Ambient Occlusion via Compressive Visibility Estimation

Wei Yang[1], Yu Ji[1], Haiting Lin[1], Yang Yang[1], Sing Bing Kang[2], Jingyi Yu[1]
[1]University of Delaware, Newark, DE, USA. [2]Microsoft Research, Redmond, WA, USA.

The problem of recovering intrinsic properties of a scene/object from images has attracted much attention in the past decade. Tremendous efforts have been focused on intrinsic properties related to shading and reflectance [1, 2]. In this paper, we explore a challenging type of intrinsic properties called the ambient occlusion map. Ambient Occlusion (AO) characterizes the visibility of a surface point due to local geometry occlusions. Given a scene point x, its AO measures the occlusion of ambient light caused by local surface geometry:

$$A(x) = \frac{1}{\pi}\int_{\Omega} v(x,\dot{w})\langle \dot{w}\cdot\dot{n}\rangle d\dot{w} \tag{1}$$

where $\dot{w}$ is the direction of incident light; $\dot{n}$ is the normal of x; and $\langle\cdot\rangle$ refers to the dot product. $v(x,\dot{w})$ is the local visibility function and is equal to 0 if the light ray from $\dot{w}$ is occluded from x.

Intuitively, we can illuminate the object using a dense set of uniform directional lights $\dot{w}_i$ and sum up images captured from all directions.

$$\sum_{i=1}^{N} I_i = \rho\sum_{i=1}^{N} v_i\langle \dot{w}_i\cdot\dot{n}\rangle = \rho\tilde{A}c \tag{2}$$

AO term $\tilde{A}$ cannot be resolved since the albedo ρ is also unknown. Hauagge et.al [3] assume the visibility function follows cone-shaped distribution centered at the normal as $A = \pi\sin^2\alpha$, α is the cone's half angle. Under uniformly distributed lighting, they show that computing $\kappa = E[I]^2/E[I^2]$ ($E[\cdot]$ stands for expectation) directly cancels the albedo. For their assumption to work, densely distributed light sources will be needed.

Instead of capturing one lighting direction at a time, we aim to enable multiple lighting directions in one shot. A downside though is that we cannot use the κ statistics to cancel out the albedo. Instead, we build our solution on compressive signal reconstruction. We use a binary vector $b = [l_1, ...l_N]$ to represent the status of N lighting directions, where $l_i = 1$ or 0 corresponds to if the lighting direction $\dot{w}_i$ is enabled or disabled. We have:

$$I = \rho\sum_{i=1}^{N} l_i v_i\langle \dot{w}_i\cdot\dot{n}\rangle \tag{3}$$

We can now use a set of M strategically coded directional lighting patterns. For each pattern b^j, $j = 1...M$, we capture an image I^j. This results in an $M\times N$ measurement matrix $B = [b^1, b^2....b^M]^T$. Rewrite Eqn. 3 as

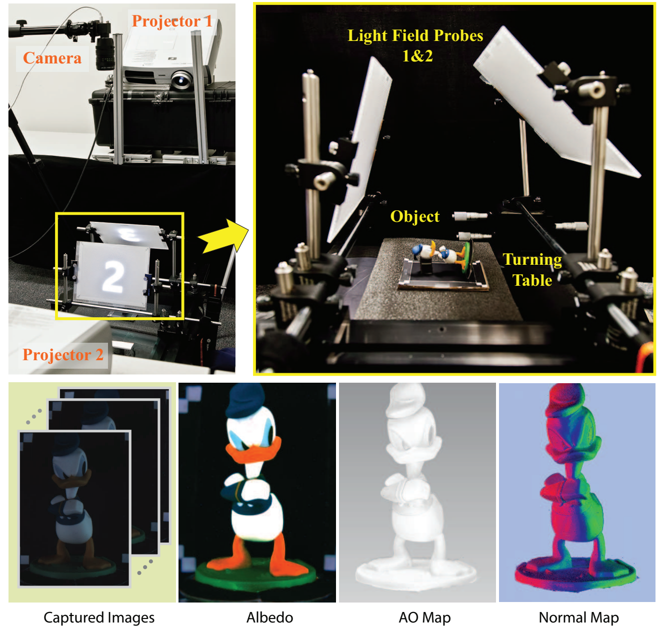
$$I = \rho B[V * W(\dot{n})] \tag{4}$$

where $W(\dot{n}) = [\langle\dot{w}_1\cdot\dot{n}\rangle, \langle\dot{w}_2\cdot\dot{n}\rangle, ..., \langle\dot{w}_N\cdot\dot{n}\rangle]$ and $[*]$ refers to the pairwise element-wise product. Given the measurements, we aim to solve for ρ, V and $\dot{n}$. Our solution is to reduce the problem to two sub-problems and solve them using iterative optimization.

Visibility Recovery Sub-problem. V is a binary pattern and solving V in this optimization is NP-hard. We reduce this problem to an ℓ_∞ regularized ℓ_1 minimization:

$$\begin{aligned}\hat{\rho},\hat{V} \leftarrow \arg\min_{\rho,V}\ &\{\|\rho B(W_0 * V) - I\|_2^2 + \lambda_1\|V\|_1 \\ &+ \lambda_1\|V-0.5\|_\infty + \lambda_2\|\nabla V\|_1\}\end{aligned} \tag{5}$$

where λ_1, λ_2 and λ_3 are weighting factors. The new objective function consists of four terms: 1) $\|\rho B(W_0 * V) - I\|_2^2$ corresponds to the fidelity term where the estimated V should be consistent with the observed pixel intensities I; 2) $\|V\|_1$ is the sparse prior term that forces the visibility of negligible light directions should be zero. With this term, the solution would favor a sparse set of visible light directions; 3) $\|V-0.5\|_\infty$ is the binary prior term. It is used to clamp the elements of V with high values to 1 and lows values to 0. Combining $\|V\|_1$ and $\|V-0.5\|_\infty$ with weighting factors allows us to obtain an *approximate* binary solution; and 4) $\|\nabla V\|_1$ is the total variation term, i.e., to bias towards a solution with compact visible areas.



Top row shows the system setup. Bottom row is the real experiment result.

Normal Recovery Sub-problem We then threshold the $\hat{V}$ to get a binary visibility vector $\widetilde{V}$. Now that we have both the visibility vector and albedo, we can refine the estimation of normal $\dot{n}$ by solving for the following least square problem:

$$\begin{aligned}\bar{\rho},\bar{\dot{n}} \leftarrow \arg\min_{\rho,\dot{n}}\ &\|\hat{\rho}B[W(\dot{n}) * \widetilde{V}] - I\|_2 \\ \text{Subject to}\ &\|\dot{n}\|_2 = 1\end{aligned} \tag{6}$$

Specifically, we relax the constraint to $\|\dot{n}\| \leqslant 1$ and solve it via constrained least square minimization. Next, we use the result $\bar{\dot{n}}$ to update W. We repeat the process to iteratively improve the visibility and normal estimation.

We construct an encodable directional light source using the light field probe [4] and validate our approach. Experiments show that our scheme produces AO estimation at comparable accuracy to [3] but with a much smaller set of images. In addition, we can recover more general visibility functions beyond the normal-centered cone-shaped models.

[1] Sean Bell, Kavita Bala, and Noah Snavely. Intrinsic images in the wild. *SIGGRAPH*, 33(4), 2014.

[2] Roger Grosse, Micah K Johnson, Edward H Adelson, and William T Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *ICCV 2009*, pages 2335–2342. IEEE, 2009.

[3] Daniel Hauagge, Scott Wehrwein, Kavita Bala, and Noah Snavely. Photometric ambient occlusion. In *CVPR 2013*, pages 2515–2522. IEEE, 2013.

[4] Gordon Wetzstein, Ramesh Raskar, and Wolfgang Heidrich. Hand-held schlieren photography with light field probes. In *ICCP 2011*, pages 1–8. IEEE, 2011.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.