

Images+Weather: Collection, Validation, and Refinement

Mohammad T. Islam, Nathan Jacobs
University of Kentucky
{mtis222, jacobs}@cs.uky.edu

Hui Wu, Richard Souvenir
University of North Carolina at Charlotte
{hwu13, souvenir}@uncc.edu

Abstract

The time and location that an image is captured indirectly defines the appearance of the scene. However, for outdoor scenes, the more directly relevant variables that affect image appearance are the scene structure, local weather conditions, and the position of the sun. In this paper, we present a large dataset of archived time-lapse imagery with associated geo-location and weather annotations collected from multiple sources. Through validation via crowdsourcing, we estimate the reliability of this automatically gathered data. We use this dataset to investigate the value of direct geo-temporal context for the problem of predicting the scene appearance and show that explicitly incorporating the sun position and weather variables significantly reduces reconstruction error.

1. Introduction

Time and location are two critical pieces of information to consider when developing outdoor image interpretation algorithms. For example, given that an image was captured in Minnesota, an algorithm for labeling the ground plane should detect white pixels in the winter, but not in the summer. Or in another setting, a recent rain shower might motivate the use of different features to localize sidewalks. Despite this, geo-temporal context is usually ignored in an effort to create algorithms that are invariant to these factors. We believe that geo-temporal context should be explicitly considered and that the lack of available training data to support learning methods has limited progress in this direction. This work is part of an effort to bridge this gap.

Image appearance varies significantly due to the weather and the position of the sun. These variations cause complicated, but often predictable, appearance changes, which most methods in outdoor imagery analysis treat as nuisance variables. This paper advocates for the alternative view: that these should be modeled and, to some extent, controlled. To organize our approach, we consider a simplified image formation model, shown in Figure 2, that describes how image appearance is related to a variety of underlying

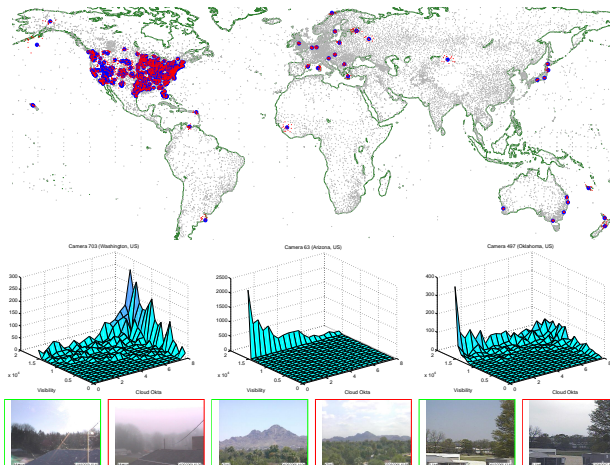


Figure 1. We created a dataset that merges outdoor time-lapse imagery with geo-temporal context (the sun position and weather data). The map shows the locations of the cameras (blue) and weather stations (red). For three example scenes from different climatic regions in the US, the histograms show the joint distribution of two weather variables (cloud okta and visibility) over a 3 year period. For each location, an image from a clear day (low okta, high visibility) is shown in green and a typical day in red.

factors.

As a key building block for this work, we extend a dataset [4] of geo-located, time-stamped images captured from outdoor webcams by adding geo-temporal metadata: the sun position and a variety of measures of the local weather. Since the sun position is a well-known function of the time and location, it is straightforward to estimate for any time-stamped geo-located image. Unfortunately, this is not so trivial for weather factors. We develop a system for merging the imagery with data from existing weather sources to construct our dataset. Figure 1 shows the locations of archived cameras and weather stations used to construct this dataset. In addition, we show how we can use a combination of crowdsourcing and machine learning techniques to improve the quality of data available for experimentation. Finally, we use show how we can use this dataset evaluate an augmented Lambertian image formation models that depends on the local weather conditions in ad-

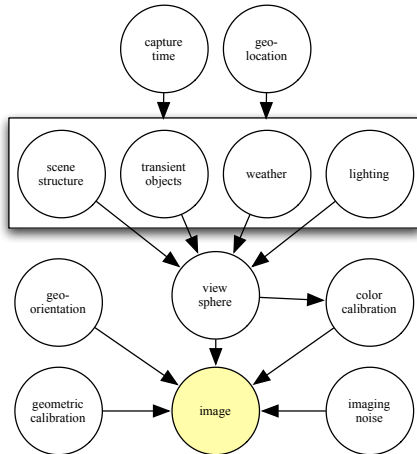


Figure 2. Images of the outdoors change for many reasons, this simplified geo-temporal image formation model describes the interactions and dependencies of the underlying factors.

dition to the sun position.

1.1. Related Work

One of the earliest and most prominent large-scale studies of scene variability from a single viewpoint is the Weather and Illumination Dataset (WILD), which captured images and weather every hour over 9 months [12]. This inspired a collection of work, including approximate physical models for removing the effects of weather and recovering clear images [11]. In this domain, data-driven approaches that make use of larger amounts of data began with the AMOS archive of webcam images [4], and continued with a set of calibrated, high resolution cameras [9]. Work on analysis of large sets of images from a single viewpoint has explored using the approach of clustering pixels that have related intensity changes [7] and of factoring the scene appearance as a function of the time of day [4]. With radiometrically calibrated images, factorization models can explicitly account for the effects of illumination direction, surface normal, and surface albedo on clear, sunny days [15, 16]. Recent work has also explored visual cues derived from passing clouds that enable scene geometry estimation [5] and methods for robust anomaly detection [1].

Previous work focused on estimating weather parameters includes efforts to use properties of the sky to calibrate the camera [10], and the use of single images to estimate the illumination map [8]. While the potential of integrating weather data with an image archive was discussed previously [3], to our knowledge there is no such current integrated archive. Relative to previous work, we present a larger and more diverse dataset that integrates weather data and imagery and make use of the weather data to evaluate methods of representing scene appearance as a function of the image capture time, sun position, and various weather

Table 1. Comparison of datasets of outdoor scenes.

Dataset	Scenes	Geo-located	Weather
WILD	1	yes	yes
Webcam Clip Art	54	yes	no
AMOS	~10K	some	no
AMOS+C	1 288	yes	yes

parameters.

2. Collection of Images+Weather

In order to explore the relationships between outdoor images and the underlying factors that affect their appearance, we extend the Archive of Many Outdoor Scenes (AMOS) [4], a publicly available image dataset. AMOS consists of images from thousands of outdoor cameras, archived every 30 minutes since March 2006. All of the images are time-stamped and many of the cameras have been manually geo-located. For the subset of cameras whose geo-location is verified, we incorporate additional geo-temporal metadata and refer to the subset of AMOS with additional context as *AMOS+C*. As with AMOS, *AMOS+C* will be made publicly available to the research community. Table 1 compares *AMOS+C* with other publicly available datasets of outdoor scenes.

To build this database we collected weather data from a variety of sources: Weather Underground [18], Weather Central [17] and the National Climatic Data Center [13], an archive of weather readings from thousands of weather stations located around the globe. The locations of the cameras and existing weather stations are shown in Figure 1. For all cameras in the dataset, local weather readings are estimated (either by us, or by the data provider) from nearby weather stations and satellites. The associated metadata includes:

- air pressure, wind velocity and dew point
- temperature and snow depth
- cloud cover, sky ceiling and visibility

Figure 3 highlights the varying appearance of a scene in relation to some of the weather parameters collected.

By imbuing the dataset with contextual information beyond the time and location of image capture, we believe that we can develop stronger scene understanding algorithms and a more accurate model of image formation. While inspired by the WILD database [12], the *AMOS+C* dataset includes more cameras in order to better capture the variability in the distribution of visual environments and how changes in the weather affects them.

3. Weather Label Validation

To support an initial evaluation of the accuracy of the weather metadata, we solicited 25 people (mostly undergraduate computer science students) who labeled approximately 140 images each from 10 of our cameras. For each

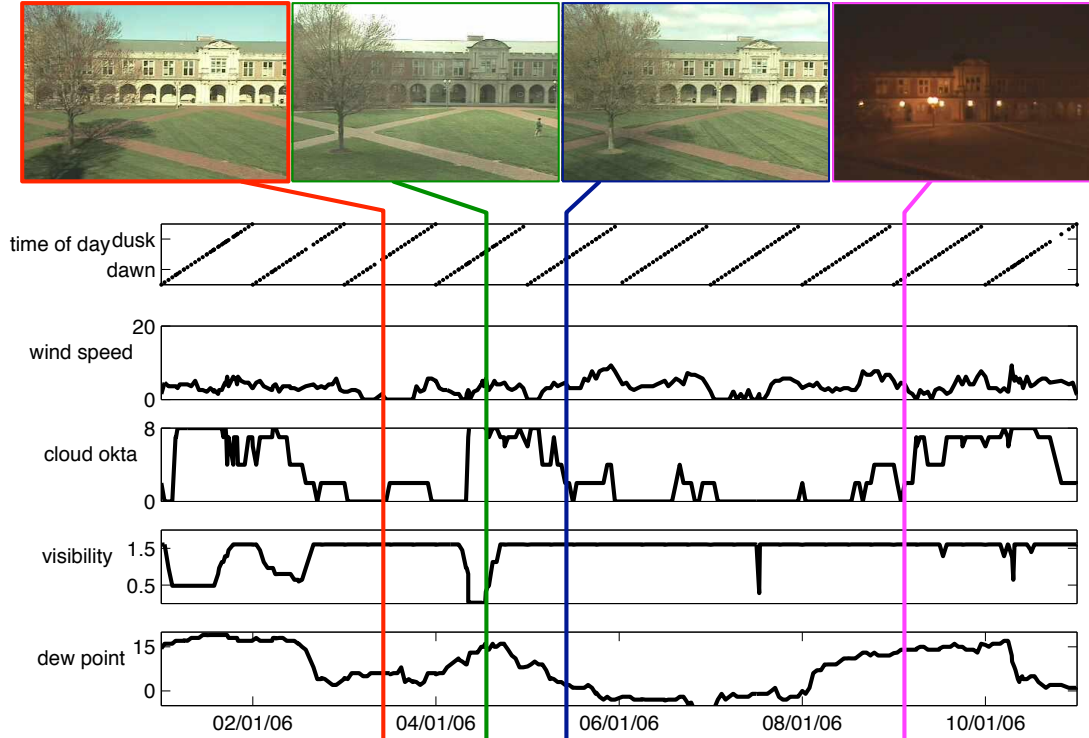


Figure 3. Image appearance depends directly on the underlying geo-temporal factors. The AMOS+C dataset enables exploring these dependencies to develop improved image analysis algorithms.

image, the user provided a label describing the weather depicted in the image.

We made several design decisions with the goal of increasing the quality of user feedback. Initial experiments showed that it was difficult for users to distinguish between certain labels. Therefore, we merged several labels, such as “Thunderstorm”, “Thunder”, and “Heavy Rain” into “Heavy Rain”. Initially, we asked users to select all (from a long list) of labels that applied to an image. Users would generally miss labels among the large set. To address this, we grouped labels into mutually exclusive groups (e.g., sky condition, precipitation) and only presented one group of labels at a time. With this approach, we could focus on specific label groups. We chose to focus on the sky condition variable (clear, cloudy, overcast, etc.) because it seemed easiest for a novice user to annotate. In total, we received between 217 and 382 annotations per camera.

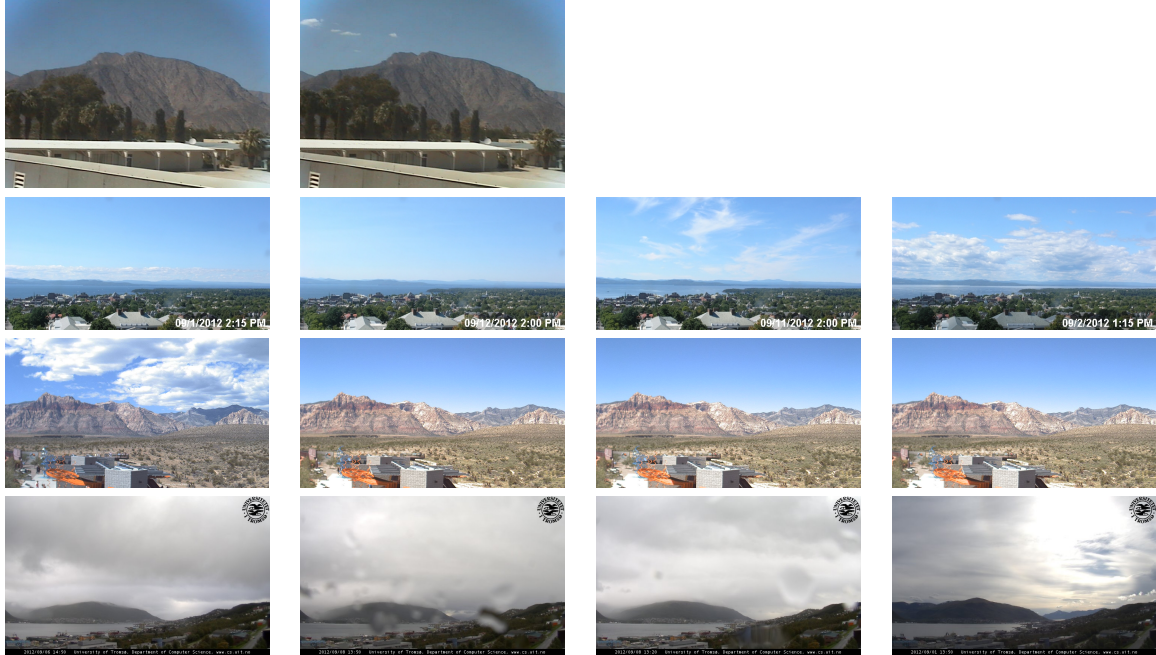
The annotations from all users were combined to produce a ranking of the degree to which the metadata from a camera agrees with the user annotations. We observed that some users were significantly more accurate in giving feedback than others, therefore we use the approach of [6] to assign a confidence to each user based on how much they agree with others. The approach has three stages: 1) assign an initial label to each image by averaging all user feedback, 2) determine the user confidence score using Kendall’s rank

correlation between the feedback submitted by the user and the average label for the same images, 3) assign a final label to each image weighted by user confidence. To determine the quality of each camera, we compute the mean number of images whose label matches the user annotation (see Figure 4 for sample images from the best and worst cameras). Figure 5 shows example images with a large discrepancy between the automatic and crowdsourced labels. Of the 10 cameras tested in the user study, the worst cameras had a significant portion with mismatched labels.

4. Weather Label Refinement

In the previous section, it was shown that qualitative weather variables such as cloud cover can be unreliable, as far as describing the weather conditions visible in the image. To determine whether or not the weather parameters are reliable for every webcam in AMOS+C would require a large amount of tedious, manual annotation. In this section, we present a simple refinement approach that takes advantage of the relationship between certain qualitative weather variables and image appearance.

Given a set of images from a single webcam, $\mathcal{I} = \{I_1, I_2, \dots, I_n\}$, we assume that the set of images, \mathcal{I} , lie on a manifold parameterized by $Y^* = \{y_1^*, y_2^*, \dots, y_n^*\}$, where y_i^* is an ordinal or real value that represents a weather parameter (e.g., cloudiness). So, given the associated (noisy)



Original Clear Partly Cloudy Mostly Cloudy Cloudy

Figure 4. Images from the two best (top) and two worst (bottom) cameras. Empty spaces denote that no images are available for that camera for the particular label.



Original Partly Cloudy Clear Mostly Cloudy Mostly Cloudy
Crowdsourced Cloudy Clear Clear

Figure 5. Example images with a large discrepancy in weather labels between the original data and the crowdsourced ground truth.

metadata $Y = \{y_1, y_2, \dots, y_n\}$, we treat weather context refinement as a regression problem where the goal is to predict Y^* .

We employ support vector regression (ν -SVR) [14]. This kernel-based method can represent complex underlying functions and the kernel function selected can be matched to particular image appearance changes. Our initial refinement task focuses on cloudiness, as changes in cloudiness lead most directly to obvious image appearance changes. To minimize the effects of slight camera motion and transient objects in the scene, the images were downsized 10x (as small as 50*25, for certain webcams) and linearized to vectors. We used the implementation of (ν -SVR) in `libsvm` [2] with the Gaussian radial basis function kernel, $\nu = 0.5$, $C = 1$, $\gamma = 0.001$.

Figure 6 shows refinement results for two scenes. Each pair of rows show images sorted by cloudiness based on the collected data (top) and after refinement (bottom). For these scenes, the automatically collected labels do not re-

fect the weather visible in the images. In both examples, the weather was reported as clear for the first two images, partly cloudy for the next two, and mostly cloudy or overcast for the last two. After refinement, the images are sorted by the new weather parameters. In both cases, this order appears to reflect the amount of cloudiness in the scene. For regression, discrete labels (clear, partly cloudy, mostly cloudy, overcast) are converted to ordinal values.

5. Applications Using Weather Context

To demonstrate the potential value of this database of weather metadata, we develop two conditional regression models for predicting individual pixel intensities. The first is an extension of the Lambertian model with a linear weather dimension; the second uses a non-parametric method. These models allow us to explore additional aspects of the relationship between appearance, sun position and weather conditions and highlight the complicated interactions between image appearance and the underlying geo-

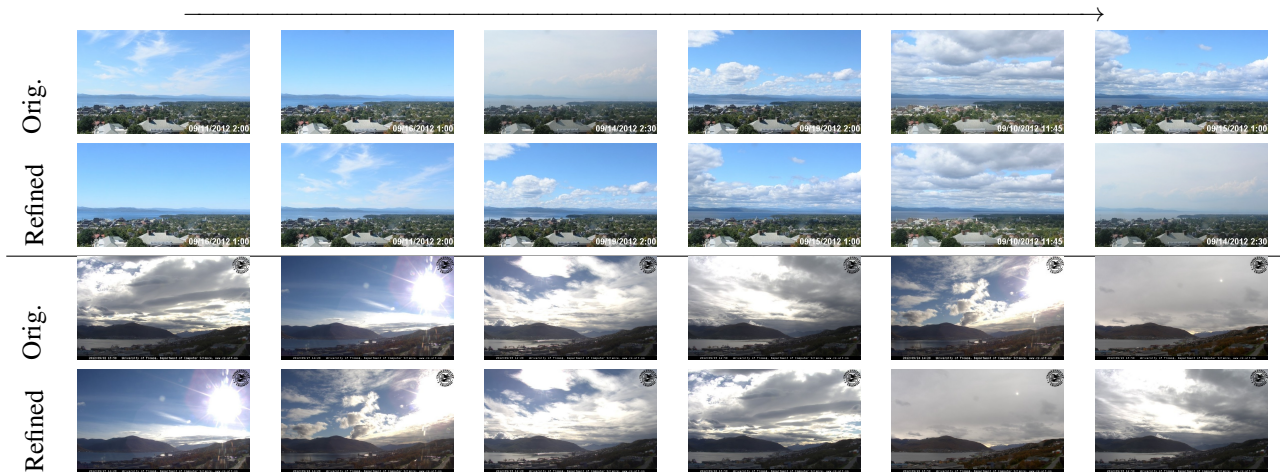


Figure 6. Each pair of rows show images sorted by cloudiness based on the collected data (top) and after refinement (bottom).

temporal factors.

5.1. Conditional Scene Prediction Methods

We propose an extension to the Lambertian model that incorporates a weather parameter. The Lambertian function defines a basis for representing our conditional linear regression model:

$$I_{p,t} = \alpha_1 w_t L(\vec{s}_t, \vec{n}_p) + \alpha_2 L(\vec{s}_t, \vec{n}_p) + \alpha_3 w_t + \alpha_4 \quad (1)$$

where $I_{p,t}$ is the color of pixel p at time t , α is a vector of weights, w_t is the weather state, \vec{s}_t is the sun direction, and \vec{n}_p is the surface normal. Note that we ignore the surface color because it is subsumed by the parameter values α .

For a given pixel, we solve for the surface normal and α values that reduce the L_1 reconstruction error (recall that w_t and \vec{s}_t are known). Notice that, for a given surface normal, we can directly compute the Lambertian contribution function $L(\vec{s}_t, \vec{n}_p)$. This leaves a straightforward linear regression problem to solve for α . We optimize by grid sampling the space of surface normals and computing the reconstruction error using the optimal α for each sample point. We then choose the grid sample point with the minimum reconstruction error and use local descent to find an optimal surface normal. For the tested scenes, we found the error surface to be well behaved and, while it is not our focus, the estimated surface normals to be close to the ground truth. It is expected that for different scenes, especially those with fewer image examples, it may be difficult to obtain accurate surface normal estimates. This, however, is not a major problem for the task of scene prediction.

For our second conditional-linear prediction model, we replace the Lambertian contribution function $L(\vec{s}_t, \vec{n}_p)$ with a non-parametric model. We build a conditional model by first sampling the solar zenith-azimuth space on a 30×120

regular grid and then building a separate linear regression model for each sun position. We train the regression models by weighting the training points based on distance in zenith/azimuth space, the weight is determined by a Gaussian function with $\sigma = 7^\circ$. To predict the intensity of a particular pixel at a particular time of day, we first compute the zenith/azimuth angle of the sun, then interpolate the linear regression model parameters from the sample grid, and use the current weather value to predict the pixel intensity. In this model, the non-parametric grid sampling of solar zenith/azimuth space replaces the basis functions that come from solving for the optimal surface normal in the conditional Lambertian model.

5.2. Analysis

We compare our conditional scene regression models on several scenes. For both models, we estimate parameters using all the images of the scene when the sun was up (zenith angle less than 90°). As before, our context parameter is cloud okta.

Figure 7 shows data from the Växjö, Sweden webcam; for three distinct pixels, we show actual pixel values, and the values predicted by both regression models, at different sun positions and cloudiness states. We see that for each pixel cloudiness has a similar effect; the left column corresponds to high cloudiness, which decreases appearance changes due to sun position. During the clear conditions shown in the rightmost column, we see that the intensity of the pixel changes significantly due to the sun position.

Figure 7 also highlights several aspects of the two conditional scene regression models. The conditional Lambertian model is able to capture the coarse scale brightness changes, but it is not sufficiently expressive to describe many subtle aspects of the appearance of a real scene. However, it is better suited than a purely Lambertian model, which roughly

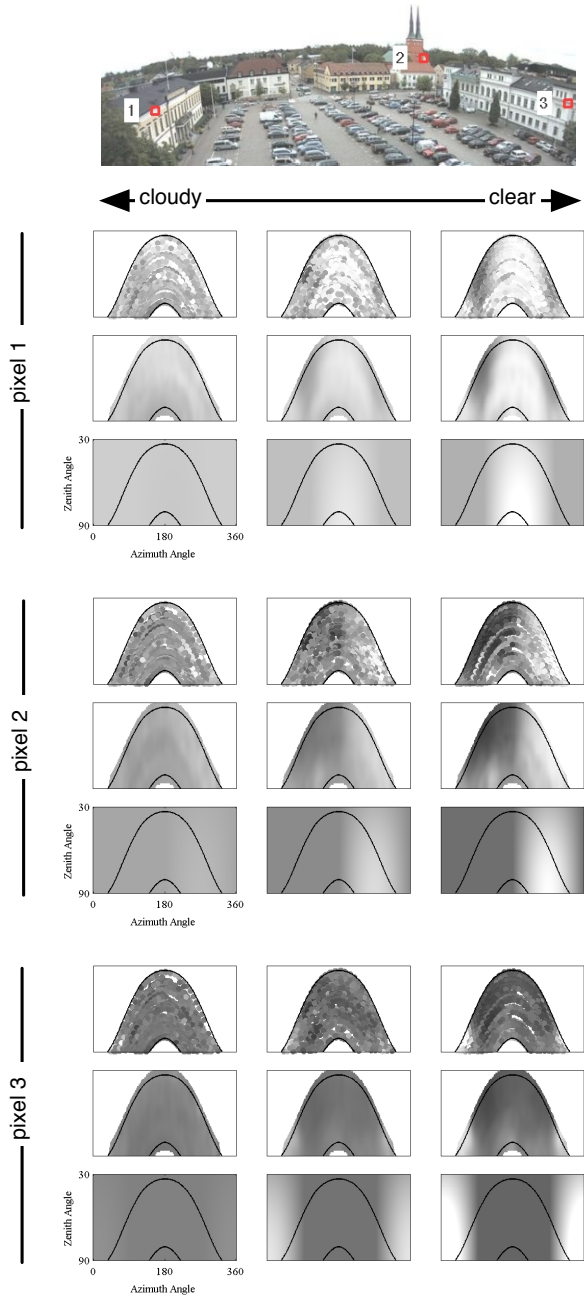


Figure 7. Actual and predicted intensities for three different pixels from a year of images. Within each group, the top row is the ground truth, the middle row is the non-parametric conditional linear model, and the bottom row is the Lambertian-based model. The Lambertian model can predict pixel values for unseen data, but smooths out significant details.

corresponds to the clear-sky column of the conditional Lambertian model. In general, we find that the non-parametric model is significantly more expressive and provides more accurate predictions in most cases, but requires more images for training and does not provide geo-oriented surface

normal estimates.

Figure 8 extends these results beyond a single pixel to predict the appearance of the full scene at an arbitrary sun position and cloudiness state. The model is able to capture subtle changes in scene appearance due to cloudiness and zenith/azimuth angle of the sun. In addition to the predicted scene appearance, the false-color images in the rightmost column depict which pixels have a strong linear dependence on the cloud okta. The figures show, as would be expected, that the cloudiness parameter makes significant contributions to scene appearance, and that the specific contributions depend on the sun position and position in the image. This further demonstrates that the model is truly capturing the complex interaction between scene appearance and the underlying weather condition.

6. Conclusion

The connection between local weather conditions and scene appearance is inextricable, but used sparingly, if at all, in surveillance and scene analysis. This work provides a catalyst for research in this direction by introducing the largest publicly available set of images of outdoor scenes with coupled weather data. Previously such work was impossible because of the lack of sufficiently diverse data to ground the research.

Using this dataset, we highlight the complex relationship between the underlying geo-temporal factors and the scene appearance and present methods for scene reconstruction, all of which explicitly incorporate geo-temporal factors. Collecting, validating, and refining this dataset and understanding the representational issues are important steps in developing methods that incorporate geo-temporal context in a range of applications from surveillance to scene parsing.

Acknowledgments We thank Chris Hawley, Austin Abrams and Robert Pless for assistance with the AMOS dataset. This work was supported by DARPA (D11AP00255).

References

- [1] M. Breitenstein, H. Grabner, and L. Van Gool. Hunting nessie - real-time abnormality detection from webcams. In *ICCV Workshop*, 2009. 2
- [2] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2011. 4
- [3] N. Jacobs, W. Burgin, R. Speyer, D. Ross, and R. Pless. Adventures in archiving and using three years of webcam images. In *CVPR Workshop*, 2009. 2

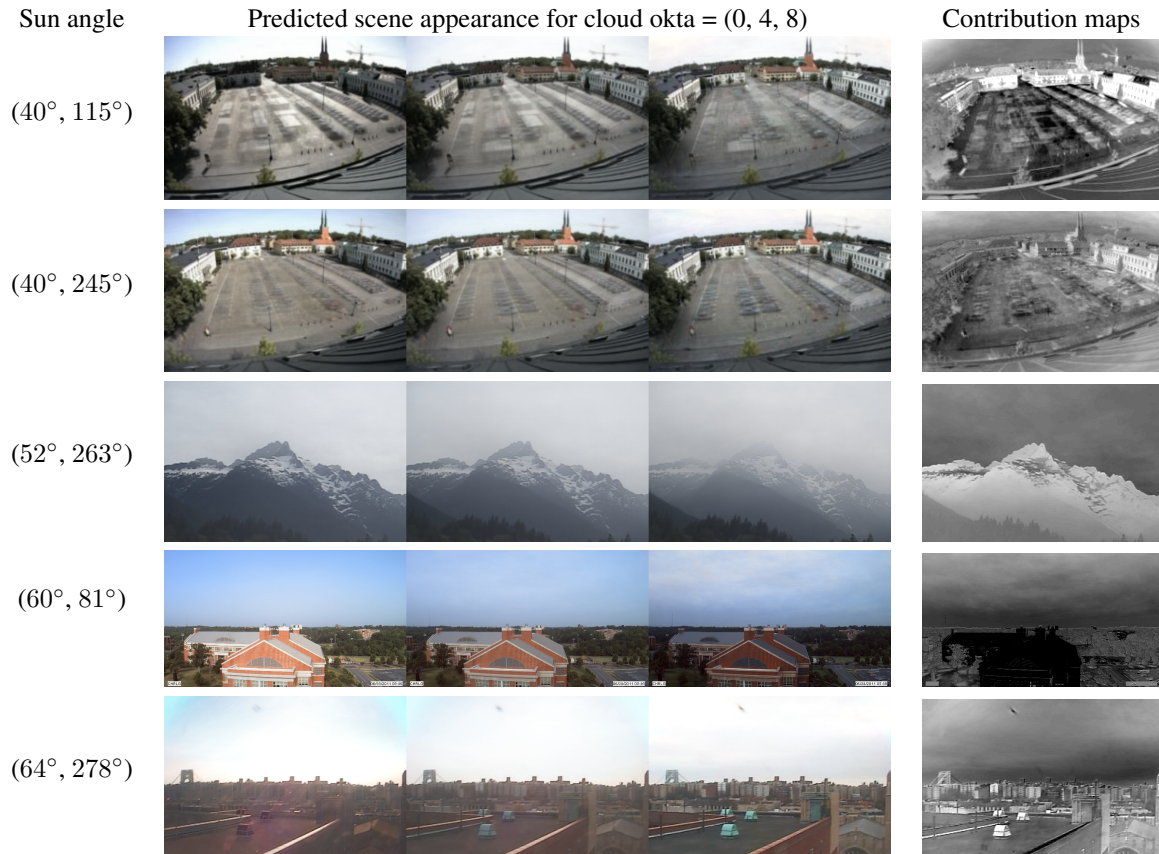


Figure 8. Predictions of scene appearance for various sun positions and cloud okta using our conditional linear model. Columns correspond to values of cloud okta from clear (left) to overcast (right). The false-color images in the right-most column highlight the contribution of cloud okta in predicting scene appearance: dark values reflect pixels that get darker as it becomes more overcast, grey values remain unchanged, and white values get brighter.

- [4] N. Jacobs, N. Roman, and R. Pless. Consistent temporal variations in many outdoor scenes. In *CVPR*, 2007. 1, 2
- [5] N. Jacobs, S. Satkin, N. Roman, R. Speyer, and R. Pless. Geolocating static cameras. In *ICCV*, 2007. 2
- [6] B. Jeong, J. Lee, and H. Cho. User credit-based collaborative filtering. *Expert Syst. Appl.*, Apr. 2009. 3
- [7] S. J. Koppal and S. G. Narasimhan. Clustering appearance for scene analysis. In *CVPR*, 2006. 2
- [8] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan. Estimating natural illumination from a single outdoor image. In *ICCV*, 2009. 2
- [9] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan. Webcam clip art: Appearance and illuminant transfer from time-lapse sequences. *ACM Trans. on Graphics (SIGGRAPH Asia)*, Dec. 2009. 2
- [10] J.-F. Lalonde, S. G. Narasimhan, and A. A. Efros. What do the sun and the sky tell us about the camera? *IJCV*, May 2010. 2
- [11] S. G. Narasimhan and S. K. Nayar. Shedding light on the weather. In *CVPR*, 2003. 2
- [12] S. G. Narasimhan, C. Wang, and S. K. Nayar. All the images of an outdoor scene. In *ECCV*, 2002. 2
- [13] NCDC. <http://www.ncdc.noaa.gov/>. 2
- [14] B. Scholkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett. New support vector algorithms. *Neural computation*, 2000. 4
- [15] K. Sunkavalli, W. Matusik, H. Pfister, and S. Rusinkiewicz. Factored time-lapse video. *ACM Trans. on Graphics (SIGGRAPH)*, 2007. 2
- [16] K. Sunkavalli, F. Romeiro, W. Matusik, T. Zickler, and H. Pfister. What do color changes reveal about an outdoor scene? In *CVPR*, 2008. 2
- [17] Weather Central, LP. <http://http://www.wxc.com/>. 2
- [18] Weather Underground. <http://www.wunderground.com/>. 2