images being used for this paper come from the Arizona Department of Environmental Quality [23]. Images are 1536x1152 pixels. This is then broken up into a 24x18 grid of 64x64 pixel blocks, where a feature is calculated for each block, resulting in a 432-dimensional feature vector. There are 6,897 images total, split into sets for 5-fold cross-validation. Sample images demonstrating various segmentation sizes can be found in the results and Appendix A.

6.4 Results

The results are categorized by both the fitting model and the type of features used.

Below is a compilation of results for both M5' and linear models across all feature types. A baseline mean fit is also included for comparison, where the output was simply the mean of the training set.

Fit	Feature	MAE	RMSE	R^2
Lin	All	7.945	11.7625	0.8449
Lin	B64	8.2146	12.2607	0.8295
Lin	B192	8.2146	12.2607	0.8295
M5'	B64	8.8751	14.207	0.776
M5'	All	8.9019	14.0892	0.7797
Lin	B128	8.9411	12.9739	0.8128
M5'	B128	9.2221	14.6384	0.762
M5'	R64	9.7219	15.525	0.7331
M5'	B192	9.8174	15.4396	0.7354
M5'	R128	10.3104	15.8499	0.7218
Lin	R64	10.4627	15.0675	0.7482
M5'	R192	10.4962	16.5134	0.6979
Lin	R128	11.0587	15.6555	0.7285
Lin	R192	11.0871	15.9387	0.7187
Mean	n/a	22.029	30.1035	0

Table 6.2: Compilation of results from M5' and Linear fits. The feature column corresponds to the features listed in Table 6.1.

Figure 6.2 shows that for the linear model, as expected, the error increases significantly for samples with a very poor visibility. While figure 6.3 shows very similar trends for the M5' model, it should be noted that the errors are much larger as indicated by the y-axis scale.

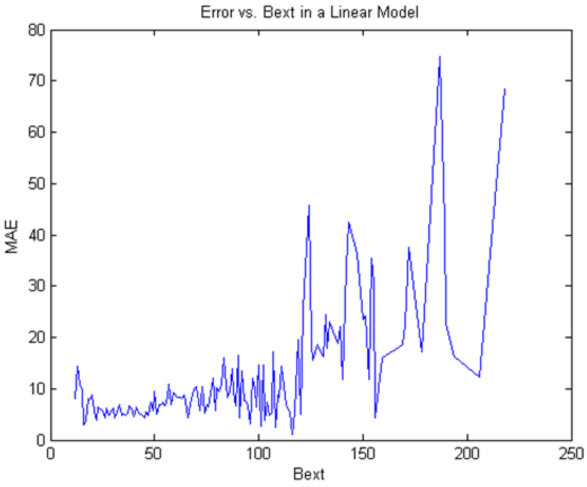


Figure 6.2: Results for the MAE in the Linear "All" Model.

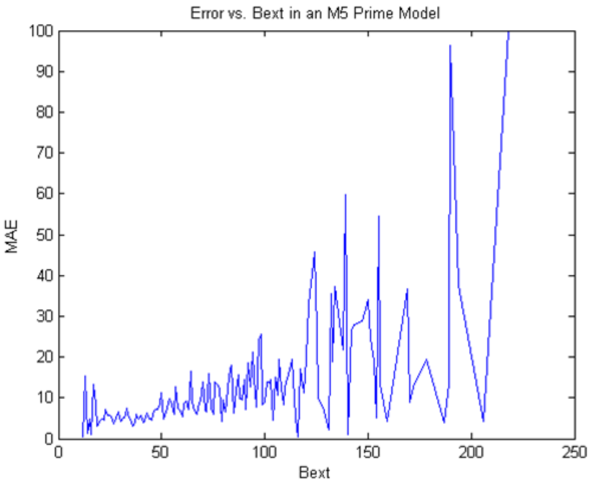


Figure 6.3: Results for the MAE in the M5' "B64" Model.

6.5 Discussion

Results show that the M5' is unable to outperform the linear system due to constraints in the tree growth process and issues with overfitting at leaf nodes. At some point during the algorithm, this creates undesirable splits that results in a larger error for the overall system. However, the use of multiscaled image features is shown to improve the ability of linear models to approximate the coefficient of extinction.

The linear models have an expected increase in error as the β_{ext} increases. While this is due in part to the lower number of samples in this range, it is also an expected complication from the exponential relationship that the transmission has with the extinction coefficient. This makes it inherently difficult to approximate large β_{ext} values due to small variations and noise in the dataset.

The leaf nodes of the M5' tree interpolate over a small subset of clustered values and may not provide accurate estimations, particularly in cases with poor visibility. This tells us that while the visibility can be modeled in a general linear system, there is significant variation in the manifestation of visibility in the dataset images. Because each leaf node may have as few as four examples, this also very heavily constrains the number of features that a piece-wise linear fit can use to model the data. The linear system, on the other hand, has the option to interpolate across all dimensions of the feature space and over a wider range of training points.

The addition of new features types may allow for a more stable model and provide for better generalization in the decision tree. This approach could benefit from salient region detection.

Chapter 7

Finding Salient Regions

This section is aimed towards detecting salient regions for visibility estimation within an image. This would allow us to eliminate pixels that add noise to our image with the goal of improving our model's fit. In addition, it could provide guidance for relative visibility estimation when ground truth training data is not available. This could ideally result in a mapping similar to Figure 7.1 without the need for ground truth data. An image registration pre-processing step was implemented, detailed in Chapter 8, because the following models rely on correlations being calculated at the pixel level.

7.1 Lambertian Surfaces

Lambertian surfaces can act as powerful indicators of salient regions for visibility applications. These surfaces reflect light in all directions such that the angle of the light source and the observer do not affect the observed luminance of the object. This approach could be particularly useful for the Phoenix dataset because it would allow us to select pixels that are robust to changes in the time of day. Previous work by Hautire et al.[27]

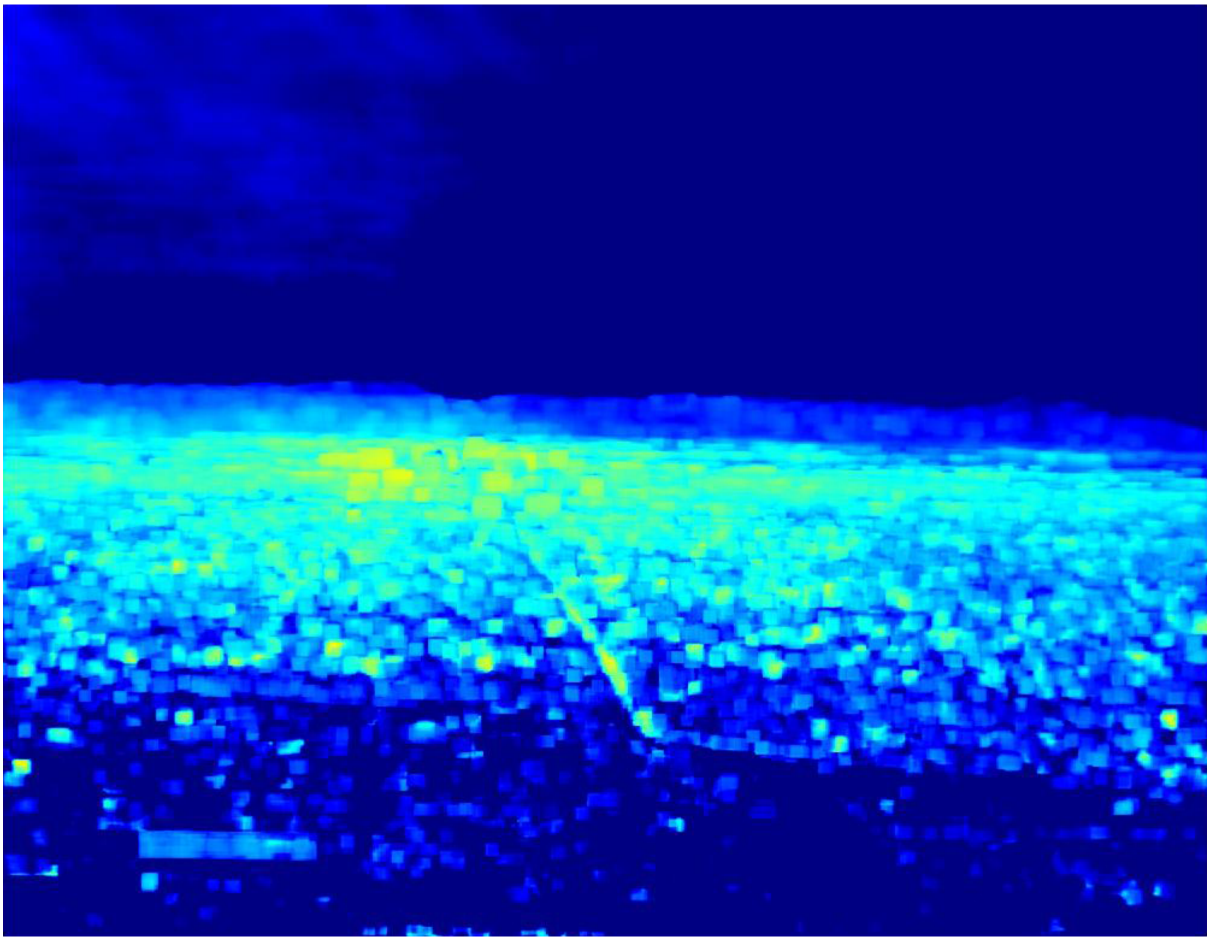


Figure 7.1: A sample weighting map generated by finding the correlation between the dark channel and the β_{ext} ground truth data.

shows that this was effective in a more controlled dataset where ground truth illumination was being measured with a luminance meter.

Hautire et al. used an identical image model to our approach, but scaled the contrast at each pixel by the probability $p(x)$ that a surface was Lambertian as shown in eq. (7.1) below. This probability was calculated by finding the correlation between pixel brightness and the scene luminance. A high correlation implies that the lighting at that point of the scene was invariant to changes in the sun's position, and is therefore likely to have Lambertian properties.

$$C_l(x) = |\Delta I(x)| \times p(x) \tag{7.1}$$

Because illumination was not being measured for the Phoenix ground truth data, the luminance needed to be estimated from the image. The dark channel prior method used in Chapter 4 was used to estimate atmospheric lighting from the image. The probability that a surface was Lambertian was then used as a weight on the Sobel gradient magnitude at each pixel in the image.

7.2 Dark Channel

The dark channel prior, proposed by He et al.[1], produces superb results for dehazing applications. It is clearly an effective tool for detecting transmission in hazy images. This section demonstrates how this feature correlates with the ground truth readings from a transmissometer.

The dark channel detects the minimum value in a specified region around each pixel. This value is an approximation of the atmospheric scattering interference outlined in eq. (3.2). By utilizing ground truth, we seek to find regions of the image that may be useful for estimating the coefficient of extinction. To do this, we find the correlation between each pixel in the dark channel and the ground truth β_{ext} data. This indicates which areas of the dark channel feature have the strongest correlation with changes in visibility.

Due to the nature of the dark channel, the foreground should be very dark. Other areas will be brighter at increasing distances due to airlight scattering. We expect these brighter areas to have a correlation with the atmospheric lighting, which should cause our algorithm to focus on hazy regions to obtain more useful features.

7.3 Dataset

We evaluate our method using a small set of images and ground truth extinction readings from the the Arizona Department of Environmental Quality which manages the PhoenixVis.net visibility web cameras website [23].

All images are in the RGB colorspace and have been JPEG compressed at an unknown quality level. The SOMT images are cropped to 1500×1161 pixels through a registration process outlined in Chapter 8. Each image is partitioned into 128×128 pixel overlapping blocks and a prediction model is trained for each block.

These images were subdivided into a training set, a validation set, and two test sets as shown in Table 5.1. Training set values were chosen across the span of three randomly

Table 7.1: Summary of results for each feature type. MAE (Mm^{-1}) values are reported from the trained model.

Feature	MAE
Gradient	11.2647
DC-weighted Gradient	**10.0913**
Lambertian-weighted Gradient	12.8777

chosen days. The validation and test sets were randomly selected from remaining labeled data.

7.4 Experimental Results

7.4.1 Lambertian Results

The results for the Lambertian weighting, listed in Table 7.1, showed reduced accuracy for the linear model. This was most likely due to the automated exposure settings for the camera. As the sky increases in brightness, the foreground automatically becomes darker. This results in a negative correlation for pixels below the horizon and renders our linear model ineffective. It would be very difficult to normalize the image without ground truth data regarding the illumination.

One potential benefit of this approach is that it provides a way to detect where the sky is for segmentation purposes. Because the luminance is being estimated from a region in the sky, the regions above the horizon will always have a strong correlation. This effect can be seen in Figure 7.2. This could be used to isolate the foreground in some scenes and focus on more relevant regions.

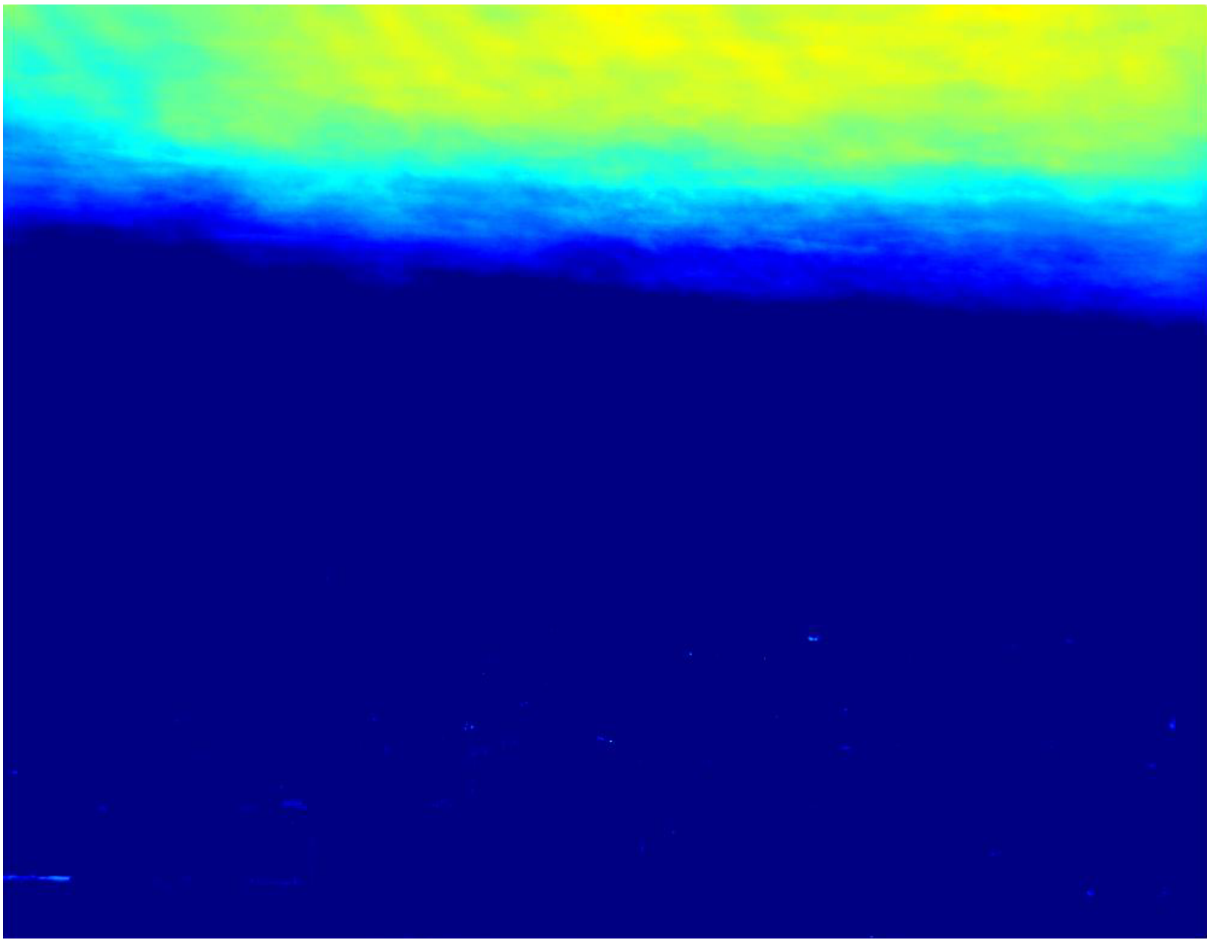


Figure 7.2: A weighting map generated by finding the correlation between the grayscale intensity and the estimated luminance.

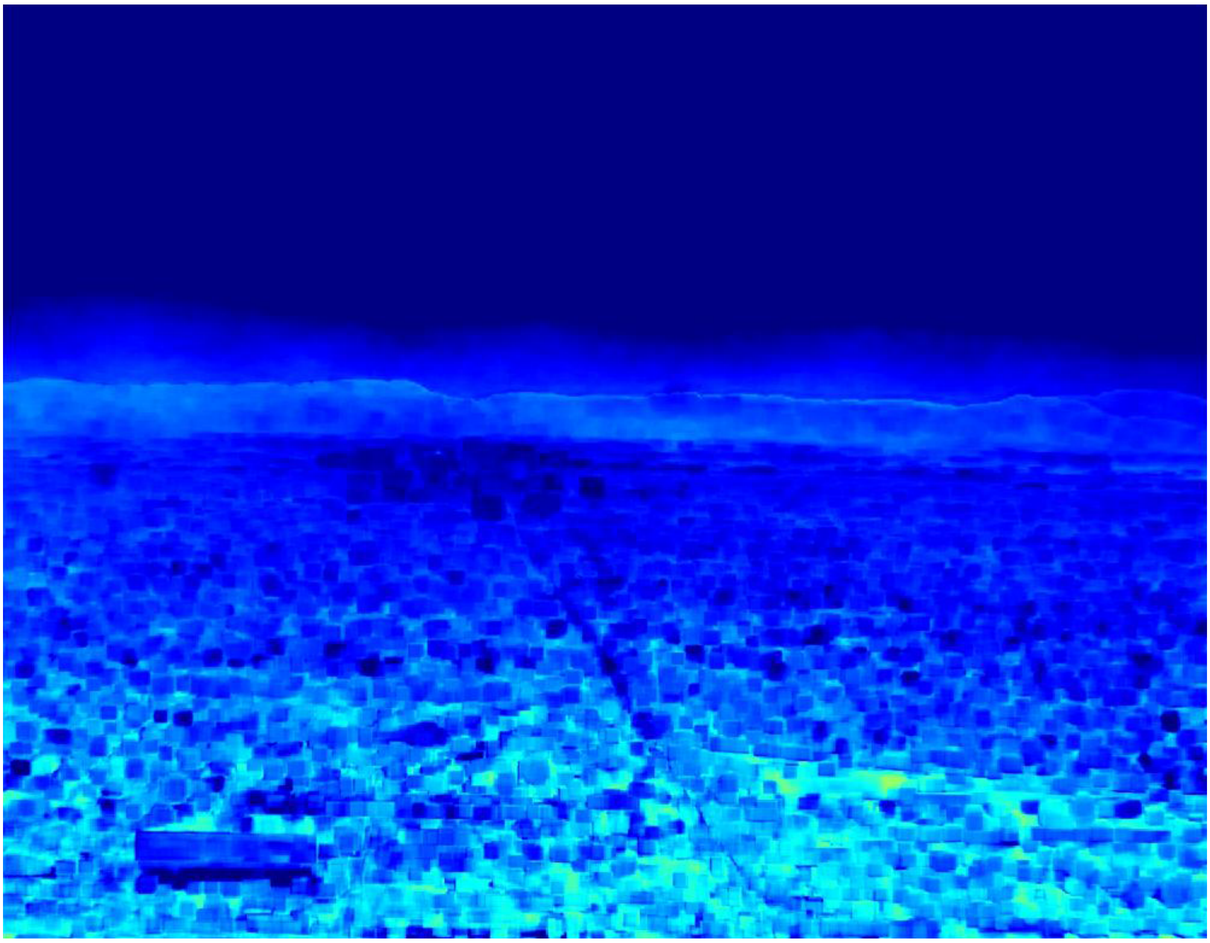


Figure 7.3: A weighting map generated by finding the negative correlation between the dark channel and the estimated luminance.

7.4.2 Dark Channel Results

The results for the dark channel weighting showed a slight improvement in the MAE over the standard gradient. The positive correlation of the dark channel yields a very similar result to the Lambertian probability, in that it emphasizes regions that are very hazy and close to the sky brightness. Instead we focused on the negative correlation between the dark channel and the estimated scene luminance. This puts a stronger weighting on the foreground and a small region along the horizon as shown in Figure 7.3.

7.4.3 Discussion

The theory behind both of our saliency models relies on a correlation between the atmospheric lighting and pixels within the image. However, that was clearly not the case from our results. We believe this was due to autoexposure settings on the Phoenixvis.net camera systems. Because they are configured to operate 24/7, they are required to automatically adjust to the brightness of the scene. While this is beneficial for qualitative analysis of images, it hampers our ability to estimate the atmospheric lighting. As the sky gets brighter, the foreground will automatically be darkened due to the decreased exposure time.

While our results were not successful for this dataset, both of these approaches are worth investigation in a more controlled system with additional ground truth readings. Detecting Lambertian pixels has been successful in related work with the addition of luminancemeters for ground truth atmospheric lighting. If this provides reliable results, it may lead to more reliable ways of tracking long-term trends in visibility for unlabeled scene images.

Chapter 8

Image Registration

The linear fits proposed in sections 4, 5, and 6 are robust to small camera movements due to the block-based approach. However, the saliency models proprosed in section 7 are calculated at each pixel value and are heavily impacted by any offsets. There is a sizeable amount of movement in the Phoenix dataset caused by wind moving the cameras. To counter this, an image registration algorithm was implemented to align images of the same scene.

Federov et al.'s automatic registration work [32] provides the basis for our implementation. After a root image has been chosen, the algorithm is applied to a batch of images in a 3-step process. First, tie points are generated automatically for each image to perform matching. Next, these points are used to solve for an affine or RST (rotation, scaling, and translation) transformation to align them with the root image. Lastly, all images are cropped to a common region so that no borders are present after the corrective transformations are applied. An example of this is shown in figure 8.1.

This method ended up requiring a lot of user intervention in order to make it feasible for our work. Due to the way that the cropping process was implemented, a single image

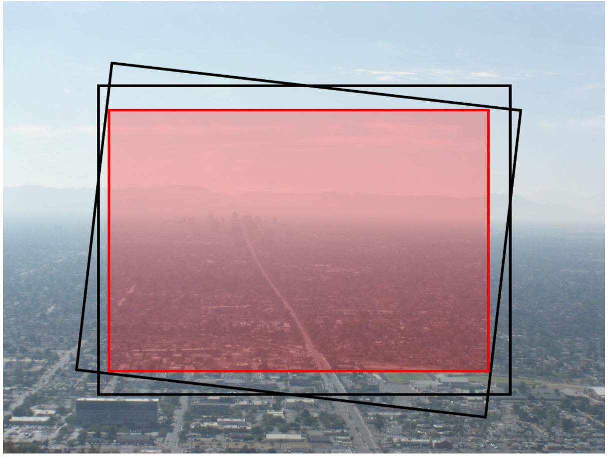


Figure 8.1: A mock-up example of a good registration batch. The final cropped (red) version will be representative of both pictures despite being offset.

could lead the algorithm to “overcrop” an entire batch if the wrong transformation was chosen. Furthermore, large amounts of haze can make it very difficult for the images to be accurately registered. This could be resolved by adding additional constraints on the rotation, scaling, and translation for individual images. Based on the assumption that our images should have similar transformations, we can disregard images that are outside of a specified threshold.

There are two additional complications with directly applying this algorithm. By default, the root image is a user-specified parameter. An algorithm that detected the image that best fit the rest of the images in the batch would produce the most reliable transformations and the largest common region. This ties in with the next problem, which is that our application would run the registration algorithm in multiple sessions. As new data is added to the training or test set, the common region of the algorithm will change. This could result in a different set of features being selected, and would then

Figure 8.2: A mock-up example of a bad registration batch. The final cropped (red) version will be very small due to the offset on the outlier image.

require that the linear fit be recomputed over the new regions. While image registration is an appealing preprocessing step, it was very difficult to produce a desirable output for the Phoenixvis cameras. Consequently, this work had limited contributions to our analysis framework.

Chapter 9

Conclusion

9.1 Visibility Estimation

As demonstrated in this paper, there are many approaches to image-based visibility estimation. While the multivariate fit worked best for the SOMT camera, our results could be highly dependent on the scene chosen. A more detailed analysis of different scenes could lead to very different results. An approach with regression trees could be expected to work well in a problem with more depth discontinuities, whereas the temporally regularized model is more suited towards datasets with large amounts of unlabeled data.

The proposed methods all rely on a very simple block-based segmentation for each region. This is based on the necessary assumption that all pixels within a block have the same depth. While the salient regions were not very effective with our dataset, continued research could make a large improvement in our visibility estimation algorithms. Additional work on properties such as Lambertian surfaces could help to find high-contrast objects at a constant depth for more accurate visibility estimation.

9.2 Future Work

The main focus for future work is on visibility detection in scenes for which there is no ground truth data for learning the model. While we would not expect to output β_{ext} values, a relative comparison of images could provide valuable information about long-term visibility and pollution trends in a given area. In Section 4.4.1 we experimented with binning visibility on a 1-5 deciview scale with good results. If we can obtain similarly accurate results for uncalibrated scenes, this system could be used to “piggyback” on top of existing webcam data and provide a detailed analysis of locations across the globe. Where ground truth data is available, our system could complement existing systems for measuring visibility. Given the accuracy of our linear model, our vision-based model could act as a backup system in the case that the visibility measurement equipment is moved to another location or is temporarily unavailable.

The features used in these experiments were fairly simple and easy to compute. Additional work could be done to investigate the effects of haze on color within the image. Likewise, our ability to test frequency analysis was hindered by JPEG compression in the images. Experimenting with high-frequency details could provide better results for other datasets. Furthermore, no knowledge of the scene geometry was considered in any of our algorithms. Additional information about the scene could be derived from satellite imagery and used to further constrain our model fitting process.

Bibliography

[1] He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2009) 1956–1963

[2] Narasimhan, S., Nayar, S.: Interactive (de)weathering of an image using physical models. In: ICCV Workshop on Color and Photometric Methods in Computer Vision (CPMCV). (2003)

[3] Narasimhan, S., Nayar, S.: Shedding light on the weather. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Volume I. (2003) 665–672

[4] Narasimhan, S., Nayar, S.: Contrast restoration of weather degraded images. IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003) 713–724

[5] Narasimhan, S., Nayar, S.: Vision and the atmosphere. International Journal on Computer Vision **48** (2002) 233–254

[6] Narasimhan, S., Nayar, S.: Removing weather effects from monochrome images. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Volume II.

(2001) 186–193

[7] Schechner, Y., Narasimhan, S., Nayar, S.: Instant dehazing of images using polarization. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Volume I. (2001) 325–332

[8] Schechner, Y., Narasimhan, S., Nayar, S.: Polarization-based vision through haze. Applied Optics, Special issue **42** (2003) 511–525

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

[10] Shwartz, S., Namer, E., Schechner, Y.Y.: Blind haze separation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Volume II. (2006) 1984–1991

[11] Namer, E., Shwartz, S., Schechner, Y.Y.: Skyless polarimetric calibration and visibility enhancement. Optics Express **17** (2009) 472–493

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

[13] Kim, K., Kim, Y.: Perceived visibility measurement using the HSI color different method. Journal of the Korean Physical Society **46** (2005) 1243–1250

[14] Baumer, D., Versick, S., Vogel, B.: Determination of the visibility using a digital panorama camera. Atmospheric Environment **42** (2008) 2593–2602

[15] Luo, C., Wen, C., Yuan, C., Liaw, J., Lo, C., Chiu, S.: Investigation of urban atmospheric visibility by high-frequency extraction: Model development and field test. Atmospheric Environment **39** (2005) 2545–2552

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

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

[18] Seinfeld, J., Pandis, S.: Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. Wiley (2006)

[19] Ruby, M.G., Waggoner, A.P.: Intercomparison of integrating nephelometer measurements. Environmental Science & Technology **15** (1981) 109–113

[20] Tan, R.: Visibility in bad weather from a single image. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2008)

[21] Fattal, R.: Single image dehazing. In: ACM SIGGRAPH. (2008)

[22] Narasimhan, S., Nayar, S.: Chromatic framework for vision in bad weather. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2000)

[23] : Phoenix visibility web camera website managed by the Arizona Department of Environmental Quality (ADEQ). (`http://www.phoenixvis.net/`)

[24] Lee, P., Hoffer, T., Schorran, D., Ellis, E., Moyer, J.: Laser transmissometer–a description. Science of The Total Environment **23** (1982) 321–335

[25] Betts, J.: The instrumental assessment of visual range. Proceedings of the IEEE **59** (1971) 1370–1371

[26] Pitchford, M.L., Malm, W.C.: Development and applications of a standard visual index. Atmospheric Environment **28** (1994) 1049–1054

[27] Hautire, N., Babari, R., Dumont, ., Brmond, R., Paparoditis, N.: Estimating meteorological visibility using cameras: A probabilistic model-driven approach. (In: Proceedings of the 10th Asian Conference on Computer Vision)

[28] Krishnakumar, V., Venkatakrishnan, P.: Determination of the atmospheric point spread function by a parameter search. Astronomy & Astrophysics Supplement Series **126** (1997) 177–181

[29] : Recommendation for a Phoenix Area Visibility Index by the Visibility Index Oversight Committee (2003)

[30] Xie, L., Carreira-Perpinan, M.A., Newsam, S.: Semi-supervised regression with temporal image sequences. In: Proceedings of the IEEE International Conference on Image Processing. (2010)

[31] : (M5primelab: M5ŕegression tree and model tree toolbox for matlab/octave) http://www.cs.rtu.lv/jekabsons/.

[32] Fedorov, D., Fonseca, L., Kenney, C., Manjunath, B.: Automatic registration and mosaicking system for remotely sensed imagery. In: SPIE 9th International Symposium on Remote Sensing. (2002)

Chapter 10

Appendix

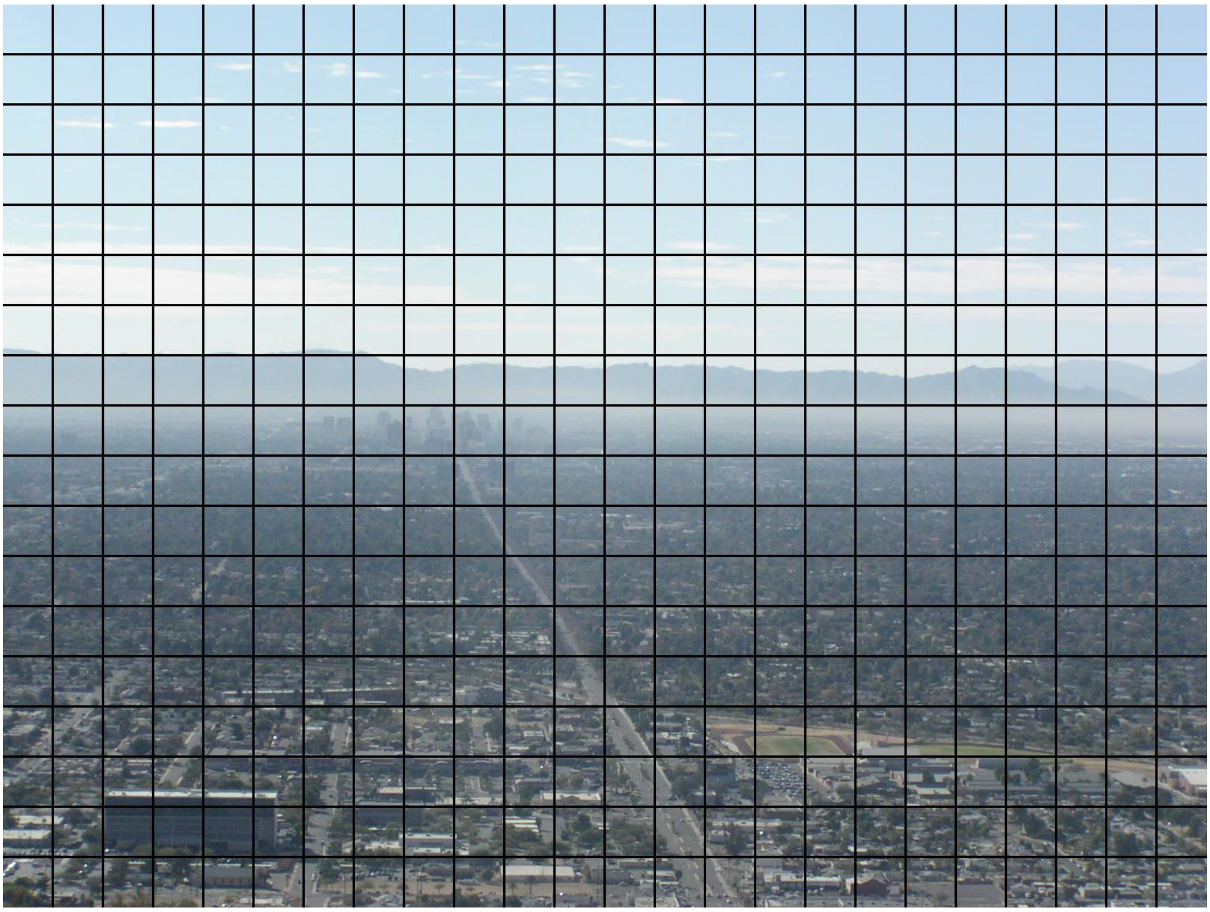


Figure 10.1: Example SOMT image with 64-pixel blocks.

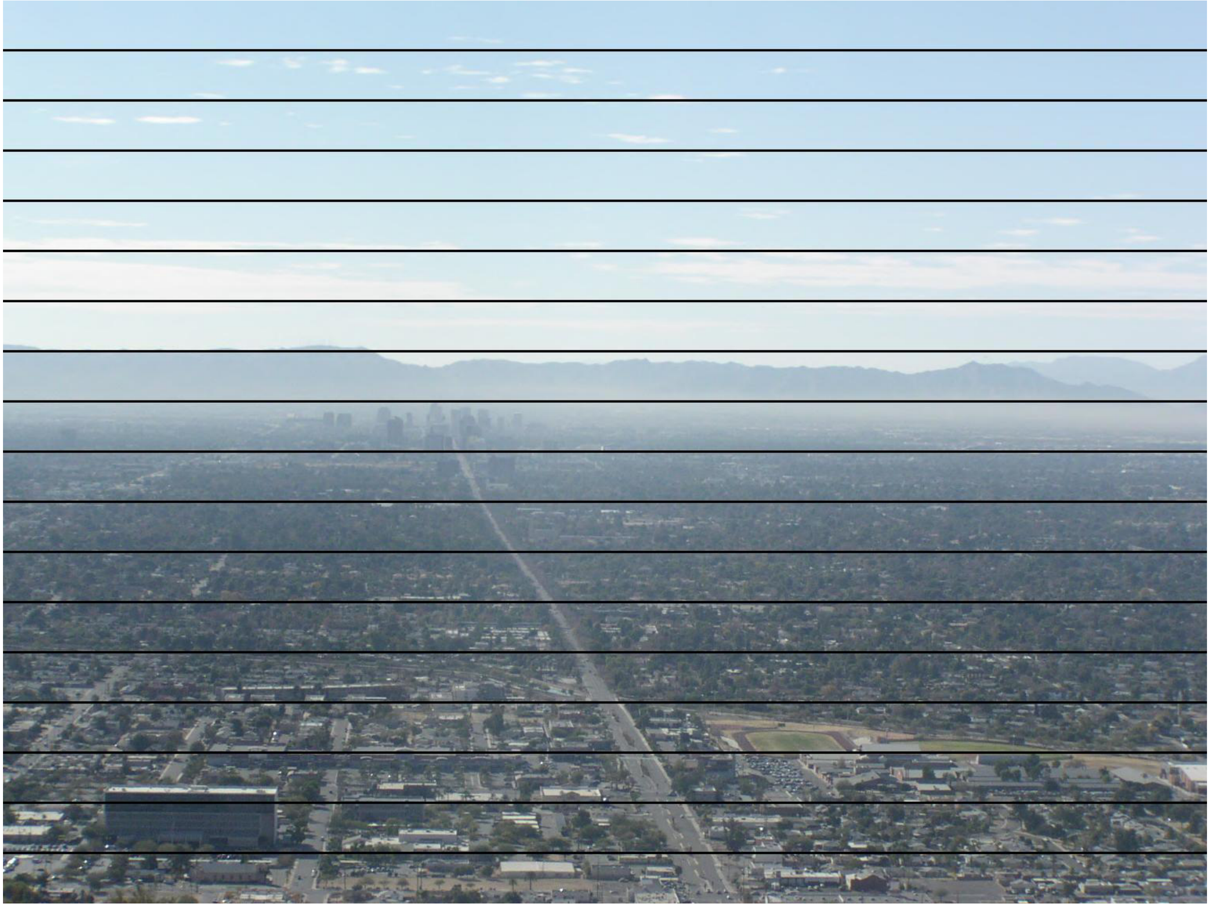


Figure 10.2: Example SOMT image with 64-pixel rows.

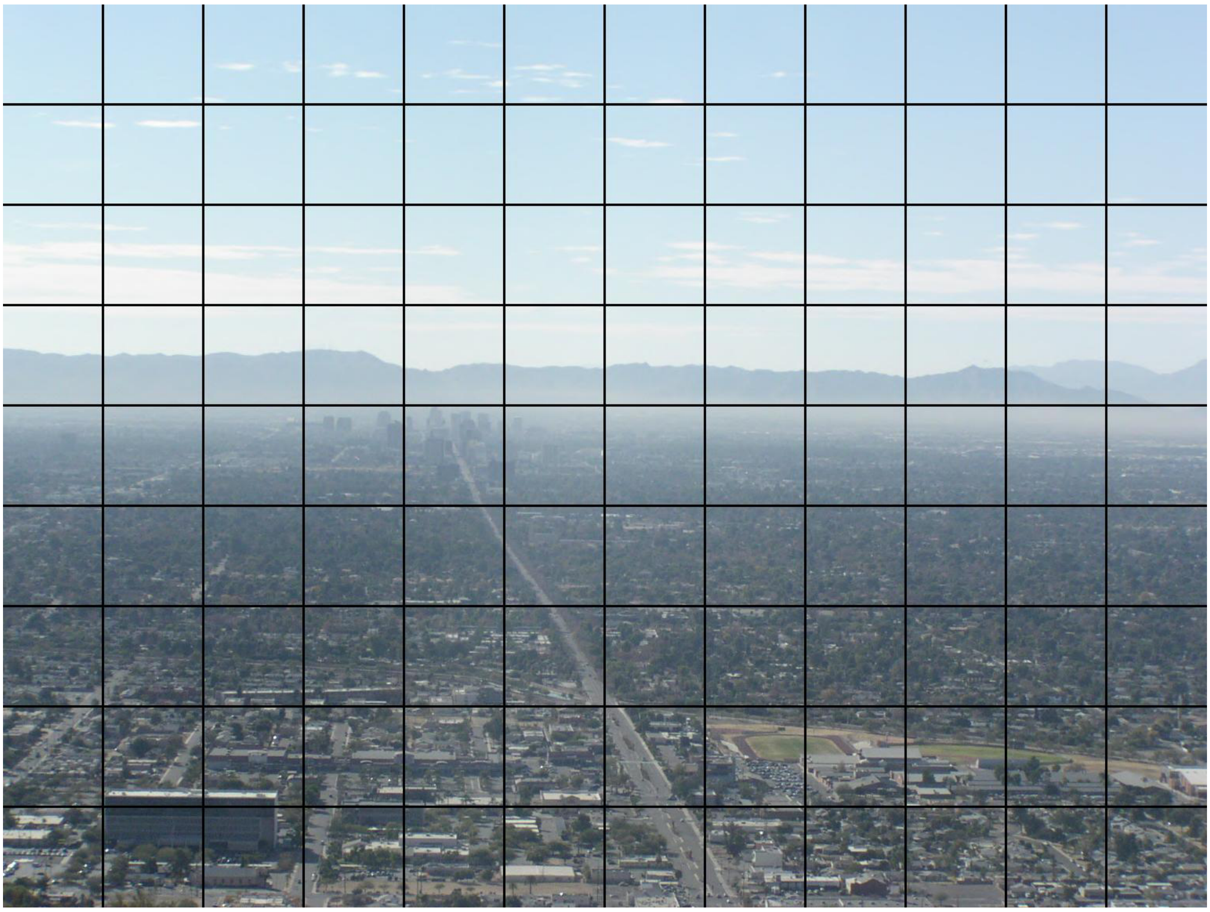


Figure 10.3: Example SOMT image with 128-pixel blocks.

Figure 10.4: Example SOMT image with 128-pixel rows.

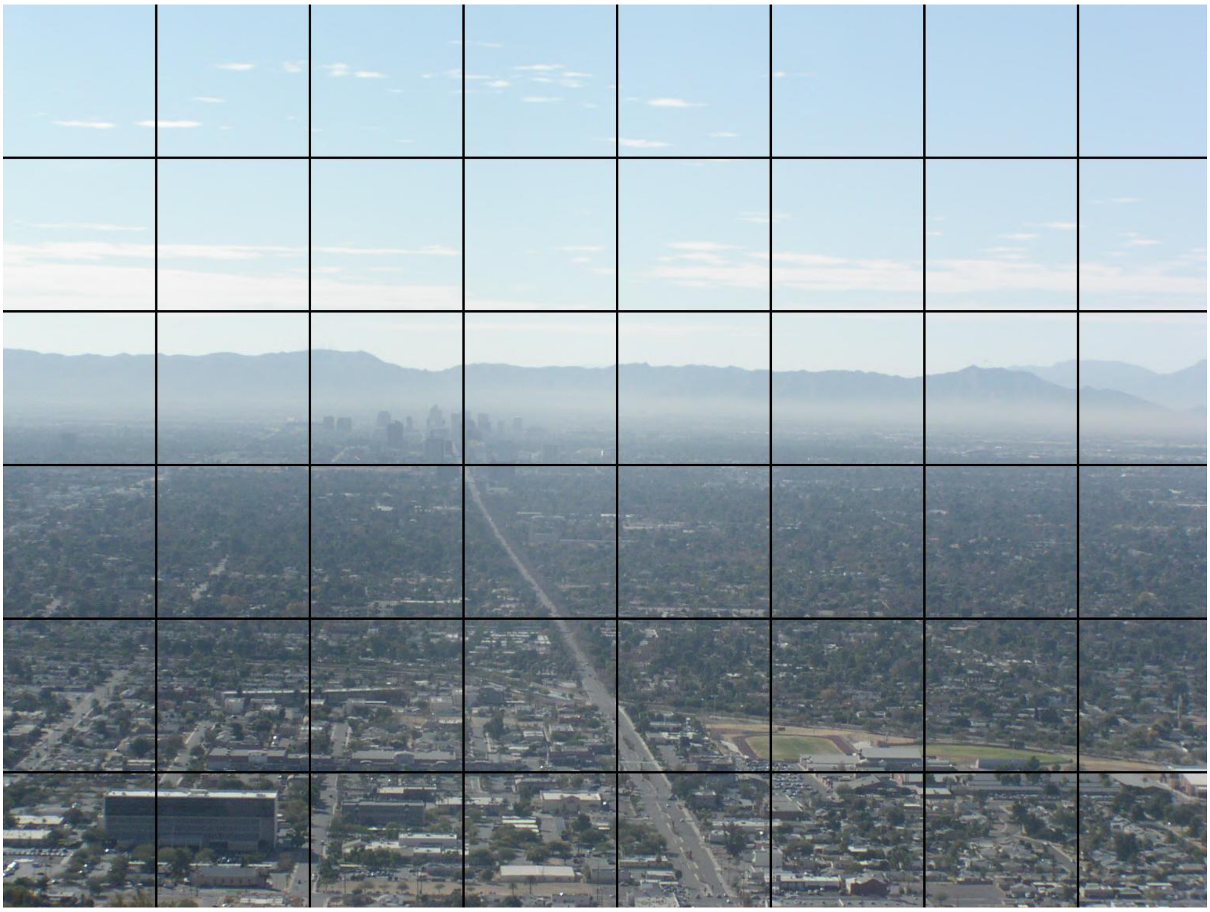


Figure 10.5: Example SOMT image with 196-pixel blocks.

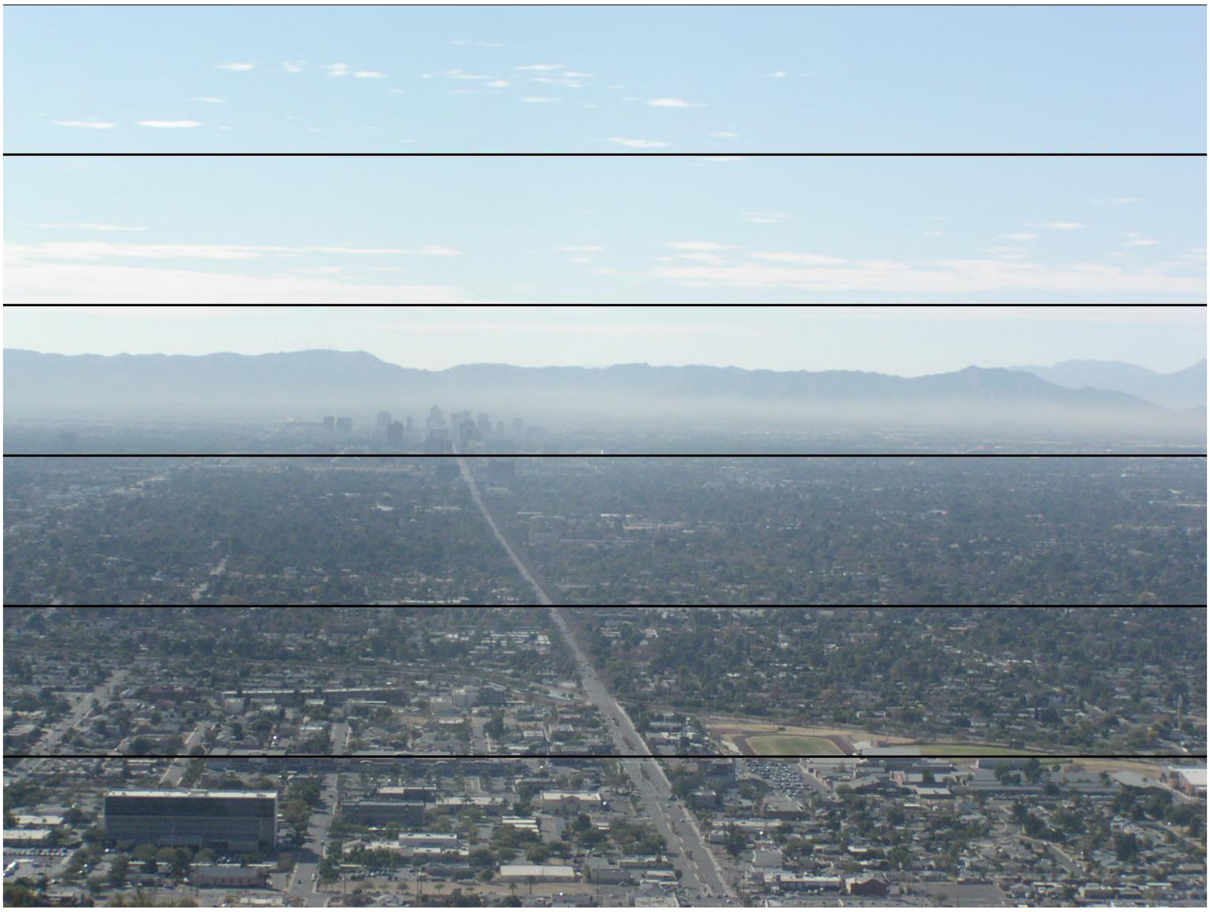


Figure 10.6: Example SOMT image with 196-pixel blocks.

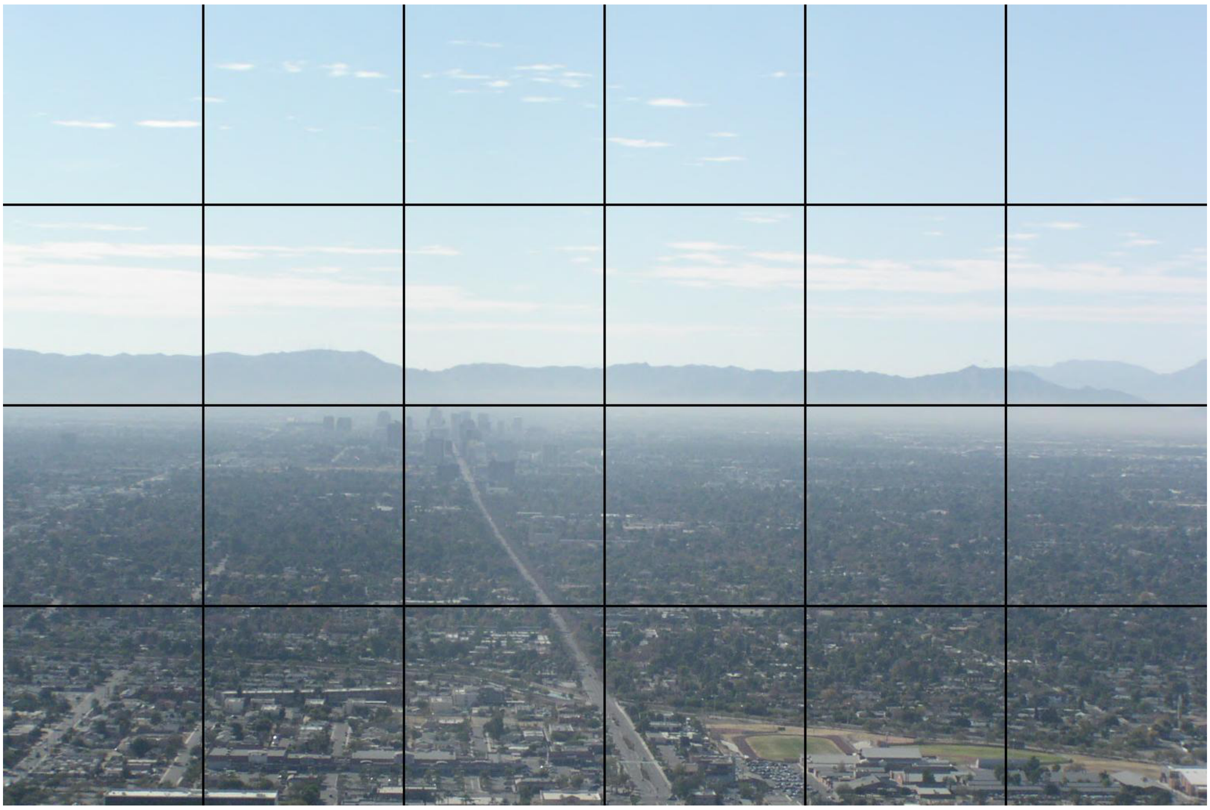


Figure 10.7: Example SOMT image with 256-pixel blocks.