Non-line-of-sight Surface Reconstruction
Using the Directional Light-cone Transform (Supplement)

Sean I. Young
Stanford University

David B. Lindell
Stanford University

Bernd Girod
Stanford University

David Taubman
UNSW Sydney

Gordon Wetzstein
Stanford University

S1. LCT Resampling

Here, we provide a derivation for the LCT resampler T^* introduced in Section 3.1 of the main paper. Recapitulating that the forward confocal model [1] is given by

$$\tau(x', y', t) = \iiint_\Omega \mathrm{d}\Omega \; \frac{\rho(x, y, z)}{r^4} \cdot \delta\Big(2\sqrt{(x'-x)^2 + (y'-y)^2 + z^2} - tc\Big), \tag{S1}$$

the objective of the LCT is to convert (S1) to a convolution equation using a suitable reparameterizer (in the continuous case) or a resampler (in the discrete case).

Absorbing $1/r^4 = (2/tc)^4$ into the definition τ, and then reparameterizing the t-z axis using the change of variables

$$z = \frac{\sqrt{u}}{2}, \;\; \frac{\mathrm{d}z}{\mathrm{d}u} = \frac{1}{2\sqrt{u}}, \;\; t = \frac{\sqrt{u'}}{c}, \;\; \frac{\mathrm{d}t}{\mathrm{d}u'} = \frac{1}{c\sqrt{u'}}, \tag{S2}$$

one can rewrite (S1) as the convolution

$$\underbrace{\frac{1}{c\sqrt{u'}}\tau\left(x', y', \frac{\sqrt{u'}}{c}\right)}_{\mathcal{T}^*_t[\tau](x',y',u')} = \iiint_\Omega \mathrm{d}\Omega \; \underbrace{\frac{1}{2\sqrt{u}}\rho\left(x, y, \frac{\sqrt{u}}{2}\right)}_{\mathcal{T}^*[\rho](x,y,u)} \cdot \underbrace{\delta\Big(4(x'-x)^2 + 4(y'-y)^2 - (u'-u)\Big)}_{h(x'-x, y'-y, u'-u)}, \tag{S3}$$

noting that $\mathcal{T}^*$ is a linear operator. Operator T^* in (4) can be seen as the discrete counterpart of $\mathcal{T}^*$, and H, a filter whose impulse response is obtained by sampling $h(\cdot)$.

S2. D-LCT Resampling

In the case of the D-LCT, the definition of the resampling operator is different from that used by the LCT. The D-LCT aims to express the directional-albedo model

$$\tau(x', y', t) = \iiint_\Omega \mathrm{d}s \frac{\langle \boldsymbol{\upsilon}(s), \mathbf{s}' - \mathbf{s} \rangle}{r^5} \cdot \delta\Big(2\sqrt{(x'-x)^2 + (y'-y)^2 + z^2} - tc\Big), \tag{S4}$$

as a sum of convolutions. Absorbing $1/r^5 = (2/tc)^5$ in the definition of τ and unwrapping $\langle \boldsymbol{\upsilon}(s), \mathbf{s}' - \mathbf{s} \rangle$ produces

$$\begin{aligned} \tau(x', y', t) = &\iiint_\Omega \mathrm{d}\Omega \; \upsilon_x(x, y, z)\delta(\cdot)(x - x') \\ + &\iiint_\Omega \mathrm{d}\Omega \; \upsilon_y(x, y, z)\delta(\cdot)(y - y') \\ + &\iiint_\Omega \mathrm{d}\Omega \; \upsilon_z(x, y, z)\delta(\cdot)(z) \end{aligned} \tag{S5}$$

in which $\delta(\cdot)$ denotes the last term in the integral of (S1).

Applying the LCT (S3) to the three integrals, we obtain

$$\begin{aligned} \mathcal{T}^*_t[\tau](x', y', u') = &\iiint_\Omega \mathrm{d}\Omega \; \mathcal{T}^*_x[\upsilon_x](x, y, u)h_x(\cdot) \\ + &\iiint_\Omega \mathrm{d}\Omega \; \mathcal{T}^*_y[\upsilon_y](x, y, u)h_y(\cdot) \\ + &\iiint_\Omega \mathrm{d}\Omega \; \mathcal{T}^*_z[\upsilon_z](x, y, u)h_z(\cdot) \end{aligned} \tag{S6}$$

in which the resampling operators

$$\begin{aligned} \mathcal{T}^*_x[\upsilon_x](x, y, u) &= (1/2\sqrt{u})\upsilon_x(x, y, \sqrt{u}/2) \\ \mathcal{T}^*_y[\upsilon_y](x, y, u) &= (1/2\sqrt{u})\upsilon_x(x, y, \sqrt{u}/2) \\ \mathcal{T}^*_z[\upsilon_z](x, y, u) &= (1/4)\upsilon_z(x, y, \sqrt{u}/2), \end{aligned} \tag{S7}$$

and note the $\mathcal{T}^*_x$ and $\mathcal{T}^*_y$ are identical to the LCT resampling operator $\mathcal{T}^*$, while $\mathcal{T}^*_z$ differs due to the presence of z in the last integral of (S5). We have also denoted by

$$\begin{aligned} h_x(\cdot) &= h(x'-x, y'-y, u'-u)(x - x') \\ h_y(\cdot) &= h(x'-x, y'-y, u'-u)(y - y') \\ h_z(\cdot) &= h(x'-x, y'-y, u'-u) \end{aligned} \tag{S8}$$

the three, shift-invariant D-LCT operators (filters) in (S6).

S3. Model Linearization

Here, we show the directional-albedo model corresponds to the linearization of a physically-based, higher-order light transport model. We can express the confocal version of the

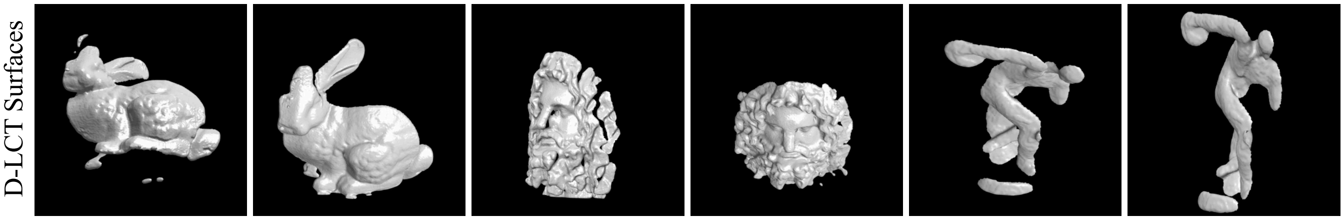

Figure S2. **Albedo and surface reconstructions using captured data:** The extracted normals (c) are used to fit surfaces (d). SU has a spatial resolution of 64×64 pixels (1min exposure), and the remaining scenes, 512×512 (180min exposure). We use $\lambda = 2^0$, 2^3 and 2^3 for SU, Discobolus and Dragon, respectively. Parameters of the other methods were optimized using grid search.

physical model proposed by Tsai *et al.* [2] as

$$\tau(x', y', t) = \iiint_\Omega \mathrm{d}\Omega \, \frac{\rho(\mathbf{s})}{r^4} \underbrace{\left\langle \mathbf{n}(s), \frac{\mathbf{s}' - \mathbf{s}}{\|\mathbf{s}' - \mathbf{s}\|} \right\rangle^2}_{f(\mathbf{n}(\mathbf{s}), \mathbf{s}' - \mathbf{s})} \cdot \delta\Big(2\sqrt{(x'-x)^2 + (y'-y)^2 + z^2} - tc\Big), \tag{S9}$$

in which the squaring of the inner product takes into account the cosine fall-offs due to both the incident and the reflected rays at $\mathbf{s} = (x, y, z)$.

Due to the squaring of the dot product in (S9), recovering $n(s)$ requires solving a nonlinear least-squares problem. We can apply the Gauss-Newton method to solve this nonlinear least-squares problem iteratively, by successively linearizing the term $f(\mathbf{n}(\mathbf{s}), \mathbf{s}' - \mathbf{s})$ as

$$\hat{f}(\mathbf{n}(\mathbf{s}), \mathbf{s}' - \mathbf{s}) = a(\mathbf{s}) \left\langle \mathbf{n}(\mathbf{s}), \frac{\mathbf{s}' - \mathbf{s}}{\|\mathbf{s}' - \mathbf{s}\|} \right\rangle + b(\mathbf{s}), \tag{S10}$$

in which $a(s)$ and $b(s)$ are respectively the slope and offset parameters of each linearized model. Observe that both $\mathbf{n}(\mathbf{s})$ and $(\mathbf{s}' - \mathbf{s})/\|\mathbf{s}' - \mathbf{s}\|$ are unit vectors, so that linearizations $\hat{f}(\mathbf{n}(\mathbf{s}), \mathbf{s}' - \mathbf{s})$ depend only on the angle between $\mathbf{n}(\mathbf{s})$ and $\mathbf{s}' - \mathbf{s}$ about which the linear approximations are formed.

Our object is to derive a linearized model in terms of the directional albedos $\boldsymbol{\upsilon}(\mathbf{s}) = \rho(\mathbf{s})\mathbf{n}(\mathbf{s})$. Substituting the linear term $\hat{f}(\mathbf{n}(\mathbf{s}), \mathbf{s}' - \mathbf{s})$ in (S9) and using the bilinearity of dot products, we can write the linearized model associated with the $k+1$th iteration of Gauss-Newton as $\tau(x', y', t)$

$$= \iiint_\Omega \left(a_k(s) \left\langle \frac{\boldsymbol{\upsilon}(\mathbf{s})}{r^4}, \frac{\mathbf{s}' - \mathbf{s}}{\|\mathbf{s}' - \mathbf{s}\|} \right\rangle - b_k(\mathbf{s})\rho_k(\mathbf{s}) \right) \cdot \delta\Big(2\sqrt{(x'-x)^2 + (y'-y)^2 + z^2} - tc\Big) \, \mathrm{d}\Omega, \tag{S11}$$

in which

$$\begin{aligned} a_k(\mathbf{s}) &= 2 \left\langle \mathbf{n}_k(\mathbf{s}), \frac{\mathbf{s}' - \mathbf{s}}{\|\mathbf{s}' - \mathbf{s}\|} \right\rangle, \\ b_k(\mathbf{s}) &= -\left\langle \mathbf{n}_k(\mathbf{s}), \frac{\mathbf{s}' - \mathbf{s}}{\|\mathbf{s}' - \mathbf{s}\|} \right\rangle^2, \end{aligned} \tag{S12}$$

and $\mathbf{n}_k(\mathbf{s})$ and $\rho_k(\mathbf{s})$ are respectively the normal and albedo at s estimated during the kth iteration. Figure S1 (left) plots the linearization of $f(\mathbf{n}(\mathbf{s}), \mathbf{s}' - \mathbf{s})$ at different values of the incident angle w between $\mathbf{n}_k(\mathbf{s})$ and $\mathbf{s}' - \mathbf{s}$.

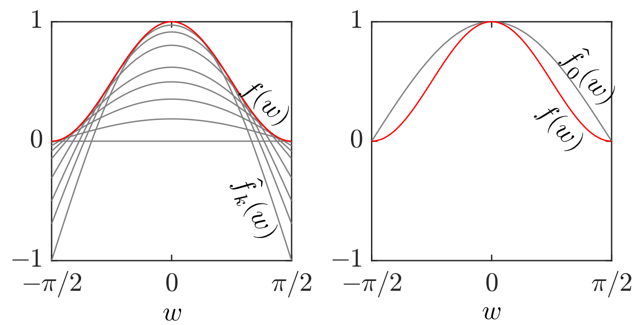

Figure S1. **Square cosine fall-off $f(w)$ and its linearizations.** We denote by w, the angle between the surface normal $n_k(s)$ at point s and the incident ray $(\mathbf{s}' - \mathbf{s})/\|\mathbf{s}' - \mathbf{s}\|$ from $\mathbf{s}'$ to $\mathbf{s}$. We show in the left subplot the first-order Taylor approximation of $f(w)$. The right subplot shows the initial linearization where we constrain the offset parameter $b_0 = 0$.

In the first iteration ($k = 0$) of Gauss-Newton, we do not have the previous estimate ρ_{-1} of the albedo. Therefore, we require a linearization $\hat{f}(\mathbf{n}(\mathbf{s}), \mathbf{s}' - \mathbf{s})$ that does not involve the offset $b(\mathbf{s})$. Constraining $b(\mathbf{s}) = 0$ for all $\mathbf{s}$, and using the fact that (S9) is invariant to a constant scaling of the normal $\mathbf{n}(\mathbf{s})$, we can choose

$$a(\mathbf{s}) = 1, \quad b(\mathbf{s}) = 0 \tag{S13}$$

for all $\mathbf{s}$, from which we obtain (5) and consequently (8). In Figure S1 (right), we plot the offset-constrained linearization $\hat{f}_0(w)$ of $f(w)$. Flatness can be more prominent for scenes that are further away from the relay wall since the light rays are almost collinear, providing less directional information in the transients, as seen with Discobolus in Figure S2.

S4. Thresholding and Masking

Having thus obtained the 3D volume of directional albedo $\boldsymbol{\upsilon} = (\boldsymbol{\upsilon}_x, \boldsymbol{\upsilon}_y, \boldsymbol{\upsilon}_z) \in \mathbb{R}^{V^3 \times 3}$, we generate foreground mask as $\mathbf{m} = \boldsymbol{\upsilon}_z > \alpha \max(\boldsymbol{\upsilon}_z) \in \mathbb{R}^{V^3}$. Using the threshold $\alpha = 0.2$ gives us good masks in practice. The masked directional-

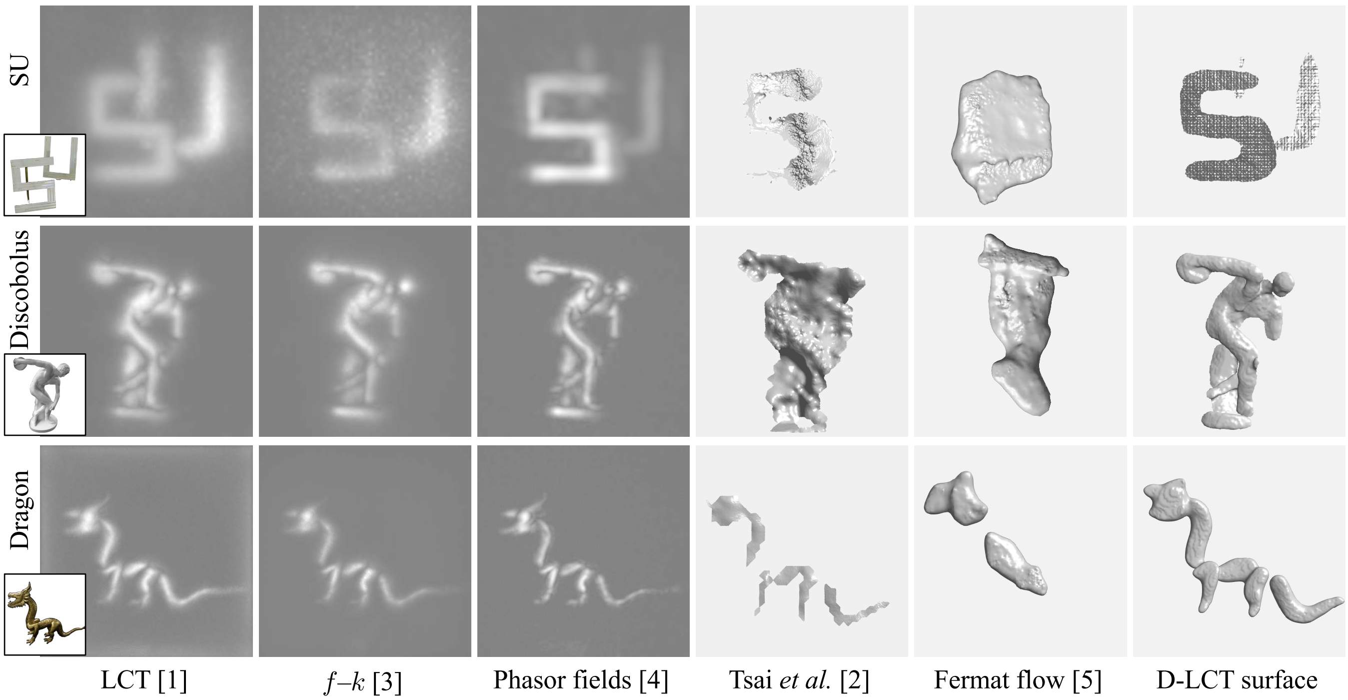

Figure S3. **Albedo and surface reconstructions using captured data:** The extracted normals (c) are used to fit surfaces (d). SU has a spatial resolution of 64×64 pixels (1min exposure), and the remaining scenes, 512×512 (180min exposure). We use $\lambda = 2^0$, 2^3 and 2^3 for SU, Discobolus and Dragon, respectively. Parameters of the other methods were optimized using grid search.

albedo is $\mathbf{m}\boldsymbol{\upsilon} = (\mathbf{m}\boldsymbol{\upsilon}_x, \mathbf{m}\boldsymbol{\upsilon}_y, \mathbf{m}\boldsymbol{\upsilon}_z)$. This is similar to the procedure of Tsai et al. [2]. For simulations, the ground truth masks $\mathbf{m} \in \mathbb{R}^{HW}$ are two-dimensional binary images, so we first replicate $\mathbf{m}$ along the depth dimension.

S5. Additional Results

We provide a more extensive set of experimental results for comparison with various baselines. We show the surface reconstructions from different viewpoints in Figure S2. We show in Figure S3, the albedo and surface reconstructions produced by a number of methods: the light-cone transform [1], f–k migration [3], phasor fields [4], Tsai *et al.* [2], and Fermat flow [5].

References

[1] Matthew O'Toole, David B. Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555(7696):338–341, 2018.

[2] Chia-Yin Tsai, Aswin C. Sankaranarayanan, and Ioannis Gkioulekas. Beyond volumetric albedo—a surface optimization framework for non-line-of-sight imaging. In *CVPR*, 2019.

[3] David B. Lindell, Gordon Wetzstein, and Matthew O'Toole. Wave-based non-line-of-sight imaging using fast f-k migration. *ACM Trans. Graph. TOG*, 38(4):116, 2019.

[4] Xiaochun Liu *et al.* Non-line-of-sight imaging using phasor-field virtual wave optics. *Nature*, :1–4, 2019.

[5] Shumian Xin, Sotiris Nousias, Kiriakos N. Kutulakos, Aswin C. Sankaranarayanan, Srinivasa G. Narasimhan, and Ioannis Gkioulekas. A theory of Fermat paths for non-line-of-sight shape reconstruction. In *CVPR*, 2019.