Photometric Ambient Occlusion

Daniel Hauagge Scott Wehrwein Kavita Bala Noah Snavely
Cornell University
{hauagge, swehrwein, kb, snavely}@cs.cornell.edu

Abstract

We present a method for computing ambient occlusion *(AO) for a stack of images of a 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.*

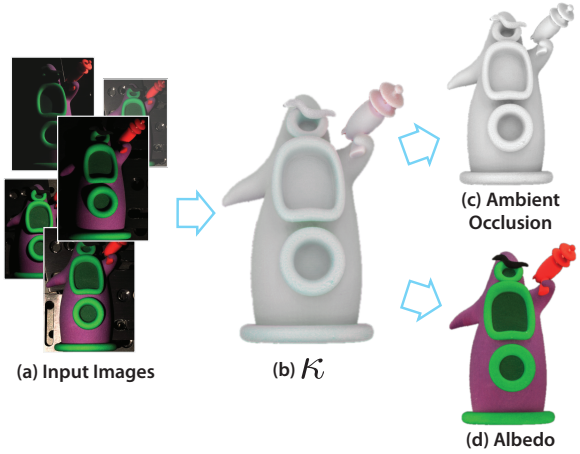


Figure 1. Our method takes as input a stack of images captured from varying, unknown illumination (a) and computes a per-pixel statistic, κ, over this stack (b). We infer both per-pixel ambient occlusion, a measure of local visibility (c), and albedo (d) for the scene by relating κ to a simple image formation model.

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 [6, 25, 2], recovering diffuse reflectance and illumination as intrinsic images [27, 15], and computing low-dimensional models of appearance of objects and scenes [26, 9]. 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, 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 directly from image observations, by introducing a simple, aggregate statistic (κ in Fig. 1(b)), 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 (Fig. 1(c)). 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 (Fig. 1(d)). The result is a *photometric* approach to estimating ambient occlusion and albedo.

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 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 contributions are:

- 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 Fig. 1. We also 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 [10]. This demonstrates the utility of reasoning about AO when measuring properties of scenes from images.

2. Ambient Occlusion

Ambient occlusion [16] 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 Fig. 6. 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 unoccluded points AO is 1. If the albedo at x is ρ, the measured radiance due to 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 does not account for inter-reflections.

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 [17].

In computer graphics, the main focus is on computing AO in 3D scenes to render images [20, 13, 19]. 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 has received relatively little attention in computer vision. Some examples of its use include early work in shape-from-shading [17], 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*. [7] and Aldrian & Smith [3] 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) [29]. 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 [14]. To account for points that are darker due to AO, they compute AO explicitly by generating and analyzing a 3D scene reconstruction. In contrast to these methods, we do not explicitly model geometry, instead reasoning 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 [27], 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 [10], 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 [24]. 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 & Malik optimize for reflectance, shape, and illumination from single images under strong priors on illumination and color of natural scenes [5]. In contrast, our method operates at a per-pixel level and does not make assumptions about the texture in the scene.

Photometric stereo techniques [28, 6] 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 source directions are unknown. A key challenge in photometric stereo is dealing with shadows, either by detecting them in some manner [8, 25] (a non-trivial problem with surfaces of varying albedo or complex self-occlusions), or treating them as a source of noise [30]. Sunkavalli *et al*. reason about lighting visibility of surface points, by clustering them into "visibility subspaces" that see a common set of lights [25]. 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.

Finally, our work is also related to methods that recover shape from AO [17, 21], and our algorithm could potentially be used to generate inputs to such methods.

4. A Model for AO 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 introduce an 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 constant ambient light source, both of which are of constant intensity over time. We assume that the distribution of directional light sources is uniform over the hemisphere. The images are radiometrically and pixel x, $I(x)$ 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.

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 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. Fig. 2 shows an example of observed PIDs in an image stack for two points. For example, a point in a deep

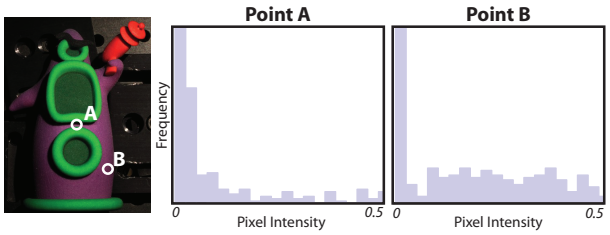


Figure 2. Histogram of pixel intensities for two points of TENTACLE over an image stack (only blue color channel). Notice that even though the two points have very similar albedos their histograms are quite different due to local visibility. Point A is mostly occluded, so values are in general lower.

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 Fig. 2.) The intuition then, is that the samples we record give us information about a pixel's PID, which in turns 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, and is related to the *coefficient of variation*, a normalized measure of variance used in statistics. Fig. 1(b) illustrates κ for an example image stack. 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.

Over our sequence of images I, $\rho(x)$ is constant and greater than zero, while $L(x)$ varies due to lighting. 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 $\mathcal{E}[L] = 0$ so $\kappa = 0$ (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 could be useful as a measure of ambient occlusion at a point.

4.2. A physical image formation model for κ

So far we have shown that κ is independent of albedo and is bounded. What 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 (Fig. 3). 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" and a "sky," 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, $\kappa(x)$ only depends on the local visibility angle $\alpha(x)$.

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:

$$I = I_d + I_a \tag{6}$$

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 point normal $\vec{n}$, and whether the light is blocked by other geometry. It is given by:

$$\begin{aligned} I_d(t) &= \rho l_d V_\alpha(\vec{n}, \vec{\omega}_d(t)) \langle \vec{n}, \vec{\omega}_d(t) \rangle & (7) \\ &= \rho l_d V_\alpha(\theta(t)) \cos\theta_d(t) & (8) \end{aligned}$$

where V_α is the visibility term (i.e., $V_\alpha(\theta) = 1$ if $\theta \leq \alpha$, $V_\alpha(\theta) = 0$ otherwise). The ambient component is constant and proportional to the projected solid angle of the local visibility angle. From Eqs. (1) and (2) we integrate the ambient illumination over the visible hemisphere at the point:

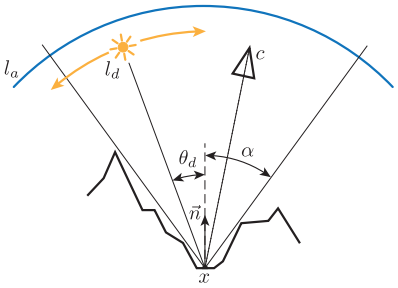


Figure 3. A point x on a Lambertian surface is observed by camera c and illuminated by a distant, moving light source with intensity l_d, and a constant ambient term of intensity l_a. The local visibility is approximated by a cone with angle α. If the light source angle θ_d with the surface normal $\vec{n}$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

$$I_a = \rho \int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\alpha} l_a \cos(\theta) \sin(\theta) d\theta d\varphi = \rho l_a \pi \sin^2 \alpha \tag{9}$$

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 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.

For the direct component, we integrate over the visible cone of angles at the point, assuming the point light is uniformly distributed over the hemisphere for the image stack:

$$\begin{aligned} \mathcal{E}[I_d] &= \frac{1}{2\pi} \int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\alpha} I_d \sin\theta d\theta d\varphi = \frac{1}{2} \rho l_d \sin^2(\alpha) \\ \mathcal{E}[I_d^2] &= \frac{1}{2\pi} \int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\alpha} I_d^2 \sin\theta d\theta d\varphi = -\frac{1}{3} \rho^2 l_d^2 \left(\cos^3(\alpha) - 1\right) \end{aligned} \tag{10}$$

Given these equations, κ can be derived in terms of α as:

$$\begin{aligned} \kappa(\alpha) &= \frac{\mathcal{E}^2[I]}{\mathcal{E}[I^2]} = \frac{\left(\mathcal{E}[I_d] + I_a\right)^2}{\mathcal{E}[I_d^2] + 2I_a\mathcal{E}[I_d] + I_a^2} & (11) \\ &= \frac{3(2\pi f + 1)^2 \sin^4(\alpha)}{4\left(3\pi f(\pi f + 1)\sin^4(\alpha) - \cos^3(\alpha) + 1\right)} \end{aligned}$$

where f is the relative intensity of l_a with respect to l_d, i.e. $l_a = f l_d$. To get a better intuition for κ we consider two special cases $l_d = 0$ and $l_a = 0$, which correspond to $f \to \infty$ and $f = 0$ respectively:

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

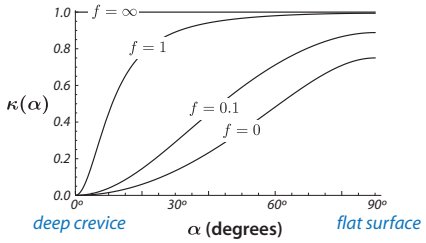


Figure 4. $\kappa(\alpha)$ for different ratios of ambient to direct light f. Note that as $f \to \infty$ ($l_d = 0$) we have a constant curve ($\kappa(\alpha) = 1$) so information about α cannot be recovered.

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.

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 Fig. 4 we show $\kappa(\alpha)$ for a few different values of f.

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 also uses additional constraints from color images.

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. (12). 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:

$$\alpha_1, f_1 \leftarrow \min_{\alpha,f} \sum \|\kappa_{obs} - \kappa(\alpha_0, f_0)\|^2 \qquad (12)$$

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 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. 12 defines a non-linear least squares problem, which we minimize using 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. (10) and (9). 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:

$$\rho = \frac{2\mathcal{E}[I]}{l_d \sin^2(\alpha)\,(1 + 2f\pi)} \qquad (13)$$

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 estimate of the albedo.

6. Results

We begin by demonstrating results of our algorithm on various datasets (Section 6.1) and exploring the different measures 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 [11]. Finally, in Section 6.4 we examine how the number of images affects our estimate of α.

6.1. Image Decomposition

Fig. 5 shows results on several datasets, including images used in prior work. For each dataset we show κ, ambient occlusion, ρ, and the illumination. More results can be found on our project webpage [1].

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.

The other datasets are public datasets that violate the assumptions of our model in various ways. FROG and SCHOLAR, from [25], contain 47–48 images lit under varying directional lights that do not cover the full hemisphere. FACE from the Yale Face Database B+ [18], contains 64 images with light positions over a range of angles. This scene violates our assumptions in that skin is not Lambertian, and exhibits significant subsurface scattering. Nevertheless we see from the images for AO and L in Fig. 5 that our technique can qualitatively separate geometry and reflectance quite well. In particular, one can see from the area on the neck close to the chin 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 5 shows that the recovered AO seem to match our expectation of local visibility for these scenes.

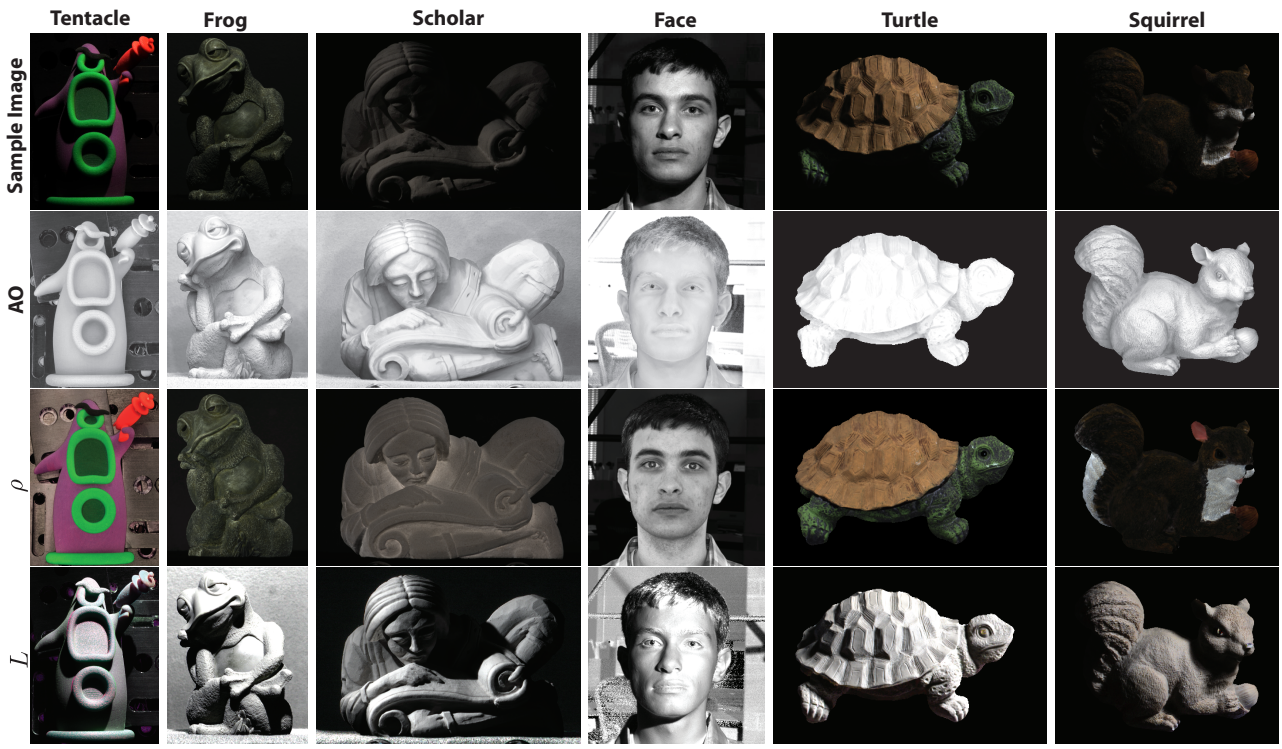


Figure 5. Results of our algorithm. Each column shows results from a different dataset. The rows show 1) sample images from the original dataset, 2) our estimated AO, 3) albedo, and 4) the illumination in the sample image.

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 neck 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.

For color images we estimate κ independently for each color channel. For a white directional light source, color casts in κ reveal the color of the ambient term, since f has a different value per channel. Observe that in Fig. 4, for a fixed α, κ increases if f increases. The same mechanism might explain local color shifts in Fig. 1, where one can see red tints on the ray gun and green casts along the mouth. The cause is likely to be subsurface scattering, where light arriving after multiple subsurface scattering events can be thought of as acting like a local ambient light term.

6.2. Estimated 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. This object is a solid block of material with a series of cylindrical holes of varying but known depth [1]. We printed this object in four colors: white (original material color), red, green, and blue to evaluate the impact of different albedos on our estimate. 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 was not perfectly diffuse, and exhibited a fair amount of subsurface scattering (see the red ray gun of TENTACLE).

Figure 6 compares our AO result for TENTACLE to the ground truth. We can see qualitatively that both are very similar. 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. Another difference is that our estimate is in general darker, meaning that our algorithm is predicting that locally the geometry is more occluded. We attribute this in part to the material roughness from the 3D printing process. At a mesolevel 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.

For a quantitative measure of error we report in Fig. 9 (a) the average error for α at the center of the crevice for LIGHTWELL compared to ground truth, as a function of the local visibility angle α. We show four curves, one for each color of LIGHTWELL. In the plot we see two trends. First, the error is larger for brighter albedos (red and white). We suspect that this is caused by the increase in light interreflections for higher albedos. Since our model does not

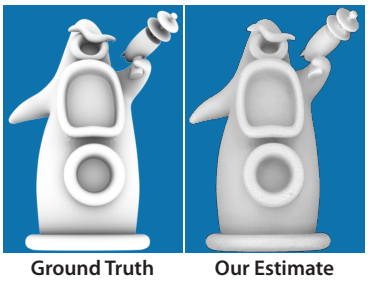


Figure 6. A comparison of ground truth (computer generated) with our estimated AO (from actual images) for TENTACLE. The background clutter is masked.

account for this effect, a patch at the bottom of a deeper hole looks brighter than our model would predict. A second trend is that deeper holes have larger errors. This can be explained by remembering that κ is the quotient of two expectations and that for these regions we expect these averages to stabilize more slowly (as we will show in Section 6.4).

6.3. Estimated Albedo

We ran our algorithm on the MIT Intrinsic Images benchmark [11] 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 [11] to evaluate performance. Some methods evaluated by the benchmark (e.g., Retinex) operate on a single image, usually by imposing priors on the illumination and albedo images. However, the best-performing reported prior method combines Retinex [15] with Weiss's method [27] 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)). Fig. 7 shows our method's performance compared to others included in the benchmark. In Fig. 8 we show a subset of results against the best algorithm in the benchmark. First, we note that our approach outperforms the competing methods. Interestingly, our initial estimate (i.e., $f = 0$) performs better than the refined one. We believe that this is a result of the setup, which indeed does not contain ambient illumination, and the fact that most objects have a very high albedo, resulting in a larger contribution due to inter reflections, which is not modeled by our algorithm. Our results also compare favorably to recent single-image algorithms [4, 22, 23] which reports results on different subsets of the benchmark datasets (a full comparison can be found on our project webpage [1]).

6.4. Rate of convergence

We now consider the impact of the number of images and the visibility angle in estimating ambient occlusion. Figure 9 (b) shows the root mean squared error (RMSE) of our ambient occlusion estimate as a function of the number of

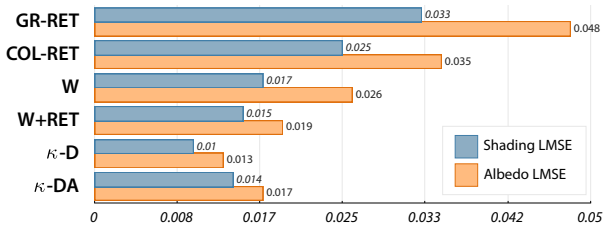


Figure 7. Comparison of LMSE error on the MIT intrinsic image dataset [11] (shorter bars are better). Compared algorithms are: Grayscale Retinex (GR-RET), Color Retinex (COL-RET), Weiss (W), Weiss+Retinex (W+RET), ours with only direct term (κ-D) and our second estimate containing direct and ambient terms (κ-DA).

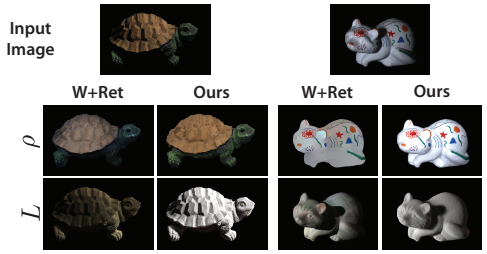
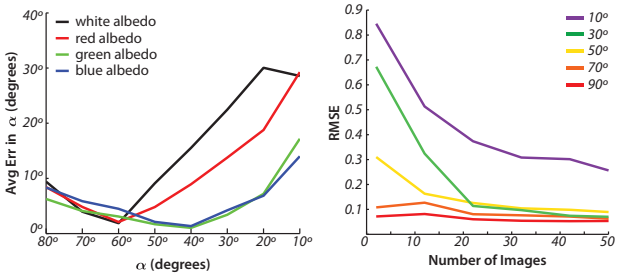


Figure 8. Comparison of our method with W+Ret from the MIT benchmark. Results are for our first estimate of the albedo (i.e., ambient illumination is assumed to be zero) as this gave us the best results on the benchmark.



(a) Error in α for different albedos

(b) Convergence rate

Figure 9. (a) 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). (b) Average Root Mean Square Error (RMSE) for our estimate of ambient occlusion vs. number of images used in the estimate. Different curves represent different crevice depths and their corresponding angles (α).

input images for different hole depths. 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 [12]). 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 small number of images on scenes where $\alpha \geq 40°$.

7. 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 approach makes a few assumptions that we would like to generalize. We assume that input images are illuminated by a point light source that moves over the entire hemisphere visible to any given point. For outdoor scenes, where the directional light is from the sun, this assumption is violated; we need improved models to account for more general distributions of lighting directions.

Our assumption of diffuse materials with no interreflections 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] Photometric Ambient Occlusion webpage. `http://www.cs.cornell.edu/projects/photoao`.

[2] J. Ackermann, F. Langguth, S. Fuhrmann, and M. Goesele. Photometric stereo for outdoor webcams. In *CVPR*, 2012.

[3] O. Aldrian and W. A. Smith. Inverse rendering of faces on a cloudy day. In *ECCV*. 2012.

[4] J. T. Barron and J. Malik. High-frequency shape and albedo from shading using natural image statistics. In *CVPR*, 2011.

[5] J. T. Barron and J. Malik. Color constancy, intrinsic images, and shape estimation. In *ECCV*, 2012.

[6] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *IJCV*, 2007.

[7] T. Beeler, D. Bradley, H. Zimmer, and M. Gross. Improved reconstruction of deforming surfaces by cancelling ambient occlusion. In *ECCV*, 2012.

[8] M. Chandraker, S. Agarwal, and D. Kriegman. Shadowcuts: Photometric stereo with shadows. In *CVPR*, 2007.

[9] R. Garg, H. Du, S. M. Seitz, and N. Snavely. The dimensionality of scene appearance. In *ICCV*, 2009.

[10] R. Grosse, M. Johnson, E. Adelson, and W. Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *ICCV*, 2009.

[11] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman. MIT Intrinsic Images, 2009. `http://people.csail.mit.edu/rgrosse/intrinsic/`.

[12] W. Jakob. Mitsuba renderer, 2010. `http://www.mitsuba-renderer.org`.

[13] J. Kontkanen and S. Laine. Ambient occlusion fields. In *Proc. Symp. on Interactive 3D Graphics and Games*. ACM, 2005.

[14] P.-Y. Laffont, A. Bousseau, S. Paris, F. Durand, and G. Drettakis. Coherent intrinsic images from photo collections. *SIGGRAPH Asia*, 2012.

[15] E. Land, J. McCann, et al. Lightness and retinex theory. *Journal of the Optical society of America*, 1971.

[16] H. Landis. Production-ready global illumination. *SIGGRAPH Course Notes*, 2002.

[17] M. S. Langer and S. W. Zucker. Shape-from-shading on a cloudy day. *J. Optical Society of America A*, 1994.

[18] K.-C. Lee, J. Ho, and D. Kriegman. The Extended Yale Face Database B, 2005. `http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html`.

[19] J. Pantaleoni, L. Fascione, M. Hill, and T. Aila. PantaRay: Fast ray-traced occlusion caching of massive scenes. In *ACM Transactions on Graphics*, 2010.

[20] M. Pharr and S. Green. Ambient occlusion. *GPU Gems*, 2004.

[21] 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.

[22] J. Shen, X. Yang, X. Li, and Y. Jia. Intrinsic image decomposition using optimization and user scribbles. *Trans. Systems, Man, and Cybernetics*, 2012.

[23] L. Shen and C. Yeo. Intrinsic images decomposition using a local and global sparse representation of reflectance. In *CVPR*, 2011.

[24] K. Sunkavalli, W. Matusik, H. Pfister, and S. Rusinkiewicz. Factored time-lapse video. In *SIGGRAPH*, 2007.

[25] K. Sunkavalli, T. Zickler, and H. Pfister. Visibility subspaces: Uncalibrated photometric stereo with shadows. In *ECCV*, 2010.

[26] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 1991.

[27] Y. Weiss. Deriving intrinsic images from image sequences. In *ICCV*, 2001.

[28] R. Woodham. Analysing images of curved surfaces. *Artificial Intelligence*, 1981.

[29] C. Wu, B. Wilburn, Y. Matsushita, and C. Theobalt. High-quality shape from multi-view stereo and shading under general illumination. In *CVPR*, 2011.

[30] T. Wu and C. Tang. Photometric stereo via expectation maximization. *PAMI*, 2010.