

# Eclipse: Disambiguating Illumination and Materials using Unintended Shadows

## Supplementary Material

### S1. Additional results

Figure S1 contains additional synthetic results, similar to Figures 1 and 4 of the main paper, and Figure S2 contains additional real results from the Stanford-ORB dataset [18], similar to Figure 8 from the main paper. As explained in the paper, the recovered illumination estimated by our model corresponds to the incident light at the object, which is typically extremely warped relative to the collection of environment maps provided by the dataset as ground truth, since those are obtained using a light probe placed at different points in the scene. We therefore compute our quality metrics relative to the mean environment map obtained from these differently-placed probes, blurred using a (normalized) spherical Gaussian kernel  $p(\theta, \phi) \propto \exp(\kappa \cos \theta)$  (we set  $\kappa = 100$ ) to fuse the misaligned light sources. See Figure S10 for an example of this process.

Figure S3 provides additional evidence that the signal provided by unintended shadows improves the recovery of environment illumination, expanding on Figure 9 in the main paper.

Figure S4 contains an additional result for recovered illumination and material properties when the object geometry is not given, as in Figure 7 of our main paper.

**Relighting.** Figure S5 shows that our recovered albedo enables convincing relighting results. See our supplemental video for rendered video results.

**Non-diffuse results.** Figure S6 shows the results obtained by our method when applied to the same microfacet BRDF model from our main paper, but with a roughness value of 0.2 (see Section S2).

**Recovered masks.** The main goal of our approach is the recovery of illumination and materials. However, it may also be informative to visualize the recovered per-image occluder masks. Figure 5 in the main paper shows the recovered masks in the case of letter-shaped occluders, and we show additional recovered masks from the same experiment in Figure S7. Similarly, Figure S8 shows the recovered masks from another experiment done on the *potatoes* scene, where the masks are the spherical caps with solid angle  $0.1 \cdot 4\pi$  steradians used in all other experiments (see description in Section 5 of the main paper). Finally, Figure S9 shows masks recovered from the Stanford-ORB dataset [18], corresponding to the recovered illumination shown in Figure 8 of the main paper. Note that although the recovered lights and albedos in Figure 8 match the ground

truth ones, the masks appear to be less plausible. This can be explained by the difficulty of recovering mask values in regions occluding dark regions of the environment. This is true almost everywhere in the case of highly-directional illumination, like the environment map shown in Figure 8. See Figure 5 in the main paper and the text in Section 5 for a demonstration of this effect in synthetic data.

### S2. BRDF model

The experiments in the paper and supplement were performed using either a standard Lambertian BRDF (where specified) or a BRDF based on Unreal Engine’s version of GGX [15, 43]:

$$f(\mathbf{x}, \hat{\omega}_i, \hat{\omega}_o) = \frac{1}{\pi} \rho(\mathbf{x}) (1 - F(\hat{\mathbf{n}} \cdot \hat{\omega}_i; \kappa)) (1 - F(\hat{\omega}_h \cdot \hat{\omega}_i, \kappa)) (\hat{\mathbf{n}} \cdot \hat{\omega}_i)_+ + \frac{D(\hat{\mathbf{n}} \cdot \hat{\omega}_h; \alpha) F(\hat{\omega}_h \cdot \hat{\omega}_i; \kappa) G(\hat{\mathbf{n}} \cdot \hat{\omega}_o, \hat{\mathbf{n}} \cdot \hat{\omega}_i; \alpha)}{4(\hat{\mathbf{n}} \cdot \hat{\omega}_i)_+ (\hat{\mathbf{n}} \cdot \hat{\omega}_o)_+}, \quad (\text{S1})$$

where:

$$F(\cos \theta; \kappa) = \kappa + (1 - \kappa)(1 - \cos \theta)^5, \quad (\text{S2})$$

$$D(\cos \theta; \alpha) = \frac{\alpha^2}{\pi(1 + (\alpha^2 - 1) \cos^2 \theta)^2}, \quad (\text{S3})$$

$$G(\cos \theta_i, \cos \theta_o; \alpha) = g(\cos \theta_i; \alpha) g(\cos \theta_o; \alpha), \quad (\text{S4})$$

$$g(\cos \theta; \alpha) = \frac{\cos \theta}{k(\alpha) + (1 - k(\alpha)) \cos \theta}, \quad (\text{S5})$$

$$k(\alpha) = \frac{(\alpha + 1)^2}{8}, \quad (\text{S6})$$

$$\hat{\omega}_h = \frac{\hat{\omega}_i + \hat{\omega}_o}{\|\hat{\omega}_i + \hat{\omega}_o\|}. \quad (\text{S7})$$

We omit the positional dependence of the surface normal vector  $\hat{\mathbf{n}}$  in the point  $\mathbf{x}$  on the surface.

The 5 parameters describing the BRDF at a given location  $\mathbf{x}$  are therefore the RGB albedo  $\rho$ , the microfacet roughness  $\alpha$ , and the specular reflectance at normal incidence  $\kappa$  which determines the strength of the Fresnel factor  $F$ .

### S3. Full Description of Renderer

We describe our renderer’s approximation of the integral in Equation 2 of the main paper. Our renderer follows well-established practices that are readily documented in public textbooks such as PBRT [32].

In order to render a specific pixel, we begin by choosing a random ray from the camera center through the square

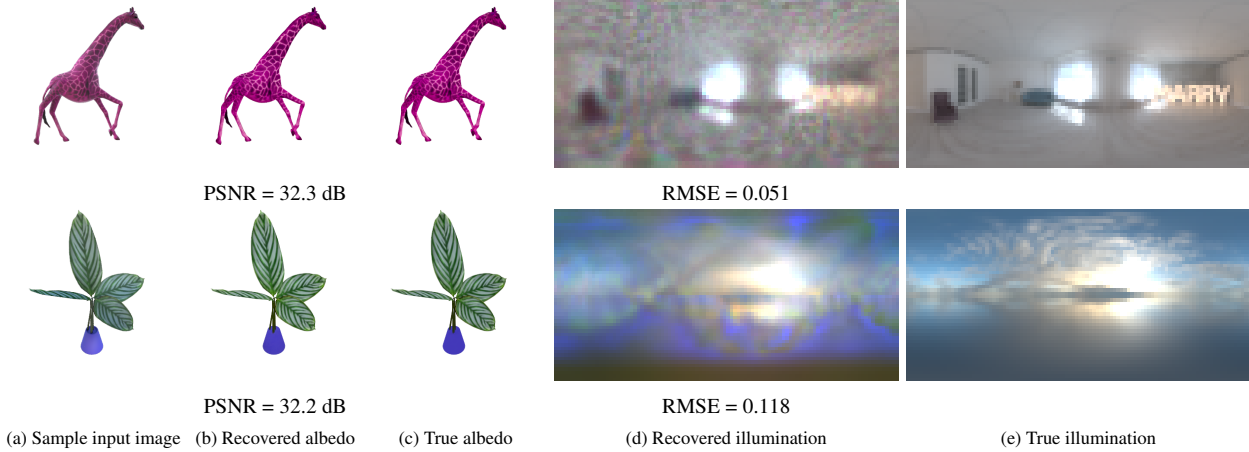


Figure S1. Additional results on diffuse objects, similar to Figures 1 and 4 of the main paper. We report the RMSE of each environment map in linear color space but plot the images after tonemapping for better evaluation of the full dynamic range. The albedo PSNR values are reported on object pixels only.

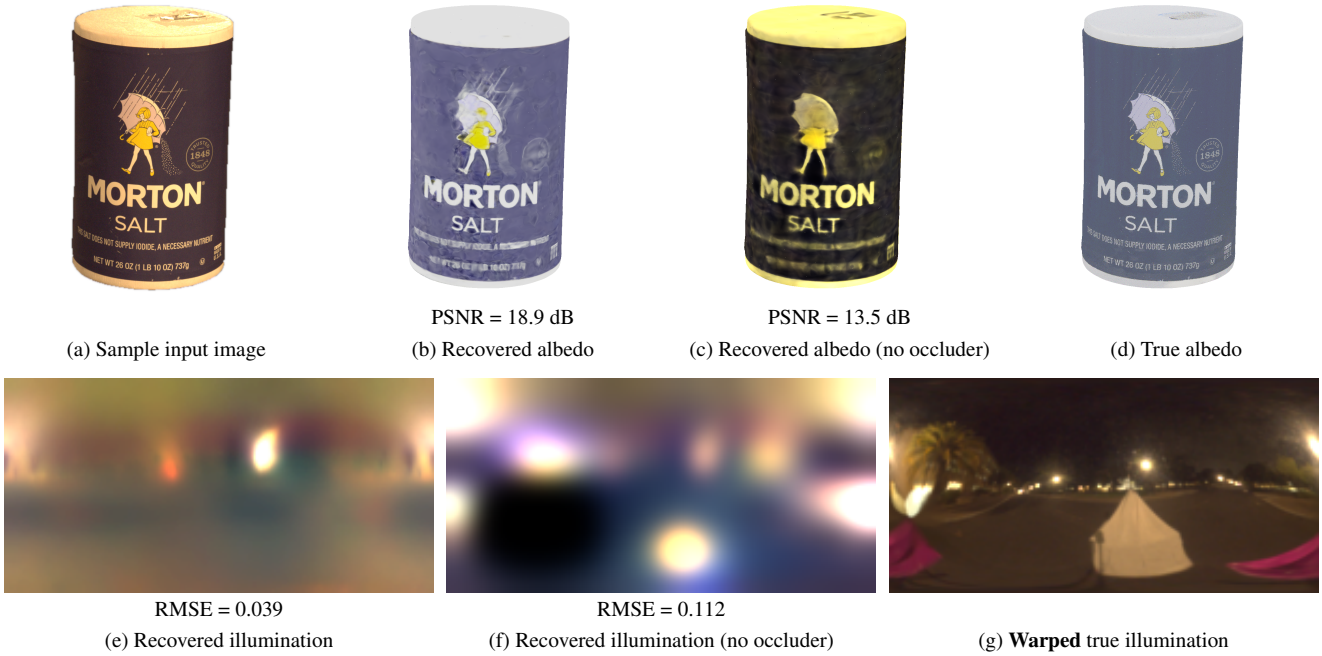


Figure S2. Results on the captured `salt007` scene from Stanford-ORB [18], extending Figure 8 of the main paper. Note that the “warped true illumination” environment map (g) was captured by a light probe that was not co-located with the object, any may therefore be significantly warped. The recovered RMSE reported is computed with respect to the average of all environment maps provided with the dataset rotated the same coordinate frame, see text and Figure S10.

footprint of the pixel. The intersection point of the ray with the object’s mesh is then computed, as well as the normal vector at that point (interpolated from per-vertex normals). Then, we use multiply importance sampling, and randomly sample  $s_\ell + s_m = 1024$  rays from this intersection point, with  $s_\ell = 512$  sampled from the lighting distribution, and  $s_m = 512$  sampled from the material distribution. The lighting distribution is simply a piecewise-constant distri-

bution proportional to the average environment map values taken across the three channels, weighted by the Jacobian of the parameterization,  $\sin \theta$ :

$$p^{(\ell)}(\hat{\omega}) \propto \sum_{c \in \{R, G, B\}} \tilde{L}_c(\hat{\omega}) \sin \theta, \quad (\text{S8})$$

where  $\tilde{L}_c$  is the  $c$ -th channel of the environment map obtained using nearest-neighbor interpolation, and  $\theta$  is the po-

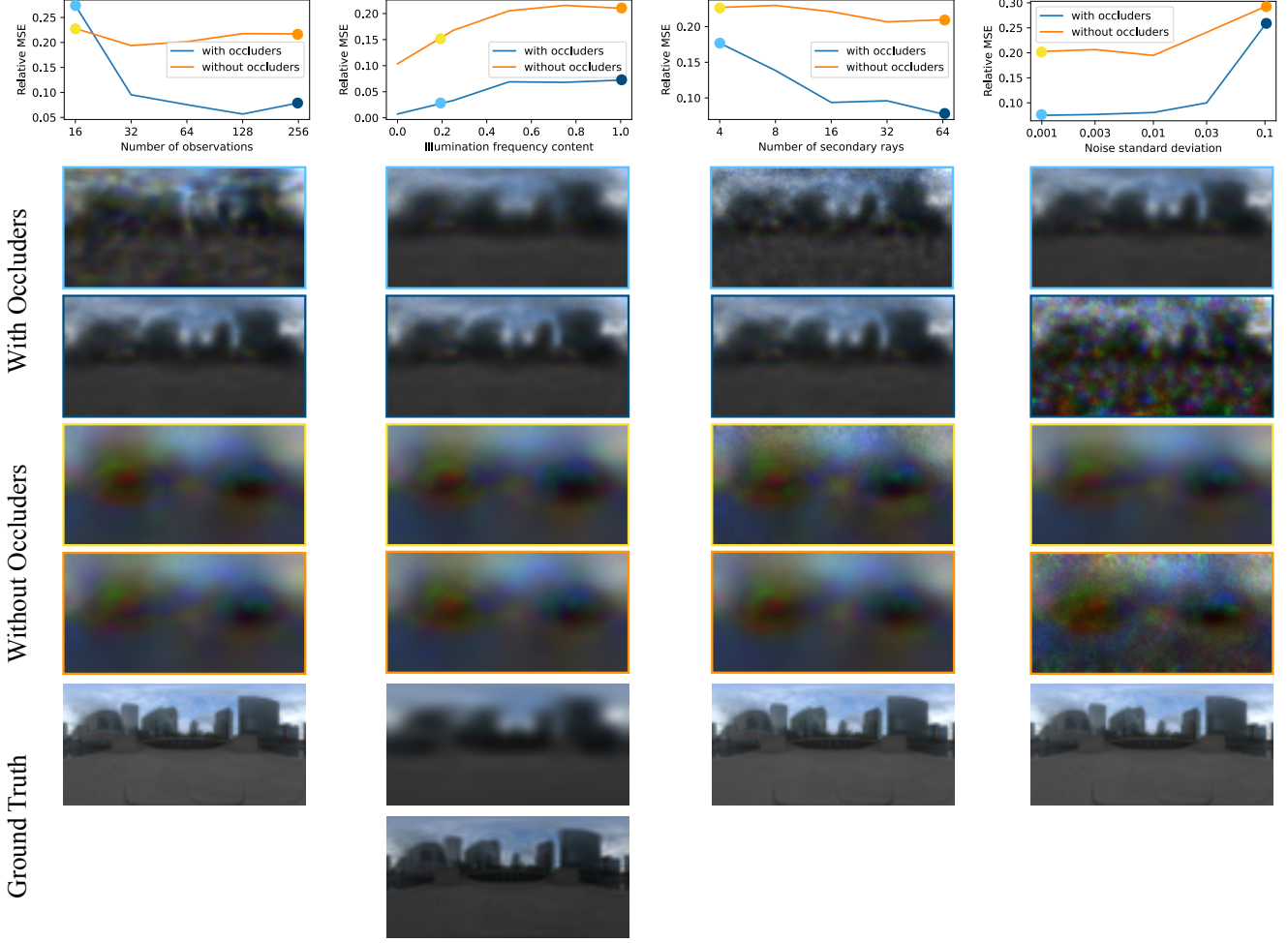


Figure S3. The signal provided by unintended shadows cast by unseen occluders improves the quality of recovered environment illumination. This is an expansion of Figure 8 in the main paper, but for an object with purely Lambertian material (we use the *potatoes* scene for these experiments). Here, we plot the relative MSE for the recovered environment maps under two scenarios: (blue) using images rendered with unobserved occluders and jointly estimating materials, illumination and occluder shape; and (orange) using images rendered without occluders and only estimating materials and illumination (this is similar to the problem setting considered by Swedish *et al.* [37], but optimized using our method). The cue of unintended shadows consistently improves the quality of estimated illumination across varying (first column) number of observed images, (second column) illumination frequency content, (third column) number of secondary rays traced, (fewer secondary rays causes increased Monte Carlo noise), and (fourth column) additive Gaussian noise. We display recovered environment maps corresponding to two points on each plot.

lar angle corresponding to  $\hat{\omega}$ , *i.e.*  $\sin \theta = \sqrt{1 - (\hat{\omega} \cdot \hat{\mathbf{z}})^2}$ , with  $\hat{\mathbf{z}}$  denoting the unit vector in the  $z$  direction.

The material distribution is simply the Trowbridge-Reitz normal distribution function reparameterized for sampling incoming light directions:

$$p^{(m)}(\hat{\omega}) = D(\hat{\mathbf{n}} \cdot \hat{\omega}_h; \alpha) \frac{|\hat{\omega}_h \cdot \hat{\mathbf{z}}|}{4(\hat{\omega}_h \cdot \hat{\omega}_o)_+}, \quad (\text{S9})$$

where  $\hat{\omega}_h$  is the half-vector defined in Equation S7, and  $D$  is the Trowbridge-Reitz distribution with roughness parameter  $\alpha$ , as defined in Equation S3.

The two sets of samples are combined using the power

heuristic [40]:

$$C(\mathbf{x}, \hat{\omega}^{(j)}, \hat{\omega}_o) = L_t(\hat{\omega}^{(j)}) f_{\mathbf{x}}(\hat{\omega}^{(j)}, \hat{\omega}_o) (\hat{\mathbf{n}}(\mathbf{x}) \cdot \hat{\omega}^{(j)})_+$$

$$\tilde{I}_t(\mathbf{u}) = \frac{1}{s_m} \sum_{j=1}^{s_m} \frac{\beta_m(\hat{\omega}^{(j)}) C(\mathbf{x}, \hat{\omega}^{(j)}, \hat{\omega}_o)}{p^{(m)}(\hat{\omega}^{(j)})}$$

$$+ \frac{1}{s_\ell} \sum_{j=s_m+1}^{s_m+s_\ell} \frac{\beta_\ell(\hat{\omega}^{(j)}) C(\mathbf{x}, \hat{\omega}^{(j)}, \hat{\omega}_o)}{p^{(\ell)}(\hat{\omega}^{(j)})}, \quad (\text{S10})$$

with  $\hat{\omega}^{(1)}, \dots, \hat{\omega}^{(s_m)} \sim p^{(m)}(\hat{\omega})$ ,

and  $\hat{\omega}^{(s_m+1)}, \dots, \hat{\omega}^{(s_m+s_\ell)} \sim p^{(\ell)}(\hat{\omega})$ ,

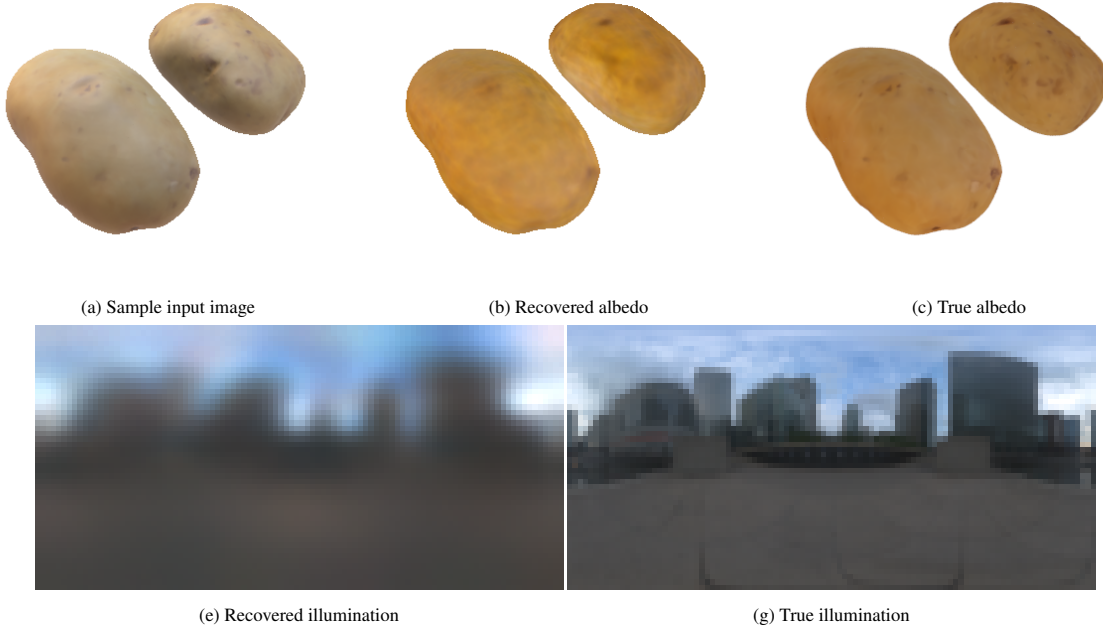


Figure S4. Expansion of Figure 7 in the main paper. We are able to recover illumination even when geometry is unknown by first optimizing a volumetric representation of geometry using a NeRF-based method. Despite this inaccurate proxy geometry, our method still recovers plausible (albeit blurry) illumination (e) and albedo (b).

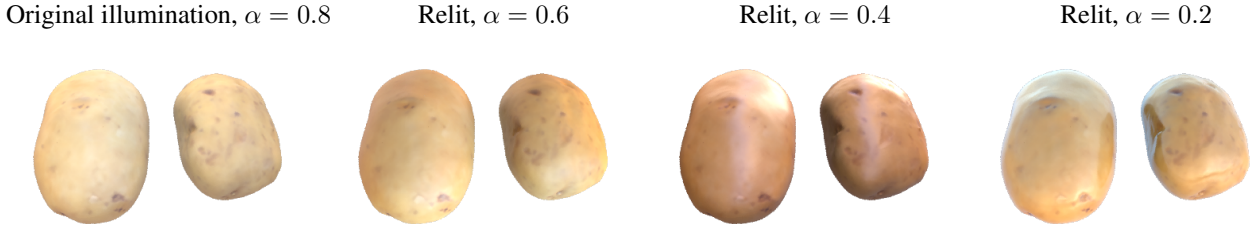


Figure S5. Our method enables modifying the illumination and material properties of the object. On the left, we show the original illumination and rough material ( $\alpha = 0.8$  in our BRDF model). The three columns to the right feature our recovered albedo, but rendered under different environment maps and progressively lower roughness values.

where  $\beta_m$  and  $\beta_\ell$  are the power heuristic weights for multiple importance sampling, with exponent 2, as in [40]:

$$\beta_m(\hat{\omega}) = \frac{(s_m p^{(m)}(\hat{\omega}))^2}{(s_m p^{(m)}(\hat{\omega}))^2 + (s_\ell p^{(\ell)}(\hat{\omega}))^2}, \quad (\text{S11})$$

$$\beta_\ell(\hat{\omega}) = \frac{(s_\ell p^{(\ell)}(\hat{\omega}))^2}{(s_m p^{(m)}(\hat{\omega}))^2 + (s_\ell p^{(\ell)}(\hat{\omega}))^2}. \quad (\text{S12})$$

Similar to the observation of Zeltner *et al.* [45], our experiments show that it is beneficial to “detach” gradients from the sampling procedure for  $\hat{\omega}^{(j)}$  as well as from the PDFs  $p^{(m)}$  and  $p^{(\ell)}$  in Equations S8 and S9.

In order to generate samples from the lighting and mate-

rial distributions, we use inverse transform sampling, where the  $n$  input pairs of samples  $(u_0, v_0), \dots, (u_{n-1}, v_{n-1}) \in [0, 1]^2$  are computed from  $2n$  i.i.d. uniform random variables  $r_0, \dots, r_{n-1}, t_0, \dots, t_{n-1} \sim \text{Uniform}[0, 1]$  using:

$$u_i = \frac{\text{mod}(i, s) + r_i}{s} \quad (\text{S13})$$

$$v_i = 2 \cdot \frac{\lfloor i/s \rfloor + t_i}{s}, \quad (\text{S14})$$

where:

$$s = 2^{\lfloor \frac{\log_2(n)+1}{2} \rfloor}. \quad (\text{S15})$$

This procedure divides the unit square into an  $\frac{s}{2} \times s$  grid,

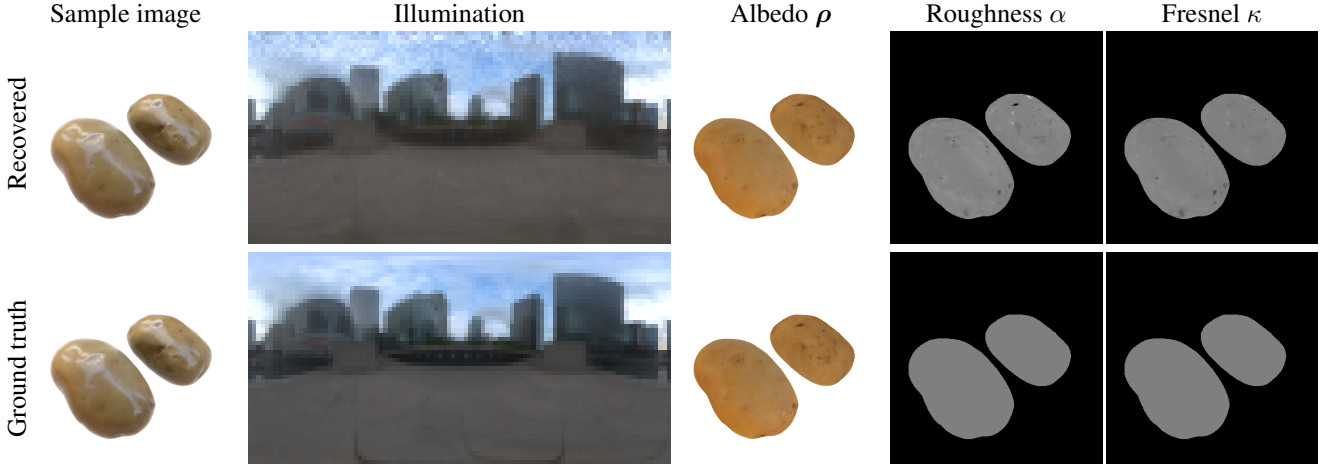


Figure S6. Our results on a shinier object. Our method is designed for diffuse objects, but it still manages to extract accurate material properties and illumination when materials are more specular ( $\alpha = 0.2$  for this figure). For visualization purposes the roughness values  $\alpha$  are linearly mapped from  $[0.15, 0.25]$  to  $[0, 1]$ , and the specular reflectance at normal incidence  $\kappa$  are mapped from  $[0.03, 0.05]$  to  $[0, 1]$  (with the true value being 0.04), in order to show small errors.

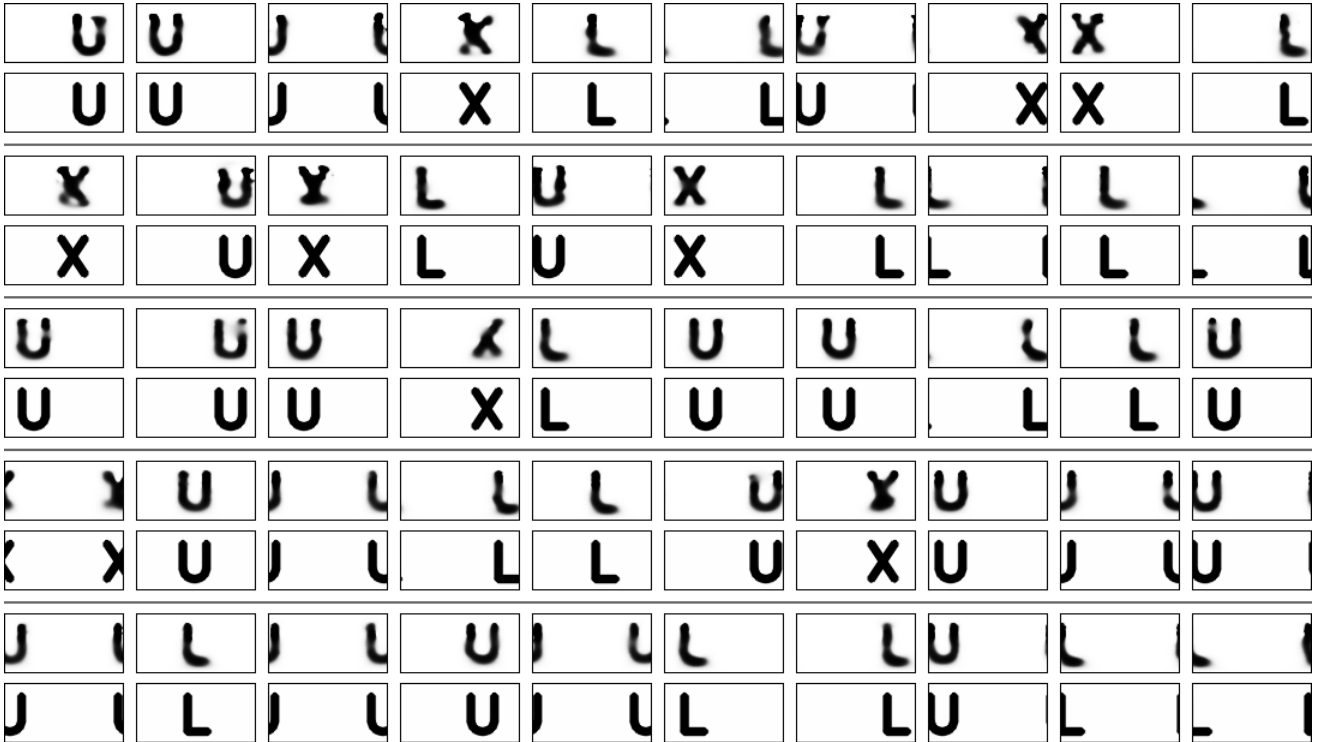


Figure S7. Expansion of Figure 5 in the main paper, showing a random selection of recovered masks (top image in each pair), and the corresponding ground truth (bottom image).

and uniformly samples a single pair  $(u_i, v_i)$  in each one of the grid cells, resulting in a stratified sampling pattern.

For generating the data, we repeat the entire rendering process described above 16 times using different primary and secondary rays, for every pixel and average the results

in order to antialias the results. Note that this corresponds to using a box reconstruction filter, which may have visible artifacts, yet we decided to use it in our experiments for simplicity.

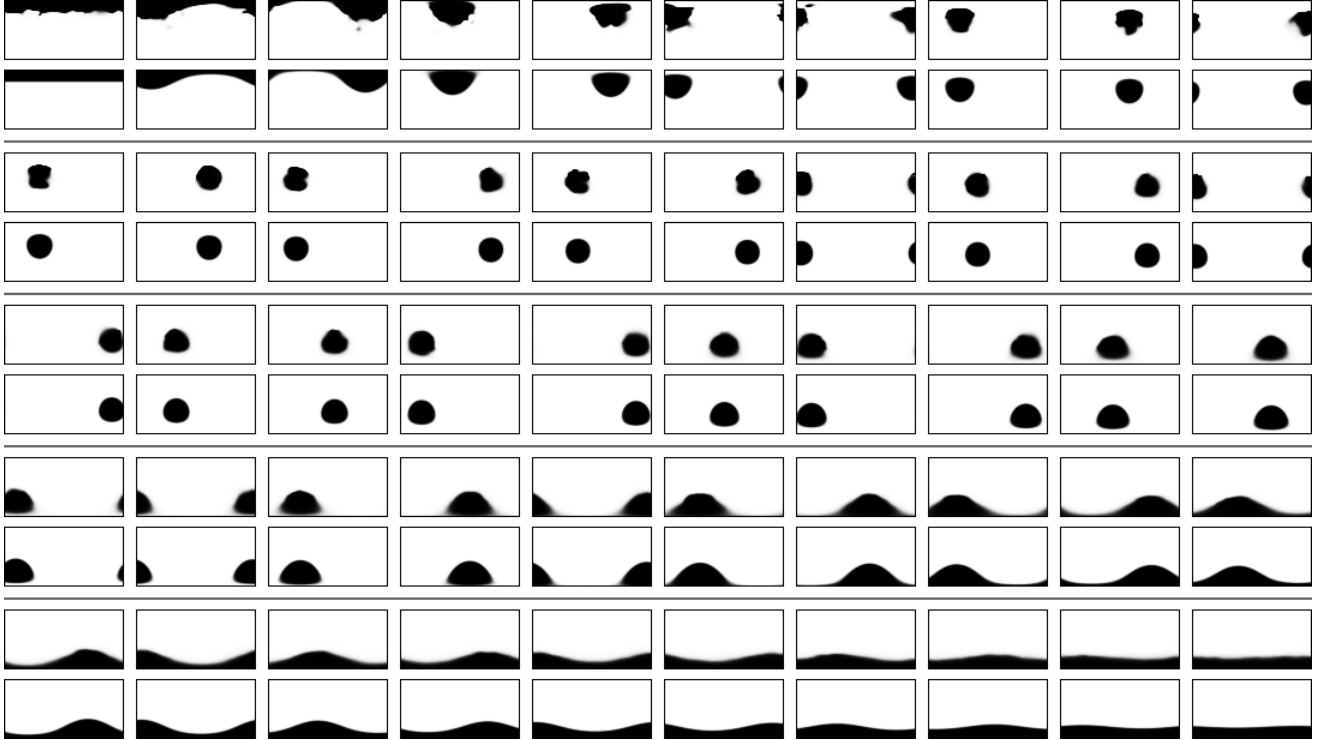


Figure S8. Masks recovered by our method on the *potatoes* scene. The top image in each pair shows the recovered masks, and the bottom one shows its corresponding true mask.

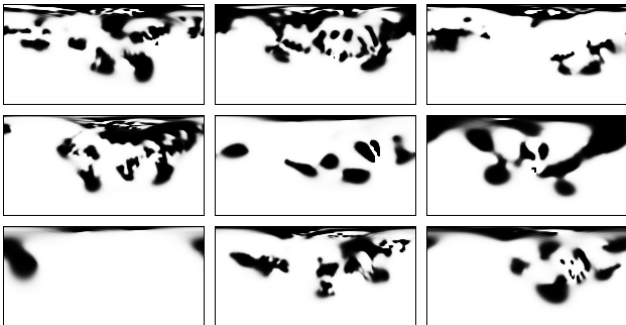


Figure S9. Masks recovered by our method on the *salt004* scene from Stanford-ORB [18]. Note that the masks are not guaranteed to be plausible, since regions which occlude small amounts of light cannot be recovered.

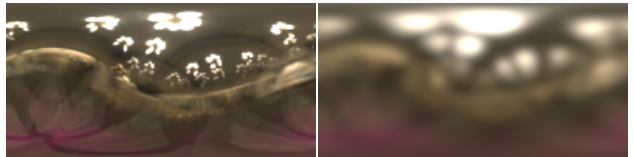
## S4. Theoretical Considerations

### S4.1. Figure 3 from Main Paper

Figure 3 from the main paper is based on a simplified “flat-land” version of our problem, where the occluders and materials are known. A circular object with radius 1 and Lambertian BRDF with albedo 1 is placed at the origin, and illuminated by a (monochromatic) environment map  $L(\phi)$ . An occluder of some angular width  $\Delta\theta$  and central angle  $\theta$  is



(a) Sample environment maps from the same captured scene



(b) Mean environment map      (c) Blurred mean environment map

Figure S10. Our blurring process for environment maps provided with the Stanford-ORB dataset [18]. The environment maps provided with the dataset are not aligned since they were captured by a light probe not co-located with the object, as shown by the two misaligned images in panel (a), which makes them unusable for computing error metrics for illumination recovery. We therefore first: (b) average all environment maps provided along with each scene, and then (c) blur them using a normalized spherical Gaussian.

then placed at some radius  $r$  from the origin, blocking all shadow rays from the object to the illumination.

We parameterize the surface of the object using a single

angle parameter  $\lambda \in [0, 2\pi)$ , *i.e.* the “surface” of the disk is the set of points  $\mathcal{S} = \{(\cos \lambda, \sin \lambda) : \lambda \in [0, 2\pi)\}$ . The “image” at the point  $\lambda$  at time  $t$  is then:

$$I_t(\lambda) = \int_0^{2\pi} L(\phi) M_t(\lambda, \phi; \theta_t, \Delta\theta_t) (\cos(\lambda - \phi))_+ d\phi, \quad (\text{S16})$$

with the “flatland” (unnormalized) BRDF set to 1, and the occluder value at time  $t$  is:

$$M_t(\lambda, \phi; \theta_t, \Delta\theta_t) = \mathbb{1}\left[|\psi(\lambda, \phi) - \theta_t| < \frac{\Delta\theta_t}{2}\right], \quad (\text{S17})$$

with:

$$\psi(\lambda, \phi) = \arctan\left(\frac{\sin \lambda + t(\lambda - \phi) \sin \phi}{\cos \lambda + t(\lambda - \phi) \cos \phi}\right), \quad (\text{S18})$$

$$t(\delta) = \sqrt{\cos^2(\delta) + r^2 - 1} - \cos(\delta).$$

We then discretize the integral in Equation S16 for every  $t = 1, \dots, T$ , and stack the resulting equations:

$$\underbrace{\begin{bmatrix} \mathbf{I}_1 \\ \vdots \\ \mathbf{I}_T \end{bmatrix}}_{\mathbf{I}} = \underbrace{\begin{bmatrix} A_1 \\ \vdots \\ A_T \end{bmatrix}}_A \boldsymbol{\ell}, \quad (\text{S19})$$

where  $\mathbf{I}_t \in \mathbb{R}^N$  is a vector of rendered values  $\{I_t(\lambda)\}$  at time  $t$ ,  $\boldsymbol{\ell} \in \mathbb{R}^M$  is a vector of illumination values  $\{L(\phi)\}$ , and  $A_t \in \mathbb{R}^{N \times M}$  contains the corresponding values of the occluders  $M(\lambda, \phi; \theta_t, \Delta\theta_t)$  at time  $t$  and the cosine lobe  $(\cos(\lambda - \phi))_+$ .

Figure 3 then shows the singular values of  $A$  (which are the square roots of the eigenvalues of  $A^\top A$ ), normalized to have maximum value 1, for a few scenarios:

- “No occlusions”: The scenario described above with a single observation ( $T = 1$ ) but without an occluder, *i.e.* where the mask is set to 1 everywhere (or, alternatively,  $\Delta\theta = 2\pi$ ).
- “1 observation”: The scenario described above, with a single occluder ( $T = 1$ ) placed at  $\theta = 0$ .
- “ $T$  observations”, for  $T \in \{2, 8, 16, 32\}$ :  $A$  contains  $T$  stacked matrices corresponding to a discretization of Equation S16, each one for a different occluder location  $\theta_t$  uniformly spread over  $[0, 2\pi)$ , and the same  $\Delta\theta$ .

The exact values used for Figure 3 are occluder width  $\Delta\theta = 0.7$  rads, placed at distance  $r = 10$  from the origin, with 512 samples per observation.

## S4.2. Spherical Harmonics Intuition

In the main paper we show that expressing the occluder masks in the basis of spherical harmonics, instead of the standard basis, leads to significantly improved results. The reasoning behind that can be illustrated in 1D using the

Fourier basis (sines and cosines), since spherical harmonics are the spherical equivalent of the Fourier basis.

Similar to Equation S19, the integral in Equation S16 can also be discretized and combined into a linear equation in the occluder values. However, the resulting matrix is block-diagonal, with each  $M_t$  only depending on  $\mathbf{I}_t$  at the same time  $t$ :

$$\mathbf{I}_t = B \mathbf{m}_t, \quad (\text{S20})$$

where  $\mathbf{I}_t$  is again a vector of image values at time  $t$ ,  $\mathbf{m}_t$  is a vector of occluder values at the same time, and  $B$  describes their linear relation according to a discretization of Equation S16.  $B$  is composed of a circulant matrix (corresponding to the BRDF’s convolution), multiplied by a diagonal matrix with the illumination values  $L$  along its diagonal. This means that when the illumination is uniform,  $B$  is also a circulant matrix, and is therefore diagonalized by the Discrete Fourier Transform (DFT) matrix. This makes the problem diagonal in the Fourier basis, which makes adaptive optimizers such as Adam [16] especially effective.

When the illumination is non-uniform,  $B$  is not generally diagonalized by the DFT matrix. However, for natural lighting, when the DFT matrix  $F$  is applied to  $B$ , most of the matrix’s energy is along the diagonal elements, *i.e.* the elements of the  $i$ th row of  $F B F^\top$  are maximized by the  $i$ th element. See Figure S11 for examples showing rows of the  $F B F^\top$  matrix for different illumination spectra. The fact that the Fourier basis (or in the original problem, the basis of spherical harmonics) “nearly-diagonalizes” the problem is the reason for its effectiveness.

Furthermore, since we are not interested in recovering the occluders in regions with low illumination (see right of Figure 5 of the main paper), it is informative to consider the problem of estimating the masked illuminant:

$$\mathbf{I}_t = C(\boldsymbol{\ell} \circ \mathbf{m}_t), \quad (\text{S21})$$

where  $\circ$  denotes elementwise multiplication, and  $C$  is a circulant matrix corresponding to the (discretized) convolution in Equation S16. The problem of estimating  $\boldsymbol{\ell} \circ \mathbf{m}_t$  given  $\mathbf{I}_t$  is in fact diagonalized by the Fourier basis, because  $C$  is circulant.

## S4.3. Image Pyramid Intuition

The set of linear equations satisfied by the illumination  $\boldsymbol{\ell}$  described in Equation S19 cannot be separately solved for each  $t$ , and therefore we must consider the entire system of equations described by the matrix  $A \in \mathbb{R}^{NT \times M}$ . Figure S12 shows the entries of  $A^\top A$ , for different choices of  $T$ . The entries of the matrix exhibit a block structure in the standard basis, *i.e.* without applying the DFT matrix to it, with the sizes of the blocks becoming smaller for larger values of  $T$ . In fact, for very large values of  $T$  the matrix approaches a circulant matrix, which is diagonalized by the

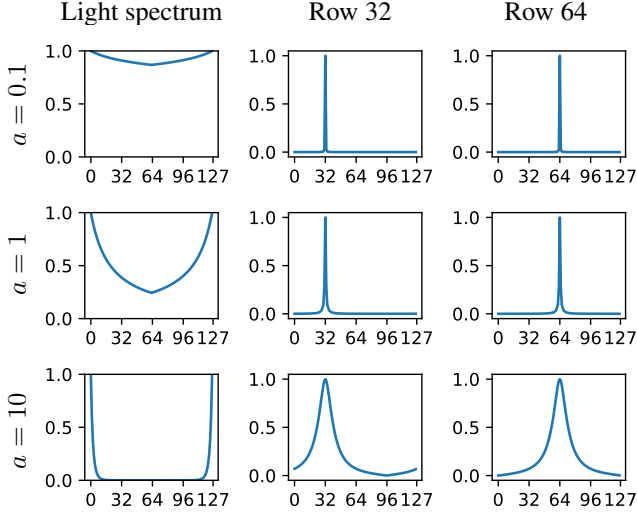


Figure S11. Most of the energy of the matrix  $B$  is concentrated around the diagonal elements in the Fourier basis. Each row shows, from left to right: the incident illumination spectrum (which modifies  $B$ ), and the magnitude of the elements of the 32nd and 64th rows of  $FBT^T$ , where  $F$  is the DFT matrix. Each row corresponds to a  $1/f^a$  spectrum for  $a = 0.1, 1, 10$ . See text for more details.

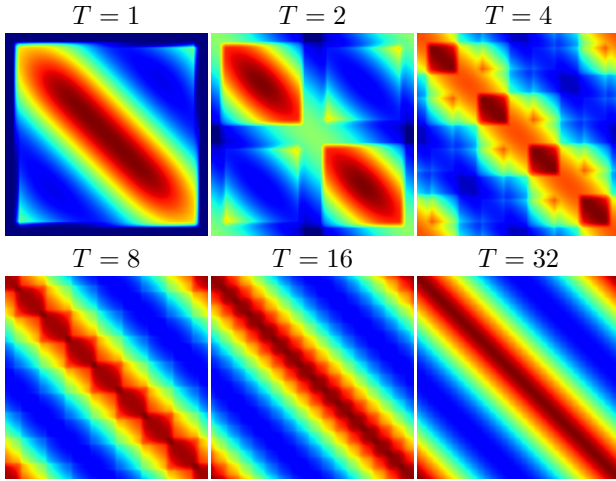


Figure S12. The structure of  $A^T A$  for the light transport matrix  $A$ , plotted for different values of the number of observations  $T$ . Large values in red, small values in blue.

DFT matrix. However, we wish our method to not make use of that fact and work even for a sparse set of observations, and therefore a natural choice for parameterizing the illumination  $\ell$  is by using an image pyramid, which matches the block structure of  $A^T A$  for general  $T$  values.

## S5. Geometry Estimation Method

We estimate unknown object geometry from input images by recovering a Neural Radiance Field (NeRF) [25] of the object, based on the Instant Neural Graphics Primitives [26] representation. To improve the recovered normal vectors (computed as the negative normalized gradient of the volume density field), we use the normals orientation regularization and MLP-predicted normals technique from Ref-NeRF [41].

## S6. Data Specification

All BRDF parameters (RGB albedo, roughness, and the specular reflectance at normal incidence), are output by a coordinate-based MLP taking in a positionally-encoded location:

$$\gamma(\mathbf{x}) = (\sin(\mathbf{x}), \cos(\mathbf{x}), \sin(2\mathbf{x}), \cos(2\mathbf{x}), \dots, \sin(64\mathbf{x}), \cos(64\mathbf{x})). \quad (\text{S22})$$

The MLP has 4 layers with 128 hidden units each, with ReLU nonlinearities. The weights are initialized around zero, and the output BRDF parameters are obtained by mapping the MLP’s output through a sigmoid function, meaning that they are all initialized around a value of 0.5.

The spherical harmonic coefficients  $\{a_{\ell m}\}$  from Equation 6 of the main paper are all set to zero, but a positive bias of 100 is added to the pre-sigmoid value (which can equivalently be done by initializing  $a_{t00}$  to a constant). We find that initializing the masks to be close to 1 everywhere improves our method’s performance and prevents it from getting stuck in a local minimum.

The environment map pyramid levels are also all initialized to zero, such that the illuminant is set to all-ones, due to the exponential nonlinearity (see Equation 7 in the main paper).

## S7. Additional Data Details

The geometry of the objects in the paper and their textures originated in the following BlendSwap models:

1. Potatoes: created by *mik1190*, CC0 license (model #15725)
2. Chair: created by *IDInc*, CC0 license (model #8261).
3. Mannequin: created by *salimrached*, CC0 license (model #27747).
4. Toad: created by *arenyart*, CC0 license (model #13078).
5. Plant: created by *New Enemy*, CC0 license (model #30071).
6. Giraffe: created by *amx360*, CC-BY license (model #29651).

The environment maps are from the following Poly Haven assets:



1. "Canary Wharf": created by *Andreas Mischok*, CC0 license.
2. "Abandoned Factory Canteen 01": created by *Sergej Majboroda*, CC0 license.
3. "Outdoor Umbrellas": created by *Sergej Majboroda*, CC0 license.
4. "Thatch Chapel": created by *Dimitrios Savva, Jarod Guest*, CC0 license.
5. "Evening Road 01 (Pure Sky)": created by *Jarod Guest, Sergej Majboroda*, CC0 license.
6. "Marry hall": created by *Sergej Majboroda*, CC0 license.