ENGN 8501

Physics based Vision-2:

Lighting Estimation and Relighting

Administrative matter

This week is Week #6 (i.e., half way through this semester).

Project proposal due: this Thursday.

Next two weeks is "teaching break", no lectures. However, it is not "study break", and not "holidays". Please make the best use of the two lecture-free weeks working on your research project, and reports.

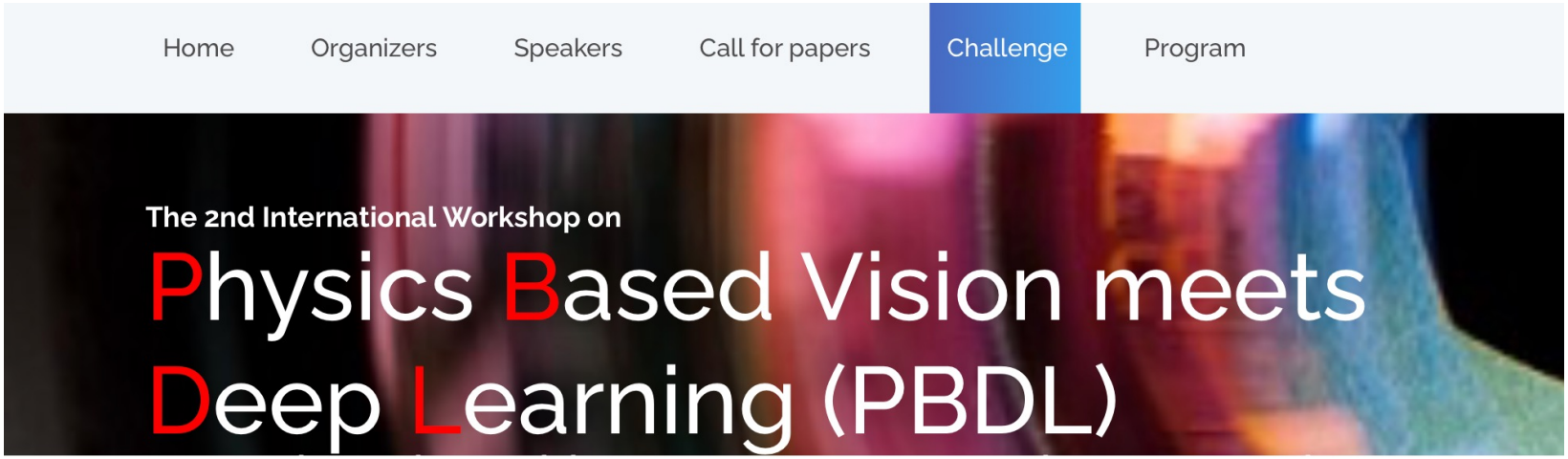
Project Proposal

Your project proposal must be no more than one-page A4 in PDF (with references inclusive), and must contain the following contents:

- Team members (student names, uni IDs)
- Paper Title.
- Project aims and main method.
- Project timeline and tentative work-load plan (i.e., who will be mainly working on which parts/aspects of the project).
- Relevant key bibliographic references.

NOTE, VERY IMPORTANT: you must submit your project proposal individually to Wattle, despite all team members of your team will necessarily have the identical proposal. This is because Wattle's marking system only allows individual marking. If Wattle does not receive your submission, the "marking" button won't show up.

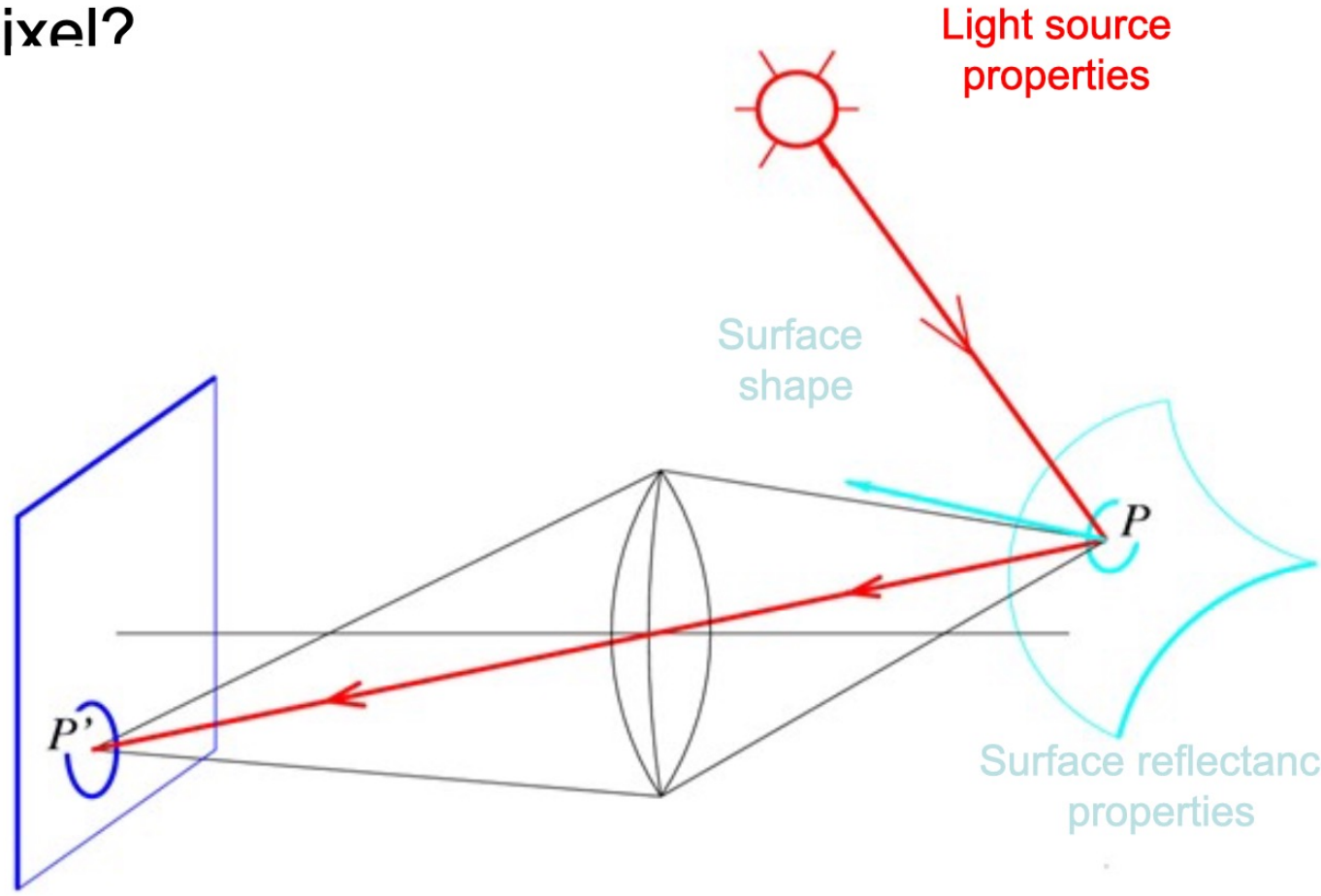
What is physics based vision ?



Light traveling in the 3D world interacts with the scene through intricate processes before being captured by a camera. These processes result in the dazzling effects like color and shading, complex surface and material appearance, different weathering, just to name a few. Physics based vision aims to invert the processes to recover the scene properties, such as shape, reflectance, light distribution, medium properties, etc., from the images by modeling and analysing the imaging process to extract desired features or information.

Last Week: Physics based image formation

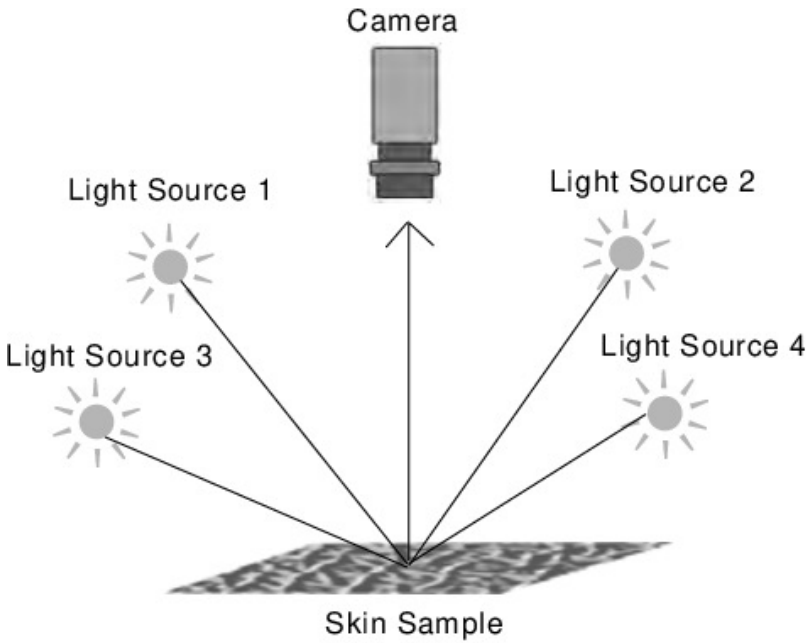
- What determines the brightness of an image pixel?



Last Week: Photometric Stereo

Key Idea: use pixel brightness changes to understand shape

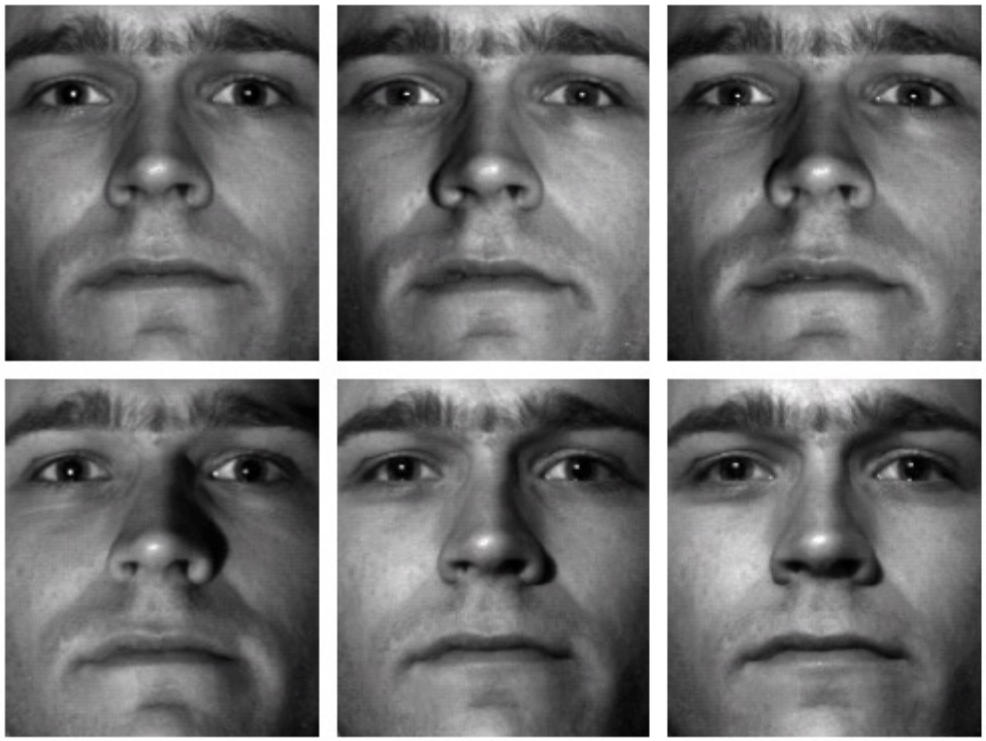
Photometric Stereo



(a)

(b)

Photometric Stereo

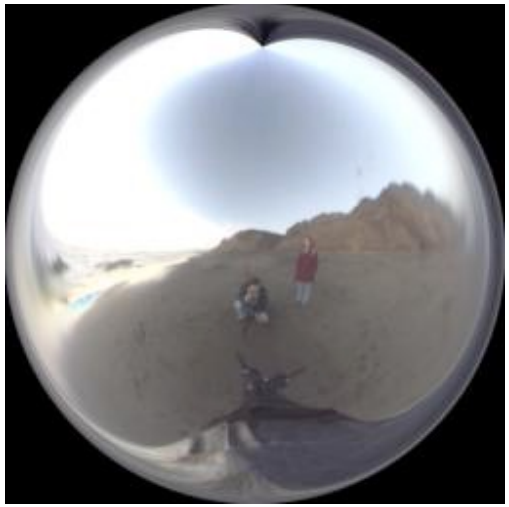


from Athos Georghiades

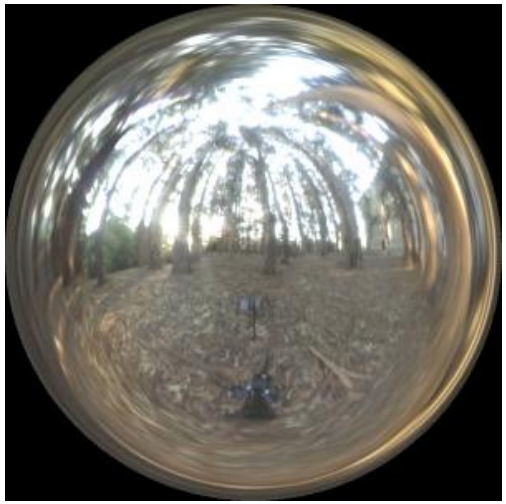
http://cvc.yale.edu/people/Athos.html

Real-World Lighting Environments

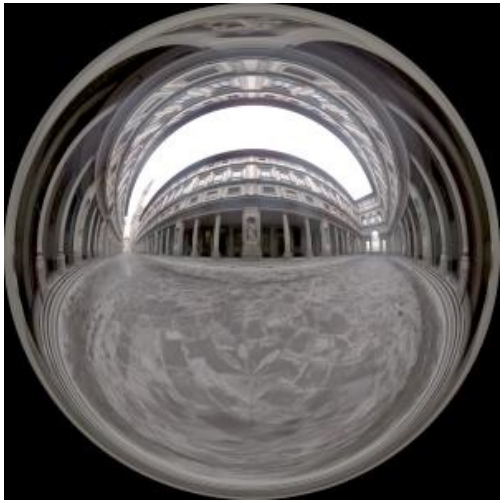
Funston Beach



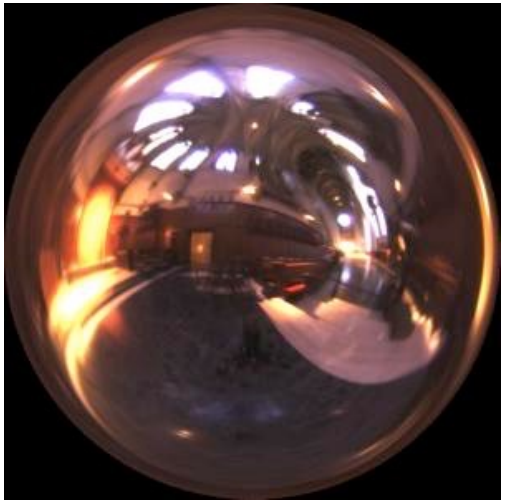
Eucalyptus Grove



Uffizi Gallery



Grace Cathedral



Lighting Environments from the Light Probe Image Gallery:
http://www.debevec.org/Probes/

Mirrored Sphere

Synthetic images

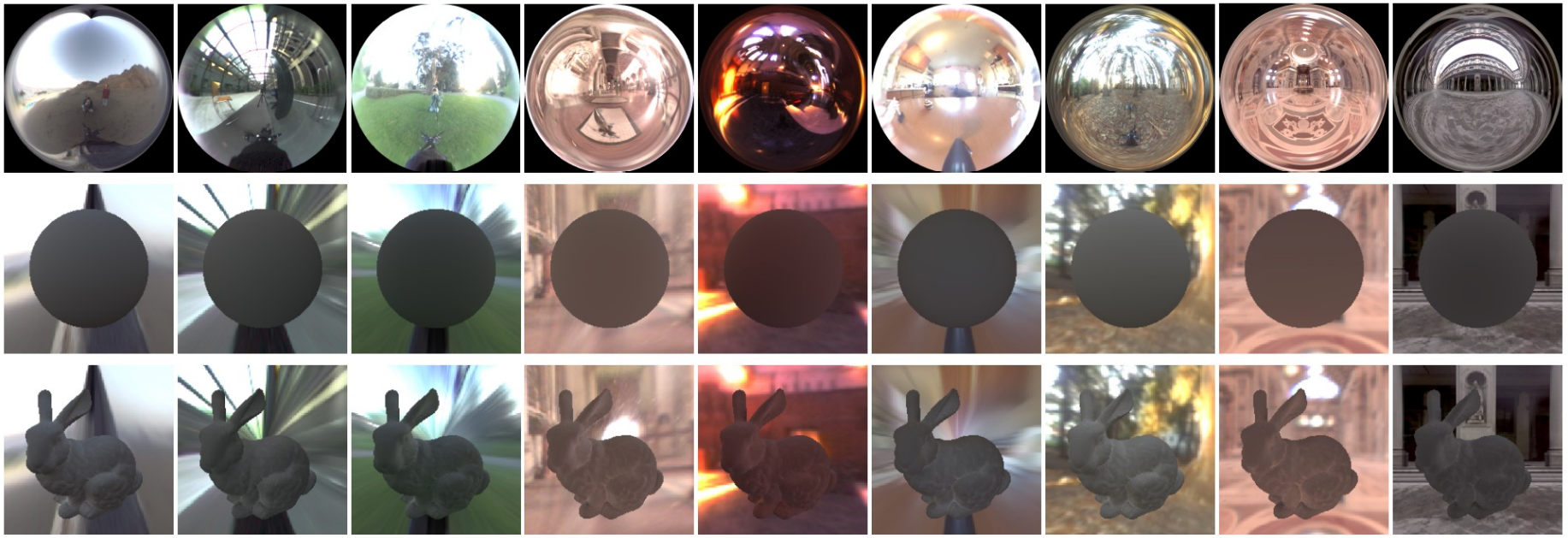
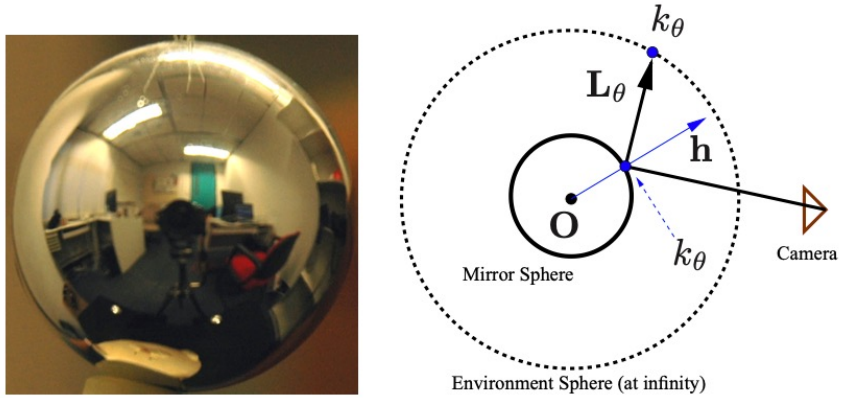


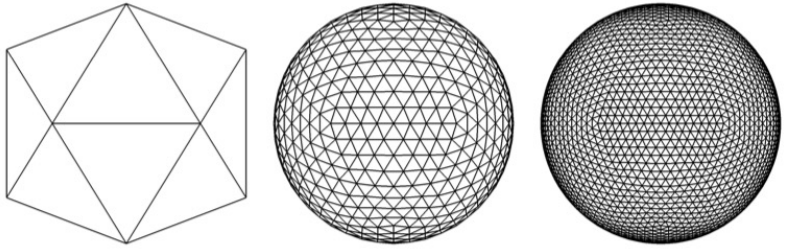
Figure 5. Input environments and images for the synthetic examples SPHERE and BUNNY.

Environment light map



(a) (b)

Figure 2. A mirror sphere for capturing environment lighting. (b) shows the relationship between the mirror sphere and environment sphere, where $\mathbf{O}$ is the *ideal* location where the object is placed, $\mathbf{h}$ is the angle bisector between the lighting direction $\mathbf{L}_\theta$ and viewing direction. k_θ is the incident intensity along $\mathbf{L}_\theta$.



(a) (b) (c)

Figure 3. Icosahedron and its subdivision. (a) A 20-face polyhedron, where the vertices are evenly distributed on a 3D unit sphere that encloses and touches the polyhedron. (b) A 3-time subdivided icosahedron. (c) A 4-time subdivided icosahedron. Each subdivision is done by splitting each face into four equilateral triangles followed by reprojecting the vertices onto a unit 3D sphere.

Tan [21] is derived based on cosine kernels and coefficients

Outdoor illumination estimation:

What if the photos were already taken, and there was not a light-probing mirror ball in the scene ?

Another motivation:

Natural Image Matting

The world is your green screen

Background Matting: The World is Your Green Screen

We capture 2 images, with and without the subject. | Captured Image and Background | Predicted Alpha Matte | Composite Image

By Soumyadip Sengupta, Vivek Jayaram, Brian Curless, Steve Seitz, and Ira Kemelmacher-Shlizerman

This paper will be presented in IEEE CVPR 2020.

Google’s Pixel Phone: matting result

Google's **Pixel** Phone: image composition

Google's Photo Relighting

(a) Input image and estimated lighting

(b) Rendered images from our method under three novel illuminations

Fig. 1. Given only a single input image taken with a standard cellphone camera of a portrait (a), our model is able to quickly (160 ms.) generate new images of our human subject as though they are illuminated under new, previously-unseen lighting environments (b).

Papers to read today

1. **Estimating Natural Illumination from a Single Outdoor Image**

Jean-François Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan
School of Computer Science, Carnegie Mellon University

2. **Deep Outdoor Illumination Estimation**

Yannick Hold-Geoffroy[1*], Kalyan Sunkavalli[†], Sunil Hadap[†], Emiliano Gambaretto[†], Jean-François Lalonde[*]
Université Laval[*], Adobe Research[†]

3. Single Image Portrait Relighting

TIANCHENG SUN, University of California, San Diego
JONATHAN T. BARRON and YUN-TA TSAI, Google Research
ZEXIANG XU, University of California, San Diego
XUEMING YU, GRAHAM FYFFE, CHRISTOPH RHEMANN, JAY BUSCH, and PAUL DEBEVEC, Google
RAVI RAMAMOORTHI, University of California, San Diego

Paper#1:

Estimating Natural Illumination from a Single Outdoor Image

Jean-François Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan
School of Computer Science, Carnegie Mellon University
http://graphics.cs.cmu.edu/projects/outdoorIllumination

Abstract

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

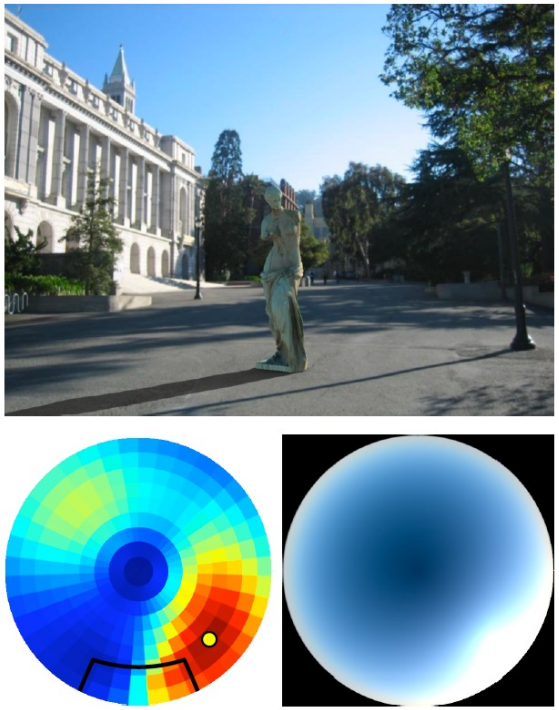


Figure 1. A synthetic 3-D statue has been placed in a photograph (top) in an illumination-consistent way. To enable this type of operation, we develop an approach that uses information from the sky, the shading and the shadows to estimate a distribution on the likely sun positions (bottom-left), and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

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 atmo-*

Image clues for illumination estimation

Sky region

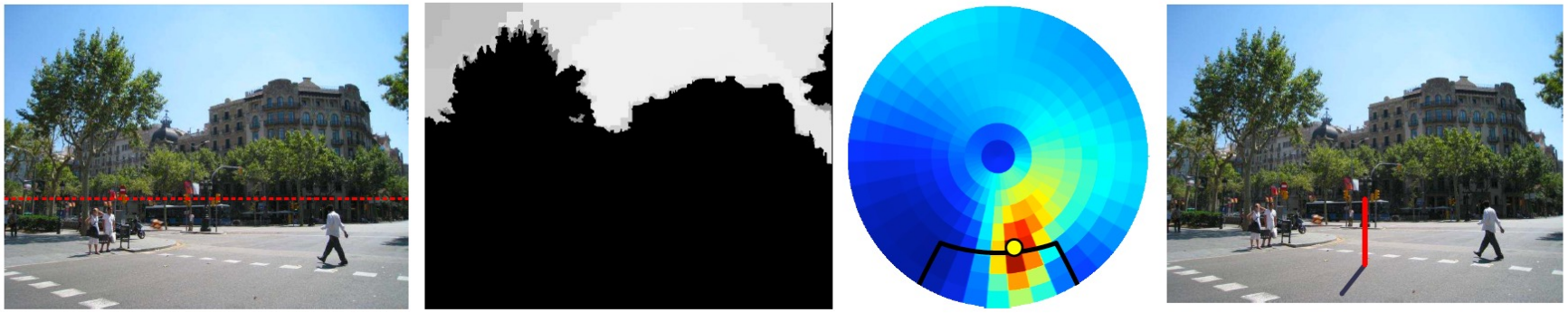
Shadow on ground

Building facade shading

Image region segmentation via the “geometric context” method (ICCV 2005)

Figure 1: Geometric context from a single image: ground (green), sky (blue), vertical regions (red) subdivided into planar orientations (arrows) and non-planar solid (’x’) and porous (’o’).

Sky clue



(a) Image and estimated horizon (b) Sky mask [9] (c) $P(\theta_s, \Delta\phi_s|\mathcal{S})$ (d) Inserted sun dial

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

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

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

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

Shadow clue

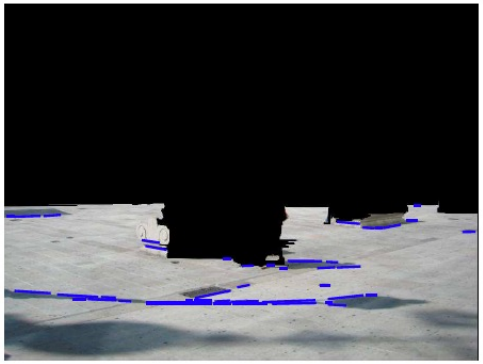
Given a potential shadow line $l_i \in \mathcal{G}$, we compute its relative orientation α_i on the ground plane by rectifying it via a homography using the camera parameters estimated in the previous section. We assume that each shadow line predicts the sun azimuth in the following way:

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

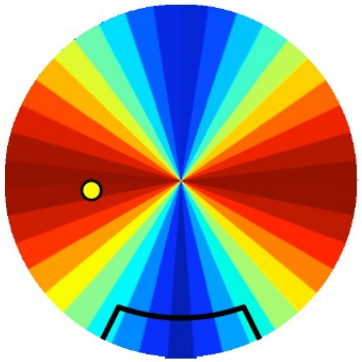
and combine all the shadow lines by making each one vote for its preferred shadow direction:

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

(a) Image and estimated horizon



(b) Ground mask [9] and shadow lines

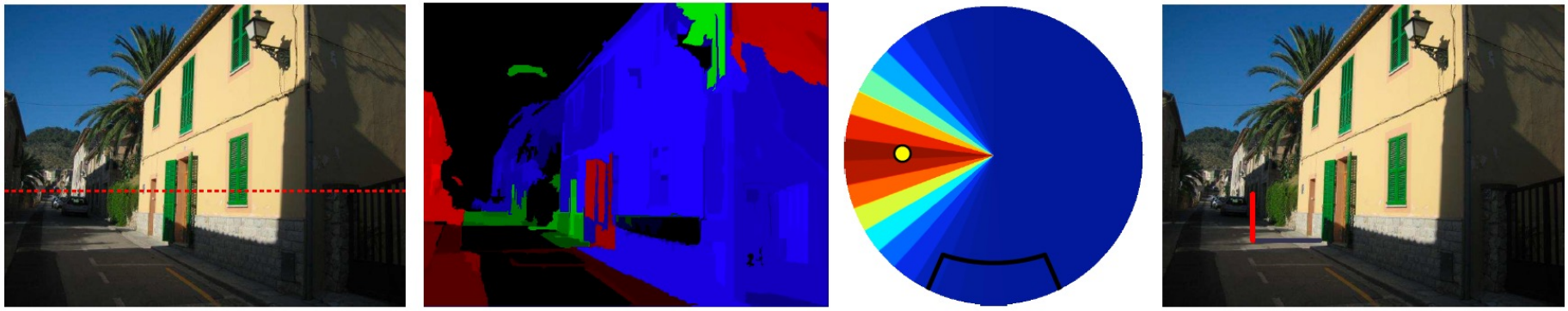


(c) $P(\theta_s, \Delta\phi_s|\mathcal{G})$

(d) Inserted sun dial

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

Building clue, and combine everything



(a) Image and estimated horizon (b) Vertical mask [9] (c) $P(\theta_s, \Delta\phi_s|\mathcal{V})$ (d) Inserted sun dial

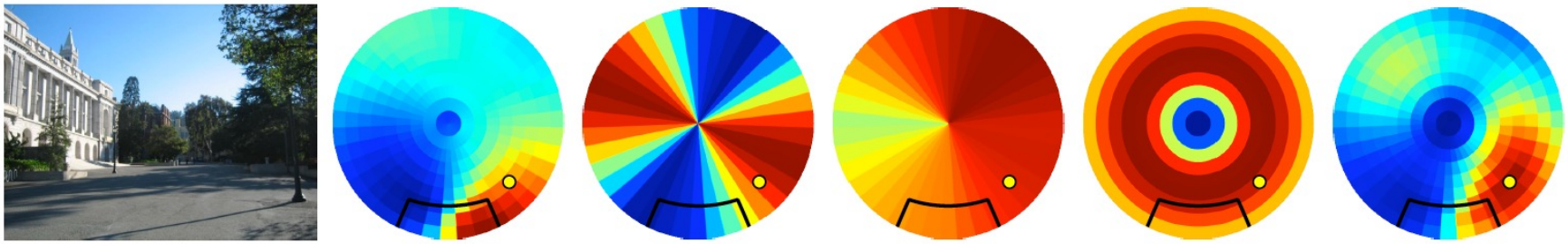
Figure 4. Illumination cue from the vertical surfaces only. Starting from the input image (a), we compute the vertical surfaces mask (b) using [9] (blue = facing left, red = facing right, green = facing forward). The distribution of pixel intensities on each of these surfaces are then used to estimate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

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

where σ_w^2 is proportional to its corresponding surface intensity. 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:

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

Combine all cues



(a) Input image (b) $P(\theta_s, \Delta\phi_s|\mathcal{S})$ (c) $P(\theta_s, \Delta\phi_s|\mathcal{G})$ (d) $P(\theta_s, \Delta\phi_s|\mathcal{V})$ (e) $P(\theta_s, \Delta\phi_s)$ (f) $P(\theta_s, \Delta\phi_s|\mathcal{P})$

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

4.1. Cue combination

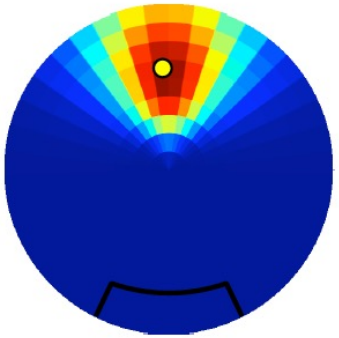
We are interested in estimating the distribution $P(I|\mathcal{P})$ over the illumination parameters $I = \{\theta_s, \Delta\phi_s, v_s\}$, given the entire image $\mathcal{P}$. We apply Bayes rule, and write

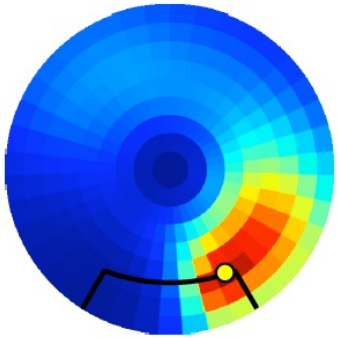
$$P(I|\mathcal{S}, \mathcal{G}, \mathcal{V}) \propto P(\mathcal{S}, \mathcal{G}, \mathcal{V}|I)P(I) \ . \quad (7)$$

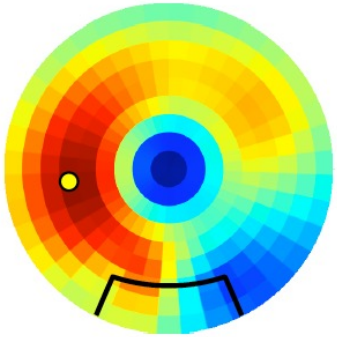
We make the assumption that the image pixels are conditionally independent given the illumination conditions, and that the priors on each region of the image ($P(\mathcal{S})$, $P(\mathcal{G})$ and $P(\mathcal{V})$) are uniform over their own respective domains. Applying Bayes rule twice, we get

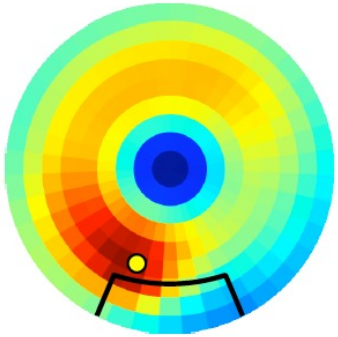
$$P(I|\mathcal{S}, \mathcal{G}, \mathcal{V}) \propto P(I|\mathcal{S})P(I|\mathcal{G})P(I|\mathcal{V})P(I) \ . \quad (8)$$

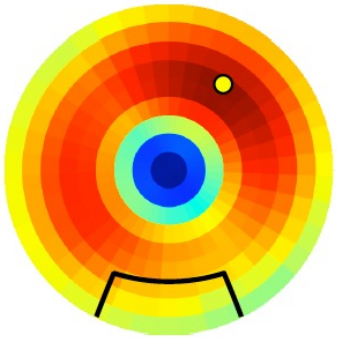
Results









Results

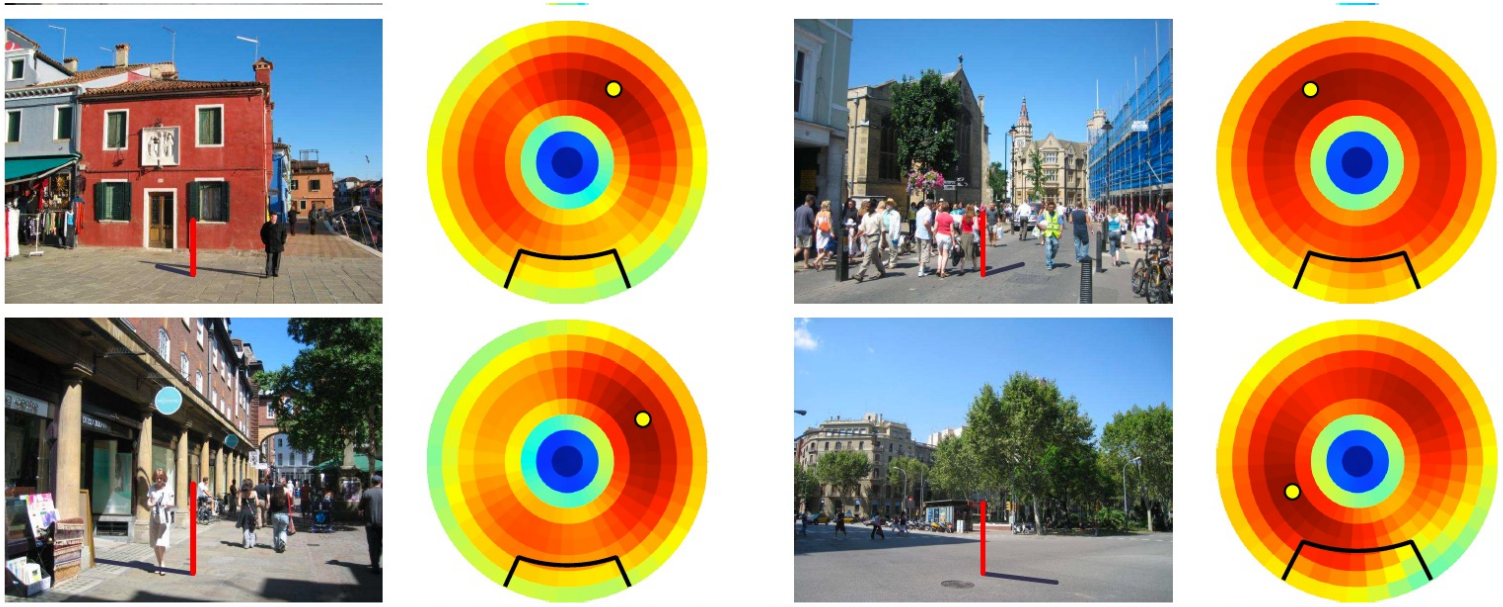
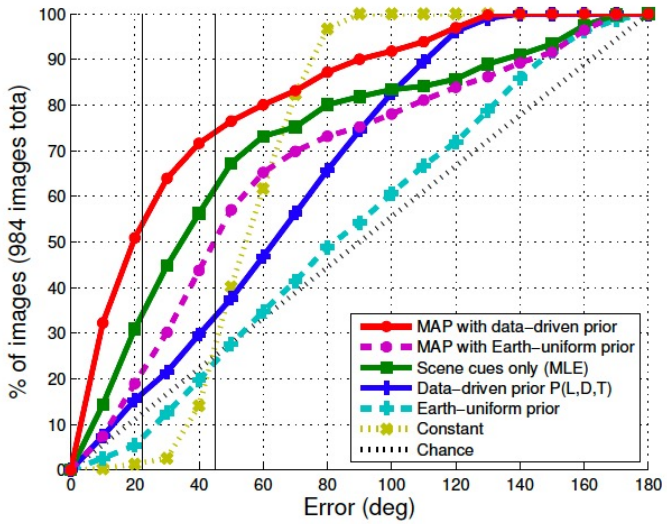


Figure 8. Sun direction estimation from a single image. A virtual sun dial is inserted in each input image (first and third columns), whose shadow correspond to the MAP sun position in the corresponding probability maps $P(\theta_s, \Delta\phi_s|\mathcal{P})$ (second and fourth columns). The rows are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

Results

(a) Madrid sequence

(b) Vatican sequence



3D Object Relighting

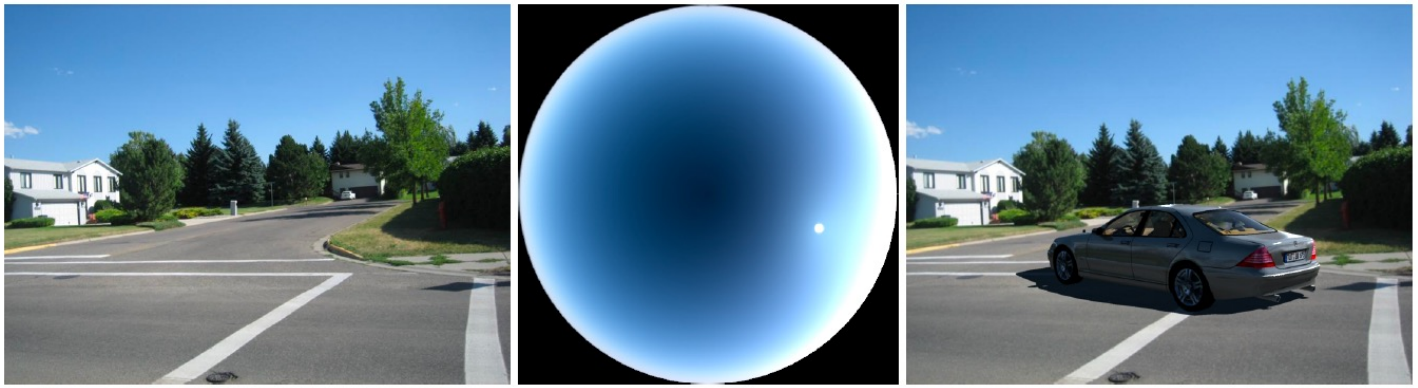


Figure 10. 3-D object relighting. From a single image (left), we render the most likely sky appearance (center) using the sun position computed with our method, and then fitting the sky parameters using [17]. We can realistically insert a 3-D object into the image (right).

Paper#2

Deep Outdoor Illumination Estimation

Yannick Hold-Geoffroy[1*], Kalyan Sunkavalli[†], Sunil Hadap[†], Emiliano Gambaretto[†], Jean-François L
Université Laval[*], Adobe Research[†]
yannick.hold-geoffroy.1@ulaval.ca, {sunkaval,hadap,emiliano}@adobe.com, jflalonde@gel.ula
http://www.jflalonde.ca/projects/deepOutdoorLight

Abstract

We present a CNN-based technique to estimate high-dynamic range outdoor illumination from a single low dynamic range image. To train the CNN, we leverage a large dataset of outdoor panoramas. We fit a low-dimensional physically-based outdoor illumination model to the skies in these panoramas giving us a compact set of parameters (including sun position, atmospheric conditions, and camera parameters). We extract limited field-of-view images from the panoramas, and train a CNN with this large set of input image–output lighting parameter pairs. Given a test image, this network can be used to infer illumination parameters that can, in turn, be used to reconstruct an outdoor illumination environment map. We demonstrate that our approach allows the recovery of plausible illumination conditions and enables photorealistic virtual object insertion from a single image. An extensive evaluation on both the panorama dataset and captured HDR environment maps shows that our technique significantly outperforms previous solutions to this problem.

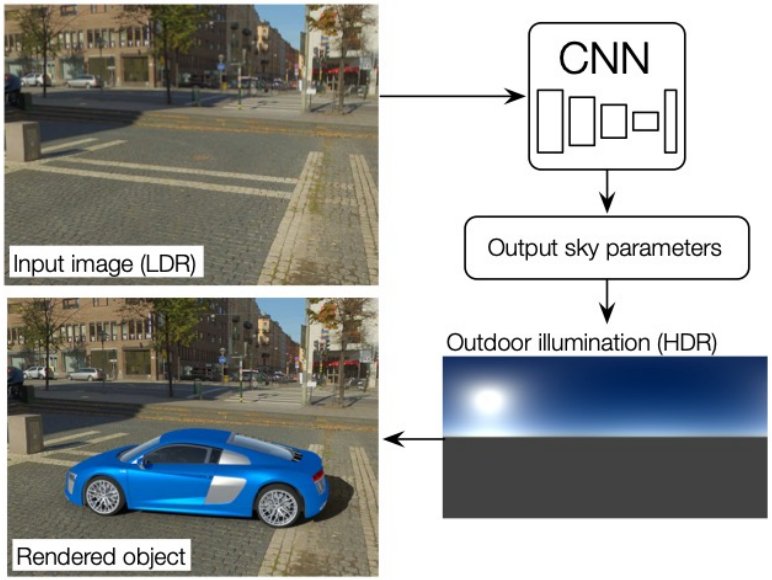


Figure 1. We present an approach for predicting full HDR lighting conditions from a single LDR outdoor image. Our prediction can readily be used to insert a virtual object into the image. Our key idea is to train a CNN using input-output pairs of LDR images and HDR illumination parameters that are automatically extracted from a large database of 360° panoramas.

Pipeline

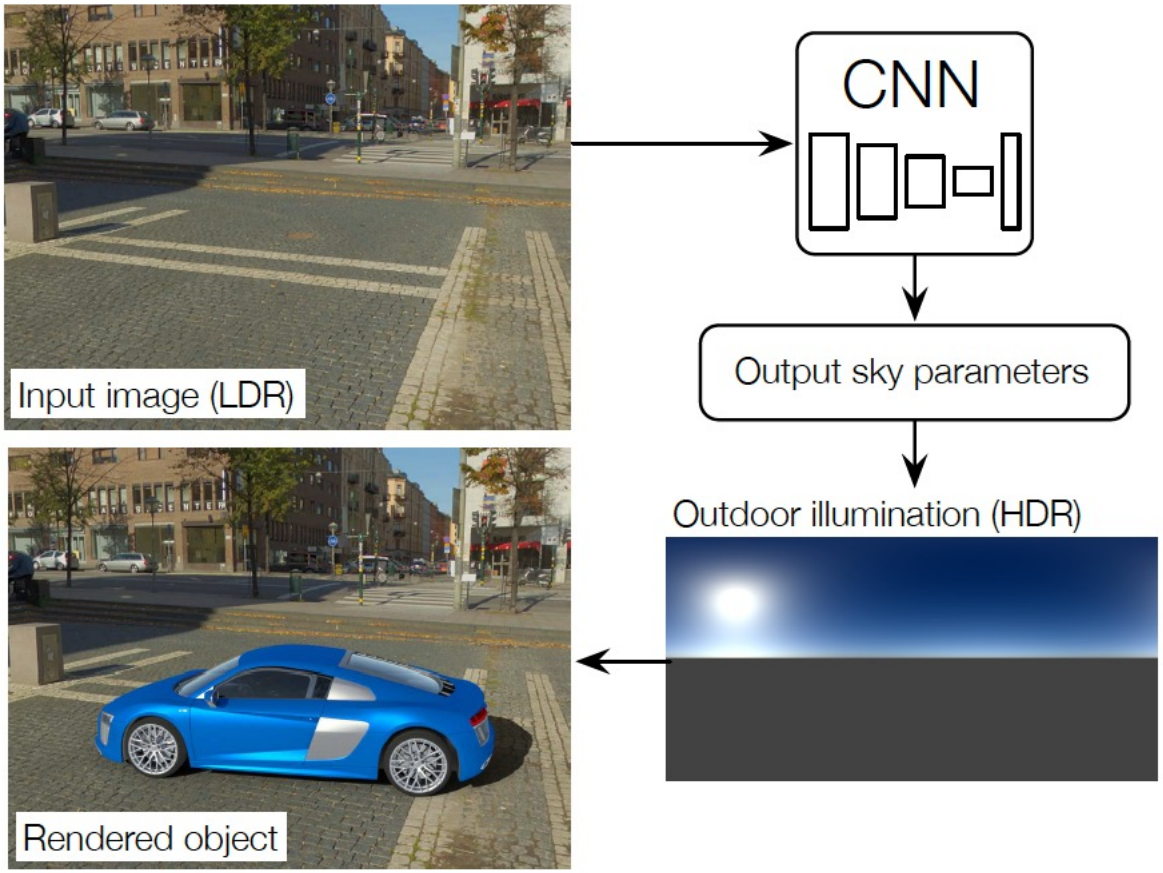


Figure 1. We present an approach for predicting full HDR lighting conditions from a single LDR outdoor image. Our prediction can readily be used to insert a virtual object into the image. Our key idea is to train a CNN using input-output pairs of LDR images and HDR illumination parameters that are automatically extracted from a large database of 360° panoramas.

Sky lighting model

4.1. Sky lighting model

We employ the model proposed by Hošek and Wilkie [16], which has been shown [21] to more accurately represent skylight than the popular Preetham model [32]. The model has also been extended to include a solar radiance function [17], which we also exploit.

In its simplest form, the Hošek-Wilkie (HW) model expresses the spectral radiance L_λ of a lighting direction along the sky hemisphere $\mathbf{l} \in \Omega_{\text{sky}}$ as a function of several parameters:

$$L_\lambda(\mathbf{l}) = f_{\text{HW}}(\mathbf{l}, \lambda, t, \sigma_g, \mathbf{l}_s)\,, \tag{1}$$

where λ is the wavelength, t the atmospheric turbidity (a measure of the amount of aerosols in the air), σ_g the ground albedo, and $\mathbf{l}_s$ the sun position. Here, we fix $\sigma_g = 0.3$ (approximate average albedo of the Earth [12]).

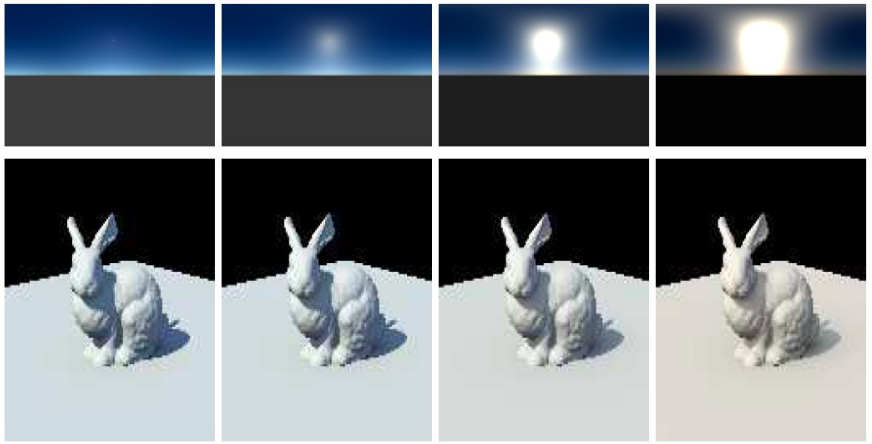


Figure 2. Impact of sky turbidity t on rendered objects. The top row shows environment maps (in latitude-longitude format), and the bottom row shows corresponding renders of a bunny model on a ground plane for varying values for the turbidity t, ranging from low (left) to high (right). Images have been tonemapped with $\gamma = 2.2$ for display.

Model fitting on training panoramas

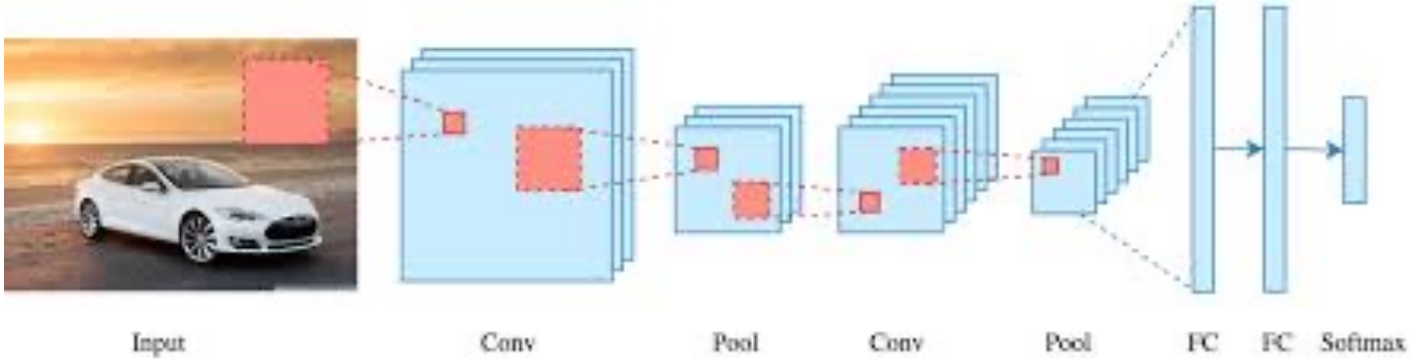
4.2. Optimization procedure

We now describe how the sky model parameters are estimated from a panorama in the SUN360 dataset. This procedure is carefully crafted to be robust to the extremely varied set of conditions encountered in the dataset which severely violates the linear relationship between sky radiance and pixel values such as: unknown camera response function and white-balance, manual post-processing by photographers and stitching artifacts.

Given a panorama P in latitude-longitude format and a set of pixels indices $p \in \mathcal{S}$ corresponding to sky pixels in P, we wish to obtain the sun position $\mathbf{l}_s$, exposure ω and sky turbidity t by minimizing the visible sky reconstruction error in a least-squares sense:

$$\mathbf{l}_s^*, \omega^*, t^* = \underset{\mathbf{l}_s, \omega, t}{\arg\min} \sum_{p \in \Omega_s} \left(P(p)^\gamma - \omega f_{\text{RGB}}(\mathbf{l}_p, t, \mathbf{l}_s) \right)^2 \quad \text{s.t.} \quad t \in [1, 10]\,, \tag{3}$$

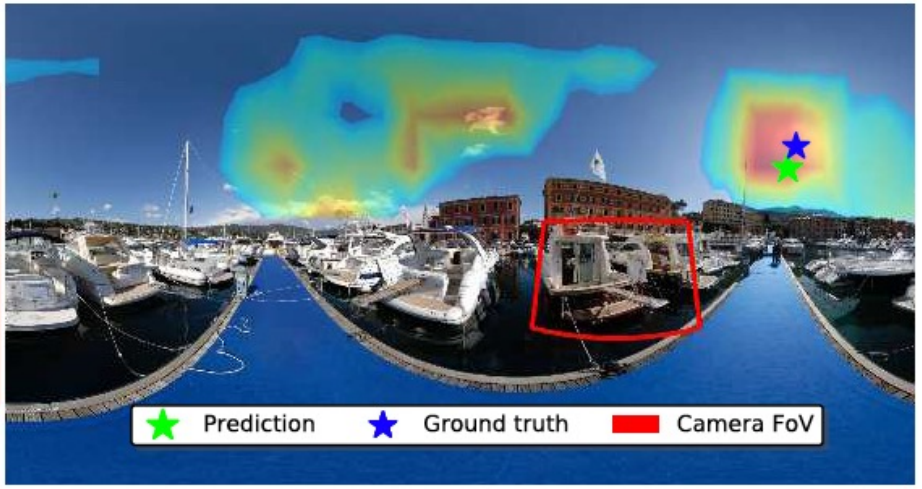
CNN Network structure

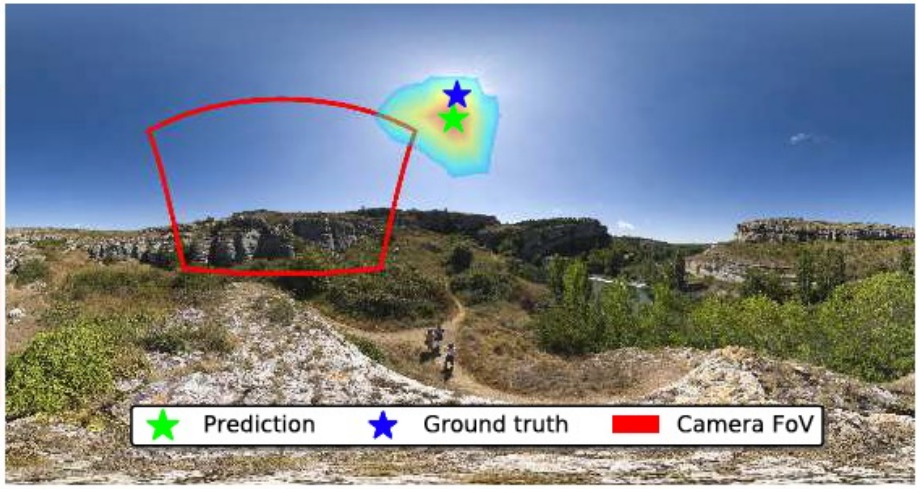


Layer	Stride	Resolution
Input		320×240
conv7-64	2	160×120
conv5-128	2	80×60
conv3-256	2	40×30
conv3-256	1	40×30
conv3-256	2	20×15
conv3-256	1	20×15
conv3-256	2	10×8
FC-2048		
FC-160 LogSoftMax		FC-5 Linear
Output: sun position distribution $\mathbf{s}$		Output: sky and camera parameters $\mathbf{q}$

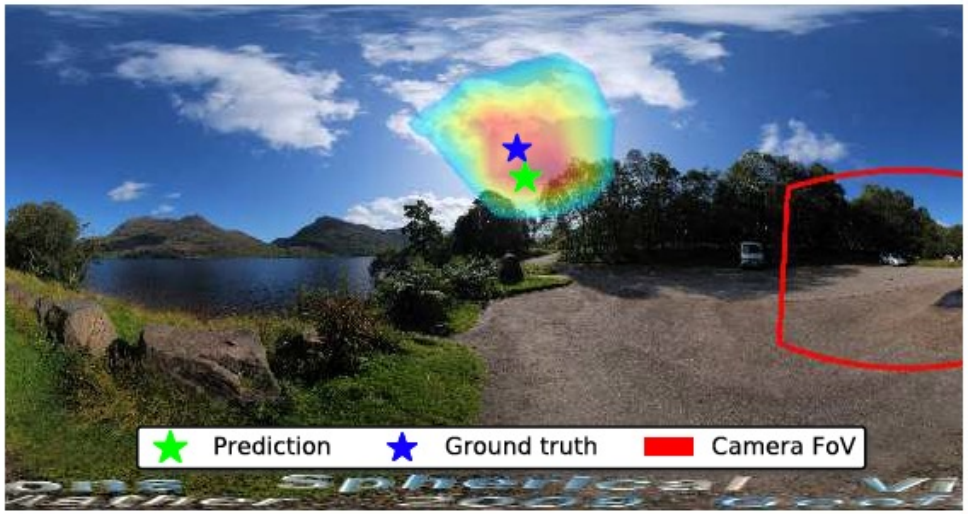
Figure 3. The proposed CNN architecture. After a series of 7 convolutional layers, a fully-connected layer segues to two heads: one for regressing the sun position, and another one for the sky and camera parameters. The ELU activation function [6] is used on all layers except the outputs.

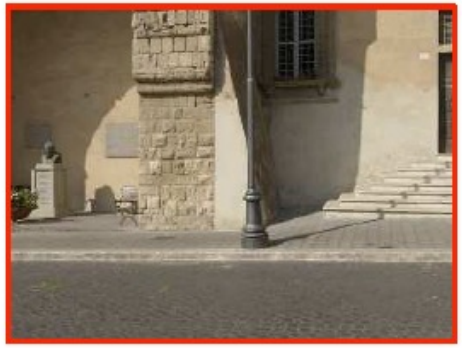
Result

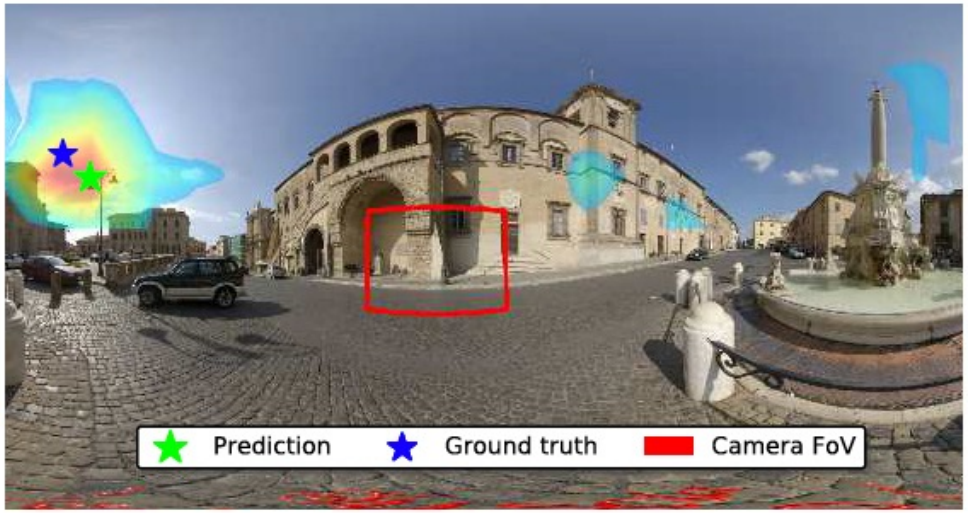

Prediction
Ground truth
Camera FoV


Prediction
Ground truth
Camera FoV

Result


Prediction
Ground truth
Camera FoV




Prediction
Ground truth
Camera FoV

Result: virtual object insertion

Figure 9. Virtual object insertion with automated lighting estimation. From a single image, the CNN predicted a full HDR sky map, which is used to render an object into the image. No additional steps are required. More results on automated object insertion are available in the supplementary materials.

Compare with ground-truth

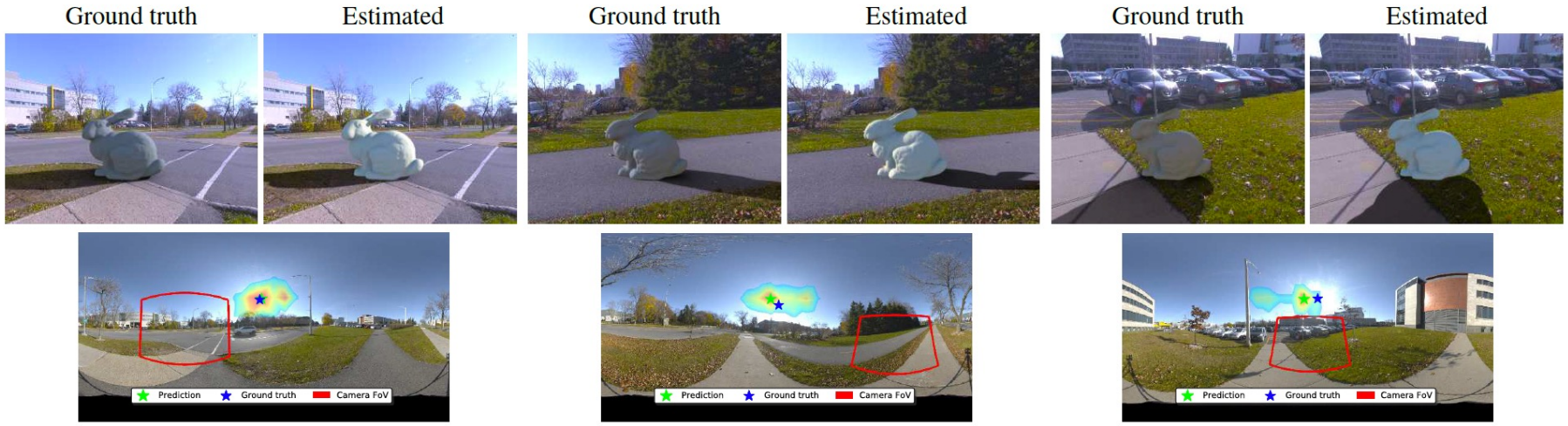


Figure 11. Object relighting comparison with ground truth illumination conditions on captured HDR panoramas. For each example, the top row shows (left) a bunny model relit by the ground truth HDR illumination conditions captured in situ; (right) the same bunny model, relit by the illumination conditions estimated by the CNN solely from the background image, completely automatically. No further adjustment (e.g. overall brightness, saturation, etc.) was performed. The bottom row shows the original environment map, field of view of the camera (in red), and the distribution on sun position estimation (as in fig. 6). Please see additional results on our project page.

Failure mode

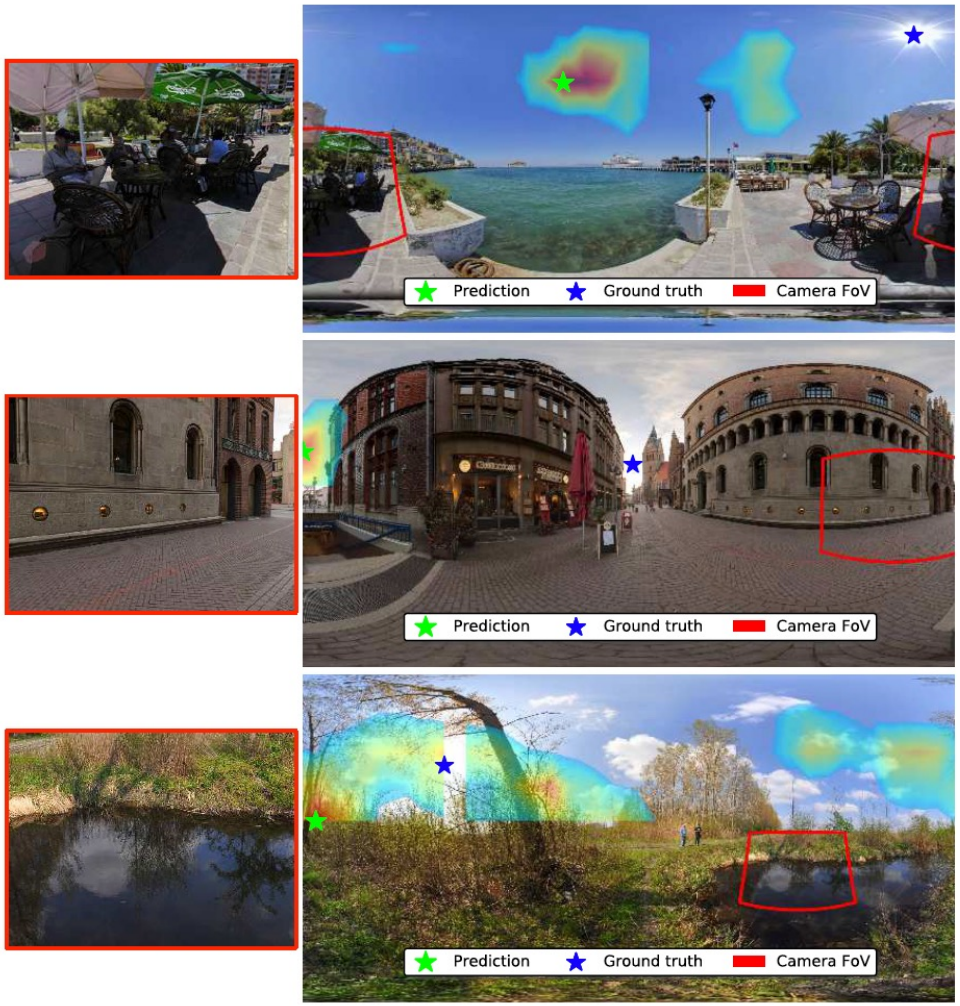


Figure 12. Typical failure cases of sun position estimation from a single outdoor image. See fig. 6 for an explanation of the annotations. Failure cases occur when illumination cues are mixed with complex geometry (top), absent from the image (middle), or in the presence of mirror-like surfaces (bottom).

Paper#3

Single Image Portrait Relighting

TIANCHENG SUN, University of California, San Diego
JONATHAN T. BARRON and YUN-TA TSAI, Google Research
ZEXIANG XU, University of California, San Diego
XUEMING YU, GRAHAM FYFFE, CHRISTOPH RHEMANN, JAY BUSCH, and PAUL DEBEVEC, Google
RAVI RAMAMOORTHI, University of California, San Diego

(a) Input image and estimated lighting

(b) Rendered images from our method under three novel illuminations

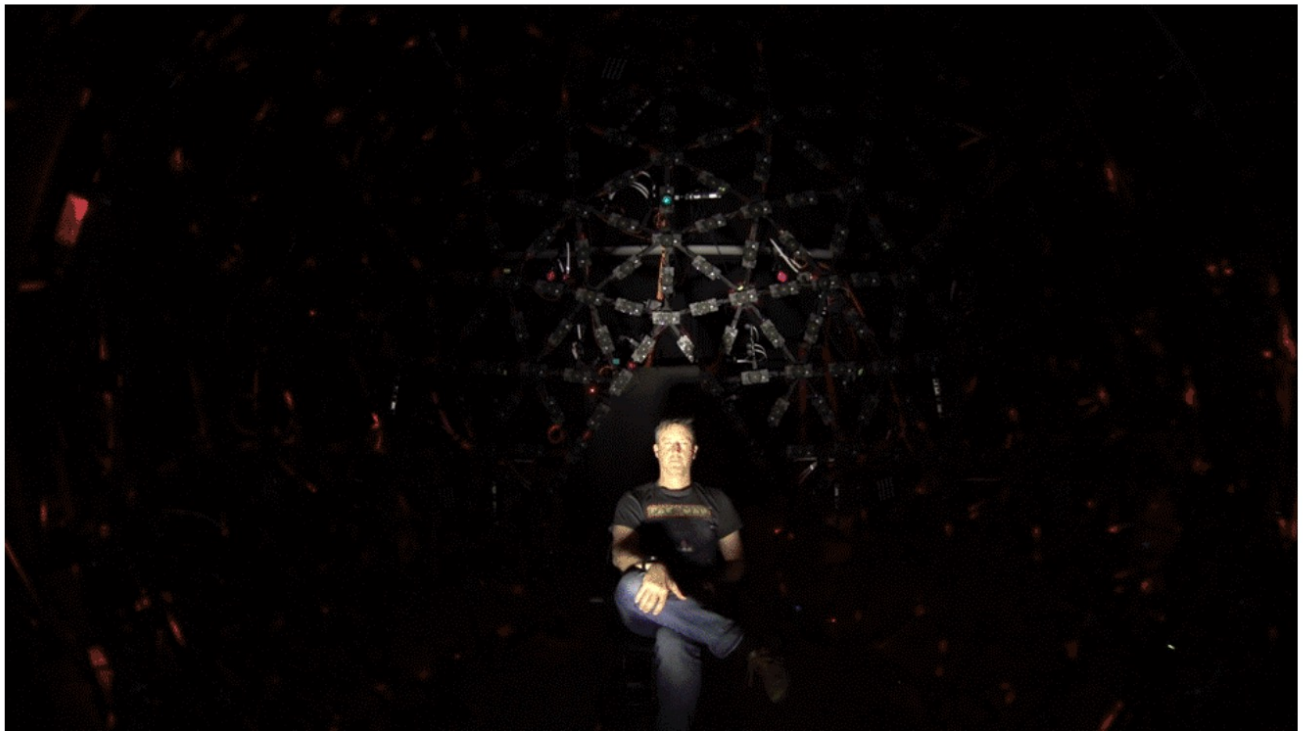
Fig. 1. Given only a single input image taken with a standard cellphone camera of a portrait (a), our model is able to quickly (160 ms.) generate new images of our human subject as though they are illuminated under new, previously-unseen lighting environments (b).

Google AI Blog

https://ai.googleblog.com/2020/12/portrait-light-enhancing-portrait.html

Estimating the high dynamic range, omnidirectional illumination profile from an input portrait. The three spheres at the right of each image, diffuse (**top**), matte silver (**middle**), and mirror (**bottom**), are rendered using the estimated illumination, each reflecting the color, intensity, and directionality of the environmental lighting.

Training process



(a) OLAT images (7 cameras)

(b) Ground-truth renderings

Network architecture

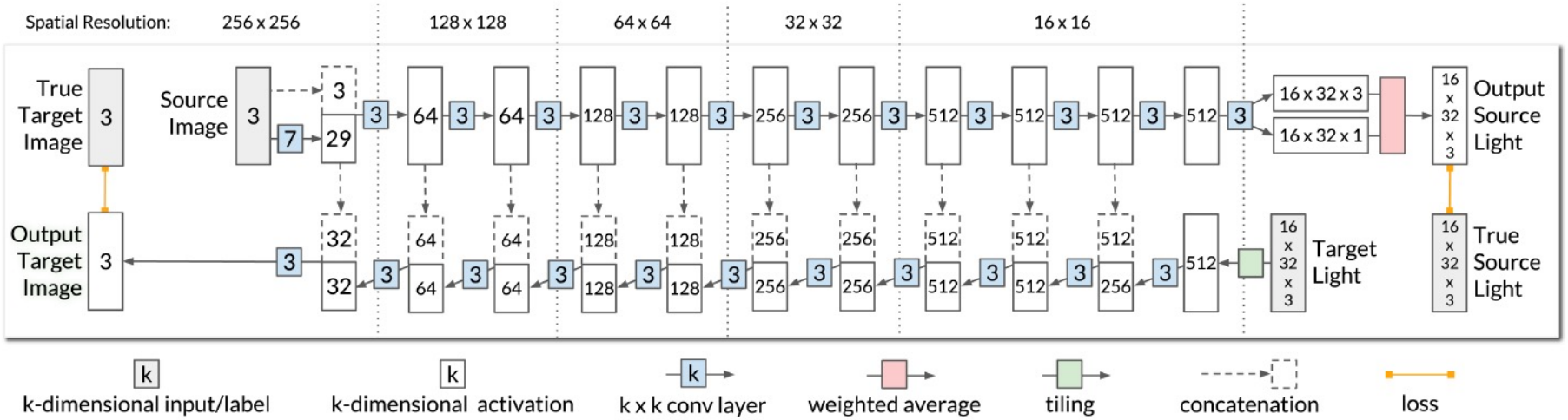


Fig. 4. The architecture of our neural network. The source input image is passed through a series of conv layers that gradually decrease spatial resolution while increasing the number of channels. After encoding, a confidence-weighted average predicts the illumination that the source image was lit by. The target light is then injected as input into the bottleneck of the network, and this target light along with the encoding of the source image is decoded (with skip connections) into the output target image. Losses are imposed to minimize the differences between the true and output target images, and between the true and output source lights. When evaluating our "self-supervision" loss, this architecture is modified such that the output source light with a certain rotation is used as the target light when decoding.

Network architecture

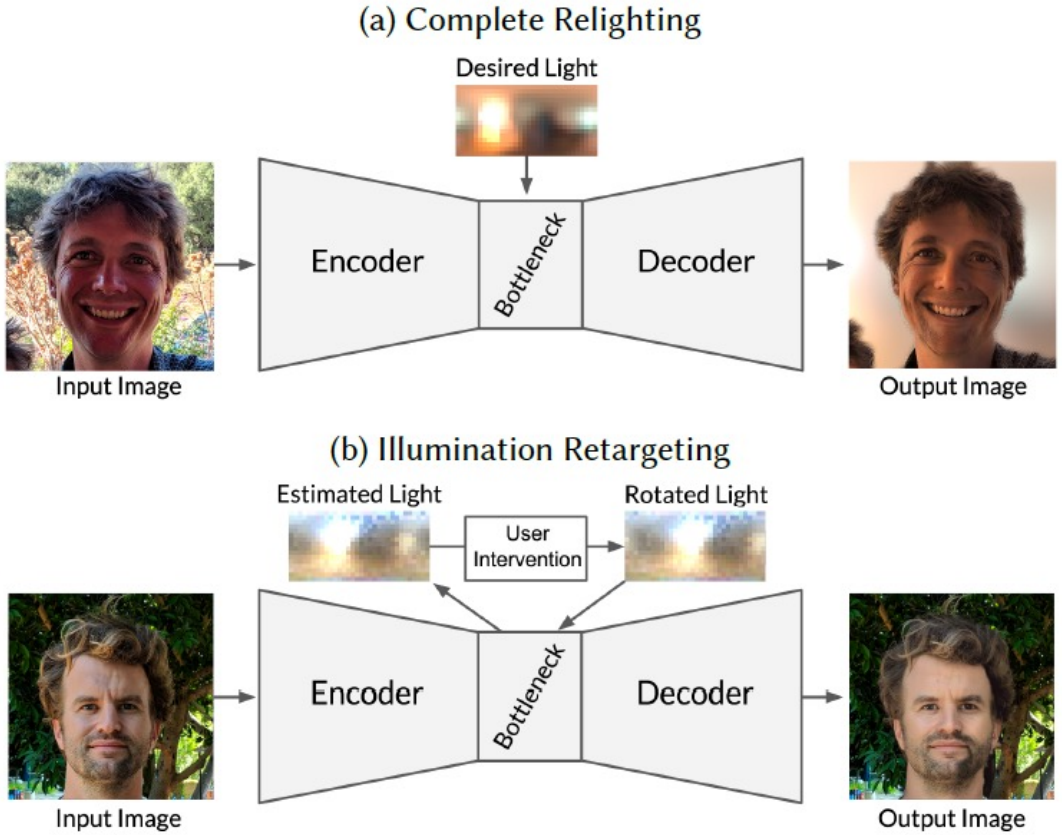


Fig. 2. Our relighting system is an encoder-decoder neural network that takes as input a single input image and a target illumination (injected into the bottleneck of the network), and produces as output a relit image (a). The encoder also predicts the illumination of the input image, thereby allowing an input image's illumination to be recovered during encoding, modified within the bottleneck of the network, and then decoded to produce a relit image corresponding to, say, a rotation of the original illumination (b).

Result: lighting estimation

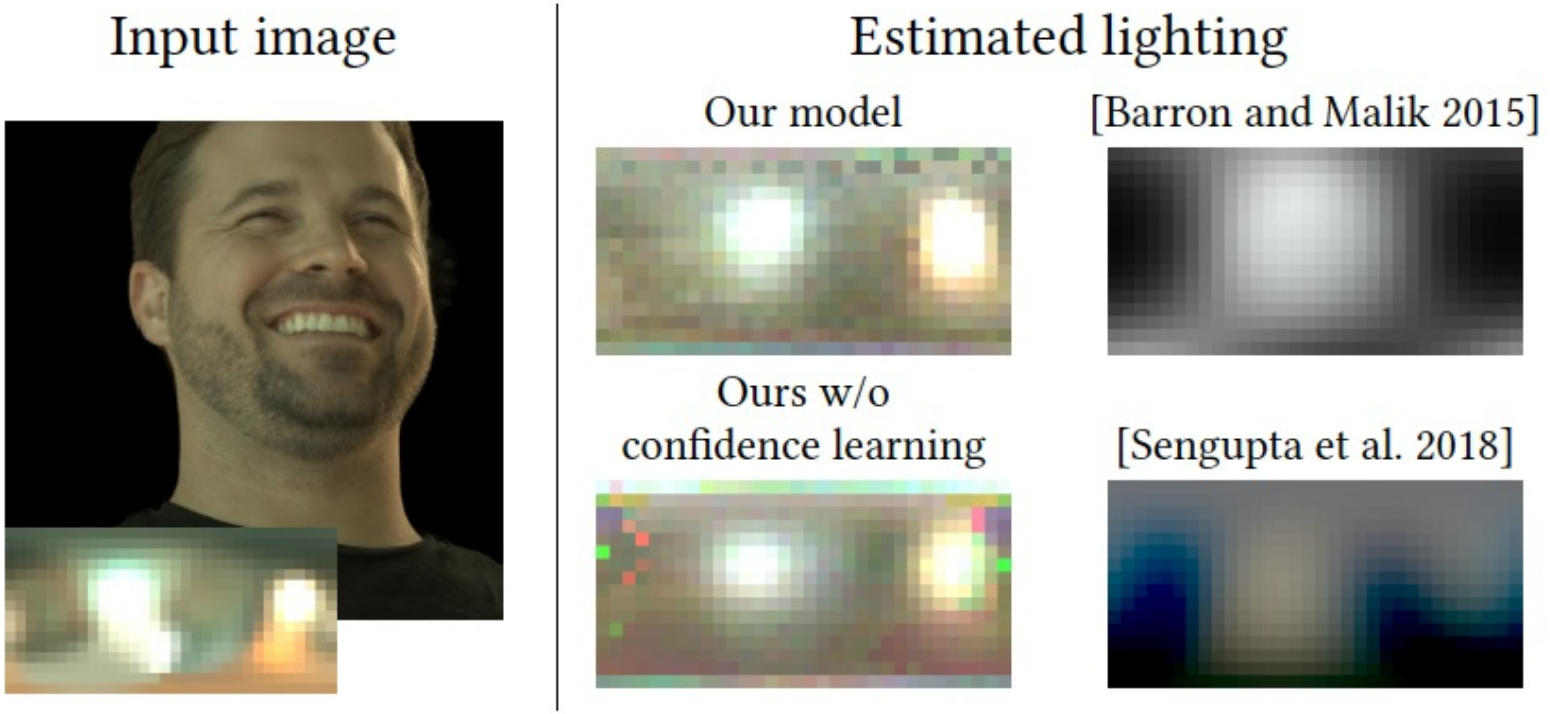
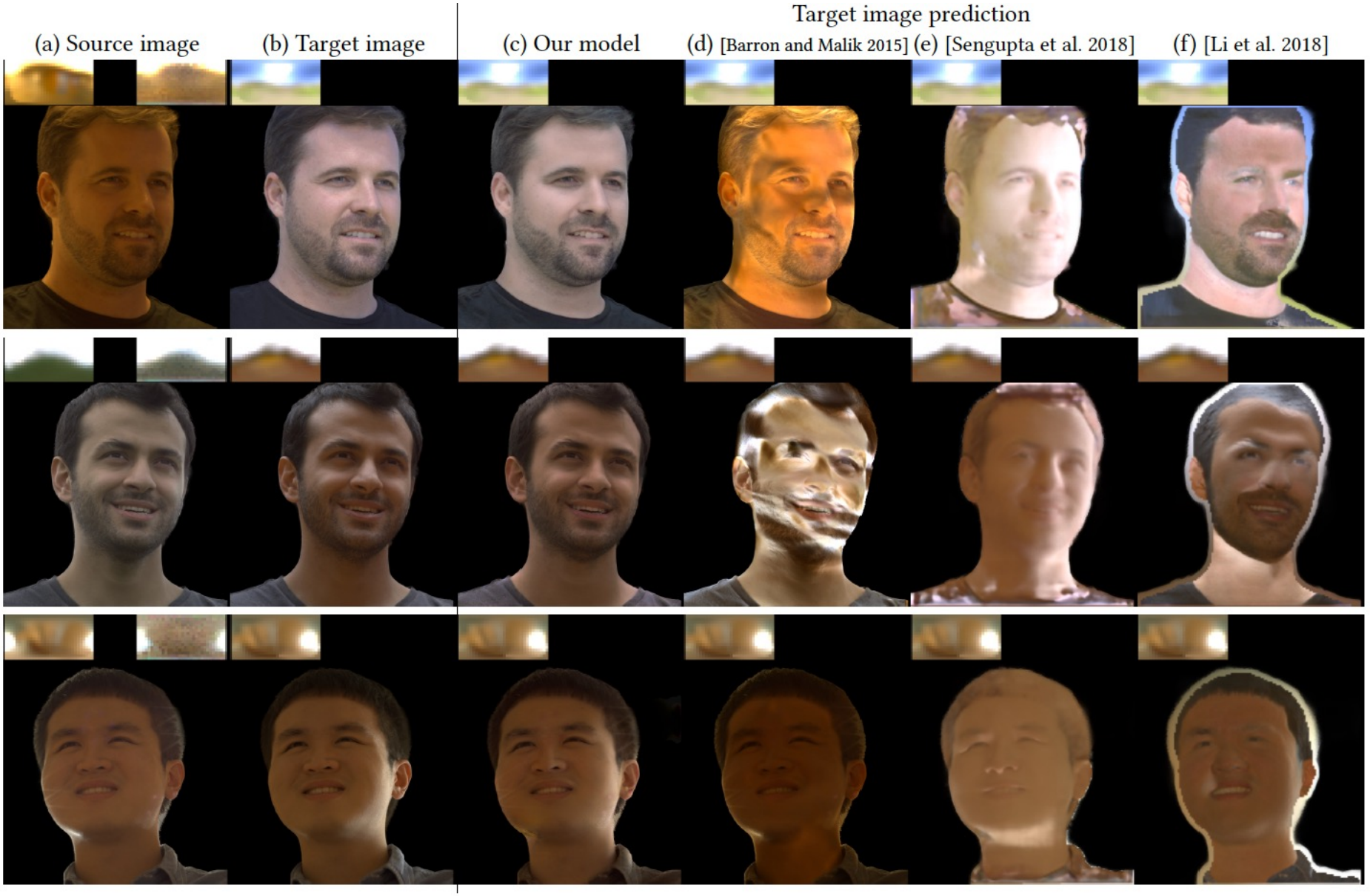


Fig. 8. Here we show lighting estimation results for our model and prior works given a single portrait image. Our model can accurately recover the locations and the colors of multiple light sources, while other methods struggle with high-frequency illumination effects or the non-Lambertian properties of human skin.

Result

Result

Comparisons



Breakout room: Please pick up 1 paper only to read and to discuss.

Breakout room group discussion & report:

Check your name on Zoom

Take note: breakout room ID, room members, “speaker”.

Your task:

- Recorded your Room ID, and room members’ names, and nominate a “Speaker”.

- Quickly re-read it for about 5 minutes.
- Discuss their limitations/drawbacks, for another 20 minutes.

- Try to answer the following “critical analysis” questions → page turn ..
- Report back in 3 minutes by the speaker.