Supplementary: Role of Transients in Two-Bounce Non-Line-of-Sight Imaging

Siddharth Somasundaram[1] Akshat Dave[2] Connor Henley[1]
Ashok Veeraraghavan[2] Ramesh Raskar[1]
[1]Massachusetts Institute of Technology [2]Rice University
{sidsoma,co24401,raskar}@mit.edu, {akshat,vashok}@rice.edu

1. Implementation Details

Experimental Setup The object of interest is flanked on each side by parallel white walls. We illuminate the right wall, and measure two-bounce light (shadows) on the left wall, as shown in Fig. 1(c) of the main text. We use a two-axis scanning galvonometer (Thorlabs GVS412) to scan a pulsed laser source (Picoquant LDH-D Series) with 640 nm wavelength across a 10×6 grid on the right wall. This results in 60 image acquisitions. We obtain the multiplexed measurement by summing these 60 measurements in post-processing. We use a single-pixel SPAD (MPD PDM Series) and another set of galvo mirrors (Thorlabs GVS012) to scan the field of view on the left wall across a 80×55 detector grid. Scanning a single pixel effectively mimics the output of a SPAD array. The laser source has a pulse repetition rate of 20 MHz, average power of 0.5 mW, and pulsewidth of 90 ps (FWHM). The SPAD has a timing bin width of 8 ps and timing jitter of 50 ps (FWHM). We use a pixel dwell time of 100 ms. The overall instrument response function was measured to be 128 ps (FWHM). A time-correlated single photon counter (TCSPC) (Picoquant Hydraharp 400) was used to synchronously record photon arrival times. The experimental setup is shown in Fig. 1. We don't include an image of the occluder for simplicity, though in a practical application there would be an occluder blocking the mannequin from the camera. We also make the following assumptions about our scene: (1) all shadows are caused by occluded object and not the occluder itself, (2) the occluded object is completely opaque, and (3) there are no interreflections or higher order bounces (e.g. three-bounce light).

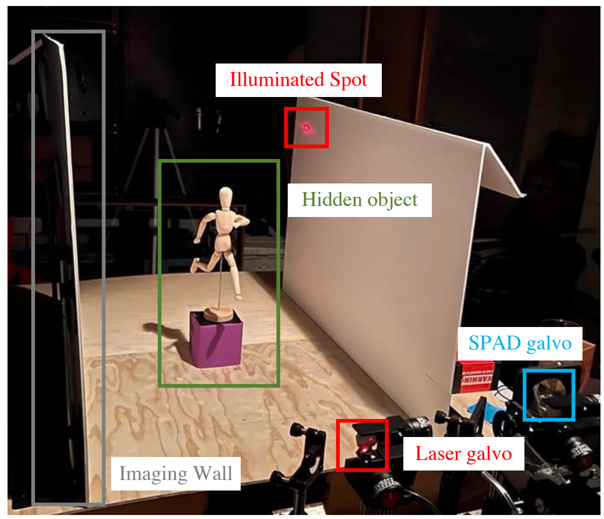

Figure 1. **Experimental Setup.** The hidden object lies between two planar walls. The right wall is the illumination wall and the left wall is the imaging wall. One set of galvo mirrors steers the laser toward a point on the illumination wall. This spot uniformly radiates light in all directions towards the imaging wall, where a cast shadow of the hidden object can be observed. The light from a point on the imaging wall returns to the single-pixel SPAD. Another set of galvos is scanned across the field of view of the right wall to mimic the behavior of a SPAD array. The multiplexed images are obtained by summing up all the per-laser spot transient images together in post-processing.

Estimating Shadow Transients In order to perform image reconstruction, it is necessary to estimate i_0. Doing so requires estimation of the time of flight and radiance throughput of every virtual source and virtual detector pair. The time of flights can be estimated as $(\|\mathbf{l} - \mathbf{g}\| + \|\mathbf{s} - \mathbf{l}\| + \|\mathbf{s} - \mathbf{h}\|)/c$ using known locations of $\mathbf{l}$ and $\mathbf{s}$ from one-bounce signals, where $\mathbf{g}$ and $\mathbf{h}$ are the laser and camera positions respectively. Estimating the radiance throughput, however, is non-trivial due to model mismatch and SPAD photon statistics. As mentioned in Eq. (1) of the main text, the radiance is modeled as

$$\alpha_{\mathbf{s},\mathbf{l}} = \frac{\alpha}{\|\mathbf{l} - \mathbf{s}\|^2}, \tag{1}$$

where α is a uniform scaling for all $\mathbf{l}$-$\mathbf{s}$ pairs. Due to model mismatch we do not directly know α. Instead, we treat α as a hyperparameter. We compute $\mathbf{A}^\top \hat{\mathbf{I}}_0$ and $\mathbf{A}^\top \mathbf{I}_l$ separately, where $\hat{\mathbf{I}}_0 = \mathbf{I}_0 / \alpha$ is known from just ToF information. We then tune $\alpha \mathbf{A}^\top \hat{\mathbf{I}}_0 - \mathbf{A}^\top \mathbf{I}_l$ based on an initial estimate of α

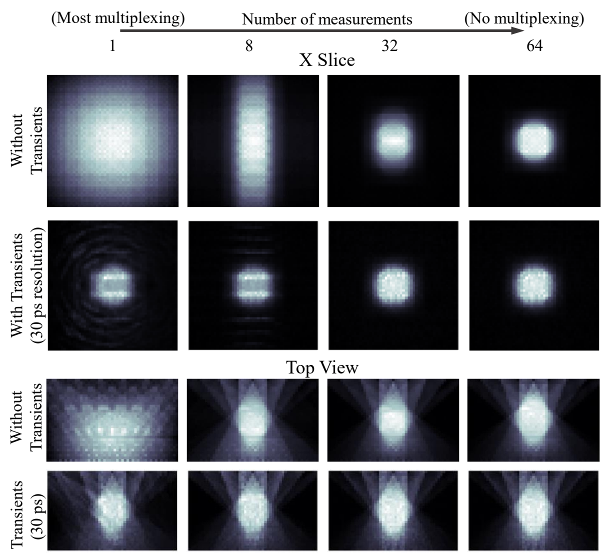

Figure 2. **Number of Measurements vs. Temporal Resolution on Simulated Data.** A fixed spatial resolution of 128×128 is used with varying temporal resolution and number of measurements. As expected, the reconstruction quality improves with number of measurements both with and without transient information. ToF-based methods are robust to multiplexing, but both methods are inherently limited to reconstructing the visual hull of the object.

from the laser power. We find that tuning α is easier in the backprojected space than in transient space $\alpha\hat{\mathbf{I}}_0 - \mathbf{I}_l$.

Backprojection Reconstruction In practice, computing backprojection $\mathbf{f}_{bp} = \mathbf{A}^\top \mathbf{I}_s$ as a matrix multiplication is impractical due to the large dimensionality of $\mathbf{A}$. Instead, we compute $\mathbf{f}_{bp}$ one entry at a time (i.e. one voxel at a time). For each voxel $\mathbf{x}_i$, we compute the dot product of the ith column of $\mathbf{A}$ and the shadow transient $\mathbf{I}_s$ ($\mathbf{x}_i = \mathbf{A}_i^\top \mathbf{I}_s$). Computing backprojection one voxel at a time ensures that the matrices can be stored in memory. Recall that the column space of $\mathbf{A}$ represents the expected space-time measurement for every voxel in the hidden scene (i.e. the space-time impulse response of individual voxels). This impulse response can be computed by using the known positions of the virtual sources and virtual detectors. If a virtual source lies along the ray connecting the voxel and a virtual detector, then an impulse would appear in the transient measurement of that particular virtual detector. Using the known positions of the virtual source and virtual detector, the specific time of flight of that impulse can be computed. Repeating this process for all virtual detectors yields the space-time impulse response of a particular voxel.

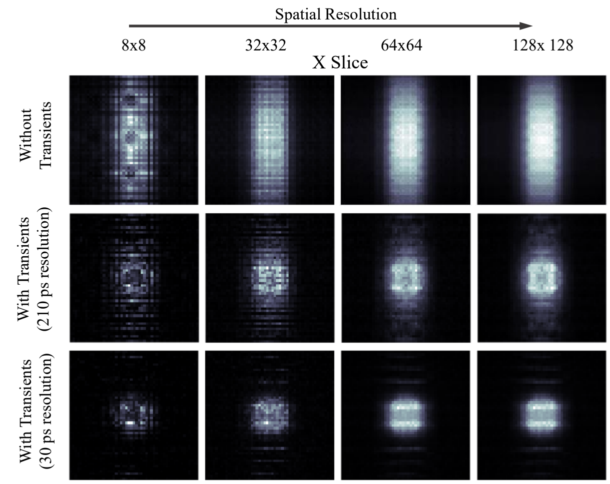

Figure 3. **Spatial vs. Temporal Resolution on Simulated Data.** The number of measurements is kept fixed at 8 and the temporal and spatial resolution are varied. We see that hole artifacts appear in reconstructions as spatial resolution is reduced.

2. Additional Simulated 3D Results

We further analyze the tradeoffs between spatial resolution, temporal resolution, and number of measurements using simulated 3D data. We render transient profiles for generated scenes using a time-of-flight extension on the Mitsuba renderer [2]. We consider a setup with a cube box, as shown in Fig. 2. We can see that for smaller number of measurements, reconstructions require projections over monotonically larger surface areas for the intensity-only case, resulting in blurrier reconstructions. We note that there are little to no changes in performance with ToF information even with more measurements, suggesting the robustness of ToF methods in the presence of multiplexing. However, we can see persisting conical artifacts in both reconstructions. These artifacts are related to the visual hull of the object, which is an inherent limitation of shadow-based reconstruction methods [1]. When multiplexing with 8 laser spots without transient information (first row, second column), we see a vertical shape. This shape occurs because the multiplexed beams are arranged horizontally. Therefore, the multiplexing artifacts are concentrated along the vertical direction. Fig. 3 compares the impact of spatial vs. temporal resolution. As spatial resolution decreases, the hidden scene is subsampled and hole artifacts appear in the reconstruction. Other than these hole artifacts, the reconstructions are very similar to that shown in Fig. 2.

References

[1] Aldo Laurentini. The visual hull concept for silhouette-based image understanding. *PAMI*, 16(2):150–162, 1994. 2

[2] Adithya Pediredla, Ashok Veeraraghavan, and Ioannis Gkioulekas. Ellipsoidal path connections for time-gated rendering. *ACM Transactions on Graphics (TOG)*, 38(4):1–12, 2019. 2