E-mail: {jlalonde,efros,srinivas}@cs.cmu.edu

Estimating the Natural Illumination Conditions from a Single Outdoor Image

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

the date of receipt and acceptance should be inserted later

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, the ground, and the convex objects in the image. 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 photos. We present quantitative results on a webcam dataset with annotated sun positions, as well as quantitative and 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, and how to transfer appearance across images while keeping the illumination consistent.

Keywords illumination estimation · data-driven methods · shadow detection · scene understanding · image synthesis

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.

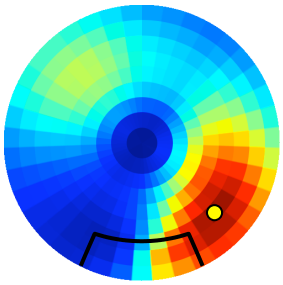
But while a lot of work has been done on modeling and using illumination in a laboratory setting, relatively little is known about it "in the wild", i.e. in a typical outdoor scene. In fact, most vision applications treat illumination more as a nuisance — something that one strives to be invariant to — rather than a source of signal. Examples include illumination adaptation in tracking and surveillance (e.g. [62]), or contrast normalization schemes in popular object detectors (e.g. [11]). Alas, the search for the ultimate illumination invariant might be in vain [8]. 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 natural illumination (sun position and visibility) from a single outdoor image. To be sure, this is an extremely difficult task, even for humans [6]. 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).

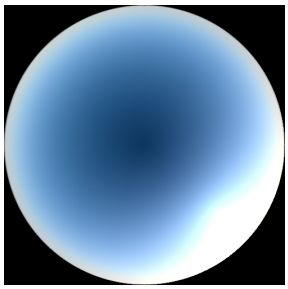
So what information about illumination is available in a single image? Unfortunately, there is no simple answer. When we humans perform this task, we look at different parts of the image for clues. The appearance of the sky can tell us if it is clear or overcast (i.e. is the sun visible?). On a clear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

School of Computer Science, Carnegie Mellon University

(a)



(b)



(c)

Fig. 1: A synthetic 3-D statue has been placed in a photograph (a) in an illumination-consistent way (the original image is shown in the inset). To enable this type of operation, we develop an approach that uses information from the sky, the shading, the shadows, and the visible pedestrians to estimate a distribution on the likely sun positions (b), and is able to generate a synthetic sky model (c)

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, and so can the shading effects on convex objects populating the scene (e.g. pedestrians, cars, poles, etc.)

Our approach is based on implementing some of these intuitions into a set of illumination cues. More precisely, the variation in sky color, cast shadows on the ground, relative shading of vertical surfaces, and intensity variation on convex objects (pedestrians) are the four cues used in this work. Of course, each one of them are rather weak and unreliable when taken individually. The sky might be completely saturated, or might not even be present in the image. The ground might not be visible, be barren of any shadow-casting structures, or be deprived of any recognizable convex objects. 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 or pedestrians, 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 sections (Secs. 3.2 and 6) will show that the sun visibility can be estimated with 83.5% accuracy, and the combined estimate is able to successfully locate the sun within an octant (quadrant) for 40% (55%) of the real-world images in our very challenging test set, and as such outperforms any of the cues taken independently. While this goes to show how hard the problem of illumination from single images truly is, we believe this 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.

1.1 Related work

The color and geometry of illuminants can be directly observed by placing probes, such as mirror spheres [63], 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 illuminant color. For this, several works either derive transformations between scene appearances under different source colors (e.g. [19]), or transform images into different color spaces that are insensitive to illuminant colors (e.g. [73]). Our work focuses on a complementary representation of outdoor illumination (sun direction and 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 [59, 1]. Of particular interest, Sun *et al.* [64] estimate illumination conditions in complex urban scenes by registering the photographs to 3-D models of the

scene. 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 [69,22,70]. The extracted shadows have also been used to estimate the sun direction in constrained settings [35] or in webcams [31]. But shadows are only weakly informative about illumination when their sizes in the image are too small or their shapes are complex or blurred. Li et al. [45] propose a technique that combine shadow, shading, and specularity information in a framework for estimating multiple illuminant directions in single images, but their approach is restricted to tabletop-like objects. Our work, for the first time, combines shadow cues with other semi-informative cues to better estimate illumination from a single image of a general outdoor scene.

Illumination estimation from time-lapse sequences Sunkavalli et al. [65] 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. [43] exploit a physically-based model of sky appearance [51] 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. [42] use cues such as multivariate histograms of color and intensity together with a rough classification of scene geometry [27] 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.

2 Representations for natural illumination

Because of its predominant role in any outdoor setting, it has been of critical importance to understand natural illumination in many fields such as computer vision and graphics, but also in other research areas such as biology [25], architecture [56], solar energy [48], and remote sensing [24]. Consequently, researchers have proposed many different representations for natural illumination adapted to their respective applications. We present here an overview of popular representations that we divide into three main categories: "physically-based", "environment maps", and "statistical". We conclude this section by describing the representation that we use in this work.

2.1 Physically-based representations

The type of representation that is probably the most popular in the literature is what we name here "physically-based" representations: mathematical expressions that model the physics of natural illumination. They typically stem from the following equation (adapted from [61]), which models the ground spectral irradiance $L(\lambda)$ as a function of the sun and sky radiances E_{sun} and E_{sky} respectively:

$$L(\lambda) = K E_{sun}(\theta_s, \phi_s, \lambda)\cos(\theta_s) + \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\frac{\pi}{2}} E_{sky}(\theta,\phi,\lambda)\cos(\theta)\sin(\theta)d\theta d\phi\,, \quad (1)$$

where K is a binary constant which accounts for occluding bodies in the solar to surface path, and E_{sky} is integrated over the entire sky hemisphere. While (1) may look simple, its complexity lies in the characterization of the sun and sky radiance components which depend on the form and amount of atmospheric scattering. As a result, a wide variety of ways to model the sun and the sky have been proposed; we summarize a few here that are most relevant to our work.

Being the dominant light source outdoors, the sun has been studied extensively. For example, in architectural design, the relative position of the sun is an important factor in the heat gain of buildings and in radiance computations that determine the quantity of natural light received inside each of the rooms [68]. Understanding the quantity of energy delivered by the sun at ground level is also critical to predict the quantity of electricy that can be produced by photovoltaic cells [48]. As such, highly precise physical models for sunlight transport through the atmosphere (E_{sun} in (1)) have been proposed [2,32].

The second main illuminant outdoors, the sky, has also long been studied by physicists. One of the most popular physically-based sky model was introduced by Perez *et al.* [51], and was built from measured sky luminances. This model has been used in graphics for relighting architectural models [71], and for developing an efficient sky rendering algorithm [53]. This is also the model we will ourselves use to understand the sky appearance (see Sec. 4.1). It has recently been shown that some of the parameters of this model can be estimated from an image sequence captured by a static camera, in which the sky is visible [43].

In computer vision, Sato and Ikeuchi [60] use a model similar to (1) along with simple reflectance models to estimate the shape of objects lit by natural illumination. This has led to applications in outdoor color rep-

resentation and classification [5], surveillance [67], and robotics [46]. A similar flavor of the same model has also been used in the color constancy domain by [47]. Because physical models possess many parameters, recovering them from images is not an easy task. Therefore, researchers have resorted to simplifying assumptions, approximations, or to the use of image sequences [65] to make the estimation problem more tractable.

2.2 Environment map representations

An alternative representation is based on the accurate measurement and direct storage of the quantity of natural light received at a point that comes in from all directions. As opposed to "physically-based" representations, here no compact formula is sought and all the information is stored explicitly. Originally introduced in the computer graphics community by Blinn and Newell [4] to relight shiny objects, the representation (also dubbed "light probe" in the community) was further developed by Debevec [13,14] to realistically insert virtual objects in real images. It is typically captured using a high-quality, high-dynamic range camera equipped with an omni-directional lens (or a spherical mirror), and requires precise calibration. Stumpfel *et al.* [63] more recently proposed to employ this representation to capture the sky over an entire day into an environment map format, and used it for both rendering and relighting [15].

In comparison to its "physically-based" counterpart, the "environment map" representation has no parameters to estimate: it is simply measured directly. However, its main drawback — aside from large memory requirements — is that physical access to the scene of interest is necessary to capture it. Thus, it cannot be used on images which have already been captured.

2.3 Statistical representations

The third type of representation for natural illumination has its roots in data mining and dimensionality reduction techniques. The idea is to extract the low-dimensional trends from datasets of measurements of natural illumination. This "statistical", or "data-driven" representation is typically obtained by gathering a large number of observations — either from physical measurements of real-world illumination captured around the world [30], or by generating samples using a physical model [61] — and performing a dimensionality reduction technique (e.g. PCA) to discover which dimensions explain most of the variance in the samples.

One of the first to propose such an idea was Judd *et al.* [30], and is still to this date considered as one of the best experimental analysis of daylight [61]. Their study considered 622 samples of daylight measured in the visible spectrum, and observed that most of them could be approximated accurately by a linear combination of three fixed functions. Subsequently, Slater and Healey [61] reported that a 7-dimensional PCA representation capture 99% of the variance of the spectral distribution of natural illumination in the visible and near-infrared spectra, based on a synthetically-generated dataset of spectral lighting profiles. Dror *et al.* [17] performed a similar study by using a set of HDR environment maps as input.

Since then, linear representations for illumination has been successful in many applications, notably in the joint estimation of lighting and reflectance from images of an object of known geometry. Famously, Ramamoorthi and Hanrahan [54] used a spherical harmonics representation (linear in the frequency domain) of reflectance and illumination and expresses their interaction as a convolution. More recently, Romeiro and Zickler [57] also use a linear illumination basis to infer material properties by marginalizing over a database of real-world illumination conditions.

2.4 An intuitive representation for natural illumination

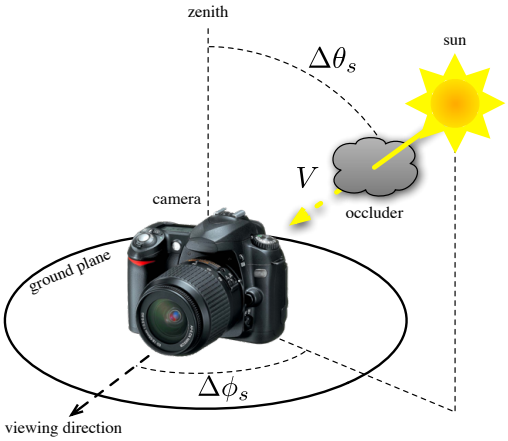
All the previous representations assume that the camera (or other sensing modalities) used to capture illumination are of very high quality — there must be a linear relationship between the illumination radiance and sensor reading, they have high dynamic range, etc. — and that the images contain very few, easily recognizable objects. However, most consumer photographs that are found on the Internet do not obey these rules, and we have to deal with acquisition process issues like non-linearity, vignetting, compression and resizing artifacts, limited dynamic range, out-of-focus blur, etc. In addition, scenes contain occlusions, depth discontinuities, wide range of scales due to perspective projection. These issues make the previous illumination representations very hard to estimate from these images.

Instead, we propose to use a representation that is more amenable to understand these types of images. The model focuses on the main source of outdoor lighting: the sun. We model the sun using two variables:

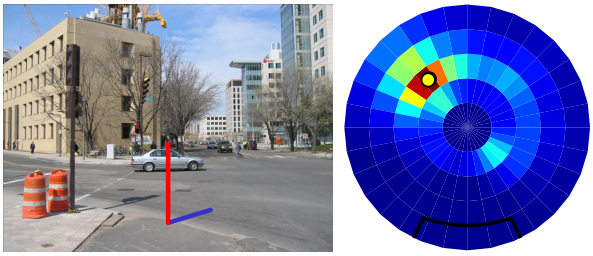
- V its visibility, or whether or not it is shining on the scene;
- S its angular position relative to the camera. In spherical coordinates, $S = (\Delta\theta_s, \Delta\phi_s)$, where $\Delta\theta_s = \theta_s - \theta_c$ and $\Delta\phi_s = \phi_s - \phi_c$ (the s and c indices

denote the sun and camera respectively). S is specified only if $V = 1$, and undefined if $V = 0$.

In contrast to existing approaches, this representation for natural illumination is more simple and intuitive, therefore making it better-suited for single image interpretation. For example, if the sun visibility V is 0 (for example, when the sky is overcast), then the sun relative position S is undefined. Indeed, if the sun is not shining on the scene, then all the objects in the scene are lit by the same area light source that is the sky. It is not useful to recover the sun direction in this case. If V is 1 however, then knowing the sun position S is very important since it is responsible for illumination effects such as cast shadows, shading, specularities, etc. which might strongly affect the appearance of the scene. Fig. 2 illustrates our model schematically.



(a)



(b) (c)

Fig. 2: An intuitive representation for natural illumination. (a) Illumination I is modeled using two variables: 1) the sun visibility V, which indicates whether the sun is shining on the scene or is blocked by an occluder (e.g. cloud); and 2) its angular position relative to the camera $S = (\Delta\theta_s, \Delta\phi_s)$. Throughout the paper, the sun position probability $P(S)$ for an input image such as (b) is displayed (c) as if the viewer is looking straight up (center point is zenith), with the camera field of view drawn at the bottom. Probabilities are represented with a color scale that vary from blue (low probability) to red (high probability). In this example, the maximum likelihood sun position (yellow circle) is estimated to be at the back-right of the camera. We sometimes illustrate the results by inserting a virtual sun dial (red stick in (b)) and drawing its shadow corresponding to the sun position.

3 Is the sun visible or not?

Using the above representation, we estimate $P(I|\mathcal{I})$, the probability distribution over the illumination parameters I, given a single image $\mathcal{I}$. Because our representation defines the sun position S only if it is visible, we propose to first estimate the sun visibility V:

$$P(I|\mathcal{I}) = P(S, V|\mathcal{I}) = P(S|V, \mathcal{I})P(V|\mathcal{I})\,. \quad (2)$$

In this section, we present how we estimate the distribution over the sun visibility variable given the image $P(V|\mathcal{I})$. The remainder of the paper will then focus on how we estimate $P(S|V = 1, \mathcal{I})$, that is, the probability distribution over the sun position if it was determined to be visible. We propose a supervised learning approach to learn $P(V|\mathcal{I})$, where a classifier is trained on features computed on a manually-labelled dataset of images. We first describe the features that were developed for this task, then provide details on the classifier used.

3.1 Image cues for predicting the sun visibility

When the sun is shining on the scene, it usually creates noticeable effects in the image. Consider for a moment the first and last images of Fig. 3. When the sun is visible (Fig. 3a), it creates hard cast shadow boundaries, bright and dark areas correponding to sunlit and shadowed regions, highly-saturated colors, and clear skies. On the other hand, when the sun is occluded (Fig. 3e), the colors are dull, the sky is gray or saturated, and there are no visible shadows.

We devise sun visibility features based on these intuitions. We first split the image into three main geometric regions: the ground $\mathcal{G}$, the sky $\mathcal{S}$, and the vertical surfaces $\mathcal{V}$ using the geometric context classifier of [27]. We use these regions when computing the following set of features:

Bright and dark regions: We compute the mean intensity of the brightest of two clusters on the ground $\mathcal{G}$, where the clustering is performed with k-means with $k = 2$; The same is done for the vertical surfaces $\mathcal{V}$;

Saturated colors: We compute 4-dimensional marginal normalized histograms of the scene $(\mathcal{G} \cup \mathcal{V})$ in the

(a) $P(V|\mathcal{I}) = 0.95$ (b) $P(V|\mathcal{I}) = 0.75$ (c) $P(V|\mathcal{I}) = 0.50$ (d) $P(V|\mathcal{I}) = 0.25$ (e) $P(V|\mathcal{I}) = 0.05$

Fig. 3: Results of applying the sun visibility classifier on single images to estimate $P(V|\mathcal{I})$, sorted by decreasing order of probability.

saturation and value color channels; To add robustness to variations in exposure, white balance, and gamma response functions across cameras, we follow [12] and also compute a 4-dimensional histogram of the scene in the normalized log-RGB space;

Sky: If the sky is visible, we compute its average color in RGB;

Contrast: We use the contrast measure proposed in [33] (a measure of the histogram spread), computed over the scene pixels only ($\mathcal{G} \cup \mathcal{V}$);

Shadows: We apply the shadow detection method of [40], and count the fraction of pixels that belong to the ground $\mathcal{G}$ and scene ($\mathcal{G} \cup \mathcal{V}$) and are labelled as shadows.

These features are concatenated in a 20-dimensional vector which is used in the learning framework described next.

3.2 Sun visibility classifier

We employ a classical supervised learning formulation, in which the image features are first pre-computed on a set of manually-labelled training images, and then fed to a classifier. We now provide more details on the learning setup used to predict whether or not the sun is visible in an image.

We selected a random subset of outdoor images from the LabelMe dataset [58], split into 965 images used for training, and 425 for testing. Because LabelMe images taken from the same folders are likely to come from the same location or taken by the same camera, the training and test sets were carefully split to avoid folder overlap. Treating V as a binary variable, we then manually labelled each image with either $V = 1$ if the sun is shining on the scene, or $V = 0$ otherwise. Notice that this binary representation of V is only a coarse approximation of the physical phenomenon: in reality, the sun may have fractional visibility (e.g. due to partially-occluding clouds, for example). In practice, however, we find that it is extremely difficult, even for humans, to reliably estimate a continuous value for the sun visibility from a single image. Additionally, precisely measuring this value requires expensive equipment that is not available in databases of existing images.

	Not visible	Visible
Not visible	87.6%	12.4%
Visible	20.5%	79.5%

Table 1: Confusion matrix for the sun visibility classifier. Overall, the class-normalized classification accuracy is 83.5%.

We then train a binary, linear SVM classifier using this training dataset, and evaluate its performance on the test set. We use the libsvm package [7], and convert the SVM scores to probabilities using the sigmoid fitting method of [52]. Overall, we have found this method to work quite well on a variety of images, with a resulting class-normalized test classification accuracy obtained is 83.5%, and the full confusion matrix is shown in Table 1. Qualitative results are also shown in Fig. 3.

Now that we have a classifier that determines whether or not the sun is visible in the image $P(V|\mathcal{I})$, we consider how we can estimate the remaining factor in our representation (2): the distribution over sun positions S given that the sun is visible $P(S|V = 1, \mathcal{I})$.

4 Image cues for predicting the sun direction

When the sun is visible, its position affects different parts of the scene in very different ways. In our approach, information about the sun position is captured from four major parts of the image — the sky pixels $\mathcal{S}$, the ground pixels $\mathcal{G}$, the vertical surface pixels $\mathcal{V}$ and the pixels belonging to pedestrians $\mathcal{P}$ — via four cues. To partition the image in this way, we use the approach of Hoiem *et al.* [27], which returns a pixel-wise labeling of the image, together with the pedestrian detector of Felszenswalb *et al.* [18] which returns the bounding boxes of potential pedestrian locations. Both these de-

tectors also include a measure of confidence of their respective outputs.

We represent the sun position $S = \{\theta_s, \phi_s\}$ using two parameters: θ_s is the sun zenith angle, and ϕ_s the sun azimuth angle *with respect to the camera*. This section describes how we compute distributions over these parameters given the sky, the shadows, the shading on the vertical surfaces, and the detected pedestrians individually. Afterwards, in Sec. 5, we will see how to combine these cues to estimate the sun position given the entire image.

4.1 Sky

In order to estimate the sun position angles from the sky, we take inspiration from the work of Lalonde *et al.* [43], which shows that a physically-based sky model [51] 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 now dealing with a single image only. If we, for now, assume that the sky is completely clear, our solution is then 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 sun position, and are distributed according to the following generative model, function of the Perez sky model [51] $g(\cdot)$, the image coordinates (u_i, v_i) of s_i, and the camera focal length f_c and zenith angle (with respect to vertical) θ_c:

$$s_i \sim \mathcal{N}(k\, g(\theta_s, \phi_s, u_i, v_i, f_c, \theta_c), \sigma_s^2)\,, \tag{3}$$

where $\mathcal{N}(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2; and k is an unknown scale factor (see [43] for details). We obtain the distribution over sun positions by computing

$$P(\theta_s, \phi_s|\mathcal{S}) \propto \exp\left(\sum_{s_i \in \mathcal{S}} \frac{-(s_i - k\, g(\theta_s, \phi_s, ...))^2}{2\sigma_s^2}\right) \tag{4}$$

for each bin in the discrete (θ_s, ϕ_s) space, and normalizing appropriately. Note that since k in (3) and (4) is unknown, we first optimize for k for each sun position bin independently using a non-linear least-squares optimization scheme (see [43] for more details on the sky model).

As it is indicated in (3), the sky model $g(\cdot)$ also requires knowledge of two important camera parameters: its zenith angle θ_c and its focal length f_c. If we assume that f_c is available via the EXIF tag of the photograph[1], then θ_c can be computed by finding the horizon line v_h in the image (assuming the camera has no roll angle). 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. Note that all the results shown in this paper have been obtained using this approximation, which we have found to work quite well in practice. Fig. 4 demonstrates the distribution over sun positions obtained using the sky cue.

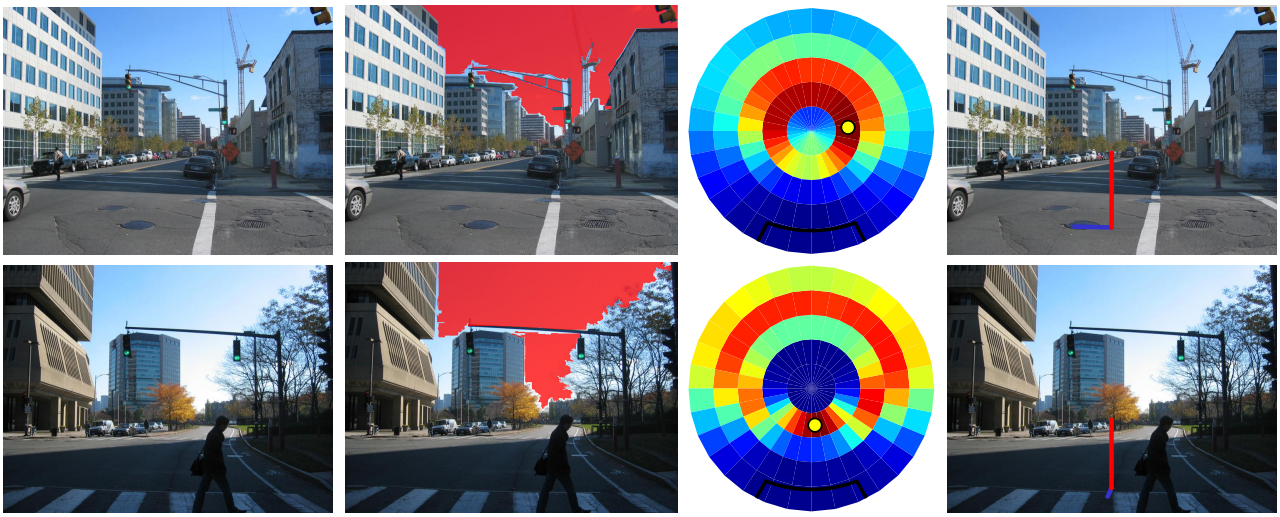
So far, we have assumed that the sky is completely clear, which might not always be the case! Even if the sun is shining on the scene, the sky might be covered with clouds, thus rendering the physical model $g(\cdot)$ useless. To deal with this problem, we classify the sky into one of three categories: clear, partially cloudy, or completely overcast. For this, we build a small database of representative skies for each category from images downloaded from Flickr, and compute the illumination context feature [42] 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 $P(\theta_s, \phi_s|\mathcal{S})$ is left uniform. For partly cloudy scenes, we remove the clouds with 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.

4.2 Ground shadows cast by vertical objects

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 [36]. 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 should be, in fact, produced by vertical things. Therefore, if we can detect a set of "long and strong" shadow lines (edges), we can use them in a probabilistic sense to determine a likely sun azimuth (up to the directional ambiguity). While shadows have also been used to estimate the sun direction with user input [35] or in webcams [31], no technique so far has proposed to do so automatically from a single image.

Most existing techniques for detecting shadows from a single image are based on computing illumination invariants that are physically-based and are functions of individual pixel values [22,20,21,47,66] or the values in a local image neighborhood [49]. Unfortunately, reliable computations of these invariants require high quality

[1] When unavailable in EXIF, the focal length f_c defaults to the equivalent of a horizontal field of view of 40° as in [26].



(a) Image and estimated horizon (b) Sky mask [27] (c) $P(\theta_s, \phi_s|\mathcal{S})$ (d) Inserted sun dial

Fig. 4: Illumination cue from the sky only. Starting from the input image (a), we compute the sky mask (b) using [27]. The resulting sky pixels are then used to estimate $P(\theta_s, \phi_s|\mathcal{S})$ (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).

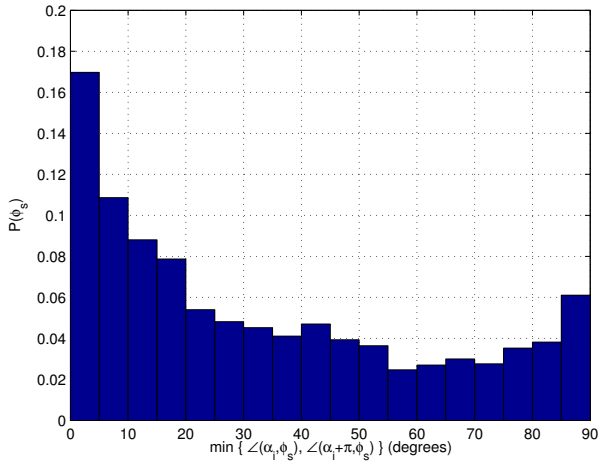
images with wide dynamic range, high intensity resolution and where the camera radiometry and color transformations are accurately measured and compensated for. Even slight perturbations (imperfections) in such images can cause the invariants to fail severely. Thus, they are ill-suited for the regular consumer-grade photographs such as those from Flickr and Google, that are noisy and often contain compression, resizing and aliasing artifacts, and effects due to automatic gain control and color balancing. Since much of current computer vision research is done on consumer photographs (and even worse-quality photos from the mobile phones), there is an acute need for a shadow detector that could work on such images. In this work, we use the shadow detection method we introduced in [40], which we summarize briefly here for completeness.

Our approach relies on a classifier trained to recognize ground shadow edges by using features computed over a local neighborhood around the edge. The features used are ratios of color intensities computed on both sides of each edge, in three color spaces (RGB, LAB, [9]), and at four different scales. As suggested by [72], we also employ a texture description of the regions on both sides of each edge. The feature distribution is estimated using a logistic regression version of Adaboost [10], with twenty 16-node decision trees as weak learners. This classification method provides good feature selection and outputs probabilities, and has been successfully used in a variety of other vision tasks [27,28]. To train the classifier, we introduced a novel dataset containing more than 130 images in which each shadow boundary on the ground has been manually labelled [41]. Finally, a Conditional Random Field (CRF) is used to obtain smoother shadow contours which lie on the ground [27].

From the resulting shadow boundaries, we detect the long shadow lines on the ground $l_i \in \mathcal{G}$ by applying the line detection algorithm of [37]. Let us consider one shadow line l_i. If its orientation on the ground plane is denoted α_i, then the angle between the shadow line orientation and the sun azimuth angle is

$$\angle(\alpha_i, \phi_s) = \min\left\{\angle(\alpha_i, \phi_s), \angle(\alpha_i + 180^\circ, \phi_s)\right\} , \qquad (5)$$

with the 180° ambiguity due to our assumption that we do not know which object is casting this shadow. Here where $\angle(\cdot,\cdot)$ denotes the angular difference. We obtain a non-parametric estimate for $P(\phi_s|\alpha_i)$ by detecting long lines on the ground truth shadow boundaries in 74 images from our shadow boundary dataset [41], in which we have manually labeled the sun azimuth. The distribution obtained from the resulting 1,700 shadow lines found is shown in Fig. 5a. The strongest peak, at $\angle(\alpha_i, \phi_s) < 5^\circ$, confirms our intuition that long shadow lines align with the sun direction. Another, smaller peak seems to rise at $\angle(\alpha_i, \phi_s) > 85^\circ$. This is explained by the roof lines of buildings, quite common in our database, which cast shadows that are perpendicular to the sun direction (Fig. 5b).



(a)

(b)

Fig. 5: How do shadow lines predict the sun azimuth? We use our shadow boundary dataset [41] to compute a non-parametric estimate for $P(\phi_s|\angle(\alpha_i, \phi_s))$ (a). A total of 1,700 shadow lines taken from 74 images were used to estimate the distribution. (b) Example image from the shadow dataset which contains shadows both aligned with and perpendicular to the sun (lamppost and roof line respectively). For visualization, the ground truth sun direction is indicated by the red stick and its shadow.

All the shadow lines are combined by making each one vote for its preferred sun direction:

$$P(\phi_s|\mathcal{G}) \propto \sum_{l_i \in \mathcal{G}} P(\phi_s|\alpha_i)\,. \tag{6}$$

Of course, computing the shadow line orientation α_i on the ground requires knowledge of the zenith angle θ_c and focal length f_c of the camera. For this we use the estimates obtained in the previous section. Fig. 6 illustrates the results obtained using the shadow cue only.

4.3 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 [27] and attempt to estimate the azimuth direction only.

Intuitively, we assume that a surface $w_i \in \mathcal{V}$ should predict that the sun lies in front of it if it is bright. On the contrary, the sun should be behind it if the surface is dark. We discover the mapping between the average brightness of the surface and the relative sun position with respect to the surface normal orientation β_i by manually labeling the sun azimuth in the Geometric Context dataset [27]. In particular, we learn the relationship between surface brightness b_i and whether the sun is in front of or behind the surface, i.e. whether $\angle(\beta_i, \phi_s)$ is less, or greater than 90°. This is done by computing the average brightness of all the vertical surfaces in the dataset, and applying logistic regression to learn $P(\angle(\beta_i, \phi_s|b_i) < 90°)$. This effectively models the probability in the following fashion:

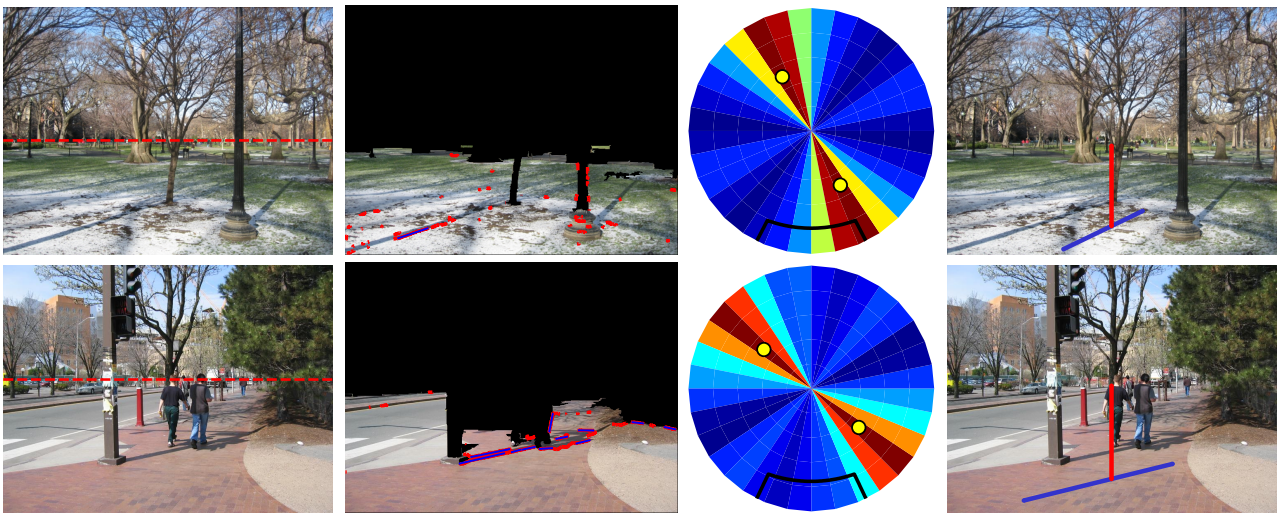
$$P(\angle(\beta_i, \phi_s) < 90°|b_i) = \frac{1}{1 + e^{-(x_1 + x_2 b_i)}}\,, \tag{7}$$

where, after learning, $x_1 = -3.35$ and $x_2 = 5.97$. Fig. 7 shows the computed data points and fitted model obtained with this method. As expected, a bright surface (high b_i) predicts that the sun is more likely to be in front of it than behind it, and vice-versa for dark surfaces (low b_i). In practice, even if the sun is shining on a surface, a large portion of it might be in shadows because of occlusions. Therefore, b_i is set to be the average brightness of the brightest cluster, computed using k-means with $k = 2$.

To model the distribution of the sun given a vertical surface of orientation β_i, we use:

$$P(\phi_s|w_i) \sim \begin{cases} \mathcal{N}(\beta_i, \sigma_w^2) & \text{if (7)} \geq 0.5 \\ \mathcal{N}(\beta_i + 180°, \sigma_w^2) & \text{if (7)} < 0.5 \end{cases}\,, \tag{8}$$

where σ_w^2 is such that the fraction of the mass of the gaussian $\mathcal{N}$ that lies in front of (behind) the surface corresponds to $P(\angle(\beta_i, \phi_s) < 90°|b_i)$ if it is greater (less)



(a) Image and estimated horizon (b) Ground [27] and shadows [40] (c) $P(\theta_s, \phi_s|\mathcal{G})$ (d) Inserted sun dial

Fig. 6: Illumination cue from the shadows only. Starting from the input image (a), we compute the ground mask (b) using [27], estimate shadow boundaries using [40] (shown in red), and finally extract long lines [37] (shown in blue). The long shadow lines are used to estimate $P(\theta_s, \phi_s|\mathcal{G})$ (c). Note that shadow lines alone can only predict the sun relative azimuth angle up to a 180° ambiguity. For visualization, the two most likely sun positions (shown with yellow circles) are used to artificially synthesize a sun dial in the scene (d).

than 0.5. Note that $\beta_i \in \{-90°, 90°, 180°\}$ since we assume only 3 coarse surface orientations. We combine each surface by making each one vote for its preferred sun direction:

$$P(\phi_s|\mathcal{V}) \propto \sum_{w_i \in \mathcal{V}} P(\phi_s|w_i)\,. \tag{9}$$

Fig. 8 shows the sun azimuth prediction results obtained using the shading on vertical surfaces only. We find that this cue can often help resolve the ambiguity arising with shadow lines.

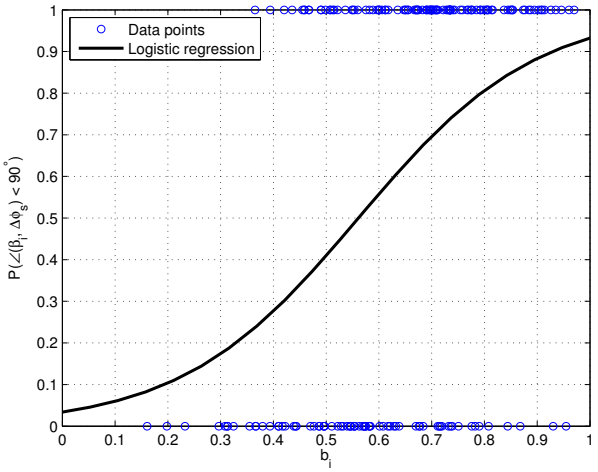
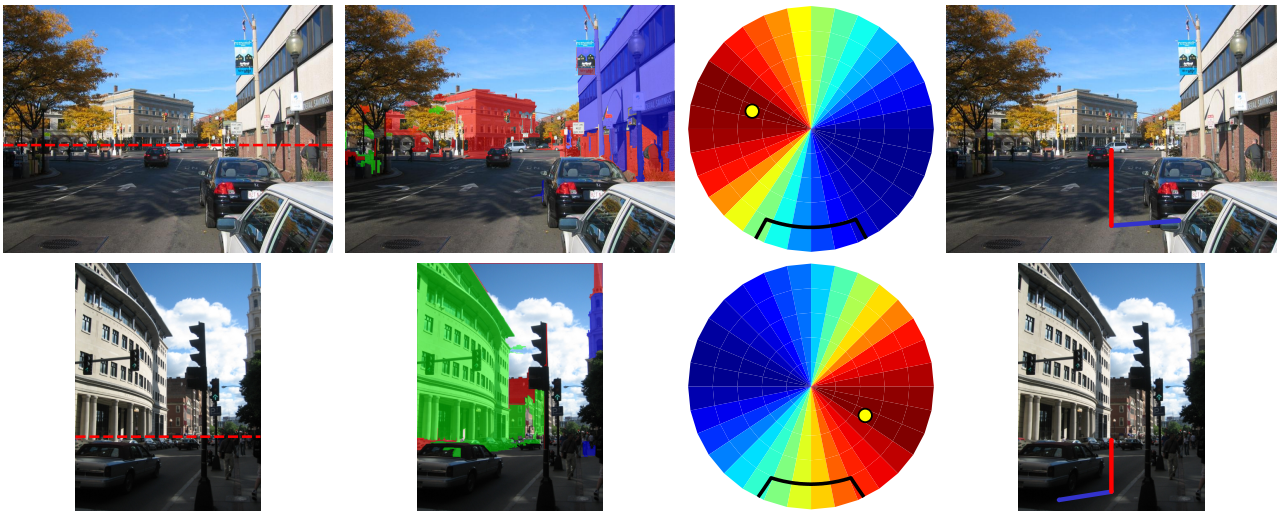


Fig. 7: How do the vertical surfaces predict the sun? From the Geometric Context dataset [27], the mapping between the probability of the sun being in front of a surface and its average brightness b_i is learned using logistic regression. Bright surfaces (high b_i) predict that the sun is in front of them (high probability), and vice-versa for dark surfaces (low b_i).

4.4 Pedestrians

When convex objects are present in the image, their appearance can also be used to predict where the sun is [44]. One notable example of this idea is the work of Bitouk *et al.* [3] which uses faces to recover illumination in a face swapping application. However, front-looking faces are rarely of sufficient resolution in the type of images we are considering (they were using mostly high resolution closeup portraits), but entire people more often are. As shown in Fig. 9, when shone upon by the sun, pedestrians also exhibit strong appearance characteristics that depend on the sun position: shadows are cast at the feet, a horizontal intensity gradient exists on the persons body, the wrinkles in the clothing are highlighted, etc. It is easy for us to say that one person is lit from the left (Fig. 9a), or from the right (Fig. 9b).

Because several applications require detecting pedestrians in an image (surveillance or safety applications



(a) Image and estimated horizon (b) Vertical surfaces [27] (c) $P(\theta_s, \phi_s|\mathcal{V})$ (d) Inserted sun dial

Fig. 8: Illumination cue from the vertical surfaces only. Starting from the input image (a), we compute the vertical surfaces mask (b) using [27] (blue = facing left, green = facing right, red = facing forward). The distribution of pixel intensities on each of these surfaces are then used to estimate $P(\theta_s, \phi_s|\mathcal{V})$ (c). Note that in our work, vertical surfaces cannot predict the sun zenith angle θ_s. In these examples, the sun zenith is set to 45° for visualization. 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).

(a) (b)

Fig. 9: Looking at these pedestrians, it is easy to say that (a) is lit from the left, and (b) from the right. In this work, we train a classifier that predicts the sun direction based on the appearance of pedestrians.

in particular), they are a class of objects that has received significant attention in the literature [11,16,50, 18], and as such many efficient detectors exist. In this section, we present a novel approach to estimate the azimuth angle of the sun with respect to the camera given detected pedestrian bounding boxes in an image. In our work, the pedestrians are detected using [18] which uses the standard Histogram of Gradients (HOG) features [11] for detection. The detector is operated at a high-precision, low-recall setting to ensure that only very confident detections are used.

We employ a supervised learning approach to the problem of predicting the sun location given the appearance of a pedestrian. In particular, a set of 2,000 random images were selected from the LabelMe dataset [58], which contain the ground truth location of pedestrians, and for which we also manually labelled the ground truth sun azimuth. To make the problem more tractable, the space of sun azimuths $\phi_s \in [-180°, 180°]$ is discretized into four bins of 90° intervals: $[-180°, -90°]$, $[-90°, 0°]$, and so forth. We then train a multi-class SVM classifier which uses the same HOG features used for detection. However, the difference now is that the classifier is *conditioned* on the presence of a pedestrian in the bounding box, so it effectively learns different feature weights that capture effects only due to illumination. In practice, the multi-class SVM is implemented as a set of four one-vs-all binary SVM classifiers. The SVM training is done with the libsvm library [7], and the output of each is normalized via a non-linear least-squares sigmoid fitting process [52] to obtain a probability for each class.

Of course, the illumination effects on pedestrians shown in Fig. 9 and captured by the classifier only

arise when a pedestrian is actually lit by the sun. However, since buildings or scene structures frequently cast large shadows upon the ground, pedestrians may very well be in the shade. To account for that, we train another binary SVM classifier to predict whether the pedestrian is in shadows or not. This binary classifier uses simple features computed on the bounding box region, such as coarse 5-bin histograms in the HSV and RGB colorspaces, and a histogram of oriented gradients within the entire bounding box to capture contrast. This simple "sunlit pedestrian" classifier results in more than 82% classification accuracy. Fig. 10 shows the sun azimuth estimation results obtained using the automatically-detected sunlit pedestrians only.

As with the previous cues, we compute the probability $P(\phi_s|p_i)$ of the sun azimuth given a single pedestrian p_i by making each one vote for its preferred sun direction:

$$P(\phi_s|\mathcal{P}) \propto \sum_{p_i \in \mathcal{P}} P(\phi_s|p_i) \tag{10}$$

5 Estimating the sun position

Now that we are equipped with several features that can be computed over an image, we show how we combine them in order to get a more reliable estimate. Because each cue can be very weak and might not even be available in any given image, we combine them in a probabilistic framework that captures the uncertainty associated with each of them.

5.1 Cue combination

We are interested in estimating the distribution $P(S|V = 1, \mathcal{I})$ over the sun position $S = \{\theta_s, \phi_s\}$, given the entire image $\mathcal{I}$ and assuming the sun is visible (see Sec. 3). We saw in the previous section that the image $\mathcal{I}$ is divided into features computed on the sky $\mathcal{S}$, the shadows on the ground $\mathcal{G}$, the vertical surfaces $\mathcal{V}$ and pedestrians $\mathcal{P}$, so we apply Bayes rule and write

$$P(S|\mathcal{S}, \mathcal{G}, \mathcal{V}, \mathcal{P}) \propto P(\mathcal{S}, \mathcal{G}, \mathcal{V}, \mathcal{P}|S)P(S)\,. \tag{11}$$

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

$$P(S|\mathcal{S}, \mathcal{G}, \mathcal{V}, \mathcal{P}) \propto P(S|\mathcal{S})P(S|\mathcal{G})P(S|\mathcal{V})P(S|\mathcal{P})P(S)\,. \tag{12}$$

The process of combining the cues according to (12) for the sun position is illustrated in Fig. 11. We have presented how we compute the conditionals $P(S|\mathcal{S})$, $P(S|\mathcal{G})$, $P(S|\mathcal{V})$, and $P(S|\mathcal{P})$ in Sec. 4.1, 4.2, 4.3 and 4.4 respectively. We now look at how we can compute the prior $P(S)$ on the sun position themselves.

5.2 Data-driven illumination prior

The prior $P(S) = P(\theta_s, \phi_s)$ captures the typical sun positions in outdoor scenes. We now proceed to show how we can compute it given a large dataset of consumer photographs.

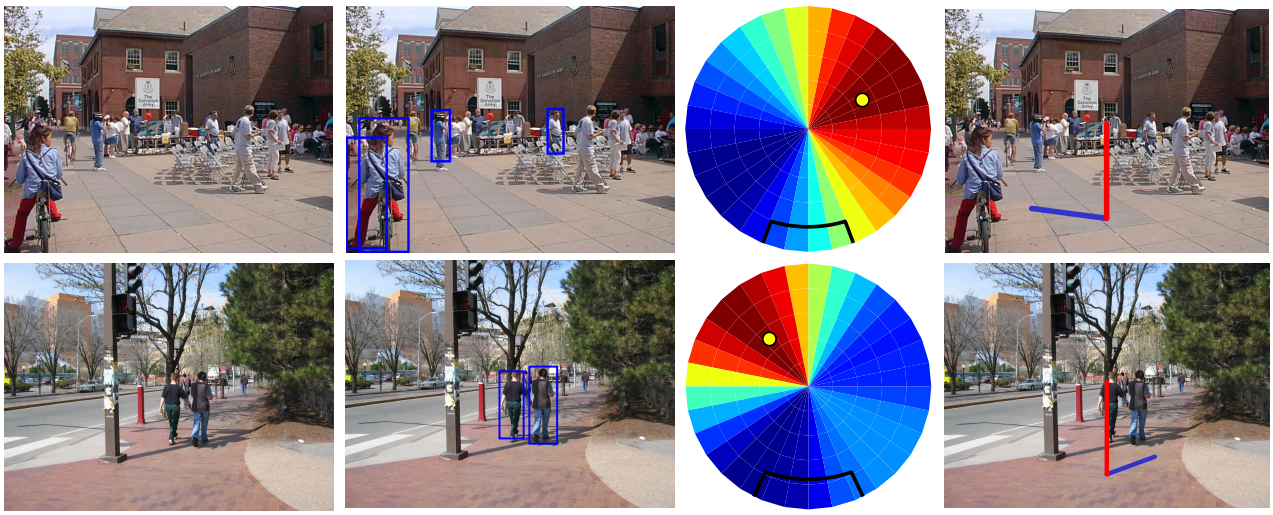
The sun position (θ_s, ϕ_s) depends on the latitude L of the camera, its azimuth angle ϕ_c, the date D and the time of day T expressed in the local timezone:

$$P(\theta_s, \phi_s) = P(f(L, D, T, \phi_c))\,, \tag{13}$$

where $f(\cdot)$ is a non-linear function defined in [55]. To estimate (13), we can sample points from $P(L, D, T, \phi_c)$, and use $f(\cdot)$ to recover θ_s and ϕ_s. But estimating this distribution is not currently feasible since it requires images with known camera orientations ϕ_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 [23] is used to compute the empirical distribution $P(L, D, T)$. We compute (13) 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 $[-180°, 180°]$ interval. As a consequence, (13) is flat along the ϕ_s dimension and is marginalized.

Fig. 12c shows 4 estimates for $P(\theta_s, \phi_s)$, 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 (θ_s close to 90°). 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 Figs. 12a and 12b). 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



(a) Image (b) Detected pedestrians [18] (c) $P(\theta_s, \phi_s|\mathcal{P})$ (d) Inserted sun dial

Fig. 10: Illumination cue from the pedestrians only. Starting from the input image (a), we detect pedestrians using the detector of [18] (b). Each of these detections is used to estimated $P(\phi_s|\mathcal{P})$ (c) by running our pedestrian-based illlumination predictor, and combining all predictions via a voting framework. The most likely sun azimuth (shown with a yellow circle) is used to artificially synthesize a sun dial in the scene (d). For display purposes, the most likely sun zenith is set to be 45°.

sensors such as compasses on digital cameras, this data will surely become available in the near future.

6 Evaluation and results

We evaluate our technique in three different ways. First, we quantitatively evaluate our sun estimation technique for both zenith and azimuth angles using images taken from calibrated webcam sequences. Second, we show another quantitative evaluation this time for the sun azimuth only, but performed on a dataset of manually labelled single images downloaded from Internet. Finally, we also provide several qualitative results on single images that demonstrate the performance of our approach. These results will show that, although far from being perfect, our approach is still able to exploit visible illumination cues to reason about the sun.

6.1 Quantitative evaluation using webcams

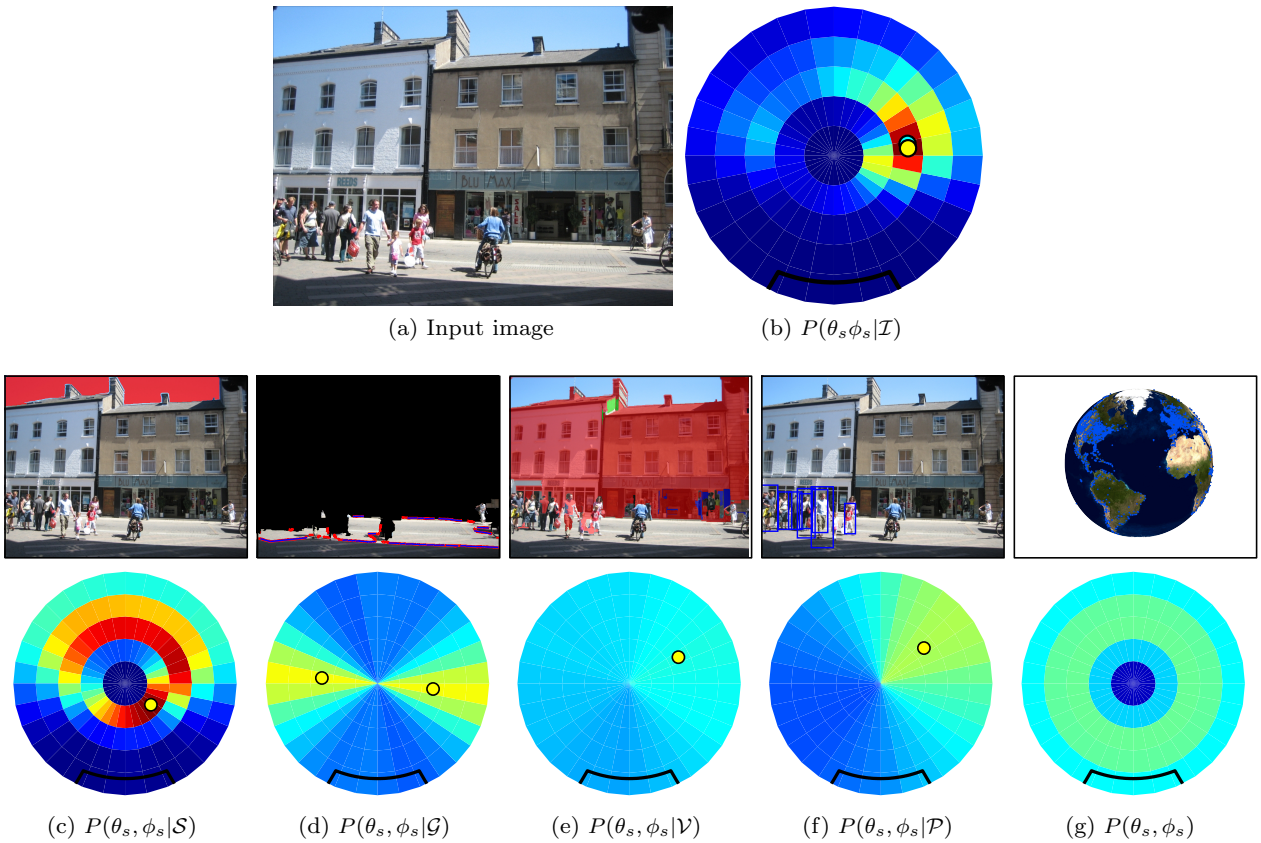
We use the technique of Lalonde *et al.* [43] to estimate the positions of the sun in 984 images taken from 15 different time-lapse image sequences, downloaded from the Internet. Two example images from our dataset are shown in Fig. 13a and 13b. Our algorithm is applied to every image from the sequences independently and the results are compared against ground truth.

Fig. 13 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. 5.2 (we tested both the data-driven and Earth uniform priors), scene cues only, and using our combined measure $P(\theta_s, \phi_s|\mathcal{I})$ with both priors as well. Fig. 13 highlights the performance at errors of less than 22.5° (50% of images) and 45° (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. 13a is not informative, since it is small and the sun is always behind the camera. It is more informative in Fig. 13b, as it occupies a larger area. Pedestrians may appear in Fig. 13a and may be useful; however they are too small in Fig. 13b to be of any use.

6.2 Quantitative azimuth evaluation on single images

When browsing popular publicly available webcam datasets [39, 29], one quickly realizes that the types of scenes captured in such sequences are inherently different than those found in single images. Webcams are typically installed at high vantage points, giving them a broad overlook on large-scale panoramas such as natural or urban landscapes. On the other hand, single images are



(a) Input image (b) $P(\theta_s \phi_s | \mathcal{I})$

(c) $P(\theta_s, \phi_s | \mathcal{S})$ (d) $P(\theta_s, \phi_s | \mathcal{G})$ (e) $P(\theta_s, \phi_s | \mathcal{V})$ (f) $P(\theta_s, \phi_s | \mathcal{P})$ (g) $P(\theta_s, \phi_s)$

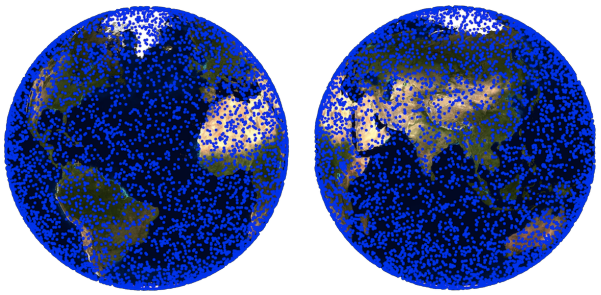
Fig. 11: Combining illumination features computed from image (a) yields a more confident final estimate (b). We show how (c) through (f) are estimated in Sec. 4, and how we compute (g) in Sec. 5.2. To visualize the relative confidence between the different cues, all probability maps in (c) through (g) are drawn on the same color scale.

most often taken at human eye level, where earthbound "objects" like cars, pedestrians, bushes or trees, occupy a much larger fraction of the image.

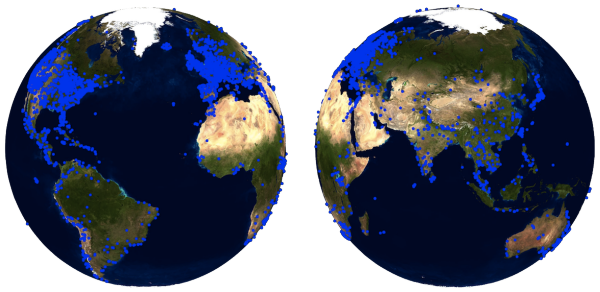
In addition, the scene in a webcam sequence does not change over time (considering static cameras only), it is the illumination conditions that vary. In single images, however, both the scenes *and* illumination conditions differ from one image to the next. In this section, we evaluate our method on a dataset of single images. Admittedly, this task is much harder than the case of webcams, but we still expect to be able to extract meaningful information across a wide variety of urban and natural scenes.

The main challenge we face here is the unavailability of ground truth sun positions in single images: as of this day, there exist no such publicly-available dataset. Additionally, as discussed in Sec. 5.2, while the GPS coordinates and time of capture of images are commonly recorded by modern-day cameras, the camera azimuth angle ϕ_c is not. We randomly selected 300 outdoor images from the LabelMe dataset [58] where the sun appears to be shining on the scene (i.e. is not occluded by a cloud or very large building), and manually labelled the sun position in each one of them. The labeling was done using an interactive graphical interface that resembled the "virtual sun dials" used throughout this paper (see Fig. 5b for example), where the task of the human labeler was to orient the shadow so that it aligned with the perceived sun direction. If the labeler judged that he or she cannot identify the sun direction with sufficient accuracy in a given image, that image was discarded from the dataset. After the labeling process, 239 images were retained for evaluation. In addition, we have found that reliably labeling the sun zenith angle θ_s is very hard to do in the absence of objects of known heights [36], so we ask the user to label the sun azimuth ϕ_s only.

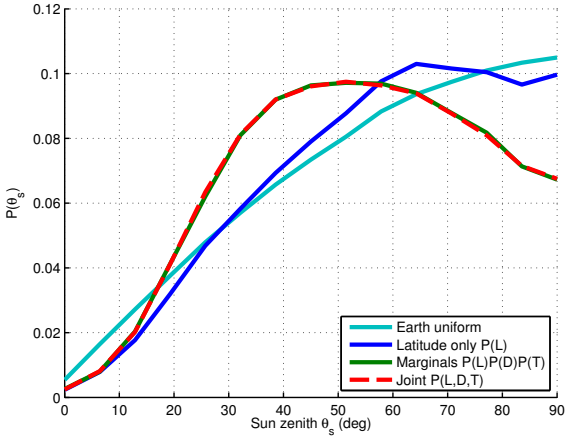
Fig. 14 reports the cumulative histograms of errors in sun azimuth estimation for each of the cues independently (Figs. 14a through 14d), and jointly (Fig. 14e). We also report the percentage of images which have less than 22.5° and 45° errors, corresponding to cor-



(a) Uniform

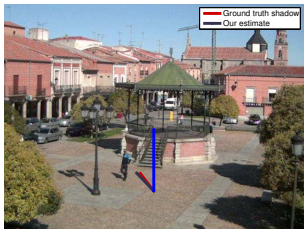


(b) Data-driven



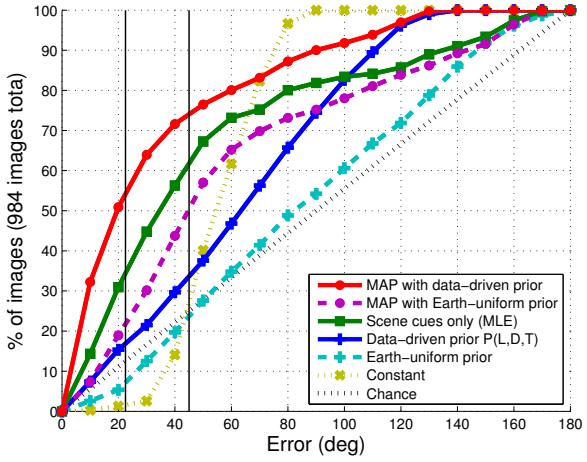
(c)

Fig. 12: Illumination priors (13) on the sun zenith angle. The Earth visualizations illustrate the (a) uniform and (b) data-driven sampling over GPS locations that are compared in this work. The data-driven prior is learned from a database of 6M images downloaded from Flickr. (c) 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.



(a) Madrid sequence

(b) Vatican sequence



(c)

Fig. 13: Quantitative evaluation using 984 images taken from 15 webcam sequences, calibrated using [43]. 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°.

rectly estimating the sun azimuth within an octant and quadrant, respectively. Since each cue is not available in all images, we also report the percentage of images for which each cue is available.

Let us point out a few noteworthy observations. First, since the shadow lines alone (Fig. 14b) have a 180° ambiguity in azimuth estimation, therefore we keep the minimum error between each predicted (opposite) direction and the ground truth sun azimuth. This explains why the minimum error is 90°, and as such should not be compared directly with the other cues. A second

note is that the vertical surfaces (Fig. 14c) are available in 99% of the images in our dataset, which indicates that our dataset might be somewhat biased towards urban environments, much like LabelMe currently is. Also observe how the pedestrians are, independently, one of the best cues when available (Fig. 14d). Their performance is the best of all cues. As a side note, we also observed that there is a direct correlation between the number of pedestrians visible in the image and the quality of the azimuth estimate: the more the better. For example, the mean error is 41° when there is only one pedestrian, 22° with two, and 20° with three or more. Finally, the overall system (Fig. 14e) estimates the sun azimuth angle within 22.5° for 40.2% of the images, and within 45° for 54.8% of the images, which is better than any of the cues taken independently.

6.3 Qualitative evaluation on single images

Fig. 15 shows several example results of applying our algorithm on typical consumer-grade images taken from our test set introduced in Sec. 6.2. The rows are arranged in order of decreasing order of estimation error. Notice how our technique recovers the entire *distribution* over the sun parameters (θ_s, ϕ_s) which also captures the degree of confidence in the estimates. High confidence cases are usually obtained when one cue is very strong (i.e. large intensity difference between vertical surfaces of different orientations in the 5th row, 2nd column), or when all 4 cues are correlated (as in Fig. 11). On the other hand, highly cluttered scenes with few vertical surfaces, shadowed pedestrians, or shadows cast by vegetation (4th row, 2nd column of Fig. 15) usually yield lower confidences, and the most likely sun positions might be uncertain.

Fig. 16 shows typical failure cases. The assumption that most shadows are cast by vertical objects from (see Sec. 4.2) is not always satistifed (Fig. 16a). In Fig. 16b, the shadows are correctly detected, but the other cues fail to resolve the ambiguity in their orientation. Misdetections, whether for pedestrians (Fig. 16c) or vertical surfaces (Fig. 16e) may also cause problems. In Fig. 16d, the "sunlit pedestrian" classifier (Sec. 4.4) incorrectly identifies a shadowed pedestrian as being in the sunlight, thus yielding erroneous sun direction estimates. Finally, light clouds can mistakenly be interpreted as being intensity variations in the sky (Fig. 16f).

6.4 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* to light virtual objects [13]. An environment map is a sample of the plenoptic function at a single point in space capturing the full sphere of light rays incident at that point. It is typically captured by either taking a high dynamic range (HDR) panoramic photograph from the point of view of the object, or by placing a mirrored ball at the desired location and photographing it. Such an environment map can then be used as an extended light source for rendering synthetic objects.

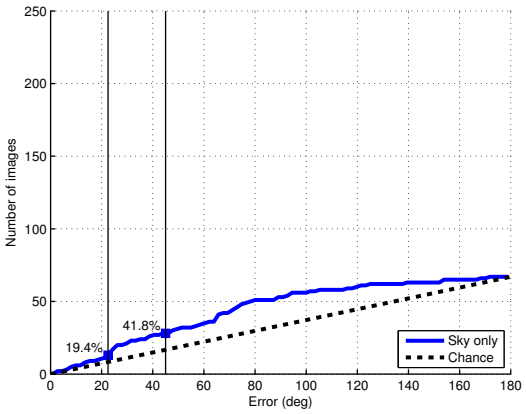
Given only an image, it is generally impossible to recover the true environment map of the scene, since the image will only contain a small visible portion of the full map (and from the wrong viewpoint besides). We propose to use our model of natural illumination to estimate a high dynamic range environment map from a single image. Since we are dealing with outdoor images, we can divide the environment map into two parts: the sky probe and the scene probe. We now detail the process of building a realistic approximation to the real environment map for these two parts.

The sky probe can be generated by using the physically-based sky model $g(\cdot)$ from Sec. 4.1. 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.* [43] to recover the most likely sky appearance by fitting its turbidity t (a scalar value which captures the degree of scattering in the atmosphere) in a least-squares fashion. This is done by taking

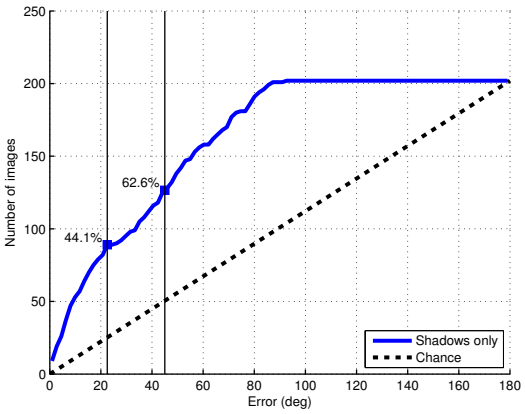
$$t^* = \arg\min_t \sum_{s_i \in \mathcal{S}'} \left(s_i - k g(\theta_s, \phi_s, t, \ldots)\right)^2 \quad , \tag{14}$$

where $\mathcal{S}'$ are clear sky pixels. The estimated turbidity t^* and camera parameters f_c and θ_c allow us to use $g(\cdot)$ and extrapolate the sky appearance on the entire hemisphere, even though only a small portion of it was originally visible to the camera. The sun is simulated by a bright circular patch (10^4 times brighter than the maximum scene brightness). An example sky probe obtained using this technique is shown in Fig. 17b.

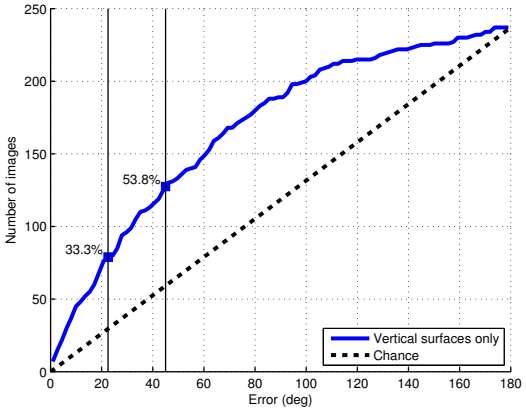
For the bottom part of the environment map (not shown), we use the spherical projection technique of Khan *et al.* [34] on the pixels below the horizon line as in [39]. A realistic 3-D model is relit using an off-the-shelf rendering software (see Fig. 17c). Notice how the shadows on the ground, and shading and reflections on the car are consistent with the image. Another example is shown in Fig. 1.



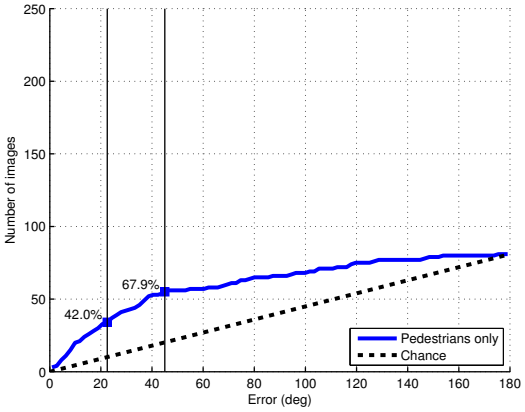
(a) Sky only (28% of images)



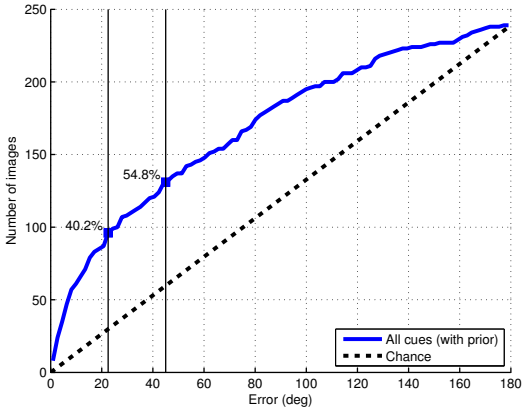
(b) Shadows only (86% of images)



(c) Vertical surfaces only (99% of images)



(d) Pedestrians only (34% of images))



(e) All cues (with prior, 100% of images)

Fig. 14: Cumulative sun azimuth error (angle between estimate and ground truth directions) on a set of 239 single images taken from the LabelMe dataset [58], for the individual cues presented in Sec. 4: (a) the sky, (b) the shadows cast on the ground by vertical objects, (c) the vertical surfaces, and (d) the pedestrians. The plot in (e) shows the result obtained by combining all cues together with the sun prior, as discussed in Sec. 5. The percentages indicated represent the fraction of images in the test set for which the cue is available.

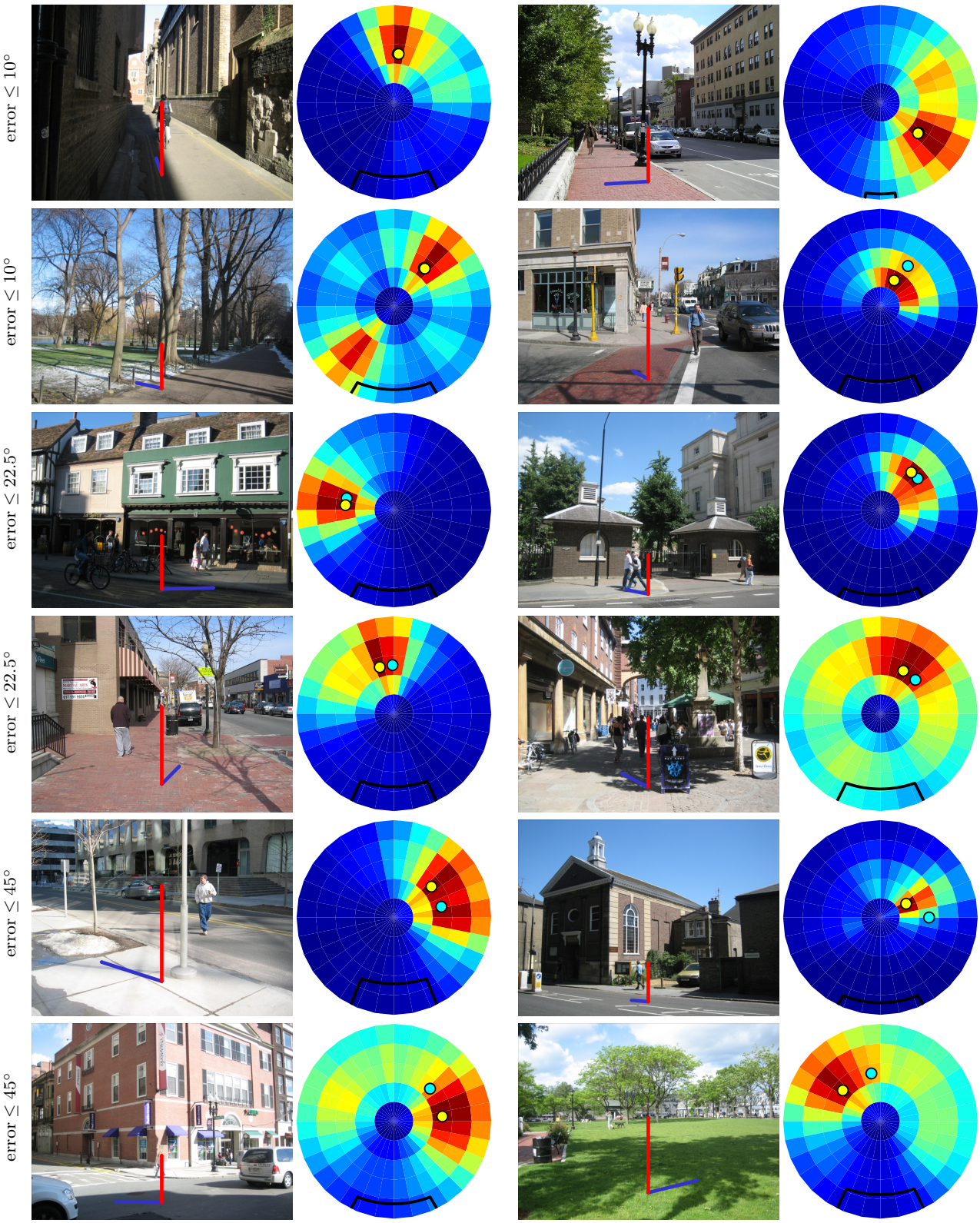
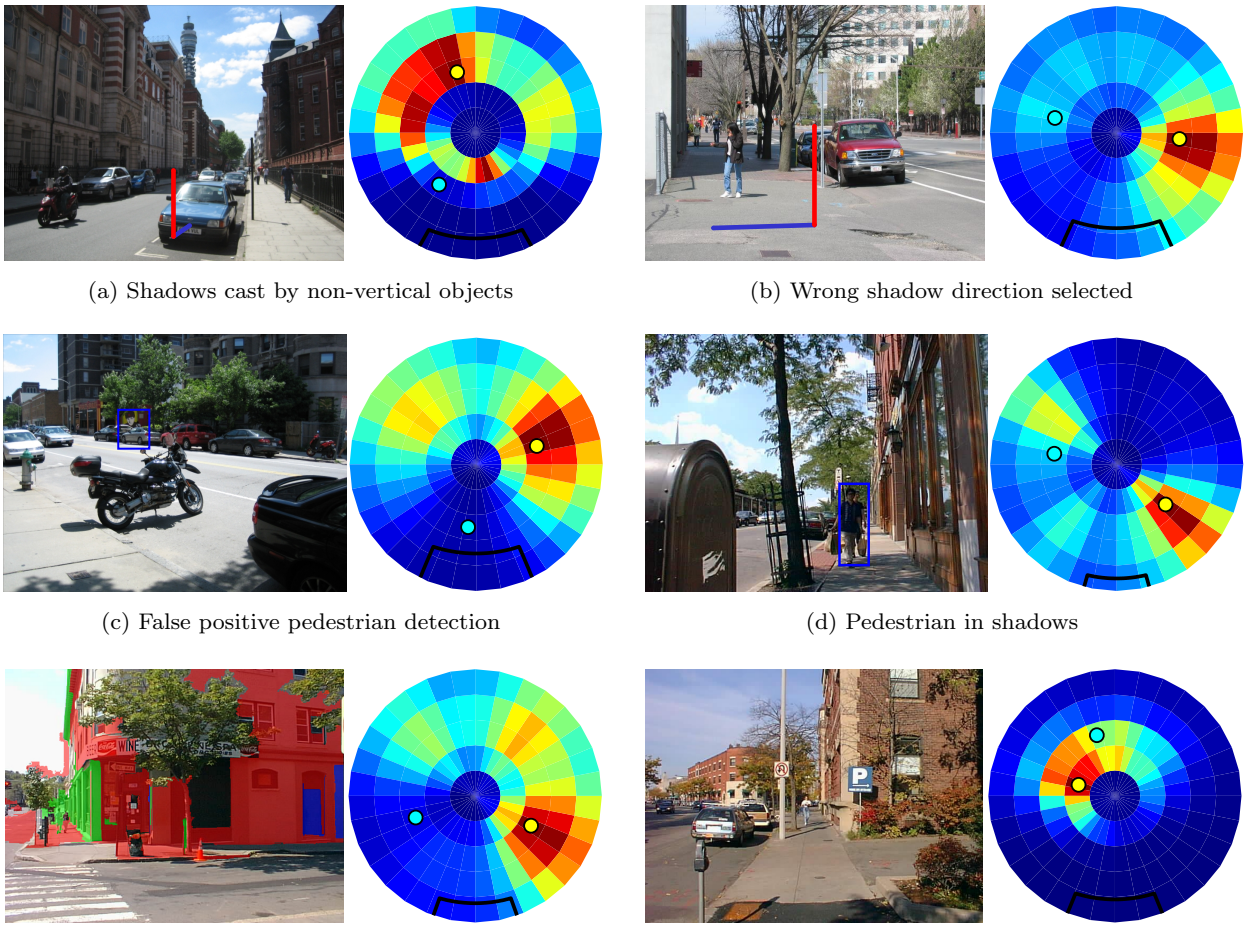


Fig. 15: Sun direction estimation from a single image. A virtual sun dial is inserted in each input image (first and third columns), whose shadow correspond to the MAP sun position in the corresponding probability maps $P(\theta_s, \phi_s|\mathcal{I})$ (second and fourth columns). The ground truth sun azimuth is shown in cyan, and since it is not available, a zenith angle of 45° is used for visualization.



(a) Shadows cast by non-vertical objects

(b) Wrong shadow direction selected

(c) False positive pedestrian detection

(d) Pedestrian in shadows

(e) Mislabeled vertical surfaces

(f) Light clouds in the sky

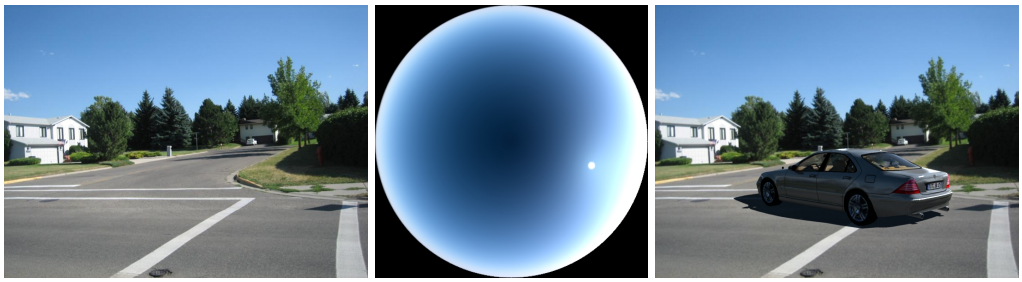
Fig. 16: Typical failure cases. (a) First, the dominating shadow lines are not cast by thin vertical objects, and are therefore not aligned with sun direction. (b) Mislabeling the vertical surfaces orientation causes the wrong shadow direction to be selected. Note that the pedestrian visible in this example was not a high-confidence detection, thus unused for the illumination estimate. (c) A false positive pedestrian detection causes the illumination predictor to fail. (d) A pedestrian is correctly detected, but incorrectly classified as being "in the sun". (e) Vertical surface (red) incorrectly encompasses left-facing side wall, leading the classifier to believe the sun is behind the camera. (f) Light clouds mistaken for clear sky variation. In all cases, the other cues were not confident enough to compensate, thus yielding erroneous estimates.

7 Discussion

We now discuss three important aspects of our approach: the distribution over the sun positions, how we can exploit additional sources of information when available, and higher-order interactions between cues.

7.1 Distribution Over Sun Positions

The quantitative evaluation presented in Sec. 6.2 revealed that our method successfully predicts the sun azimuth within 22.5° for 40% of the images in our test set, within 45° for 55% of them, and within 90° for 80% (see Fig. 14e). Admittedly, this is far from perfect. However, we believe this is still a very useful result for applications which might not require a very precise estimate. For instance, all that might be required in some cases is to correctly identify the sun quadrant ($< 45°$ error), or distinguigh between left and right, or front and back ($< 90°$ error). In these cases, our method obtains very reasonable results. But these most likely sun positions used to generate these results actually hide an important result of our method which is not captured by this evaluation.



(a) (b) (c)

Fig. 17: 3-D object relighting. From a single image (a), we render the most likely sky appearance (b) using the sun position computed with our method, and then fitting the sky parameters using [43]. We can realistically insert a 3-D object into the image (c).

While it makes intuitive sense to report the results in terms of angular errors in sun position estimation, the real "output" of our system is the *probability distribution* over the sun position. Throughout this paper, we have displayed those using colored circles, indicating the likelihood of the sun being at each position, but without paying much attention to them. A closer look at the distributions themselves reveal very interesting observations about the degree of certainty of the estimate, which itself should be a very useful result.

Fig. 18 shows examples of common scenarios that arise in practice. In Fig. 18a, we observe that strong cues in the scene—bright surfaces, strong cast shadows, visible illumination effects on pedestrians—result in an estimate that is peaked around a sun position which aligns well with the ground truth. When the scene is cluttered as in Fig. 18b, the cues become harder to detect, and the resulting estimate is less confident, as evidenced by the bluer colors in the probability map (all the colors in the figure are on the same scale, making it easy to compare them).

When there are strong cast shadows but the other cues are weaker, the shadow ambiguity remains present in the sun probability distribution as in Figs 18c and 18d. Finally, when the cues are altogether too difficult to detect or simply uninformative as in Figs 18e and 18f, the resulting estimate is of more or less constant probability. In this case, the maximum likelihood sun position, used to generate the quantitative evaluation plots of Sec. 6.2 and the virtual sun dials used for visualization, is meaningless. A more representative way of evaluating the results could employ a measure of the confidence of the distribution (e.g., variance).

7.2 Complementary Sources of Information

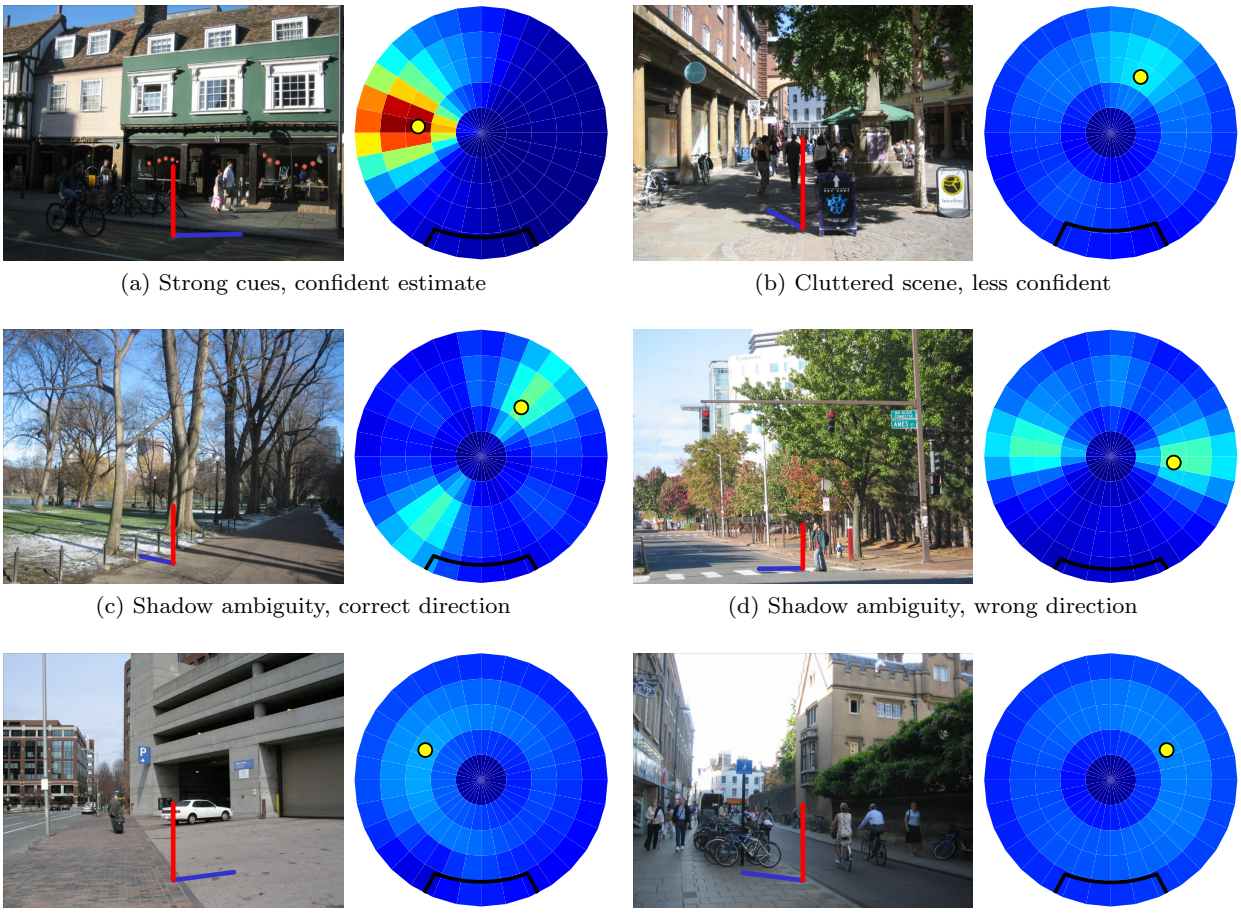
It is sometimes the case that additional sources of information about the camera are available. For example, the EXIF header in image files commonly contain information like the focal length (used in this work), the date and time of capture of the image, and the GPS location. We now discuss what the availability of this information means in the context of our work.

If the date and time of capture of the image as well as the GPS location are available, then it is possible to compute the zenith angle of the sun by using atmospherical formulas [55]. Since the camera azimuth is unknown, this effectively restricts the sun to be in a band of constant zenith, so the probability of the sun to be anywhere else can readily be set to zero.

Recent smartphone models now sport an additional sensor which readings are also available via EXIF: a digital compass. This compass records the absolute orientation of the camera, and has been shown to be useful in rudimentary augmented reality applications. In recording the camera azimuth angle, we now have everything we need to actually compute the sun position with respect to the camera. Of course, this does not indicate whether the sun is visible or not (Sec. 3), nor does it provide information about the weather conditions, but could be a tremendous tool to capture datasets of images with ground truth sun positions. The availability of large amounts of images with annotated information will surely play an important role in improving our understanding of the illumination in real outdoor images.

7.3 Higher-order Interactions Between Cues

In Sec. 5, we saw how we can combine together the predictions from multiple cues to obtain a final, more confident estimate. To combine the cues, we rely on the



(a) Strong cues, confident estimate

(b) Cluttered scene, less confident

(c) Shadow ambiguity, correct direction

(d) Shadow ambiguity, wrong direction

(e) Weak cues, low confidence

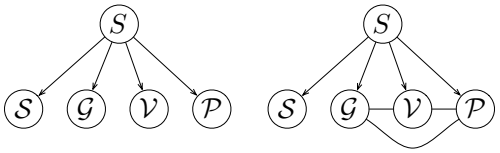
(f) Weak cues, very low confidence

Fig. 18: Different scenarios result in different confidences in the illumination estimate. When the cues are strong and extracted properly, the resulting estimate is highly confident (a). A scene with more clutter typically results in lower confidence (b). The ambiguity created by shadows is sometimes visible when the other cues are weaker (c)–(d). When the cues are so weak (e.g., no bright vertical surfaces, no strong cast shadows, pedestrians in shadows, etc.), the resulting estimate is not confident, and the maximum likelihood sun position is meaningless (e)–(f). All probability maps are drawn on the same color scale, so confidences can be compared directly.

Naive Bayes assumption, which states that all cues *conditionally* independent given the sun direction. While this conditional independence assumption makes intuitive sense for many cues—for example, the appearance of the sky is independent from the shadow directions on the ground if we know the sun direction—it does not apply for all cues. Here we discuss a few dependencies that arise in our framework, and how we could leverage them to better capture interactions across cues.

Even if the sun direction is known, there is still a strong dependency between objects and their shadows. Knowing the position of vertical objects (e.g., pedestrians) tell us that cast shadows should be near their point of contact with the ground. Similarly, knowing the location of shadow boundaries on the ground constrain the possible locations of pedestrians, since they must cast a shadow (if they are in sunlight). Capturing this interaction between pedestrians and shadows would be very beneficial: since we know pedestrians are vertical objects, simply finding their shadows would be enough to get a good estimate of the sun position, and the ambiguity in direction from Sec. 4.2 could even be resolved.

There is also a dependency, albeit a potentially weaker one, between vertical surfaces and cast shadows on the ground. The top, horizontal edge of vertical surfaces (e.g., roof of buildings) also cast shadows on the ground. Reasoning about the interaction between buildings and shadows would allow us to discard their shadows, which typically point away from the sun (Sec. 4.2).



(a) Naive Bayes model (b) Higher-order model

Fig. 19: Capturing higher-level interactions across cues. The current approach uses a Naive Bayes model (a), which assumes that all cues are conditionally independent given the sun position. Capturing higher-order interactions across cues would require a more complex model (b), with less restrictive independence assumptions..

Another interesting cross-cue dependency arises between pedestrians and vertical surfaces. Since large surfaces may create large shadow regions, if the sun comes from behind a large wall, and a pedestrian is close to that wall, it is likely that this pedestrian is in shadows, therefore unpredictive of the sun position.

Capturing these higher-level interactions across cues, while beneficial, would also increase the complexity in the probabilistic model used to solve the problem. Fig. 19 shows a comparison between the graphical model that corresponds to our current approach (Fig. 19a) and a new one that would capture these dependencies (Fig. 19b). The caveat here is that the complexity of the model is exponential in the clique size, which is determined by the number of cues in the image (e.g., number of shadow lines, number of pedestrians, etc). Learning and inference in such a model will certainly be more challenging.

8 Conclusion

Outdoor illumination affects the appearances of scenes in complex ways. Untangling illumination from surface 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 sun visibility and, if deemed visible, its position relative to the camera. 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 a Microsoft Fellowship to J.-F. Lalonde, and 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. Parts of the results presented in this paper have previously appeared in [38].

References

1. R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal on Computer Vision*, 72(3):239–257, May 2007. 3
2. R. E. Bird. A simple spectral model for direct normal and diffuse horizontal irradiance. *Solar Energy*, 32:461–471, 1984. 4
3. D. Bitouk, N. Kumar, S. Dhillon, P. N. Belhumeur, and S. K. Nayar. Face swapping: automatically replacing faces in photographs. *ACM Transactions on Graphics (SIGGRAPH 2008)*, 27(3), 2008. 11
4. J. F. Blinn and M. E. Newell. Texture and reflection in computer generated images. In *Proceedings of ACM SIGGRAPH 1976*, 1976. 5
5. S. D. Buluswar and B. A. Draper. Color models for outdoor machine vision. *Computer Vision and Image Understanding*, 85(2):71–99, 2002. 5
6. P. Cavanagh. The artist as neuroscientist. *Nature*, 434:301–307, 2005. 2
7. C.-C. Chang and C.-J. Lin. Libsvm: a library for support vector machines. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`, 2001. 7, 12
8. H. Chen, P. Belhumeur, and D. Jacobs. In search of illumination invariants. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2000. 2
9. H. Y. Chong, S. J. Gortler, and T. Zickler. A perception-based color space for illumination-invariant image processing. *ACM Transactions on Graphics (SIGGRAPH 2008)*, 2008. 9
10. M. Collins, R. Shapire, and Y. Singer. Logistic regression, adaboost and Bregman distances. *Machine Learning*, 48(1), 2002. 9
11. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005. 2, 12
12. K. Dale, M. K. Johnson, K. Sunkavalli, W. Matusik, and H. Pfister. Image restoration using online photo collections. In *International Conference on Computer Vision*, 2009. 7
13. P. Debevec. Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of ACM SIGGRAPH 1998*, 1998. 5, 17

14. P. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of ACM SIGGRAPH 1997*, August 1997. 5
15. P. Debevec, C. Tchou, A. Gardner, T. Hawkins, C. Poullis, J. Stumpfel, A. Jones, N. Yun, P. Einarsson, T. Lundgren, M. Fajardo, and P. Martinez. Estimating surface reflectance properties of a complex scene under captured natural illumination. Technical Report ICT-TR-06.2004, USC ICT, 2004. 5
16. P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: a benchmark. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009. 12
17. R. O. Dror, A. S. Willsky, and E. H. Adelson. Statistical characterization of real-world illumination. *Journal of Vision*, 4:821–837, 2004. 5
18. P. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010. 7, 12, 14
19. G. Finlayson, M. Drew, and B. Funt. Diagonal transforms suffice for color constancy. In *International Conference on Computer Vision*, 1993. 3
20. G. D. Finlayson, M. S. Drew, and C. Lu. Intrinsic images by entropy minimization. In *European Conference on Computer Vision*, 2004. 8
21. G. D. Finlayson, C. Fredembach, and M. S. Drew. Detecting illumination in images. In *IEEE International Conference on Computer Vision*, 2007. 8
22. G. D. Finlayson, S. D. Hordley, and M. S. Drew. Removing shadows from images. In *European Conference on Computer Vision*, 2002. 4, 8
23. J. Hays and A. A. Efros. im2gps: estimating geographic information from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008. 13
24. G. Healey and D. Slater. Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6):2706–2717, November 1999. 4
25. R. Hill. Theory of geolocation by light levels. In B. J. LeBouef and R. M. Laws, editors, *Elephant Seals: Population Ecology, Behavior, and Physiology*, chapter 12, pages 227–236. University of California Press, 1994. 4
26. D. Hoiem, A. A. Efros, and M. Hebert. Automatic photo pop-up. *ACM Transactions on Graphics (SIGGRAPH 2005)*, 24(3), August 2005. 8
27. D. Hoiem, A. A. Efros, and M. Hebert. Recovering surface layout from an image. *International Journal of Computer Vision*, 75(1):151–172, October 2007. 4, 6, 7, 9, 10, 11, 12
28. D. Hoiem, A. Stein, A. A. Efros, and M. Hebert. Recovering occlusion boundaries from a single image. In *IEEE International Conference on Computer Vision*, 2007. 9
29. N. Jacobs, N. Roman, and R. Pless. Consistent temporal variations in many outdoor scenes. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007. 14
30. D. B. Judd, D. L. Macadam, G. Wyszecki, H. W. Budde, H. R. Condit, S. T. Henderson, and J. L. Simonds. Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am. A*, 54(8):1031–1036, 1964. 5
31. I. N. Junejo and H. Foroosh. Estimating geo-temporal location of stationary cameras using shadow trajectories. In *European Conference on Computer Vision*, 2008. 4, 8
32. F. Kasten and A. T. Young. Revised optical air mass tables and approximation formula. *Applied optics*, 28, 1989. 4
33. Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2006. 7
34. E. A. Khan, E. Reinhard, R. Fleming, and H. Büelthoff. Image-based material editing. *ACM Transactions on Graphics (ACM SIGGRAPH 2006)*, August 2006. 17
35. T. Kim and K.-S. Hong. A practical single image based approach for estimating illumination distribution from shadows. In *IEEE International Conference on Computer Vision*, 2005. 4, 8
36. J. J. Koenderink, A. J. van Doorn, and S. C. Pont. Light direction from shad(ow)ed random gaussian surfaces. *Perception*, 33(12):1405–1420, 2004. 8, 15
37. J. Košecká and W. Zhang. Video compass. In *European Conference on Computer Vision*, 2002. 9, 11
38. J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan. Estimating natural illumination from a single outdoor image. In *IEEE International Conference on Computer Vision*, 2009. 23
39. J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan. Webcam clip art: Appearance and illuminant transfer from time-lapse sequences. *ACM Transactions on Graphics (SIGGRAPH Asia 2009)*, 28(5), December 2009. 14, 17
40. J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan. Detecting ground shadows in outdoor consumer photographs. In *European Conference on Computer Vision*, 2010. 7, 9, 11
41. J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan. Ground shadow boundary dataset. `http://graphics.cs.cmu.edu/projects/shadows`, September 2010. 9, 10
42. J.-F. Lalonde, D. Hoiem, A. A. Efros, C. Rother, J. Winn, and A. Criminisi. Photo clip art. *ACM Transactions on Graphics (SIGGRAPH 2007)*, 2007. 4, 8
43. J.-F. Lalonde, S. G. Narasimhan, and A. A. Efros. What do the sun and the sky tell us about the camera? *International Journal on Computer Vision*, 88(1):24–51, May 2010. 4, 8, 14, 16, 17, 21
44. M. S. Langer and H. H. Büelthoff. A prior for global convexity in local shape-from-shading. *Perception*, 30(4):403–410, 2001. 11
45. Y. Li, S. Lin, H. Lu, and H.-Y. Shum. Multiple-cue illumination estimation in textured scenes. In *IEEE International Conference on Computer Vision*, 2003. 4
46. R. Manduchi. Learning outdoor color classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1713–1723, November 2006. 5
47. B. A. Maxwell, R. M. Friedhoff, and C. A. Smith. A bi-illuminant dichromatic reflection model for understanding images. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008. 5, 8
48. D. Mills. Advances in solar thermal electricity and technology. *Solar Energy*, 76:19–31, January 2004. 4
49. S. G. Narasimhan, V. Ramesh, and S. K. Nayar. A class of photometric invariants: Separating material from shape and illumination. In *IEEE International Conference on Computer Vision*, 2005. 8
50. D. Park, D. Ramanan, and C. C. Fowlkes. Multiresolution models for object detection. In *European Conference on Computer Vision*, 2010. 12
51. 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. 4, 8
52. J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, 1999. 7, 12

53. A. J. Preetham, P. Shirley, and B. Smits. A practical analytic model for daylight. In *Proceedings of ACM SIGGRAPH 1999*, August 1999. 4
54. R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. In *Proceedings of ACM SIGGRAPH 2001*, 2001. 5
55. I. Reda and A. Andreas. Solar position algorithm for solar radiation applications. Technical Report NREL/TP-560-34302, National Renewable Energy Laboratory, November 2005. 13, 21
56. C. F. Reinhart, J. Mardaljevic, and Z. Rogers. Dynamic daylight performance metrics for sustainable building design. *Leukos*, 3(1):1–25, 2006. 4
57. F. Romeiro and T. Zickler. Blind reflectometry. In *European Conference on Computer Vision*, 2010. 5
58. B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. LabelMe: a database and web-based tool for image annotation. *International Journal of Computer Vision*, 77(1-3), 2008. 7, 12, 15, 18
59. I. Sato, Y. Sato, and K. Ikeuchi. Illumination from shadows. *IEEE Transactions on Pattern Matching and Machine Intelligence*, 25(3):290–300, March 2003. 3
60. Y. Sato and K. Ikeuchi. Reflectance analysis under solar illumination. In *Proceedings of the IEEE Workshop on Physics-Based Modeling and Computer Vision*, pages 180–187, 1995. 4
61. D. Slater and G. Healey. Analyzing the spectral dimensionality of outdoor visible and near-infrared illumination functions. *Journal of the Optical Society of America*, 15(11):2913–2920, November 1998. 4, 5
62. C. Stauffer. Adaptive background mixture models for real-time tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1999. 2
63. J. Stumpfel, A. Jones, A. Wenger, C. Tchou, T. Hawkins, and P. Debevec. Direct HDR capture of the sun and sky. In *Proceedings of AFRIGRAPH*, 2004. 3, 5
64. M. Sun, G. Schindler, G. Turk, and F. Dellaert. Color matching and illumination estimation for urban scenes. In *IEEE International Workshop on 3-D Digital Imaging and Modeling*, 2009. 3
65. K. Sunkavalli, F. Romeiro, W. Matusik, T. Zickler, and H. Pfister. What do color changes reveal about an outdoor scene? In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008. 4, 5
66. J. Tian, J. Sun, and Y. Tang. Tricolor attenuation model for shadow detection. *IEEE Transactions on Image Processing*, 18(10), October 2009. 8
67. Y. Tsin, R. T. Collins, V. Ramesh, and T. Kanade. Bayesian color constancy for outdoor object recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2001. 5
68. G. Ward. The RADIANCE lighting simulation and rendering system. In *Proceedings of ACM SIGGRAPH 1994*, 1994. 4
69. Y. Weiss. Deriving intrinsic images from image sequences. In *IEEE International Conference on Computer Vision*, 2001. 4
70. T.-P. Wu and C.-K. Tang. A bayesian approach for shadow extraction from a single image. In *IEEE International Conference on Computer Vision*, 2005. 4
71. Y. Yu and J. Malik. Recovering photometric properties of architectural scenes from photographs. In *Proceedings of ACM SIGGRAPH 1998*, July 1998. 4
72. J. Zhu, K. G. G. Samuel, S. Z. Masood, and M. F. Tappen. Learning to recognize shadows in monochromatic natural images. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 9
73. T. Zickler, S. P. Mallick, D. J. Kriegman, and P. N. Belhumeur. Color subspaces as photometric invariants. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2006. 3