Photometric Ambient Occlusion for Intrinsic Image Decomposition

Daniel Hauagge, *Student Member, IEEE,* Scott Wehrwein, *Student Member, IEEE,*
Kavita Bala, *Member, IEEE,* and Noah Snavely, *Member, IEEE*

Abstract—We present a method for computing *ambient occlusion* (AO) for a stack of images of a Lambertian scene from a fixed viewpoint. Ambient occlusion, a concept common in computer graphics, characterizes the local visibility at a point: it approximates how much light can reach that point from different directions without getting blocked by other geometry. While AO has received surprisingly little attention in vision, we show that it can be approximated using simple, per-pixel statistics over image stacks, based on a simplified image formation model. We use our derived AO measure to compute reflectance and illumination for objects without relying on additional smoothness priors, and demonstrate state-of-the art performance on the MIT Intrinsic Images benchmark. We also demonstrate our method on several synthetic and real scenes, including 3D printed objects with known ground truth geometry.

Index Terms—Ambient occlusion, intrinsic images, image stacks, pixel statistics

1 Introduction

Many vision methods estimate physical properties of a scene from images taken under varying illumination. Some notable examples include recovering surface normals using photometric stereo [1], [2], [3], recovering diffuse reflectance and illumination as intrinsic images [4], [5], and computing low-dimensional models of appearance of objects and scenes [6], [7]. However, these methods typically disregard the effect of the *local visibility* of illumination in determining shading. Further, many of these methods require calibrated setups (e.g., known lighting directions), special priors (e.g., smoothness of surface reflectance), or limiting assumptions (e.g., no cast shadows).

In our work[1], we revisit such estimation problems by posing the following question: what can we tell about a scene point simply by observing its appearance under many different, unknown illumination conditions? The appearance of a point over such an image stack depends on many factors, such as the point's albedo and the distribution of illuminations. However, a key observation is that the local visibility of a point—i.e., its accessibility to light from different directions, often modeled as ambient occlusion (AO) in computer graphics—is also an important property in determining its appearance in images. We show that we can estimate ambient occlusion in Lambertian scenes directly from image observations, by introducing a simple pixel-wise, aggregate statistic (κ in Figure 1), and relating this statistic to ambient occlusion. To do so, we consider a physical model of a point with a cone of visibility to the hemisphere, lit by a moving point light and constant ambient light over the image stack. We then combine this model with our statistic to infer ambient occlusion for each scene point. This kind of lighting visibility is often treated as a nuisance in computer vision methods, and in many cases is simply ignored. In contrast, we explicitly model such visibility for each scene point, and use it to aid in estimating other physical parameters, such as surface albedo. The result is a *photometric* approach to estimating ambient occlusion and albedo.

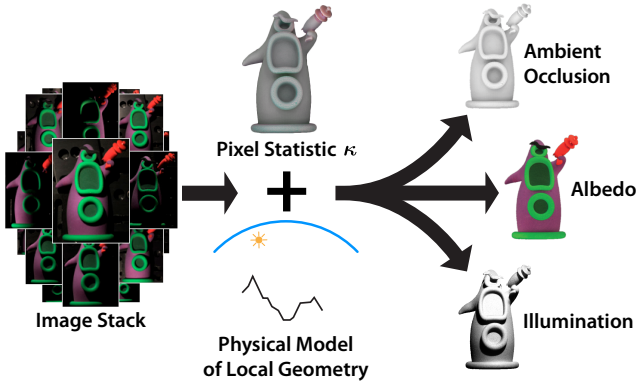

Fig. 1. Our method takes as input a stack of images captured with varying, unknown illumination and computes a per-pixel statistic (κ) over this stack. This statistic is then combined with simple physical model of the local geometry at each point and illumination to obtain an estimate of the local visibility. Local visibility is then used together with the average image to obtain an estimate for per-point albedo (reflectance), which itself can be used to compute illumination for the original input images.

Our method has several key properties: we do not require knowledge of light positions, explicit scene geometry, or surface normals. The setup for acquisition is simple, requiring a point light source and a camera. However, we do assume that light source positions vary

1. This publication is an expanded version of [8].

uniformly over the full hemisphere, although in practice we achieve good results even when this assumption does not hold. Note that we use the term image "stack" to refer to a set of images of the same scene lit under varying illumination, but captured from the same viewpoint. No frame-by-frame coherence or ordering is implied.

Our work has two main contributions:

- A per-pixel, image-space approach to estimating ambient occlusion that does not require information about the underlying geometry.
- A new method for intrinsic image decomposition using our model of ambient occlusion, accounting for local visibility at each point.

We demonstrate our method in experiments on several scenes. These include artificially generated images from a physically based renderer, as well as real objects captured in a laboratory environment. Our experiments on real objects include a validation on 3D printed objects with known geometry, including the TENTACLE dataset in Figure 1. In addition, we show that our method—despite its simplicity and its per-pixel analysis of a scene, without additional smoothness priors—outperforms current approaches on the MIT intrinsic images benchmark [9]. This demonstrates the utility of reasoning about AO when measuring properties of scenes from images. We also present a detailed analysis of several aspects of our model, including its performance with respect to albedo color, number of input images, and the effects of inter-reflections. Finally, we discuss limitations and avenues for future work such as generalizing the model to handle outdoor illumination.

2 AMBIENT OCCLUSION

Ambient Occlusion (AO) [10] is a measure of light accessibility commonly used in computer graphics to properly account for ambient illumination. Formally, for a single scene point x, AO is the integral over the hemisphere:

$$AO(x) = \frac{1}{\pi} \int_{\Omega} V(x, \vec{\omega}) \langle \vec{n}, \vec{\omega} \rangle d\omega \tag{1}$$

of the local visibility function $V(x, \vec{\omega})$ (i.e. $V(x, \vec{\omega}) = 1$ if there are no occluders between point x and the environment in direction $\vec{\omega}$, $V(x, \vec{\omega}) = 0$ otherwise) weighted by the dot product $\langle \vec{n}, \vec{\omega} \rangle$ between direction $\vec{\omega}$ and the point normal $\vec{n}$. For an example, see Figure 8. At points where most of hemisphere is occluded, e.g., in a deep crevice, V is mostly 0 and so AO is close to 0, while for points whose visibility of the hemisphere is unoccluded, AO is 1. If the albedo at x is ρ, the measured radiance due to constant, ambient illumination with intensity l_a can be expressed as:

$$I_a = \rho \pi l_a AO \tag{2}$$

Note that this only considers the first bounce of light (direct illumination), and as such does not account for interreflections.

Two properties of ambient occlusion that are useful in computer vision are: (1) it is independent of surface albedo, and so variation and discontinuities are due only to scene geometry, and (2) it explains in a simple way why regions with same albedo can have different intensities even when lit with uniform illumination [11].

In computer graphics, the main focus is on computing AO in 3D scenes to render images [12], [13], [14]. In contrast, we are interested in *estimating* AO from a set of images illuminated by a varying, unknown light source.

3 RELATED WORK

Ambient occlusion in computer vision. Ambient occlusion has received relatively little attention in computer vision. Some examples of its use include early work in shape-from-shading [11], where it was used in models of images under diffuse illumination, as well as more recent work that considers AO in various applications. In the context of high-quality face capture, Beeler et al. [15] and Aldrian and Smith [16] model AO by assuming a uniform, constant, light source, and require an initial estimate of the geometry. In the area of multi-view stereo, Wu et al. assume that a scene consists of a single albedo, and so the scene brightness under uniform area lighting is itself a good approximation to AO (e.g., darker regions are more occluded) [17]. Our work is also related to methods that recover shape from AO [11], [18], and our algorithm could potentially be used to generate inputs to such methods.

Intrinsic image decomposition. For the problem of intrinsic image decomposition from large photo collections, Laffont et al. require accurate estimates of the albedo for a sparse set of 3D scene points [19]. To account for points that are darker due to AO, they compute AO explicitly by generating and analyzing a 3D scene reconstruction. Similarly, Kyong et al. compute light accessibility in order to decompose frames from video sequences, in their case geometry is recovered by using a depth camera [20].

In contrast to these methods, we do not explicitly model geometry, and instead reason about AO purely from observed pixel values. This yields a very simple approach that could be used as a pre-process to account for light visibility in other vision algorithms.

Our work is also related to methods that analyze pixel intensity variation in images under varying illumination. Weiss proposed a method for intrinsic images from image sequences [4], derived from a model of edge intensities. In that work, a final step involves integrating a gradient field to compute a reflectance image. In our experience, and in agreement with other reports [9], this integration performs poorly in the presence of soft and persistent shadows (exactly the kind caused by AO), and we find that it can also propagate noise across the image. In contrast, our method explicitly models one cause of soft shadows (namely AO), and does not require a final integration step, which we find makes the algorithm

more robust. For outdoor scenes illuminated by the sun, Sunkavalli et al. recover albedo and normals by directly tracking the intensity of pixel values over time [21]. While they use heuristics to determine whether a pixel is in shadow, our method makes no such hard decisions, instead reasoning about statistics over the entire image sequence. In more recent work, Barron and Malik optimize for reflectance, shape, and illumination from single images under strong priors on illumination and color of natural scenes [22]. In contrast, our method operates at a per-pixel level and does not make assumptions about the texture in the scene.

Photometric stereo. Photometric stereo techniques [23] are similar to our method in their setup and the fact that they estimate albedo, but differ in that they recover different information about shape (surface normals), compared to our work. Our approach is especially related to uncalibrated photometric stereo, in which the light sources are unknown [24], [1]. A key challenge in photometric stereo is dealing with shadows, either by detecting them in some manner [25], [2] (a non-trivial problem with surfaces of varying albedo or complex self-occlusions), or treating them as a source of noise or outliers [26], [27]. Photometric stereo methods have become increasingly robust to shadows and other sources of error (such as specularities) through more sophisticated methods. In contrast, we show that by adopting a simple explicit model of shadowing through AO, we can achieve robustness using a very simple approach. We provide a more detailed discussion our method in relation to photometric stereo in Section 8.

Sunkavalli et al. reason about lighting visibility of surface points, by clustering them into "visibility subspaces" that see a common set of lights [2]. However, they use an implicit model of lighting visibility that grows in complexity as the number of lighting conditions increases. In contrast, our method relies on a simple per-pixel measure of ambient occlusion that becomes more robust as more images are added. In addition, our model incorporates ambient illumination as well as directional lighting.

4 A Model for Ambient Occlusion in Image Stacks

We now describe how to obtain a simple approximation to ambient occlusion (AO) by observing pixel intensities in multiple images under varying directional lighting. We first introduce a physically-based image formation model for our measure of AO, then use this model to derive AO and albedo from image sequences.

4.1 Inputs and image formation model

Our method takes as input a set of images, $I_1, I_2, \ldots, I_n$, captured from a fixed camera observing a static, Lambertian scene. The scene is lit by an unknown, directional light source that changes from image to image, together with a uniform ambient light source; both have constant intensity over time. We further assume that the distribution of directional light sources across the image stack is uniform over the hemisphere. The images are radiometrically calibrated and so the image intensity $I(x)$ at each pixel x is proportional to the radiance at a given scene point under a particular illumination. Because the camera is static, the same pixel x records radiance for the same scene point in each image. In the following derivation the images are treated as monochromatic without loss of generality.

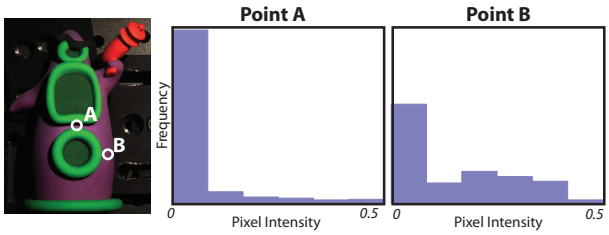

Fig. 2. Histogram of pixel intensities for two points of TENTACLE over an image stack (only blue color channel). Notice that even though the two points have very similar albedos their histograms are quite different due to local visibility. Point A is mostly occluded with respect to the light source, so its intensity values are in general lower.

A key idea in our work is that for a given pixel x, the measured radiances over all images are drawn from an underlying distribution that we refer to as its *pixel intensity distribution* (PID). This distribution of pixel intensities at a point is related to the distribution of illuminations over the image stack, as well as to the albedo of that point and to the surrounding geometry (which can occlude the light source from the point of view of that point). Figure 2 shows an example of observed PIDs in an image stack for two points. For example, a point in a deep concavity will very often appear dark, because light rarely reaches it (only when the light is shining straight down into the hole). Such a point will have a PID with mostly low intensity values. (For example, consider point A in Figure 2.) The intuition then, is that the samples we record give us information about a pixel's PID, which in turn reveals information about surface albedo and ambient occlusion. As we capture images lit under more and more possible directions, we begin to capture the actual underlying PID of a pixel.

As a useful summary of a PID, we introduce a statistic for a single pixel x over time, which we denote κ:

$$\kappa(x) = \frac{\mathcal{E}[I(x)]^2}{\mathcal{E}[I(x)^2]} \tag{3}$$

where $\mathcal{E}[\cdot]$ is the expectation operator over the set of images. That is, κ is the square of the expected (average) intensity value for that pixel, divided by the expected squared pixel intensity; this quantity is related to the *coefficient of variation*, a normalized measure of variance

used in statistics.[2] Fig. 8 (leftmost image) illustrates κ for an example image stack. In what follows, we show that this simple ratio of statistics over recorded intensities yields an approximation to ambient occlusion; to understand this relationship between κ and ambient occlusion, we first describe our image formation model, then relate this to a physical model of local scene geometry.

For a Lambertian scene, an image formation model commonly used in intrinsic images literature is:

$$I(x) = \rho(x)L(x) \tag{4}$$

where $I(x) \in \mathbb{R}^+$ is the observed radiance at point x in the image, $\rho(x) \in [0, 1)$ is the diffuse albedo, and $L(x) \in \mathbb{R}^+$ is a factor that depends on both light and geometry.[3]

Over our sequence of images I, $\rho(x)$ is constant and greater than zero, while $L(x)$ varies due to changes in illumination. Under these assumptions, we can substitute Eq. (4) into the definition of our κ statistic in Eq. (3) to obtain:

$$\kappa = \frac{\mathcal{E}[\rho L]^2}{\mathcal{E}[\rho^2 L^2]} = \frac{\cancel{\rho^2}\mathcal{E}[L]^2}{\cancel{\rho^2}\mathcal{E}[L^2]} \tag{5}$$

(for simplicity, we do not explicitly write the dependence on x, but as before κ is a statistic defined per pixel across the image stack). Thus, κ depends only on the lighting factors L, and not on albedo.

What range of values can κ take on? Because κ is the quotient of non-negative numbers, it follows that $\kappa \geq 0$. By observing that $\text{Var}(I) = \mathcal{E}[L^2] - \mathcal{E}[L]^2 \geq 0$ we can also show that $\kappa \leq 1$. For points that *never* receive light $\mathcal{E}[L] = 0$ and $\kappa = 0$ (one can arrive at this via a limit analysis). For points whose illumination term never changes we have that $\text{Var}[I] = \mathcal{E}[L^2] - \mathcal{E}[L]^2 = 0$, which implies $\mathcal{E}[L^2] = \mathcal{E}[L]^2$ and therefore $\kappa = 1$. This behavior suggests that κ might be useful as a measure of ambient occlusion at a point.

4.2 A physical model explaining κ

So far we have shown that κ is independent of albedo and is bounded. But what exactly does κ tell us about a scene point? As a statistic, κ relates to the geometry and visibility at a point; to show this, we introduce a simplified geometry and lighting model to connect κ to a physical measure of local visibility.

Our model assumes that the visibility at a point can be approximated by a cone of angle α, resulting in a very simple relation between the local geometry paramter α and AO:

$$AO(x) = \sin^2(\alpha(x)) \tag{6}$$

This idea, along with our illumination model, is illustrated in Figure 3, where a point x on a Lambertian surface, is observed by camera c while illuminated by two light sources: a directional light with intensity l_d, and a background ambient illumination with constant intensity l_a. One can think of these two components as roughly similar to a "sun" (directional) and a "sky," (ambient) light source, respectively. Surface geometry around the point blocks all light outside the cone with angle α from reaching x. We refer to this angle $\alpha(x)$ as the *local visibility angle* for point x. Further, across our input images, we assume that the directional light uniformly samples the full hemisphere, so each measure of the radiance of x captured by the camera corresponds to a different (unknown) position for the light l_d. Given these assumptions, and sufficient samples of images under different illumination conditions, $\kappa(x)$ only depends on the local visibility angle $\alpha(x)$, and we can derive the relationship between κ and α.

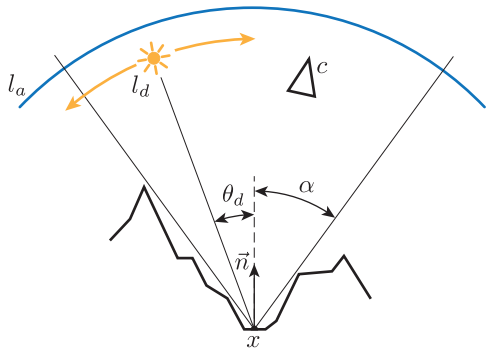

Fig. 3. A point x on a Lambertian surface is observed by camera c and illuminated by a distant, moving light source with intensity l_d, and a constant ambient term of intensity l_a. The local visibility is approximated by a cone with angle α. If the light source angle θ_d with the surface normal $\vec{n}$ is larger than α, light is blocked and does not reach point x at the bottom of the valley.

We now derive the relationship between κ and α given our model. To begin, each image I is the sum of the contributions from both light sources:

$$I = I_d + I_a \tag{7}$$

The directional component I_d varies from image to image and depends on the angle $\theta_d(t)$ between the light source direction $\vec{\omega}_d(t)$ and the surface normal $\vec{n}$ at a point, and whether the light is blocked by other geometry. This component is given by:

$$I_d(t) = \rho l_d V_\alpha(\vec{n}, \vec{\omega}_d(t))\langle \vec{n}, \vec{\omega}_d(t)\rangle \tag{8}$$

$$= \rho l_d V_\alpha(\theta(t)) \cos\theta_d(t) \tag{9}$$

where V_α is the light visibility term: $V_\alpha(\theta) = 1$ if $\theta \leq \alpha$ (i.e., the light enters the visibility cone), ands $V_\alpha(\theta) = 0$ otherwise. The ambient component I_a is constant across the image stack for a given point, and is proportional to the projected solid angle of the local visibility angle α. In particular, from Eqs. (1) and (2) we can integrate the ambient illumination over the visible potion of the

2. The coefficient of variation, c_v, is defined as $\frac{\sigma}{\mu}$, so the statistic $\kappa = \frac{1}{1+c_v^2}$.

3. This image formation model should not be confused with the model classically used in photometric stereo, where L is a vector describing the light direction; here, the scalar $L(x)$ is simply the incident illumination at point x.

hemisphere to derive a closed form relationship between I_a and α at a given point:

$$I_a = \rho \int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\alpha} l_a \cos(\theta) \sin(\theta) d\theta d\varphi = \rho l_a \pi \sin^2 \alpha \quad (10)$$

Given this model for I_d and I_a, to relate κ to our physical parameter α, we compute the expectations in Eq. (5) over images under varying light source positions:

$$\mathcal{E}[I] = \mathcal{E}[I_d] + \mathcal{E}[I_a] = \mathcal{E}[I_d] + I_a$$
$$\mathcal{E}[I^2] = \mathcal{E}[(I_d + I_a)^2] = \mathcal{E}[I_d^2] + 2I_a\mathcal{E}[I_d] + I_a^2$$

where we use the linearity of expectation, $\mathcal{E}[\cdot]$, and the assumption that I_a does not change over the image stack.

Finally, we can compute $\mathcal{E}[I_d]$ and $\mathcal{E}[I_d^2]$ in closed form by integrating over the visible cone of angles at the point, assuming the point light is uniformly distributed over the hemisphere for the image stack:

$$\mathcal{E}[I_d] = \frac{1}{2\pi} \int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\alpha} I_d \sin\theta d\theta d\varphi = \frac{1}{2} \rho l_d \sin^2(\alpha) \quad (11)$$

$$\mathcal{E}[I_d^2] = \frac{1}{2\pi} \int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\alpha} I_d^2 \sin\theta d\theta d\varphi = -\frac{1}{3} \rho^2 l_d^2 \left(\cos^3(\alpha) - 1\right)$$

Given these equations, we can derive κ in terms of α as:

$$\kappa(\alpha) = \frac{\mathcal{E}^2[I]}{\mathcal{E}[I^2]} = \frac{(\mathcal{E}[I_d] + I_a)^2}{\mathcal{E}[I_d^2] + 2I_a\mathcal{E}[I_d] + I_a^2}$$
$$= \frac{3}{4} \frac{(2\pi l_a + l_d)^2 \sin^4(\alpha)}{3\pi l_a(\pi l_a + l_d)\sin^4(\alpha) - l_d^2\cos^3(\alpha) + l_d^2} \quad (12)$$

which can be further simplified by noting that κ actually depends on the ratio of light source intensities $l_a/l_d = f$ and not their absolute values. After substituting $l_a = f l_d$ into Eq. (12) and simplifying we arrive at:

$$\kappa(\alpha) = \frac{3}{4} \frac{(2\pi f + 1)^2 \sin^4(\alpha)}{1 + 3\pi f(\pi f + 1)\sin^4(\alpha) - \cos^3(\alpha)} \quad (13)$$

To get a better intuition for κ, we consider its behavior under two special cases, $l_d = 0$ and $l_a = 0$, which correspond to $f \to \infty$ and $f = 0$ respectively:

$$\kappa|_{l_d=0} = 1 \qquad \kappa|_{l_a=0} = \frac{3\sin^4(\alpha)}{4 - 4\cos^3(\alpha)}$$

In other words, if there is no directional illumination component (i.e., $l_d = 0$) then $\kappa(\alpha)$ is always 1, and α cannot be recovered from pixel measurements alone. This case corresponds to all images in our stack being identical, with the scene lit only by an ambient term, so there is no variation in intensity for each point. In this case there is no way of directly disambiguating between shading and reflectance.

If there is no ambient component (i.e., $l_a = 0$) then κ increases monotonically in the valid range for α and is independent of l_d (as long as $l_d > 0$). In Figure 4 we show $\kappa(\alpha)$ for a few different values of f.

One interesting property of the curves in Figure 4 is that they have different κ values for $\alpha = 90°$, ranging from 0.75 to 1 as f goes from 0 to ∞. This means that if we know that a given point in our scene is not occluded by any other geometry (i.e., $\alpha = 90°$, then we can recover f directly from the value of κ for that point:

$$f(\kappa)|_{\alpha=90°} = \frac{\sqrt{3}\sqrt{\kappa - \kappa^2} + 3\kappa - 3}{6\pi(1 - \kappa)} \quad (14)$$

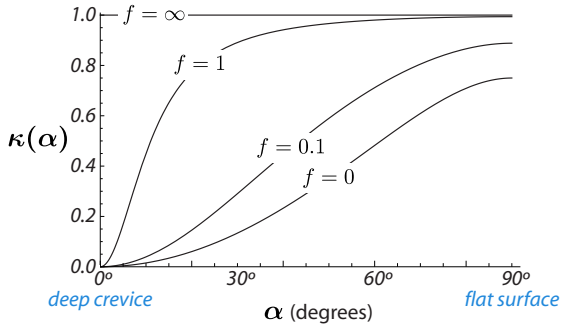

Fig. 4. $\kappa(\alpha)$ for different ratios of ambient to direct light f. Note that as $f \to \infty$ ($l_d = 0$) we have a constant curve ($\kappa(\alpha) = 1$) so information about α cannot be recovered.

In summary, we have derived a relation between the statistic κ, and the ambient occlusion at a point, using a physical model of a crevice (with a cone of visibility characterized by α) lit by a varying directional light, and a constant ambient light over a stack of images. No assumptions of smoothness or geometric reconstruction are required to derive this parameter. As we show later, this physical model, though simple and an approximation of real scenarios, works surprisingly well in characterizing the visibility at points in a scene.

5 Algorithm

In this section we use our model to compute a per-pixel local visibility angle $\alpha(x)$ and albedo $\rho(x)$ given a stack of images of the same scene under varying illumination. While our derivation has assumed grayscale images, our algorithm makes use of additional constraints from the three different color channels; while we solve for a color albedo and a per-color-channel value for f, α is constant for a given point, and f is assumed constant over the image stack and across pixels. Our full algorithm is described below, and illustrated in Figure 5.

We first compute κ using Eq. (3) by assuming $f_0 = 0$ (i.e., ambient lighting is negligible) to derive an initial α_0 using Eq. (13). We then refine $\alpha(x)$ (one value per pixel) and f (one value per color channel, but constant across pixels) by minimizing the objective function:

$$\alpha_1, f_1 \leftarrow \min_{\alpha,f} \sum \|\kappa_{obs} - \kappa(\alpha_0, f_0)\|^2 \quad (15)$$

where the subscript obs stands for "observed". In other words, we compute α and f so as to best explain the observed statistic κ. In total we have $n_c \times n_p$ equations, where n_c is the number of color channels and n_p the

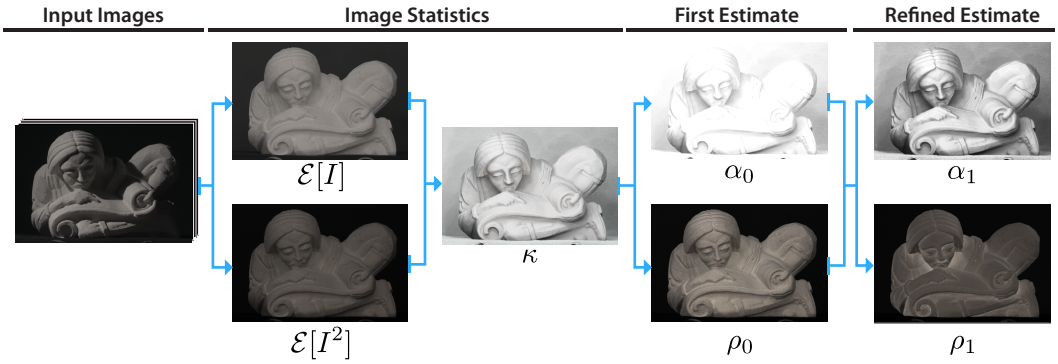

Fig. 5. A depiction of the full algorithm for computing the local visibility angle α and the reflectance ρ. Arrows show how information flows in our pipeline. Starting with an image stack we compute $\mathcal{E}[I]$ and $\mathcal{E}[I^2]$, which are used to compute κ. We then proceed to obtain a first estimate of the local visibility angle and reflectance, which are then refined using a non-linear optimization.

number of pixels, and $n_p + n_c$ variables, one α per pixel and n_c variables corresponding to the direct to ambient illumination ratios f. Eq. (15) defines a non-linear least squares problem, which we minimize using the `trust-region-reflective` mode of MATLAB's `lsqnonlin` function.

Given our final estimates α_1 and f_1, we compute estimates for the albedo $\rho(x)$ at each point from Eqs. (11) and (10). We express albedo as a function of the expected pixel value, the ratio f, the local visibility angle α, and the intensity l_d of the direct component:

$$\rho = \frac{2\mathcal{E}[I]}{l_d \sin^2(\alpha)\,(1 + 2f\pi)} \tag{16}$$

Note that there is an inherent ambiguity between light source intensity l_d and the scene albedo, so we can only estimate albedo up to a scale factor. Therefore, we assume that $l_d = 1$ to obtain ρ_1, our final estimate of the albedo.

6 Results

We begin by demonstrating results for our algorithm on various datasets (Section 6.1) and exploring the different outputs the algorithm produces. In Section 6.2 we use an object with known geometry to measure the error in our estimate of ambient occlusion. In Section 6.3 we evaluate our estimate of albedo by comparing our algorithm with others using the MIT Intrinsic Images benchmark [28]. Finally, Section 7 provides a detailed analysis of various aspects of our algorithm on a specially manufactured test object with crevices of varying (and known) depth; this includes an analysis of convergence rate as the number of images grows, and the impact of error factors such as interreflections.

6.1 Image Decomposition

Figure 6 shows image decomposition results on several datasets, including image stacks used in prior work. For each dataset we show ambient occlusion, reflectance ρ, and the illumination. More results can be found on our project webpage [29].

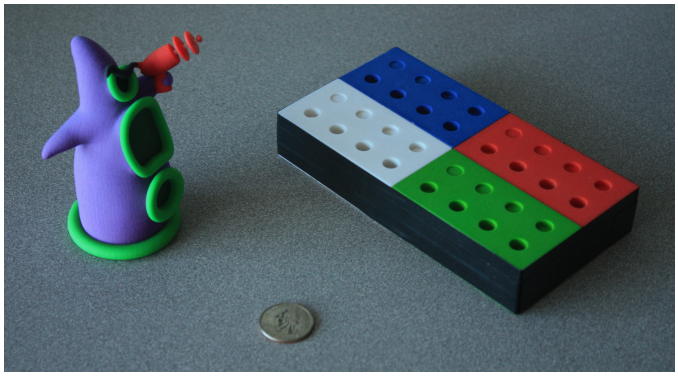

Fig. 7. 3D printed test objects TENTACLE and LIGHTWELL, together with a quarter dollar coin (for scale). Black tape on the sides of LIGHTWELL was added to reduce sub-surface scattering that resulted from light shining on the side of the object.

Datasets. The first dataset, TENTACLE, contains 350 images of a 3D printed object with known geometry. The light source position in TENTACLE was precisely controlled by a mechanical gantry allowing us to sample uniformly random positions over the full hemisphere. The known geometry lets us compare against ground truth. Note that, although we use a spherical gantry for capturing this dataset, this is for the purposes of evaluation, and our method is motivated by much more unstructured capture setups. The datasets below are more representative of the datasets we would expect to use in practice.

The other datasets are public datasets that violate the assumptions of our model in various ways. FROG and SCHOLAR, from [2], contain 47–48 images lit under varying directional lights that do not cover the full hemisphere. FACE from the Yale Face Database B+ [30], contains 64 images with light positions over a range of angles. This scene violates our assumptions in that

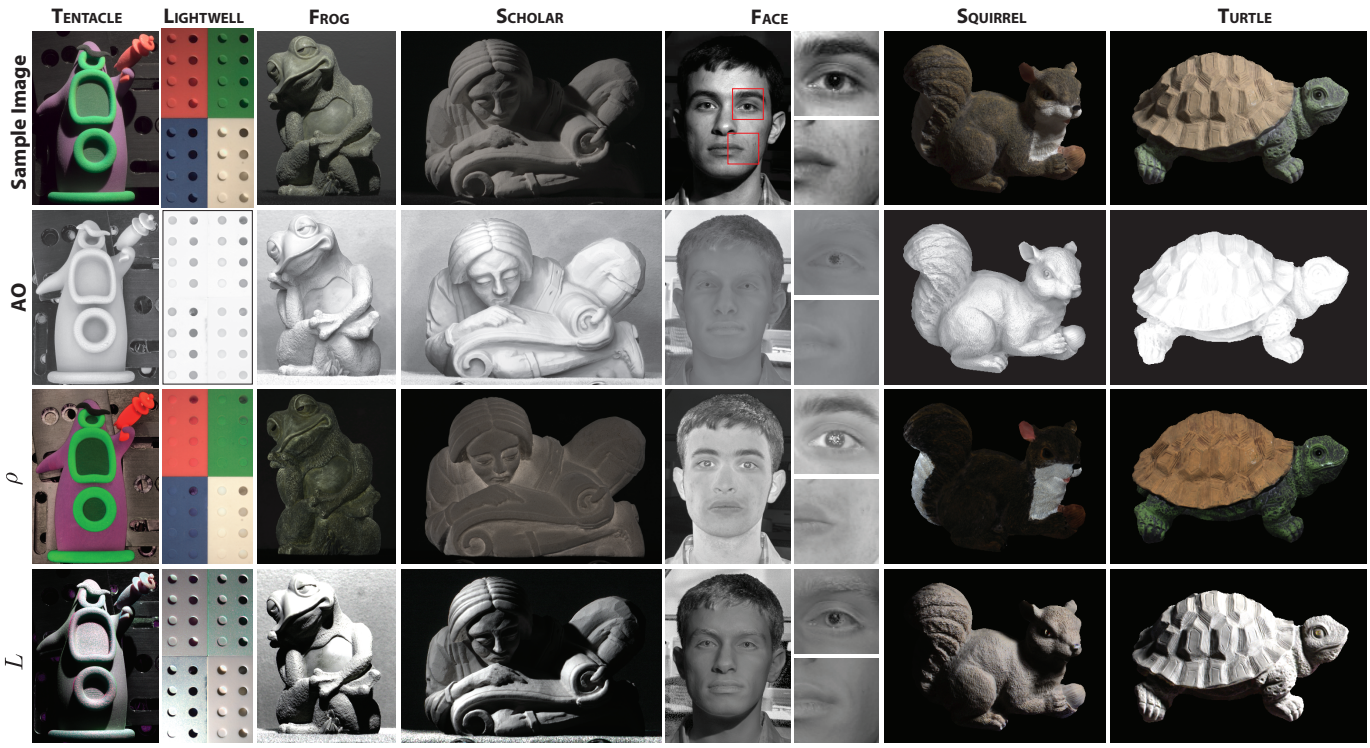

Fig. 6. Results of our algorithm (2nd estimate). Each column shows results from a different dataset. The rows show 1) sample images from the original dataset, 2) our estimated AO, 3) albedo, and 4) the illumination in the sample image.

skin is not strictly Lambertian, and exhibits significant subsurface scattering. Nevertheless we see from the images for AO and L in Figure 6 that our technique can qualitatively separate geometry and reflectance quite well. In particular, one can see from the area around the mouth that our AO image does not contain texture due to facial hair. Finally, we show results for TURTLE and SQUIRREL, from the MIT Intrinsic Image Dataset. Here the main challenge is that there are only 10 images of each object lit by a point light source.

Discussion. Figure 6 shows that the recovered AO seems to match our expectation of local visibility for these scenes. The recovered albedos are mostly free of shading and the ambient occlusion map is mostly free of albedo (e.g., the frog's nose and the mouth in FACE). It is also interesting that the pupil in the FACE dataset is black in the AO image and a light gray in the albedo.

6.2 Ambient Occlusion

We validated our estimate of AO using two objects of known geometry. In addition to TENTACLE, we 3D printed another object with a more regular shape, which we refer to as LIGHTWELL (Figure 7). This object is a solid block of material with a series of cylindrical holes of varying but known depth [29]. We printed this object in four colors: white (original material color), red, green, and blue to evaluate the impact of different albedos on our estimates. The acquisition setup for LIGHTWELL is the same as for TENTACLE (see Section 6.1). It is worth mentioning that although 3D printing offers good control over the geometry, material properties cannot be fully specified. The selected material (sandstone) was the most diffuse of the available materials, but is still not perfectly diffuse, and exhibits a fair amount of subsurface scattering (see the red ray gun of TENTACLE).

Figure 8 compares our AO result for TENTACLE to the ground truth (a more detailed analysis for the LIGHTWELL object is presented in Section 7). We can see qualitatively that our estimate of AO is very similar to ground truth. One difference is that our estimate appears smoother; we believe that this is caused in part by subsurface scattering, as the effect is most noticeable in the thin areas of the gun. Interreflections likely also have a smoothing effect in deep crevices. Another difference is that our estimate is in general darker, meaning that our algorithm is predicting that locally the geometry is more occluded than is really is. We attribute this in part to the material roughness from the 3D printing process. At a meso-scale level the structure can be thought of as being composed of many small crevices, and a single pixel in our κ image is an average of all these contributions.

6.3 Albedo

We ran our algorithm on the MIT Intrinsic Images benchmark [28] to measure the quality of our albedo estimates. This benchmark consists of 16 objects each with 11 images, and uses the local mean squared error (LMSE) defined in [28] to evaluate performance. Some methods evaluated by the benchmark (e.g., Retinex)

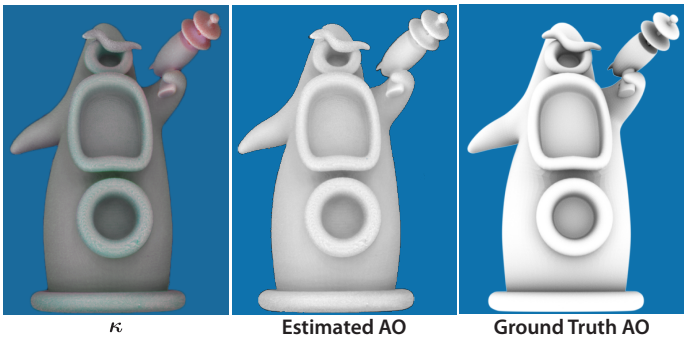

Fig. 8. Left: the statistic κ computed for TENTACLE. Right two images: Comparison of estimated AO with ground truth (computer generated). The background clutter is masked.

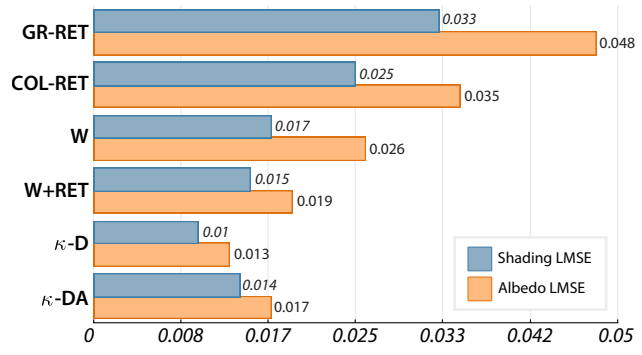

Fig. 9. Comparison of LMSE error on the MIT intrinsic image dataset [28] (shorter bars are better, indicating less error). Compared algorithms are: Grayscale Retinex (GR-RET), Color Retinex (COL-RET), Weiss (W), Weiss+Retinex (W+RET), ours with only direct term (κ-D) and our second estimate containing direct and ambient terms (κ-DA).

operate on a single image, usually by imposing priors on the illumination and albedo images or by using heuristics to classify gradients. However, the best-performing reported prior method operating on multiple images combines Retinex [5] with Weiss's method [4] which, like our own, requires a stack of images.

We obtain the shading image for each of the input images by simply dividing the input image by our estimated albedo (see Eq. (4)). Figure 9 shows our method's performance compared to others included in the benchmark. In Figure 10 we show a subset of results against the Weiss+Retinex multi-image method. We note that our approach outperforms the competing methods. Interestingly, our initial estimate (i.e., $f = 0$) performs better than our refined one. We believe that this is a result of the setup, which indeed contains no ambient illumination (as assumed by the first estimate of our algorithm, but not by our refined estimate, leading to over fitting), and the fact that most objects have a very high albedo, resulting in a larger contribution due to interreflections, which are not modeled by our algorithm. For completeness, in Table 1 we also compare our method to recent single-image algorithms [31], [32], [33], and report results on the different subsets of the benchmark dataset used in each prior evaluation. Our method compares favorably to these methods (but also uses more than a single image).

7 ANALYSIS

In this section we present a more in-depth analysis of various aspects of our algorithm on the specially created LIGHTWELL object (Fig. 7). This object has a very regular shape, with cylindrical holes of various depths that match our physical model, allowing us to evaluate in more detail how different aspects of our model impact the performance of our algorithm.

7.1 Impact of Albedo on AO Estimates

Because the 3D printed LIGHTWELL object consists of four different albedos, we can obtain a quantitative error measure of the local visibility angle α for different albedos. We report this error in Figure 11, computed by measuring the average error for α at the center of the crevice for LIGHTWELL compared to ground truth, for varying α angles corresponding to the crevice depths for the printed object. This figure shows four curves, one for each color of LIGHTWELL. In the plot three trends are evident. First, the error is larger for brighter albedos (red and white, in this case). We suspect that this is due to the increase in light interreflections for higher albedos. Since our model does not account for this effect, a patch at the bottom of a deeper hole looks brighter than our model would predict.

Second, we note that error increases for the more shallow crevices. We suspect this is due to roughness in the printed object as discussed in Section 6.2.

A third trend is that deeper holes have the largest errors. This can be explained by the fact that κ is the quotient of two expectations, and that for regions that receive light less frequently, we expect these averages to stabilize more slowly, a property we examine next.

7.2 Convergence Rate

We now consider the impact of the number of images and the visibility angle in estimating ambient occlusion. Figure 12 shows the root mean squared error (RMSE) of our ambient occlusion estimate as a function of the number of input images for different crevice depths (and hence local visibility angles). For each hole depth, we estimate AO at the center of the hole using rendered images of the blue LIGHTWELL (generated using a physically based renderer [35]). We compare our estimate to the ground truth AO in that hole using MSE, and repeat this process 100 times to compute an average RMSE. We observe that rate of convergence is strongly dependent on the depth of the crevice, but our method performs well even with a relatively small number of images on scenes where $\alpha \geq 40°$.

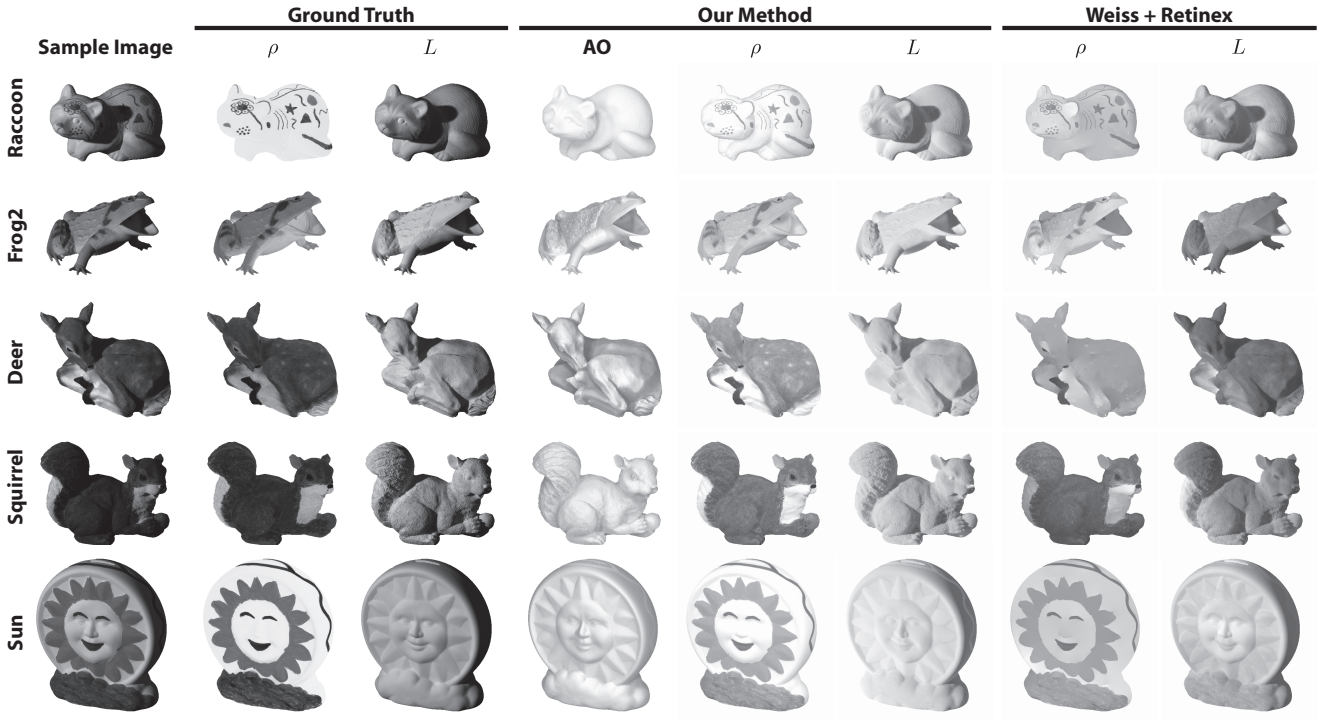

Fig. 10. Comparison of our method with W+Ret from the MIT benchmark. Results are for our first estimate of the albedo (i.e., ambient illumination is assumed to be zero) as this gave us the best results on the benchmark. We show here grayscale images as the benchmark uses grayscale versions of the decomposed images in its evaluation metric.

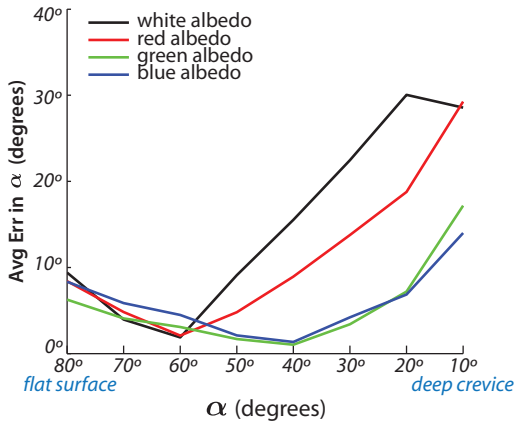

Fig. 11. Error in the estimated local visibility angle α vs. the true local visibility angle for the LIGHTWELL object printed in different colors (shown in the left).

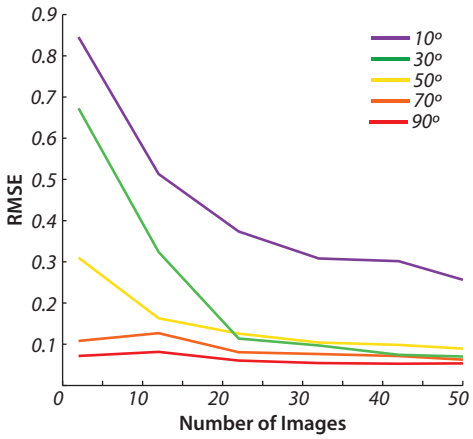

Fig. 12. Average Root Mean Square Error (RMSE) for our estimate of ambient occlusion vs. number of images used in the estimate. Each curve represents a different crevice depth and a corresponding local visibility angle α.

7.3 Impact of Interreflections

To understand how global illumination affects our method, we conducted a simple experiment with computer generated images of LIGHTWELL. The object, which was rendered with a perfect diffuse material and a reflectance of 0.5, was captured with an orthographic camera that aims directly towards the crevices, while illuminated with an ideal directional light source plus an ambient term (with $f = 0.25$). We used a physically based renderer [35] to produce 1000 images sampling the light direction uniformly at random over the hemisphere. Two sets of images were rendered: one with only the direct (single-bounce) component of light (that is, with interreflections and other indirect effects disabled in the rendering) and the other with both direct and indirect illumination components.

The error in the estimate of AO at the center of each crevice is shown in Figure 13. For the ideal case (only direct component of illumination) the ambient occlusion is in general very close to ground truth, with a max absolute error of 0.0172 (where the max possible error

TABLE 1
Local Mean Squared Error (LMSE) for individual images of our algorithm for the 1st (only direct light) and 2nd estimates (direct and ambient term), together with results from other work when available. In the last four columns of the last row we show our average on the same subset of images as reported by [28], [31], [33], [32], [34]. On all cases our algorithm outperforms these prior methods. Note that the numbers marked with † are the geometric mean as reported in [34], other averages are arithmetic means.

	Ours (1st estimate)			**Ours (2nd estimate)**							
	refl	**shading**	**avg**	**refl**	**shading**	**avg**	**Weiss + Retinex [28]**	**Barron and Malik [31]**	**Shen and Yeo [33]**	**Shen, Yang, Li, and Jia [32]**	**Barron and Malik [34]**
apple	0.006	0.0060	0.006	0.006	0.006	0.006	0.016			0.010	
box	0.004	0.0040	0.004	0.005	0.005	0.005	0.010		0.002	0.011	
cup1	0.003	0.0020	0.002	0.003	0.002	0.002	0.005		0.004	0.005	
cup2	0.003	0.0010	0.002	0.003	0.001	0.002	0.002	✓	0.005	0.007	✓
deer	0.027	0.0160	0.021	0.037	0.021	0.029	0.043	✓		0.032	✓
dinosaur	0.015	0.0120	0.014	0.016	0.007	0.012	0.015			0.021	
frog1	0.020	0.0180	0.019	0.029	0.026	0.027	0.043		0.053	0.029	
frog2	0.056	0.0120	0.034	0.053	0.017	0.035	0.053	✓	0.043	0.024	✓
panther	0.008	0.0060	0.007	0.024	0.014	0.019	0.005		0.008	0.005	
paper1	0.004	0.0040	0.004	0.010	0.008	0.009	0.003		0.001	0.013	
paper2	0.007	0.0040	0.006	0.009	0.006	0.008	0.005	✓	0.003	0.016	✓
pear	0.006	0.0050	0.005	0.006	0.004	0.005	0.006	✓		0.010	✓
phone	0.011	0.0080	0.010	0.035	0.013	0.024	0.008			0.011	
potato	0.011	0.0080	0.009	0.006	0.006	0.006	0.010	✓		0.014	✓
raccoon	0.011	0.0090	0.010	0.015	0.011	0.013	0.005	✓	0.005	0.008	✓
squirrel	0.019	0.0240	0.022	0.020	0.025	0.023	0.027			0.037	
sun	0.004	0.0050	0.005	0.007	0.005	0.006	0.003	✓	0.002	0.007	✓
teabag1	0.007	0.0160	0.012	0.012	0.033	0.023	0.014	✓	0.027	0.063	✓
teabag2	0.003	0.0110	0.007	0.012	0.020	0.016	0.006		0.015	0.031	
turtle	0.017	0.0200	0.019	0.020	0.026	0.023	0.015	✓	0.017	0.025	✓
average	0.012	0.0095	0.011	0.016	0.013	0.015	0.015	0.019	0.015	0.019	$0.021^{\dagger}$
						our avg on same subset	0.011	0.012	0.010	0.011	$0.009^{\dagger}$

would be 1.0). When indirect illumination is present the local visibility angle is overestimated for the holes in the center of the range, reaching an error of 0.0753 for $\alpha = 30°$. This happens because at the middle of the α range the contribution from the indirect light is closest to that of direct light, which means that the discrepancy between our model and what is observed is at its maximum. At the extremities of the range two different phenomena decrease the effect of global illumination. For the shallower crevices many direct light paths reach the bottom of the crevice decreasing the relative effect of indirect illumination. For deeper crevices, on the other hand, most paths that reach the bottom of the crevice do go through multiple bounces of light. Nevertheless, at each bounce the light is attenuated by the cosine factor (due to the angle of incidence and the surface normal) multiplied by the albedo, so by the time it reaches the bottom of the crevice it is attenuated so much that the total contribution from all indirect paths is still smaller than that of the direct ones.

Albedo also plays a role in the error when global illumination is present. In this experiment we used $\rho = 0.5$. For larger values of albedo the mode of the error will shift towards deeper crevices, because the light is attenuated less after each bounce.

8 Discussion

In this section we discuss our method in relation to other photometric methods in computer vision. We also discuss limitations of our work, including aspects of our method that we believe would be interesting avenues for future work.

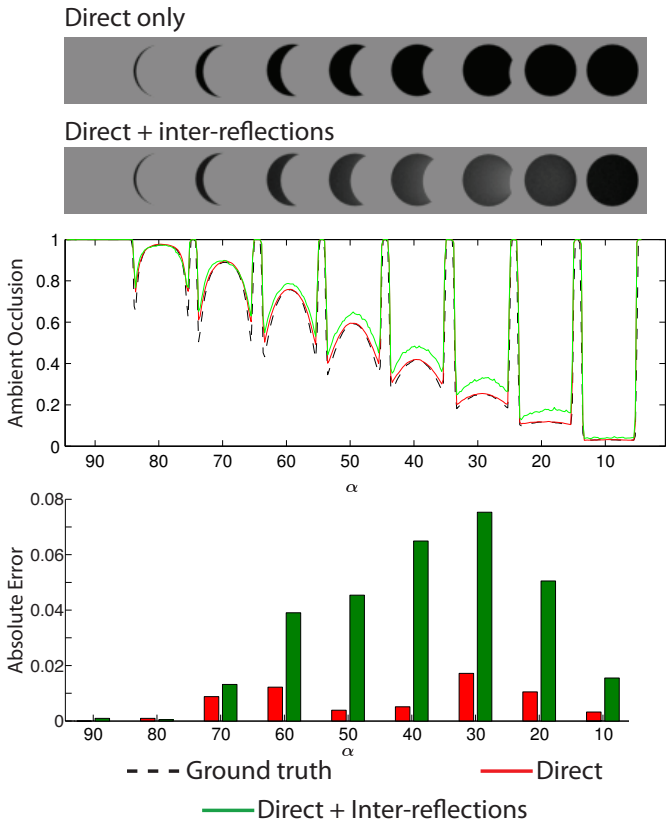

Fig. 13. Impact of interreflections in our estimates. Top: two renderings of LIGHTWELL, only the direct component of light and then with interreflections included (notice that shadows are brighter due to reflected light). Middle plot shows a cut through the rendered LIGHTWELL ambient occlusion estimate versus ground truth. Interreflections cause our method to over estimate ambient occlusion. Bottom bar plot shows the absolute error in the estimated ambient occlusion at the center of the crevices.

Relation to photometric stereo. An alternative to our method is to run uncalibrated photometric stereo on an image stack (uncalibrated because our method is designed for the case when light directions are unknown). One way to think of such a method is as a version of the multi-image intrinsic image decomposition problem, one that sets up an explicit physical model involving geometry and lighting. This is in contrast to typical intrinsic images methods (such as that of Weiss [4]) that reason about lighting as an image layer that modulates a reflectance layer. Like photometric stereo, our method uses a physical model, but one that parameterizes local geometry using a simple crevice model, rather than by surface normal. This difference yields an interesting set of strengths and weaknesses relative to photometric stereo. Our method:

- ✓ explicitly models non-convexity per-pixel without any smoothness or neighborhood reasoning,
- ✓ generalizes to color albedo in a straightforward way,
- ✓ is simple and robust,
- ✗ does not yield surface normals,
- ✗ requires a sufficient sampling of light source directions (although our experiments show that it can work well even with a small number of images).

There are likely an interesting set of new techniques to be discovered that combine aspects of our method, intrinsic images, and photometric stereo. For instance, a generalization of our technique could include an explicit representation of surface normal. On a related note, we believe it would be useful to unify the datasets used to evaluate intrinsic images (such as the MIT Intrinsic Images dataset) and photometric stereo, so that different types of multi-image methods can be compared directly.

Relation to multi-view stereo. Another alternative approach to our problem would be to reconstruct explicit geometry using multi-view stereo. At that point, ambient occlusion can be reasoned about explicitly using the known geometry. This approach could be used if we are given multiple views of a scene, or have the freedom to capture such views (e.g., with a gantry). In contrast, our method works from a static viewpoint and in fact was designed for more uncontrolled capture setups. Moreover, multi-view stereo techniques can have difficulty in capturing fine details close to the resolution of the sensor, whereas our method operates on each pixel independently to compute a simplified, per-pixel representation of local geometry. In fact, it would be interesting to try to use our approach to derive an explicit representation of detailed geometry (i.e., *geometry from ambient occlusion*).

Specularities. In deriving our algorithm one of our key assumptions is that the scene is Lambertian. Although the objects we used to evaluate our algorithm are mostly diffuse, specularities are occasionally present, leading to artifacts in the estimated AO and albedo. It is worth analyzing these artifacts in order to draw insights on how to improve our method. One such specular region is the right eyelid of FROG, shown in Figure 14. We observe that in the region where specular highlights occur (indicated by an arrow) the corresponding AO is darker than it should be (the surface on the actual eyelid is smooth, without any depressions) while the albedo is brighter. To understand why this happens, imagine how the physical model of Figure 3 can explain the observations we are given for this region. Since the point is not in fact heavily occluded, in most images the pixel color is roughly the albedo of the underlying material attenuated by the cosine factor of the light direction and the point normal. A small fraction of the observations are bright white due to the specular reflection. This can be explained with the crevice model by having (1) albedo that is the color of the direct light, (2) a deep crevice, and (3) an ambient term that is roughly the color of the albedo. In this situation the bright white (which corresponds in reality to a specular reflection) occurs when the direct light reaches the bottom of the crevice and reflects back. Otherwise, the color recorded

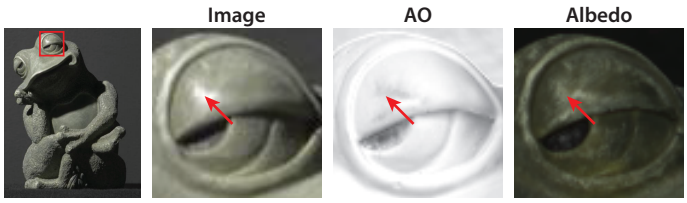

Fig. 14. Artifacts caused by specularities can be seen in the eyelid of FROG. On the left we show an image from the image stack with the inset highlighted. The arrow shows a region where specular highlights occur. Note that on the AO image that region appears darker than it should and on the albedo image it appears brighter.

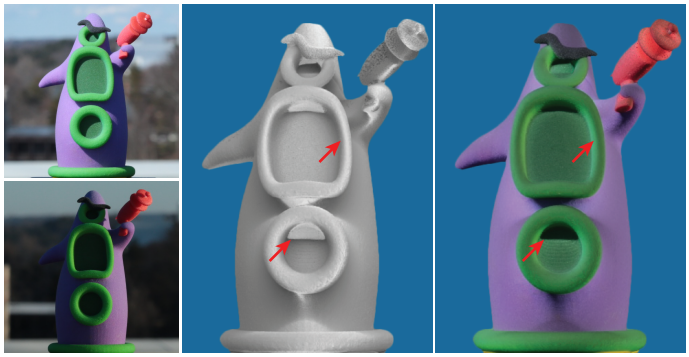

Fig. 15. Test with outdoor illumination for TENTACLE. Left: sample images, Middle: estimated ambient occlusion, Right: estimated albedo. Arrows highlight artifacts due to non-uniform coverage of the sky by the light source, causing shadows and highlights to get "baked" into the albedo.

by the camera is due to the ambient light. In practice, the ambient light color is also constrained by the rest of the data, but nonetheless we see effects like those illustrated in Figure 3.

Outdoor Illumination. All the results we have shown up to this point are for images acquired in a lab setting. To investigate how well our method performs on outdoor images illuminated by the sun, we placed TENTACLE on the roof of a building, oriented so that it faced south, and acquired 100 images over the course of a day in early September at a latitude of $42°26'36''$N. The estimated AO and albedo are shown in Fig. 15. Arrows in the figure highlight some of the artifacts we observed, which we believe are caused by the fact that the sun does not cover the entire hemisphere around the object. This results in some points being illuminated by the sun much less often than our model expects. In particular, points with normals that face away from the sun path are never sunlit, which results in estimated albedos that are much darker than they should be; the persistent shadows on these points become effectively "baked" into the albedo.

For image stacks that span more time (this dataset had images from only a single day) we expect the problem to be less pronounced, as the sun covers a wider band of the sky. Because this band changes with latitude, the location of the artifacts will also change as a function of the location of the object on the Earth's surface.

Aside from light source distribution, outdoor illumination breaks several of the assumptions made in our model. For instance, the color and relative intensity of the direct (sun) and indirect (sky) terms vary over time. For a more in depth exploration of the effects of outdoor illumination and an extension to the algorithm presented in this work for natural illumination we refer the reader to Hauagge et al. [36].

Color in κ. When introducing κ in Section 4 we focused on monochromatic images; we now discuss a property of the statistic κ that arises when dealing with color images. In this case κ is computed independently for each channel resulting in one f per channel: f_R, f_G, and f_B for the red, green, and blue color channels. In Eq. (5) we showed that the statistic κ is independent of the albedo ρ, yet a red tint appears on the ray gun of TENTACLE (Fig. 8 left). What does this color reveal about lighting? Because the direct component of light in our setup was white, color in κ is due to an ambient term.

Let's focus on a single point in the scene. Geometry at a point is the same across color channels so we can restrict our analysis to a single angle α without loss of generality. For instance, in Fig. 4 let's focus on $\alpha = 90°$. If the ambient term is blue, with $f_R = 0$, $f_G = 0$, and $f_B = 0.1$, what we will see in κ is that $\kappa_R = \kappa_G < \kappa_B$, so the color in the κ image reflects the color of the ambient term. This allows us to estimate the hue of the ambient term directly from κ. In the acquisition setup for Fig. 8 there was no ambient light, only a direct source was present. We believe that the red tint in the κ image on the ray gun is due to interreflections and subsurface scattering. Even thought this source of light is not constant, it varies much slower that the direct term so it can be thought of as a "local ambient" term.

We believe that this analysis could be useful in extending our method to account for interreflections and subsurface scattering.

9 CONCLUSIONS

Ambient occlusion, a measure of local visibility at a point, plays an important role in the shading of surfaces. We introduce an image-space approach to estimating ambient occlusion from a set of images under varying, unknown illumination. Our method analyzes the scene in terms of a physical model of a visibility cone, lit by a varying point light over the image stack. We propose a simple, per-pixel statistic, κ, based on observed intensities over the set of images; from κ, we recover per-pixel ambient occlusion and albedo values by relating our physical model to this measured statistic. Despite its simplicity, we show that this statistical approach works well in practice for a range of real-world image stacks. In

the future, it would be worth considering other statistics that might correlate to other physical properties.

Our assumption of diffuse materials with no inter-reflections is surprisingly effective. However, in the presence of specularities, subsurface scattering, or significant interreflections, our albedo estimates are less accurate. While our per-pixel statistic does not propagate errors, it would be interesting to couple our approach with sparsity or smoothness priors, or to incorporate models of inter-reflection. Our crevice model assumes a conical visibility model; in the future, we could extend this to include anisotropy so as to better match more general visibility scenarios.

Acknowledgments. This work was supported in part by the NSF (IIS-0963657, IIS-1149393, and IIS-1111534) and the Intel Science and Technology Visual Computing Center. We also thank the following people for their help and advice: Wenzel Jakob, Sean Bell, Pramook Khungurn, Steve Marschner, and Albert Liu.

REFERENCES

[1] R. Basri, D. Jacobs, and I. Kemelmacher, "Photometric stereo with general, unknown lighting," *IJCV*, 2007.
[2] K. Sunkavalli, T. Zickler, and H. Pfister, "Visibility subspaces: Uncalibrated photometric stereo with shadows," in *ECCV*, 2010.
[3] J. Ackermann, F. Langguth, S. Fuhrmann, and M. Goesele, "Photometric stereo for outdoor webcams," in *CVPR*, 2012.
[4] Y. Weiss, "Deriving intrinsic images from image sequences," in *ICCV*, 2001.
[5] E. Land, J. McCann *et al.*, "Lightness and retinex theory," *Journal of the Optical society of America*, 1971.
[6] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, 1991.
[7] R. Garg, H. Du, S. M. Seitz, and N. Snavely, "The dimensionality of scene appearance," in *ICCV*, 2009.
[8] D. Hauagge, S. Wehrwein, K. Bala, and N. Snavely, "Photometric Ambient Occlusion," in *CVPR*, 2013.
[9] R. Grosse, M. Johnson, E. Adelson, and W. Freeman, "Ground truth dataset and baseline evaluations for intrinsic image algorithms," in *ICCV*, 2009.
[10] H. Landis, "Production-ready global illumination," *SIGGRAPH Course Notes*, 2002.
[11] M. S. Langer and S. W. Zucker, "Shape-from-shading on a cloudy day," *J. Optical Society of America A*, 1994.
[12] M. Pharr and S. Green, "Ambient occlusion," *GPU Gems*, 2004.
[13] J. Kontkanen and S. Laine, "Ambient occlusion fields," in *Proc. Symp. on Interactive 3D Graphics and Games*. ACM, 2005.
[14] J. Pantaleoni, L. Fascione, M. Hill, and T. Aila, "PantaRay: Fast ray-traced occlusion caching of massive scenes," in *ACM Transactions on Graphics*, 2010.
[15] T. Beeler, D. Bradley, H. Zimmer, and M. Gross, "Improved reconstruction of deforming surfaces by cancelling ambient occlusion," in *ECCV*, 2012.
[16] O. Aldrian and W. A. Smith, "Inverse rendering of faces on a cloudy day," in *ECCV*, 2012.
[17] C. Wu, B. Wilburn, Y. Matsushita, and C. Theobalt, "High-quality shape from multi-view stereo and shading under general illumination," in *CVPR*, 2011.
[18] E. Prados, N. Jindal, and S. Soatto, "A non-local approach to shape from ambient shading," in *Scale Space and Variational Methods in Computer Vision*. Springer, 2009.
[19] P.-Y. Laffont, A. Bousseau, S. Paris, F. Durand, and G. Drettakis, "Coherent intrinsic images from photo collections," *SIGGRAPH Asia*, 2012, http://www-sop.inria.fr/reves/Basilic/2012/LBPDD12.
[20] K. J. Lee, Q. Zhao, X. Tong, M. Gong, S. Izadi, S. U. Lee, P. Tan, and S. Lin, "Estimation of intrinsic image sequences from image+depth video," in *ECCV*, 2012.
[21] K. Sunkavalli, W. Matusik, H. Pfister, and S. Rusinkiewicz, "Factored time-lapse video," in *SIGGRAPH*, 2007.
[22] J. T. Barron and J. Malik, "Color constancy, intrinsic images, and shape estimation," in *ECCV*, 2012.
[23] R. Woodham, "Analysing images of curved surfaces," *Artificial Intelligence*, 1981.
[24] R. J. Woodham, Y. Iwahori, and R. A. Barman, "Photometric stereo: Lambertian reflectance and light sources with unknown direction and strength," Vancouver, BC, Canada, Canada, Tech. Rep., 1991.
[25] M. Chandraker, S. Agarwal, and D. Kriegman, "Shadowcuts: Photometric stereo with shadows," in *CVPR*, 2007.
[26] T. Wu and C. Tang, "Photometric stereo via expectation maximization," *PAMI*, 2010.
[27] S. Ikehata, D. Wipf, Y. Matsushita, and K. Aizawa, "Robust photometric stereo using sparse regression," in *CVPR*, 2012.
[28] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "MIT Intrinsic Images," 2009, http://people.csail.mit.edu/rgrosse/intrinsic/.
[29] "Photometric Ambient Occlusion webpage," http://www.cs.cornell.edu/projects/photoao.
[30] K.-C. Lee, J. Ho, and D. Kriegman, "The Extended Yale Face Database B," 2005, http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html.
[31] J. T. Barron and J. Malik, "High-frequency shape and albedo from shading using natural image statistics," in *CVPR*, 2011.
[32] J. Shen, X. Yang, X. Li, and Y. Jia, "Intrinsic image decomposition using optimization and user scribbles," *Trans. Systems, Man, and Cybernetics*, 2012.
[33] L. Shen and C. Yeo, "Intrinsic images decomposition using a local and global sparse representation of reflectance," in *CVPR*, 2011.
[34] J. T. Barron and J. Malik, "Shape, illumination, and reflectance from shading," EECS, UC Berkeley, Tech. Rep. UCB/EECS-2013-117, May 2013.
[35] W. Jakob, "Mitsuba renderer," 2010, http://www.mitsuba-renderer.org.
[36] D. Hauagge, S. Wehrwein, P. Upchurch, K. Bala, and N. Snavely, "Reasoning about Photo Collections using Models of Outdoor Illumination," in *BMVC*, 2014.

Daniel Hauagge is a postdoctoral associate at Cornell Tech in New York City. He received his Ph.D. from the Computer Science Department at Cornell University. Prior to joining Cornell he received a B.Sc. degree in Computer Egineering and a M.Sc. degree in Computer Science from University of Campinas (UNICAMP) in Brazil. His interests include symmetry detection, and estimation of material properties and light transport from large photo collections.

Scott Wehrwein received his B.A. in Computer Science from Middlebury College in 2010. After two years with The MITRE Corporation's Computational Imaging group, he began work on his PhD at Cornell University in 2012. His research interests include computational photography and illumination modeling, especially in the context of internet photo collections.

Kavita Bala is a Professor in the Computer Science Department at Cornell University. She received her S.M. and Ph.D. from the Massachusetts Institute of Technology (MIT), and her B.Tech. from the Indian Institute of Technology (IIT, Bombay).

Bala specializes in material recognition and understanding, physically-based rendering in computer graphics, material perception and acquisition, and image-based lighting design, modeling and texturing. She is the Editor-in-Chief of the ACM Transactions on Graphics (TOG). She has chaired SIGGRAPH Asia 2011, and has served on the Papers Advisory Boards for SIGGRAPH and SIGGRAPH Asia. Her work on 3D Mandalas was featured at the Rubin Museum of Art, New York.

Noah Snavely received the B.Sc. degree in computer science from the University of Arizona in 2003, and the Ph.D. degree in computer science and engineering from the University of Washington in 2008. He is an associate professor of computer science at Cornell University, where he has been on the faculty since 2009. He works in computer graphics and computer vision, with a particular interest in using vast amounts of imagery from the Internet to reconstruct and visualize our world in 3D. He is the recipient of a Microsoft New Faculty Fellowship, a PECASE, and a SIGGRAPH Significant New Researcher Award. He is a member of the IEEE.