

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221111628>

# Estimating Natural Illumination from a Single Outdoor Image

**Conference Paper** in *Proceedings / IEEE International Conference on Computer Vision. IEEE International Conference on Computer Vision* · September 2009

DOI: 10.1109/ICCV.2009.5459163 · Source: DBLP

CITATIONS

116

READS

252

3 authors, including:



**Jean-François Lalonde**

Laval University

113 PUBLICATIONS 2,607 CITATIONS

[SEE PROFILE](#)



**Alexei Efros**

University of California, Berkeley

164 PUBLICATIONS 55,388 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Real-Time User-Guided Image Colorization with Learned Deep Priors [View project](#)



Create new project "Split-Brain Autoencoders: Unsupervised Learning by Cross-Channel Prediction" [View project](#)

# Estimating Natural Illumination from a Single Outdoor Image

Jean-François Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan

School of Computer Science, Carnegie Mellon University

<http://graphics.cs.cmu.edu/projects/outdoorIllumination>

## Abstract

Given a single outdoor image, we present a method for estimating the likely illumination conditions of the scene. In particular, we compute the probability distribution over the sun position and visibility. The method relies on a combination of weak cues that can be extracted from different portions of the image: the sky, the vertical surfaces, and the ground. While no single cue can reliably estimate illumination by itself, each one can reinforce the others to yield a more robust estimate. This is combined with a data-driven prior computed over a dataset of 6 million Internet photos. We present quantitative results on a webcam dataset with annotated sun positions, as well as qualitative results on consumer-grade photographs downloaded from Internet. Based on the estimated illumination, we show how to realistically insert synthetic 3-D objects into the scene.

## 1. Introduction

The appearance of a scene is determined to a great extent by the prevailing illumination conditions. Is it sunny or overcast, morning or noon, clear or hazy? Claude Monet, a fastidious student of light, observed: “A landscape does not exist in its own right (...) but the surrounding atmosphere brings it to life... For me, it is only the surrounding atmosphere which gives subjects their true value.” Within the Grand Vision Problem, illumination is one of the key variables that must be untangled in order to get from pixels to image understanding.

But while a lot of work has been done on modeling and using illumination in a laboratory setting, relatively little is known about it “in the wild”, i.e. in a typical outdoor scene. In fact, most vision applications treat illumination more as a nuisance – something that one strives to be invariant to – rather than a source of signal. Examples include illumination adaptation in tracking and surveillance (e.g. [21]), or contrast normalization schemes in popular object detectors (e.g. [4]). Alas, the search for the ultimate illumination invariant might be in vain [3]. Instead, we believe there is much to be gained by embracing illumination, even in the challenging, uncontrolled world of consumer photographs.

In this paper, we propose a method for estimating full sky dome illumination (sun position and clear sky appear-

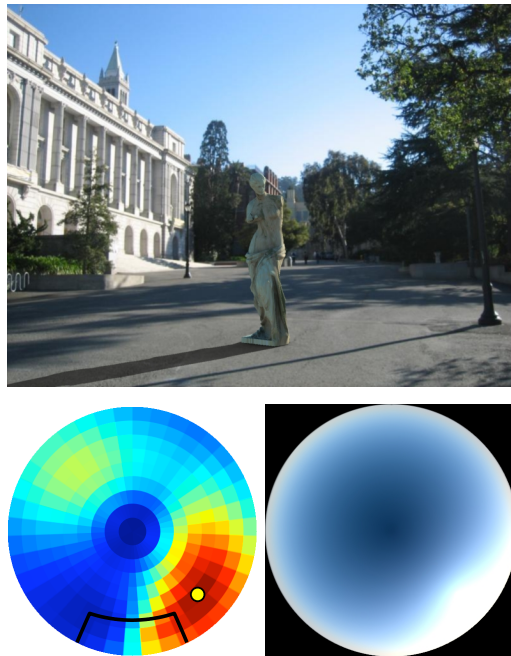


Figure 1. A synthetic 3-D statue has been placed in a photograph (top) in an illumination-consistent way. To enable this type of operation, we develop an approach that uses information from the sky, the shading and the shadows to estimate a distribution on the likely sun positions (bottom-left), and is able to generate a synthetic sky model (bottom-right) (see Fig. 5 for the original image).

ance) from a single outdoor image. To be sure, this is an extremely difficult task, even for humans [2]. In fact, the problem is severely underconstrained in the general case – while some images might have enough information for a reasonably precise estimate, others will be completely uninformative. Therefore, we will take a probabilistic approach, estimating illumination parameters using as much information as may be available in a given image and producing the maximum likelihood solution (see Fig. 1).

So what information about illumination is available in a single image? Unfortunately, there is no simple answer. When we humans perform this task, we look at different parts of the image for clues. The appearance of the sky can tell us if it’s clear or overcast (i.e. is the sun visible?). On a clear day, the sky might give some weak indication about the sun position. The presence of shadows on the ground plane can, again, inform us about sun visibility, while the

direction of shadows cast by vertical structures can tell us about sun direction. The relative shading of surfaces at differing orientations (e.g. two building facades at a right angle), can also give a rough indication of sun direction.

Our approach is based on implementing some of these intuitions into a set of illumination cues. Of course, each one of these cues by itself is rather weak and unreliable. The sky might be completely saturated, or might not even be present in the image. The ground might not be visible, or be barren of any shadow-casting structures. Shading information might, likewise, be inaccessible due to lack of appropriate surfaces or large differences between surface reflectances. Furthermore, computing these cues will inevitably lead to more noise and error (misdetected shadows, poor segmentation, incorrect camera parameters, etc). Hence, in this work, we combine the information obtained from these weak cues together, while applying a data-driven prior computed over a set of 6 million Internet photographs. The result section will show that the combined estimate yields surprisingly good performance on a wide range of very difficult real-world images.

## 2. Related work

The color and geometry of illuminants can be directly observed by placing probes, such as mirror spheres [22], color charts or integrating spheres, within the scene. But, alas, most of the photographs captured do not contain such probes and thus, we are forced to look for cues within the scene itself. There is a long and rich history in computer vision about understanding the illumination from images. We will briefly summarize relevant works here.

**Color constancy** These approaches strive to extract scene representations that are insensitive to the illumination color. For this, several works either derive transformations between scene appearances under different source colors (e.g. [6]), or transform images into different color spaces that are insensitive to source colors (e.g. [26]). Our work focuses on a complementary representation of outdoor illumination (sun direction, sky appearance, and sun visibility).

**Model based reflectance and illumination estimation** Several works estimate illumination (light direction and location), in conjunction with model-based estimation of object shape and reflectances (Lambertian, Dichromatic, Torrance-Sparrow), from one or more images of the scene [20, 1]. Our work does not rely on specific reflectance models of outdoor surfaces or exact estimation of 3-D geometry.

**Shadow extraction and analysis** Many works detect and remove shadows using one or more images [24, 7, 25]. The extracted shadows have also been used to estimate the sun direction in constrained settings [13] or in webcams [10]. But shadows are only weakly informative about illumination when their sizes in the image are too small or their

shapes are complex or blurred. Our work, for the first time, combines such weak cues with other semi-informative cues to better estimate illumination from a single image.

**Illumination estimation from time-lapse sequences** Sunkavalli et al. [23] develop techniques to estimate sun direction and scene geometry by fitting a photometric model of scene reflectance to a time-lapse sequence of an outdoor scene. Lalonde et al. [16, 17] exploit a physically-based model of sky appearance [18] to estimate the sun position relative to the viewing direction from a time-lapse sequence. We will use the same model of the sky but recover the most likely representation of the complete sky dome (sky appearance, sun position, and sun visibility) from a single image.

Finally, Lalonde et al. [15] use cues such as multi-variate histograms of color and intensity together with a rough classification of scene geometry [9] to match illuminations of different scenes. However, their cues are global in nature and cannot be used to match sun directions. This makes their approach ill-suited for 3-D object insertion.

## 3. Illumination cues from a single image

Illumination affects different parts of the scene in very different ways. In our approach, information about illumination is captured from three major parts of the image — the sky pixels  $\mathcal{S}$ , the ground pixels  $\mathcal{G}$ , and vertical surface pixels  $\mathcal{V}$  — via three separate cues. To partition the image in this way, we use the approach of Hoiem et al. [9], which returns a pixel-wise labeling of the image together with confidences (we use the latter to weight the confidence of our computed features).

We represent illumination  $I = \{\theta_s, \Delta\phi_s, v_s\}$  using three parameters, where  $\theta_s$  is the sun zenith angle,  $\Delta\phi_s$  the sun azimuth angle with respect to the camera, and  $v_s$  a binary variable for sun visibility. This section describes how we compute distributions over these parameters given the sky, the shadows on the ground, and the shading on the vertical surfaces individually. Later we combine these cues to estimate the illumination from the entire image.

### 3.1. Sky

In order to estimate illumination parameters from the sky, we take inspiration from the work of Lalonde et al. [16, 17], which shows that a physically-based sky model [18] can be used to estimate the maximum likelihood orientation of the camera with respect to the sun from a sequence of sky images. However, we are dealing with only a single image. Our solution is to discretize the parameter space and try to fit the sky model for each parameter setting. For this, we assume that the sky pixel intensities  $s_i \in \mathcal{S}$  are conditionally independent given the illumination parameters, and are distributed according to the generative model:

$$s_i \sim \mathcal{N}(k g(\theta_s, \Delta\phi_s), \sigma_s^2) , \quad (1)$$

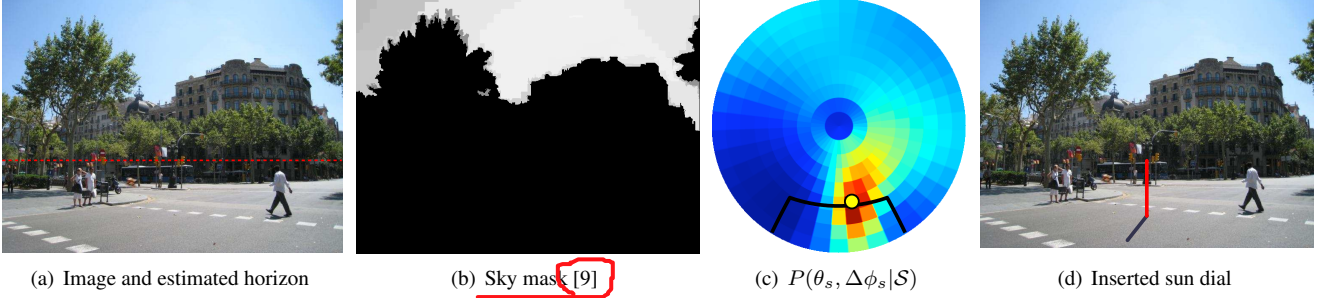


Figure 2. Illumination cue from the sky only. Starting from the input image (a), we compute the sky mask (b) using [9]. The resulting sky pixels are then used to estimate  $P(\theta_s, \Delta\phi_s | \mathcal{S})$  (c). The maximum likelihood sun position is shown with a yellow circle. We use this position to artificially synthesize a sun dial in the scene (d). Throughout the paper, the sun position probability is displayed as if the viewer is looking straight up (center point is zenith), with the camera field of view drawn at the bottom. In this example, the sun (yellow circle) appears to be to the top-right of the camera.

where,  $g(\cdot)$  is the Perez sky model [18],  $\mathcal{N}(\mu, \sigma^2)$  is the normal distribution with mean  $\mu$  and variance  $\sigma^2$ , and  $k$  is an unknown scale factor (see [17] for details). We obtain the distribution over sun positions by computing

$$P(\theta_s, \Delta\phi_s | \mathcal{S}) \propto \exp \left( \sum_{s_i \in \mathcal{S}} \frac{-(s_i - k g(\theta_s, \Delta\phi_s))^2}{2\sigma_s^2} \right) \quad (2)$$

for each bin in the discrete  $(\theta_s, \Delta\phi_s)$  space, and normalizing appropriately. Note that since  $k$  in (1) is unknown, we also discretize that space, and take the maximum value for each color channel independently.

The above function  $g(\cdot)$  requires knowledge of two important camera parameters: its zenith angle  $\theta_c$  (with respect to vertical), and its focal length  $f_c$ . If we assume that  $f_c$  is available via the EXIF tag of the photograph, then  $\theta_c$  can be computed by finding the horizon line  $v_h$  in the image. We circumvent the hard problem of horizon line estimation by making a simple approximation: select the row midway between the lowest sky pixel and the highest ground pixel as horizon. Fig. 2 demonstrates the sun position result obtained using the sky cue.

In order to determine sun visibility  $v_s$ , we classify the sky into one of three categories: clear, partially cloudy, or completely overcast. For this, we build a small database of representative skies for each category from images downloaded from the Internet, and compute the illumination context feature [15] on each. We then find the  $k$  nearest neighbors in the database, and assign the most common label (we use  $k = 5$ ). If the sky is found to be overcast, the sun position distribution is left uniform. For partly cloudy scenes, we remove the clouds by a simple binary color segmentation of the sky pixels (keeping the cluster that is closer to blue) and fit the sky model described earlier only to the clear portion of the sky. We found that setting the sun visibility probability  $P(v_s)$  to 0.8, 0.6, and 0.4 for the three classes worked well for our data.

### 3.2. Cast shadows on the ground

Shadows cast on the ground by vertical structures can essentially serve as "sun dials" and are often used by humans to determine the sun direction. Unfortunately, it is extremely hard to determine if a particular shadow was cast by a vertical object. Luckily, it turns out that due to the statistics of the world (gravity makes a many things stand-up straight), the majority of long shadows are, in fact, produced by vertical things. Therefore, if we can detect a set of "strong and long" shadow lines (edges), we can use them in a probabilistic sense to determine a likely sun azimuth (up to the directional ambiguity).

To detect shadows, we apply a series of simple heuristics. First, we compute the image derivatives in  $x$  and  $y$  both in the  $L$  and  $a$  channels of the CIELAB color space. As observed in [11], while strong *reflectance* gradients are present in both the  $L$  and  $a$  channels, strong *shadow* gradients appear mainly in the  $L$  channel. We are therefore able to reject many strong reflectance edges by subtracting the two. Of the remaining edges, we retain long shadow lines by using the approach in [14]. Finally, we observe that shadow edges have similar intensity gradients *across the image*. Thus, we choose only those edges whose intensity gradients can be clustered together. While these steps do not eliminate all reflectance edges, the number of false positives is greatly reduced.

Given a potential shadow line  $l_i \in \mathcal{G}$ , we compute its relative orientation  $\alpha_i$  on the ground plane by rectifying it via a homography using the camera parameters estimated in the previous section. We assume that each shadow line predicts the sun azimuth in the following way:

$$P(\Delta\phi_s | l_i) \sim \max(\mathcal{N}(\alpha_i, \sigma_g^2), \mathcal{N}(\alpha_i + \pi, \sigma_g^2)) \quad , \quad (3)$$

and combine all the shadow lines by making each one vote for its preferred shadow direction:

$$P(\Delta\phi_s | \mathcal{G}) \propto \sum_{l_i \in \mathcal{G}} P(\Delta\phi_s | l_i) \quad . \quad (4)$$

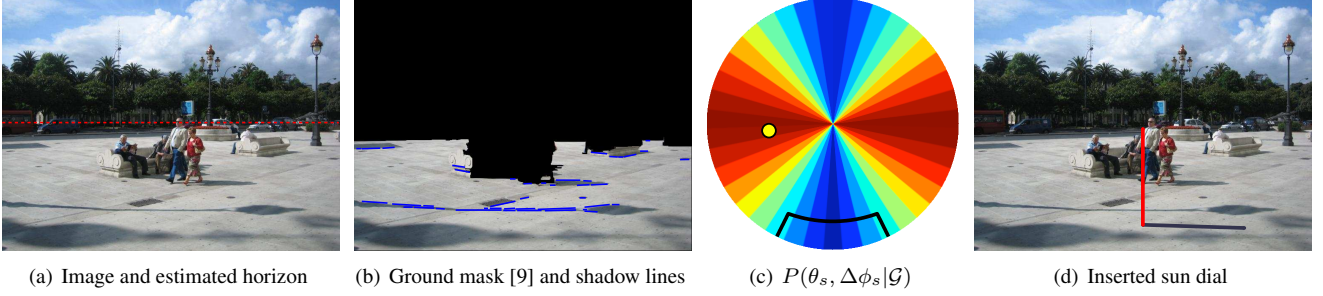


Figure 3. Illumination cue from the ground only. Starting from the input image (a), we compute the ground mask (b) using [9] and extract shadow lines, which are then used to estimate  $P(\theta_s, \Delta\phi_s|\mathcal{G})$  (c). Note that shadow lines alone can only predict the sun relative azimuth angle up to a  $180^\circ$  ambiguity. The subtle horizontal sky gradient is however able to disambiguate between the two hypotheses and select a realistic sun position, shown with a yellow circle. The most likely sun position is used to artificially synthesize a sun dial in the scene (d).

Fig. 3 shows the results obtained using the shadow cues only.

For  $v_s$ , a large number of shadow edges on the ground indicate that the sun should be visible, but it is uninformative if none are found. Therefore, we set  $P(v_s|\mathcal{G}) = 0.3 \min(n, 30)/30 + 0.5$  ( $n$  is the number of edges).

### 3.3. Shading on vertical surfaces

If the rough geometric structure of the scene is known, then analyzing the shading on the main surfaces can often provide an estimate for the possible sun positions. For example, a brightly lit surface indicates that the sun may be pointing in the direction of its normal, or at least in the vicinity. Of course, this reasoning also assumes that the albedos of the surfaces are either known or equal, neither of which is true. However, we have experimentally found that, within a given image, the albedos of the major *vertical* surfaces are often relatively similar (e.g. different sides of the same house, or similar houses on the same street), while the ground is quite different. Therefore, we use the three coarse vertical surface orientations (front, left-facing, and right-facing) computed by [9] and attempt to estimate the azimuth direction only. We assume that a bright surface  $w_i \in \mathcal{V}$  predicts that the sun is roughly in the direction of its normal,  $\beta_i$ . Therefore, each surface predicts:

$$P(\Delta\phi_s|w_i) \sim \mathcal{N}(\beta_i, \sigma_w^2) , \quad (5)$$

where  $\sigma_w^2$  is proportional to its corresponding surface intensity. Note that  $\beta_i \in \{-90^\circ, 90^\circ, 180^\circ\}$  since we assume only 3 coarse surface orientations. We combine each surface by making each one vote for its preferred sun direction:

$$P(\Delta\phi_s|\mathcal{V}) \propto \sum_{w_i \in \mathcal{V}} P(\Delta\phi_s|w_i) . \quad (6)$$

Fig. 4 shows the results obtained using the shading on vertical surfaces only. Although this is a weaker cue, we find that it can often help resolve the ambiguity arising with

shadow lines. Since bright vertical surfaces should indicate that the sun is visible, we set  $P(v_s|\mathcal{V}) = w_{max}$ , where  $w_{max} \in [0, 1]$  is the mean intensity of the brightest surface.

The three variances mentioned in this section were obtained automatically by cross-validation on a training set of webcam images (see Sec. 5.1). Their values are  $\sigma_s^2 = 0.18$ , and  $\sigma_g^2 = \sigma_w^2 = 25$ .

## 4. Estimating illumination

Now that we are equipped with several weak features that can be computed over an image, we show how we combine them in order to get a more reliable estimate in a probabilistic framework.

### 4.1. Cue combination

We are interested in estimating the distribution  $P(I|\mathcal{P})$  over the illumination parameters  $I = \{\theta_s, \Delta\phi_s, v_s\}$ , given the entire image  $\mathcal{P}$ . We apply Bayes rule, and write

$$P(I|\mathcal{S}, \mathcal{G}, \mathcal{V}) \propto P(\mathcal{S}, \mathcal{G}, \mathcal{V}|I)P(I) . \quad (7)$$

We make the assumption that the image pixels are conditionally independent given the illumination conditions, and that the priors on each region of the image ( $P(\mathcal{S})$ ,  $P(\mathcal{G})$  and  $P(\mathcal{V})$ ) are uniform over their own respective domains. Applying Bayes rule twice, we get

$$P(I|\mathcal{S}, \mathcal{G}, \mathcal{V}) \propto P(I|\mathcal{S})P(I|\mathcal{G})P(I|\mathcal{V})P(I) . \quad (8)$$

We further assume the sun visibility  $v_s$  to be independent of its position, which means (8) can be split into two independent equations: one for  $(\theta_s, \Delta\phi_s)$  and for  $v_s$ . Note that while these two independence assumptions greatly reduces the complexity of the problem, they might not always be true. For instance, the sun being blocked entirely by a scene occluder violates this assumption, but it is a rather rare event in practice.

The process of combining the cues according to (8) for the sun position is illustrated in Fig. 5. We have presented how we compute the conditionals  $P(I|\mathcal{S})$ ,  $P(I|\mathcal{G})$



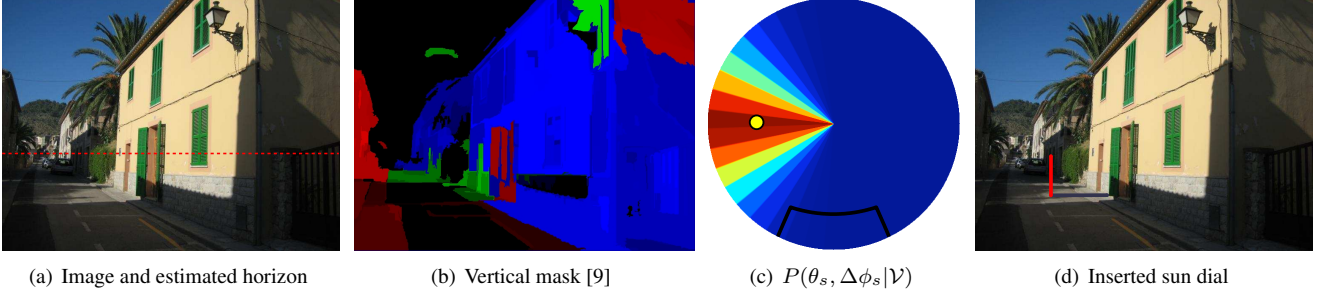


Figure 4. Illumination cue from the vertical surfaces only. Starting from the input image (a), we compute the vertical surfaces mask (b) using [9] (blue = facing left, red = facing right, green = facing forward). The distribution of pixel intensities on each of these surfaces are then used to estimate  $P(\theta_s, \Delta\phi_s | \mathcal{V})$  (c). Note that in our work, vertical surfaces cannot predict the sun zenith angle  $\theta_s$ . In this example, the sky is used to estimate the sun zenith. We then find the most likely sun position (shown with a yellow circle), which is used to artificially synthesize a sun dial in the scene (d).

and  $P(I | \mathcal{V})$  in Sect. 3.1, 3.2 and 3.3 respectively. Let us now look at how we can compute the prior  $P(I)$  on the illumination conditions themselves.

#### 4.2. Data-driven illumination prior

The prior  $P(I) = P(\theta_s, \Delta\phi_s, v_s)$  captures the typical illumination conditions in outdoor scenes. Since we assumed  $v_s$  to be independent of the sun position, we can compute  $P(\theta_s, \Delta\phi_s)$  and  $P(v_s)$  independently. Because labelled visibility data is not currently available, we conservatively take  $P(v_s) = 0.5$ . We now proceed to show how we can compute  $P(\theta_s, \Delta\phi_s)$  given a large dataset of consumer photographs.

The sun position  $(\theta_s, \Delta\phi_s)$  depends on the latitude  $L$  of the camera, its azimuth angle  $\phi_c$ , the date  $D$  and the time of day  $T$  expressed in the local timezone:

$$P(\theta_s, \Delta\phi_s) = P(f(L, D, T, \phi_c)) \quad , \quad (9)$$

where  $f(\cdot)$  is a non-linear function defined in [19]. To estimate (9), we can sample points from  $P(L, D, T, \phi_c)$ , and use  $f(\cdot)$  to recover  $\theta_s$  and  $\Delta\phi_s$ . But estimating this distribution is not currently feasible since it requires images with known camera orientations  $\phi_c$ , which are not yet available in large quantities. On the other hand, geo- and time-tagged images do exist, and are widely available on photo sharing websites such as Flickr. The database of 6 million images from [8] is used to compute the empirical distribution  $P(L, D, T)$ . We compute (9) by randomly sampling 1 million points from the distribution  $P(L, D, T)P(\phi_c)$ , assuming  $P(\phi_c)$  to be uniform in the  $[-180^\circ, 180^\circ]$  interval. As a consequence, (9) is flat along the  $\Delta\phi_s$  dimension and is marginalized.

Fig. 6 shows 4 estimates for  $P(\theta_s, \Delta\phi_s)$ , computed with slightly different variations. First, a uniform sampling of locations on Earth and times of day is used as a baseline comparison. The three other priors use data-driven information. Considering non-uniform date and time distributions

decrease the likelihood of having pictures with the sun taken close to the horizon ( $\theta_s$  close to  $90^\circ$ ). Interestingly, the red and green curve overlap almost perfectly, which indicates that the three variables  $L$ ,  $D$ , and  $T$  seem to be independent.

This database captures the distribution of where photographs are most likely to be taken on the planet, which is indeed very different than considering each location on Earth as equally likely (as shown in Fig. 6). We will show in the next section that this distinction is critical to improve our estimation results. Finally, note that the assumption of uniform camera azimuth is probably not true in practice since a basic rule of thumb of good photography is to take a picture with the sun to the back. With the advent of additional sensors such as compasses on digital cameras, this data will surely become available in the near future.

### 5. Evaluation and results

We evaluate our technique in two different ways: quantitatively using images taken from webcam sequences, and qualitatively on images downloaded from Internet.

#### 5.1. Quantitative evaluation using webcams

We use the technique of Lalonde et al. [16, 17] to estimate the positions of the sun in 984 images taken from 15 different time-lapse image sequences, downloaded from the Internet. Two example images from our dataset are shown in Fig. 7(a) and 7(b). Our algorithm is applied to every image from the sequences independently and the results are compared against ground truth. Accordingly, the sun visibility was estimated correctly 98% of the time.

Fig. 7 reports the cumulative histogram of errors in sun position estimation for different scenarios: chance, making a constant prediction of  $\theta_s = 0$  (straight up), using only the priors from Sec. 4.2 (we tested both the data-driven and Earth uniform priors), scene cues only, and using our combined measure  $P(\theta_s, \Delta\phi_s | \mathcal{P})$  with both priors as well. The

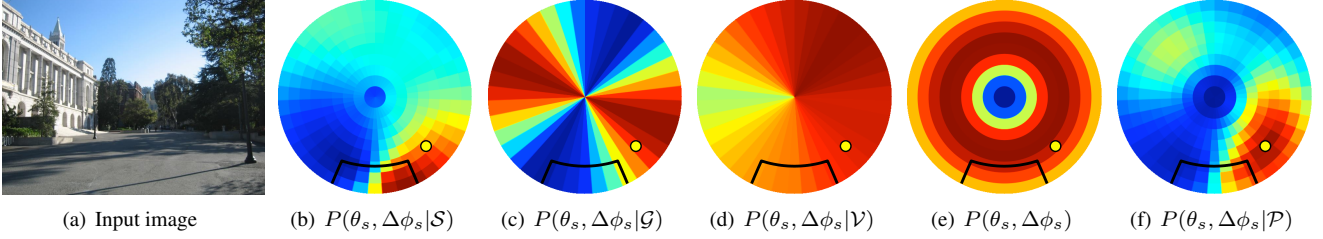


Figure 5. Combining illumination features computed from image (a) yields a more confident final estimate (f). We show how (b) through (d) are estimated in Sect. 3, and how we compute (e) in Sect. 4.2.

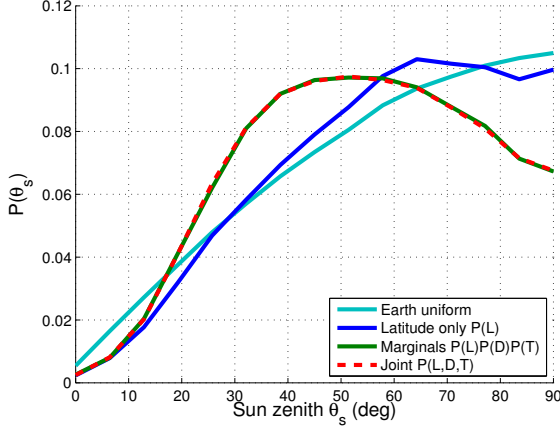


Figure 6. Illumination priors (9) on the sun position, learned from a database of 6M images downloaded from Flickr. The priors are obtained by sampling 1 million points (1) uniformly across GPS locations and time of day (cyan); from (2) the marginal latitude distribution  $P(L)$  only (blue); (3) the product of independent marginals  $P(L)P(D)P(T)$  obtained from data (green); and (4) the joint  $P(L, D, T)$ , also obtained from data (red). The last two curves overlap, indicating that the three variables  $L$ ,  $D$ , and  $T$  indeed seem to be independent.

first three curves were obtained by optimizing for the variances from Sec. 3 using 5-fold cross-validation, and taking the mean test error for each image across folds. Fig. 7 highlights the performance at errors of less than  $22.5^\circ$  (50% of images) and  $45^\circ$  (71% of images), which correspond to accurately predicting the sun position within an octant (e.g. North vs. North-West), or a quadrant (e.g. North vs West) respectively.

The cues contribute to the end result differently for different scenes. For instance, the sky in Fig. 7(a) is not informative, since it is small and the sun is always behind the camera. It is more informative in Fig. 7(b), as it occupies a larger area.

## 5.2. Qualitative evaluation on single images

Fig. 8 shows several example results of applying our algorithm on typical consumer-grade images, downloaded from the Internet. The rows are arranged in order of decreasing confidence in the result. High confidence cases are



(a) Madrid sequence

(b) Vatican sequence

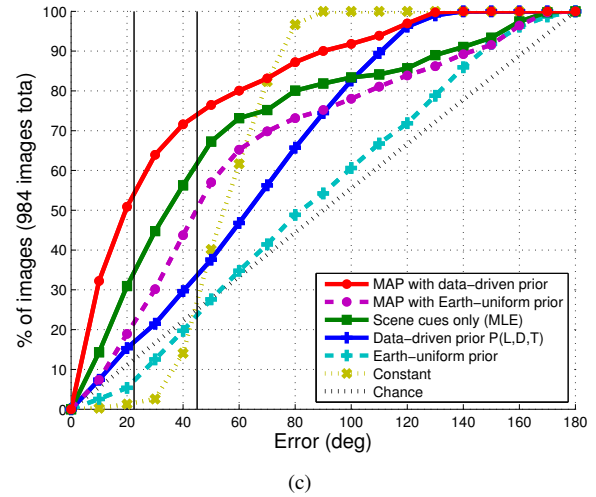


Figure 7. Quantitative evaluation using 984 images taken from 15 webcam sequences, calibrated using [16, 17]. Two example images, from (a) Madrid and (b) Vatican City, compare sun dials rendered using our estimated sun position (gray), and the ground truth (red). (c) Cumulative sun position error (angle between estimate and ground truth directions) for different methods. Our result, which combines both the scene cues and the data-driven illumination prior, outperforms the others. The data-driven prior surpasses the Earth-uniform one (both alone and when combined with the scene cues), showing the importance of being consistent with likely image locations. Picking a constant value of  $\theta_s = 0$  has an error of at least  $20^\circ$ .

usually obtained when one cue is very strong (i.e. bright facing wall in the top-left example, sky gradient in top-right), or when all 3 cues are strongly correlated (as in Fig. 1). On the other hand, few shadows and surfaces are present (bottom-left), or highly cluttered scenes (bottom-

right) usually yield lower confidences, and the most likely sun positions might be uncertain. This is emphasized in Fig. 9, which shows two typical failure cases.

### 5.3. Application: 3-D object insertion

We now demonstrate how to use our technique to insert a 3-D object into a single photograph with realistic lighting. This requires generating a plausible environment map [5] from the image. To do so, we must estimate the illumination intensity in all directions, even those not seen by the camera! Consider the image Fig. 10. We first estimate the illumination conditions using our approach. Given the most likely sun position and clear sky pixels, we use the technique of Lalonde et al. [16, 17] to recover the most likely sky parameters, and synthesize its appearance in all directions (Fig. 10, center). The sun is simulated by a bright circular patch ( $10^4$  times brighter than the maximum scene brightness). For the bottom part of the environment map (not shown), we use the spherical projection technique of Khan et al. [12] on the pixels below the horizon line. A realistic 3-D model is relit using an off-the-shelf rendering software (see Fig. 10, right). Notice how the shadows on the ground, and shading and reflections on the car are consistent with the image. Another example is shown in Fig. 1.

## 6. Discussion

Outdoor illumination affects the appearances of scenes in complex ways. Untangling illumination from surface geometry and material properties is a hard problem in general. Surprisingly, however, numerous consumer-grade photographs captured outdoors contain rich and informative cues about illumination, such as the sky, the shadows on the ground and the shading on vertical surfaces. Our approach extracts the “collective wisdom” from these cues to estimate the complete sky dome (sun position, if it is visible, and sky appearance) from a single image. Even when the lighting information within an image is minimal, and the resulting estimates are weak, we believe it can still be a useful result for a number of applications. For example, just knowing that the sun is somewhere on your left might be enough for a point-and-shoot camera to automatically adjust its parameters, or for a car detector to be expecting cars with shadows on the right. Several additional pieces of information can also be exploited to help in illumination estimation. For instance, GPS coordinates, time of day and camera orientation are increasingly being tagged in images. Knowing these quantities can further constrain the position of the sun and increase confidences in the probability maps that we estimate. We will explore these avenues in the future.

**Acknowledgements:** This work has been partially supported by NSF grants CCF-0541230, IIS-0546547, IIS-

0643628 and ONR grant N00014-08-1-0330. A. Efros is grateful to the WILLOW team at ENS Paris for their hospitality.

## References

- [1] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *IJCV*, 72(3):239–257, May 2007.
- [2] P. Cavanagh. The artist as neuroscientist. *Nature*, 434:301–307, 2005.
- [3] H. Chen, P. Belhumeur, and D. Jacobs. In search of illumination invariants. In *CVPR*, 2000.
- [4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005.
- [5] P. Debevec. Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *SIGGRAPH*, 1998.
- [6] G. Finlayson, M. Drew, and B. Funt. Diagonal transforms suffice for color constancy. In *ICCV*, 1993.
- [7] G. D. Finlayson, S. D. Hordley, and M. S. Drew. Removing shadows from images. In *ECCV*, 2002.
- [8] J. Hays and A. A. Efros. im2gps: estimating geographic information from a single image. In *CVPR*, 2008.
- [9] D. Hoiem, A. A. Efros, and M. Hebert. Recovering surface layout from an image. *IJCV*, 75(1):151–172, Oct 2007.
- [10] I. N. Junejo and H. Foroosh. Estimating geo-temporal location of stationary cameras using shadow trajectories. In *ECCV*, 2008.
- [11] E. A. Khan and E. Reinhard. Evaluation of color spaces for edge classification in outdoor scenes. In *ICIP*, September 2005.
- [12] E. A. Khan, E. Reinhard, R. Fleming, and H. Büelthoff. Image-based material editing. *SIGGRAPH*, August 2006.
- [13] T. Kim and K. Hong. A practical single image based approach for estimating illumination distribution from shadows. In *ICCV*, 2005.
- [14] J. Koščeká and W. Zhang. Video compass. In *ECCV*, 2002.
- [15] J.-F. Lalonde, D. Hoiem, A. A. Efros, C. Rother, J. Winn, and A. Criminisi. Photo clip art. *SIGGRAPH*, 2007.
- [16] J.-F. Lalonde, S. G. Narasimhan, and A. A. Efros. What does the sky tell us about the camera? In *ECCV*, 2008.
- [17] J.-F. Lalonde, S. G. Narasimhan, and A. A. Efros. What do the sun and the sky tell us about the camera? Technical Report CMU-RI-TR-09-04, Robotics Institute, Carnegie Mellon University, January 2009.
- [18] R. Perez, R. Seals, and J. Michalsky. All-weather model for sky luminance distribution – preliminary configuration and validation. *Solar Energy*, 50(3):235–245, March 1993.
- [19] I. Reda and A. Andreas. Solar position algorithm for solar radiation applications. Technical Report NREL/TP-560-34302, November 2005.
- [20] I. Sato, Y. Sato, and K. Ikeuchi. Illumination from shadows. *IEEE Transactions on PAMI*, 25(3):290–300, March 2003.
- [21] C. Stauffer. Adaptive background mixture models for real-time tracking. In *CVPR*, 1999.
- [22] J. Stumpfel, A. Jones, A. Wenger, C. Tchou, T. Hawkins, and P. Debevec. Direct HDR capture of the sun and sky. In *AFRIGRAPH*, 2004.
- [23] K. Sunkavalli, F. Romeiro, W. Matusik, T. Zickler, and H. Pfister. What do color changes reveal about an outdoor scene? In *CVPR*, 2008.
- [24] Y. Weiss. Deriving intrinsic images from image sequences. In *ICCV*, 2001.
- [25] T.-P. Wu and C.-K. Tang. A bayesian approach for shadow extraction from a single image. In *ICCV*, 2005.
- [26] T. Zickler, S. P. Mallick, D. J. Kriegman, and P. N. Belhumeur. Color subspaces as photometric invariants. In *CVPR*, 2006.





Figure 8. Sun direction estimation from a single image. A virtual sun dial is inserted in each input image (first and third columns), whose shadow correspond to the MAP sun position in the corresponding probability maps  $P(\theta_s, \Delta\phi_s | \mathcal{P})$  (second and fourth columns). The rows are ordered from most (top) to least (bottom) confidence.



Figure 9. Failure cases. First, the dominating shadow lines are not cast by thin vertical objects, and are not aligned with sun direction. In the second case, the strong reflectance edges are being misclassified as shadow lines. In both cases, none of the other cues were confident enough to compensate, yielding erroneous estimates.

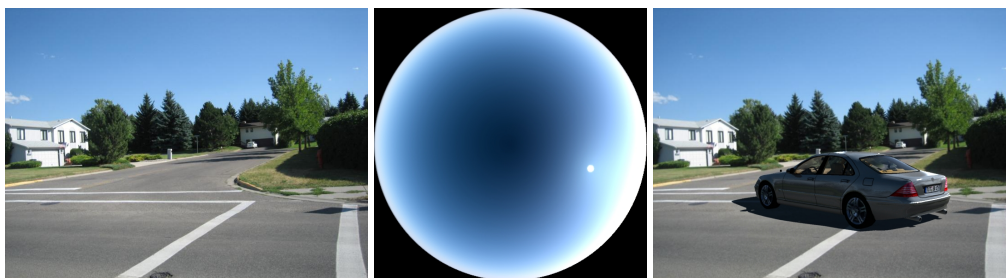


Figure 10. 3-D object relighting. From a single image (left), we render the most likely sky appearance (center) using the sun position computed with our method, and then fitting the sky parameters using [17]. We can realistically insert a 3-D object into the image (right).