Computational 3D Imaging with Position Sensors

Jeremy Klotz
Columbia University

Mohit Gupta
University of Wisconsin-Madison

Aswin C. Sankaranarayanan
Carnegie Mellon University

Abstract

Underlying many structured light systems, especially those based on laser scanning, is a simple vision task: tracking a light spot. To accomplish this, scanners use conventional CMOS sensors to capture, transmit, and process millions of pixel measurements. This approach, while capable of achieving high-fidelity 3D scans, is wasteful in terms of (often scarce) sensing and computational resources. We present a structured light system based on position sensing diodes (PSDs), an unconventional sensing modality that directly measures the centroid of the spatial distribution of incident light, thus enabling high-resolution 3D laser scanning with a minimal amount of sensor data. We develop theory and computational algorithms for PSD-based structured light under a variety of light transport effects. We demonstrate the benefits of the proposed techniques using a hardware prototype on several real-world scenes, including optically-challenging objects with long-range interreflections and scattering.

1. Computational Position Sensors

Traditional photodetectors measure the intensity of light incident on them. To measure the spatial intensity distribution of a scene (i.e. an image), a pixelated array of detectors such as a CMOS sensor is used, which requires sophisticated readout mechanisms and invariably results in large bandwidth, compute, and memory requirements. There are several downstream vision applications that do not require the entire image, instead relying on some aggregate statistics (e.g. histograms [13, 5]) of the intensity distribution. For such applications, it is often desirable to create *compact*, memory- and compute-efficient summary statistics of the images. This raises a natural question: Why capture high-resolution images, only to compress it later? Can we design imaging modalities that *directly* capture compact and accurate statistics of the scene's intensity distribution?

We consider position sensing diodes (PSDs), a sensing modality that can directly characterize the spatial distribution of incident light. A PSD, in addition to measuring the amount of incident flux, also computes the centroid of the image formed on it. While the centroid is a rather coarse description of the spatial intensity distribution, there are numerous applications where this is sufficient. For example, consider a 3D scanning system where a laser spot scans an object. In conventional scanning, the camera captures an image for each spot location, often resulting in prohibitively large amounts of data. In an ideal scenario, where the image consists of a single illuminated spot, the centroid measured by the PSD tracks the light spot location on the sensor, which can be used to establish projector-camera correspondences and obtain 3D scans via triangulation at high-speeds, while maintaining low power and compute budgets.

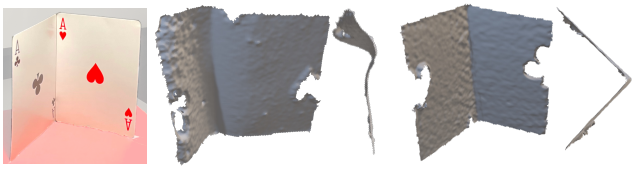


Figure 1: **Structured light 3D scanning using a position sensing diode.** Computational position sensors can enable high-speed, high-resolution 3D imaging at low power and compute budgets. However, global illumination effects such as interreflections in a V-groove significantly bias the measurements, as seen in the poor surface reconstruction in the middle. We propose methods that are robust to such effects, enabling the high-fidelity reconstruction on the right.

Position sensing and global illumination. Most real-world scenes often have global illumination effects such as interreflections and subsurface scattering. As a result, even a single laser beam can create an intensity image consisting of a large number of illuminated spots. This would bias the centroid from the "direct" location of the laser spot, which in turn will distort the estimated 3D shape resulting in large, systematic errors. An example with a V-groove is shown in Fig. 1. Such effects could limit the scope of this otherwise powerful technology.

This paper proposes a technique for 3D scanning with PSDs that is robust to global illumination. Our primary contribution is a computational imaging design that reduces

the structured biases induced by global illumination on the centroid measurements. Our proposed method accurately recovers shape in scenes with strong global illumination effects, such as the V-groove in Fig. 1. We develop theoretical foundations for PSD-based structured light (SL) imaging under a wide range of global illumination effects and develop computationally lightweight, practical techniques that mitigate global illumination effects in PSD-based SL.

Contributions. This paper develops theory and algorithms for a novel 3D scanning approach based on PSDs.

- *3D scanning with PSDs.* We present a SL system for 3D scanning with PSDs and analyze its performance under noise and defocus.
- *Handling global light.* Under global effects such as interreflections, the centroid measurements differ significantly from the illuminated (direct) point. To resolve this, we propose a direct-global separation method for PSDs that operates *without having access to images* captured by conventional cameras. This proposed approach extends prior work by Nayar *et al*. [28] using duality of light transport and a robust estimation unique to our setup.
- *Hardware prototype.* We design and implement a prototype of the proposed scanner and demonstrate its performance across different scenes and operating modes.

The software and data are available on our project page [21].

A PSD-based 3D scanning system does not have pixelation at either the sensor or the projector (which is typically implemented using 2D galvos). The resolution limits for the proposed approach are *atypical* in that they are determined by the size of the laser spot and its intensity. The system also lacks the constraints imposed by readout in traditional image sensors and simultaneously overcomes the fill factor issues endemic to asynchronous readout sensors. Furthermore, due to lack of pixelation, defocus at the sensor does not adversely affect the resulting 3D scans.

Limitations. The noise and errors in centroid measurements obtained with a PSD are affected significantly by the total incident flux. Hence, in its current implementation, it is not conducive to scanning large scenes where light falloff due to $1/r^2$-losses can be significant. Furthermore, the acquisition time for a single 3D scan is limited by the analog-to-digital converter's (ADC) sampling rate and the PSD amplifier's bandwidth. While these bandwidths may be pushed to hundreds of MHz or a few GHz, our lab prototype has modest specifications, resulting in long scan times.

2. Related Work

Position sensing diodes. The lateral photoeffect, the phenomenon that enables an analog computation of the image centroid, was first discovered in 1930 by Schottky. Subsequent work [37] provided a theoretical explanation [23] and analyzed various PSD configurations [38]. Research in PSDs now includes applications spanning microscopic dimensioning [33], object tracking [2, 35, 32], ego-motion sensing [20], and closed-loop beam steering [34]. Idesawa proposed a triangulation-based range estimation systems using PSDs [15]. This method assumed only direct reflection, and was designed to measure the depth of a *single* point. Our goal is to develop a practical 3D scanning system that can operate under real-world conditions.

Sparse event-driven sensors. Event cameras [22] support high-speed, sparse, and asynchronous data readout, which has been used for high-speed SL 3D imaging systems [24, 26]. The proposed PSD SL approach shares the same motivation in reducing the data bandwidth requirement, while achieving high-fidelity 3D scans. When compared to PSDs, event sensors have high circuit complexity for each pixel, thus lowering the fill-factor and low-light sensitivity, and limiting the overall spatial resolution.

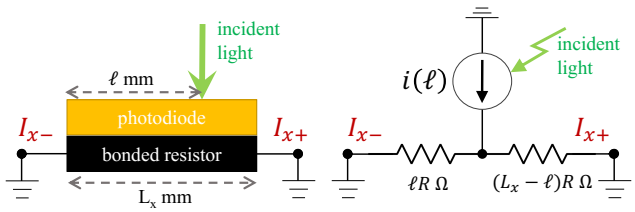
Global illumination. Active 3D imaging techniques such as SL and time-of-flight (ToF) largely assume only direct reflection, i.e., scene points receiving illumination only directly from the light source. This assumption is violated in the presence of global illumination effects [28], which can result in large, systematic errors in the recovered 3D geometry [6, 11, 12, 9, 10, 8, 29, 1, 31, 4, 7, 16, 30]. There is an extensive body of work devoted to developing SL 3D imaging techniques that are robust to global illumination effects. This includes methods based on projected pattern design [3, 25, 9, 11, 8, 4], as well as hardware approaches that use a digital micromirror device [31] or a camera's rolling shutter [29] synchronized with a projector to optically block a large portion of global light.

3. 3D Scanning with Position Sensing Diodes

For simplicity in exposition, we consider only an ideal PSD whose lateral photocurrents are linear with the light spot position; we refer the interested reader to [38] for a detailed analysis of common diode geometries.

3.1. Measurement Model Underlying PSDs

A PSD can be modelled as a traditional photodiode that is bonded on top of a resistor. Figure 2(a) shows a schematic of a 1D PSD of length L_x mm, where a laser spot is incident at a distance ℓ mm from its left terminal. The equivalent circuit of this system is shown in Fig. 2(b). The photocurrents measured at the two terminals are given by the current di-



(a) Schematic of a PSD (b) Equivalent circuit

Figure 2: **A schematic of a PSD and its equivalent circuit.**

vider model [17, 18]:

$$I_{x-} = \frac{L_x - \ell}{L_x}\, i(\ell), \quad I_{x+} = \frac{\ell}{L_x}\, i(\ell), \tag{1}$$

where $i(\ell)$ is a current that models the stream of photoelectrons produced by the incident light. We can now estimate the location of the incident spot using the measurements as:

$$\widehat{\ell} = L_x \frac{I_{x+}}{I_{x+} + I_{x-}}. \tag{2}$$

When multiple light sources are incident on the PSD—say, an image is formed on the diode—the model in Fig. 2(b) includes multiple current sources along the resistor. The photocurrents follow from linear superposition:

$$I_{x-} = \frac{1}{L_x} \int (L_x - \ell) i(\ell) d\ell, \tag{3}$$

$$I_{x+} = \frac{1}{L_x} \int \ell\, i(\ell) d\ell, \tag{4}$$

and the expression for the spot location in (2) evaluates to:

$$\widehat{\ell} = \frac{\int \ell\, i(\ell) d\ell}{\int i(\ell) d\ell}. \tag{5}$$

Since $i(\ell)$, the strength of the current source at the location ℓ, is linear in the incident flux, $\widehat{\ell}$ in (5) evaluates to the weighted centroid of the image formed on the PSD.

Extension to 2D PSDs. The expressions above extend mutatis mutandis to the 2D scenario. In practice, a transimpedance amplifier converts the photocurrents at each terminal to voltages. After some simple analog filtering that performs appropriate subtractions and additions on the terminal voltages, we can access the following voltages:

$$\begin{bmatrix} V_x \\ V_y \\ V_s \end{bmatrix} = G \iint_{x,y} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} i(x,y)\, dx\, dy, \tag{6}$$

where G is a constant. Here, V_s is proportional to the total light incident on the diode, the only measurement a traditional photodiode provides. Hence, the weighted centroids

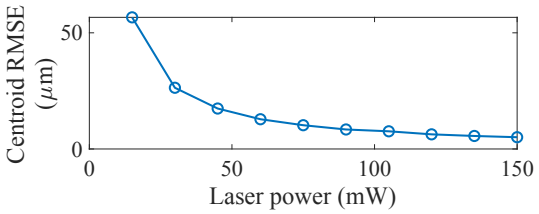


Figure 3: **Centroid resolution and incident light power.** Here, a laser with varying power is projected on a whiteboard approximately 30cm using the prototype described in Sec. 5.1. At 150mW, the centroid resolution attains 5μm.

(C_x, C_y) are estimated as:

$$C_x = L_x \frac{V_x}{V_s}, \quad C_y = L_y \frac{V_y}{V_s}, \tag{7}$$

where L_x and L_y are the dimensions of the PSD.

3.2. Properties of Estimated Centroids

Noise modeling. In the presence of read noise, we can characterize V_x, V_y, and V_s as random variables. Using basic statistical analysis, which we present in the supplemental material, we can approximate the variance of C_x. If we denote the noise-free centroid as μ_x, then the measured centroid variance along the x dimension can be estimated by:

$$\frac{\sigma^2}{E(V_s)^2} \left(\mu_x^2 + L_x^2\right), \tag{8}$$

where σ^2 is the variance of the read noise on each channel (V_x, V_s), and $E(V_s)$ is the expected value of V_s. The centroid variance along the y dimension follows symmetrically.

We can use the (inverse of the) variance to characterize the resolution of the centroid estimate. This suggests that the resolution is position-dependent; refer to the supplemental material for empirical evidence supporting this. This analysis of position-dependent resolution is derived from signal-independent noise on the measurement channels and differs from prior work that examines the diode geometry [38] or a phase-based positioning method [27],

Dependence on incident flux. Figure 3 empirically measures the centroid resolution with varying incident flux, which depends on the laser power and object albedo. This is in agreement with (8), which suggests that the standard deviation of the centroid is inversely proportional to V_s.

Defocus invariance. The estimated centroids are largely invariant to defocus blur. Specifically, let $I(x,y)$ be the image formed on the PSD surface and $b(x,y)$ be a spatially-invariant point spread function. The centroid of the blurred

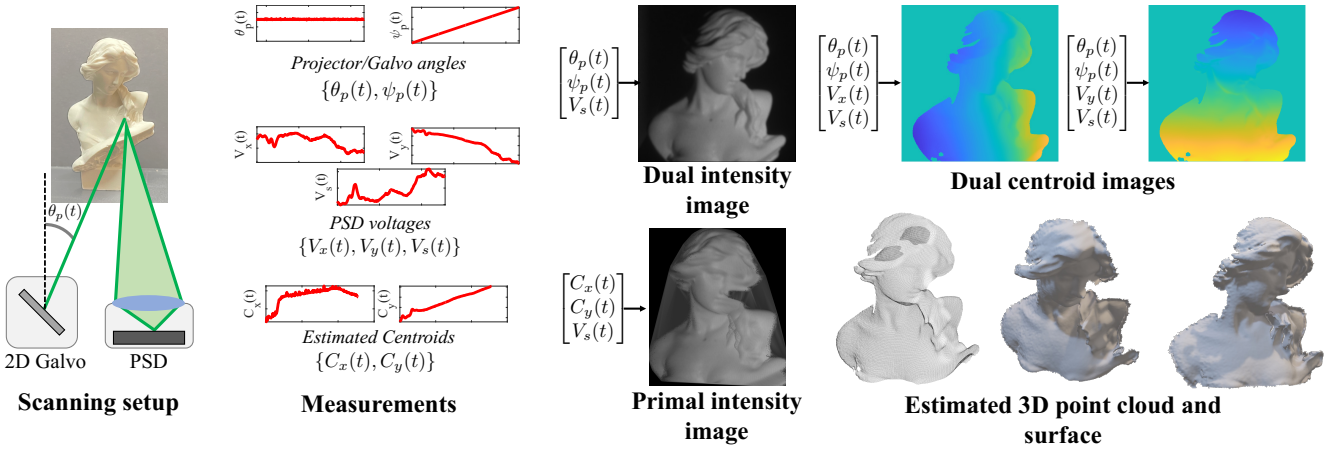


Figure 4: **3D scanning using PSDs.** On the left, a laser projector scans a light spot across the scene. We synchronously measure the projector angles and PSD voltages, producing light spot positions (centroids) at each time interval. These measurements are sufficient to recover the dual image (an image from the projector's view) and the primal image (an image from the PSD's view) by resampling the measurements on a discrete grid. At each projector location, we compute the scene depth by triangulating the projector and PSD rays to obtain a point cloud and complete surface reconstruction.

image $(I * b)(x, y)$, on a sufficiently large sensor is:

$$\begin{bmatrix} C'_x \\ C'_y \end{bmatrix} = \begin{bmatrix} C_x + C^b_x \\ C_y + C^b_y \end{bmatrix}, \tag{9}$$

where (C^b_x, C^b_y) is the centroid of the blur kernel. Hence, the centroid is unaffected by any symmetric blur kernel, which is often the case with defocus blur. This simple result, however, breaks down near the boundary of the sensor as the defocus kernel is clipped, and that resulting asymmetry causes biased centroids.

3.3. 3D Scanning using PSDs

To obtain 3D scans, we envision a setup with a point-scanning projector, a 2D galvo that scans a collimated laser beam, and a PSD with a lens focused on the scene. The 2D galvo continuously steers the beam over its field-of-view, while the PSD reports a triplet of voltage values as described in (6). We now have a time-synchronized pentuplet set of measurements:

$$\{\theta_p(t), \psi_p(t), V_x(t), V_y(t), V_s(t)\}, \tag{10}$$

where $\{\theta_p(t), \psi_p(t)\}$ describe the 2D angles scanned by the galvo, in the projector's reference, and $\{V_x(t), V_y(t), V_s(t)\}$ are the analog voltages measured by the PSD. Using (7), we can compute the set of associated centroids $\{C_x(t), C_y(t)\}$ as well.

Primal and dual images. One way to interpret the measurements from our device is via the mindset of dual photography [36]. Consider a laser scanning system and a traditional photodiode, where we measure only the incident flux for each projected ray. We can create the so-called "dual image" by associating the intensity/flux measured with the corresponding projector ray; this dual image is a view of the scene from the projector's view, illuminated from the PSD's location. Similar dual images can be estimated for the centroids as well, where we visualize the centroids in the projector's viewpoint. This remapping of measurements in the projector's view is useful both as a visualization tool and for subsequent processing techniques that we will develop.

Since we have access to the centroid of the light intensity on the sensor, we can also create a "primal image" of the scene observed from the viewpoint of the PSD by associating the measured flux with the centroid. Figure 4 shows the processing pipeline including the primal-dual image sets for a simple scene. Note that these stereo images provide sufficient information to compute depth. However, unlike traditional stereo where correspondences need to be estimated, our system outputs corresponding pairs of projector and sensor locations.

Obtaining 3D scans. Our measurement set gives us corresponding pairs of projector and sensor locations in $\{\theta_p(t), \psi_p(t)\}$ and $\{C_x(t), C_y(t)\}$, respectively. With knowledge of the projector and camera intrinsics and extrinsics, we can obtain a set of 3D points—one for each t—by triangulating the pre-images of the respective points. Since the centroid estimates are noisier than the projector rays, we enforce the epipolar constraint on the centroids using the known PSD-projector geometry prior to triangulation. The depth map is then smoothed in the projector viewpoint using a bilateral filter. Finally, we obtain surface normals on the estimated point cloud using simple local plane fitting

techniques and generate a surface using screened Poisson surface reconstruction [19].

The effect of global illumination. A complicating factor in obtaining reliable 3D scans with PSDs is the presence of global illumination. While the projector *directly* illuminates only a single scene point, that point can *indirectly* illuminate other scene points. The image formed on the PSD, hence, is no longer a single spot but an extended image depending on the specific global light transport effect.

Figure 5 illustrates the effect of global light on the measurements and the final 3D scans. For example, convex opaque objects have little global effects, which results in unbiased centroids; even in the presence of sensor defocus, our scans remain fairly reliable due to properties discussed in Sec. 3.2. Similarly, short-range global effects like subsurface scattering generally do not bias the centroid measurements, especially for dense scatterers like wax; here, while the directly illuminated spot does spread out, the spread is often symmetric except near geometric edges, leading to little change in the centroids. Long-range global effects, however, can significantly bias the centroid measurements. This is easily seen in scenes with strong interreflections, such as a V-groove. Here, the presence of global illumination produces significant errors in an object's estimated shape.

4. Handling Global Illumination

To reduce the effects of global illumination, we design a computational imaging system around the PSD.

Imaging setup. We propose a novel imaging system that collocates an amplitude spatial light modulator (SLM) with the PSD. Figure 6 illustrates the imaging setup. The scene is focused on to the SLM using an imaging lens through a polarizing beamsplitter, which linearly polarizes the light. The SLM rotates the polarization state; in a binary operating mode, the pixels selected as "white" on the SLM change their polarization state by 90°; such light reflects off the beamsplitter towards the sensors. We use two lenses, first to collimate the light from the SLM, and the second to focus it on the PSD. A helper camera is introduced between the two lenses via a non-polarizing beamsplitter; the lens in front of this camera is focused at infinity, ensuring that it observes the SLM plane. To maximize light levels on the PSD, we use a 90:10 beamsplitter.

Suppose we denote the amplitude mask introduced by the SLM as a binary image $m(x, y)$; then the measurement model of (6) is changed to:

$$\begin{bmatrix} V_x \\ V_y \\ V_s \end{bmatrix} = G \iint_{x,y} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} i(x, y)\, m(x, y)\, dx\, dy. \quad (11)$$

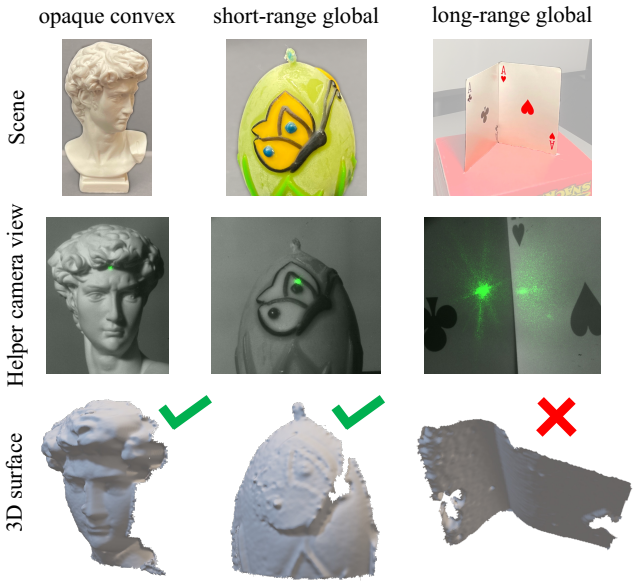


Figure 5: **Global illumination significantly affects the surface reconstructions.** The second row shows the scene from a helper camera view to illustrate such global effects. In the first two columns, the bust and candle are faithfully reconstructed, despite the candle's subsurface scattering as that effect does not bias the measurements. However, long-range interreflections between the playing cards significantly bias the measurements, causing an erroneous reconstruction.

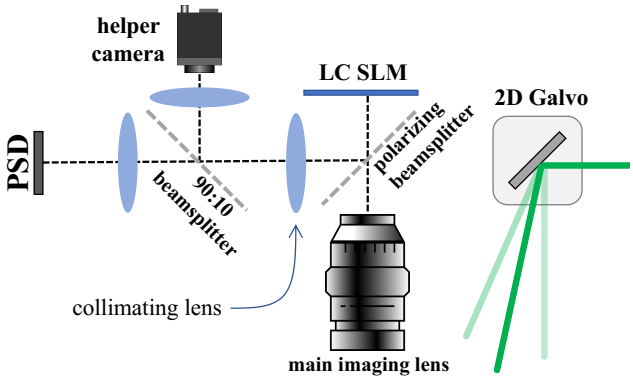


Figure 6: **Optical schematic.** The main imaging lens focuses the scene onto the SLM. The PSD images the masked scene on the SLM. A helper camera views the scene for calibration and ground truth depth experiments.

As we will see next, this spatial blocking of light allows us to remove the global illumination effects by capturing multiple sets of measurements under different masks.

Connections to prior work. The optical setup above can be interpreted as the *optical dual* of Nayar *et al.* [28], where

instead of projecting high-frequency patterns, we introduce a high-frequency spatial mask on the sensor. However, there are significant differences in the additional centroid measurements that the PSD provides; as we will see next, we can design a noise-robust technique that avoids the traditional min-max processing of [28].

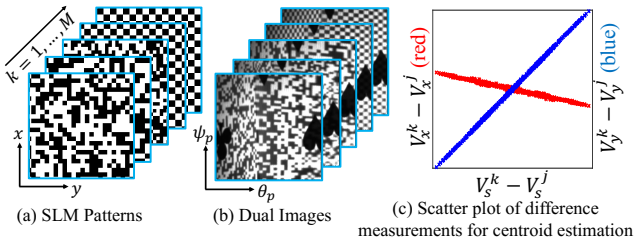
Direct-global separation. Our direct global separation procedure repeatedly raster scans the object, cycling through a set of mask patterns on the SLM. Suppose that we display M masks, denoted $\{m^k(x,y)\}_{k=1}^M$, on the SLM; for each mask, we obtain the pentuplet of measurements from the devices as in (10). We can then construct a set of M dual images, denoted as $\{V_x^k, V_y^k, V_s^k\}$, corresponding to the PSD measurements at each point on the projector plane. This dual image stack is visualized in Fig. 7. Note that each dual image is constructed by resampling the PSD measurements on a discrete grid in the projector's view. This discrete grid defines the "projector pixels" to which our subsequent analysis will refer.

One approach to perform direct-global separation on these images is to use the processing framework suggested in Nayar *et al.* [28], which we refer to as the min-max technique. Specifically, if the global light transport is only comprised of low-frequency effects, then the global component measured at a projector pixel is roughly the same, regardless of the mask (provided the mask is comprised of high frequency patterns). The measured global component in this case would be approximately half the amount of global light we would receive if we did not mask any pixels. Hence, the minimum at each pixel across the stack of M images provides this half-global component, and the maximum across the stack provides the sum of the direct and half-global component. This strategy can be applied to our measurements as well. However, it discards all but two of the captured measurements at each pixel. As we see next, we can exploit the specifics of our setup to use the entire stack of M measurements to extract the centroid at a higher fidelity.

Robust centroid estimation. Consider now the intensity $V_s^k[m,n]$ measured on the PSD that is associated with a projector pixel $[m,n]$ in the dual image and the k-th mask pattern on the SLM. We can write

$$V_s^k[m,n] = \alpha^k[m,n]d[m,n] + g[m,n], \tag{12}$$

where $d[m,n]$ and $g[m,n]$ are the direct and half-global intensity components associated with this projector pixel. The term $\alpha^k[m,n]$ is the attenuation of the direct component due to the k-th mask, and since the directly illuminated spot can be partially blocked by the SLM, we assume $\alpha^k[m,n] \in [0,1]$. Note that, as with Nayar *et al.* [28], the half-global term is assumed to be independent of the specific pattern shown on the SLM. Since we will process each



(a) SLM Patterns (b) Dual Images (c) Scatter plot of difference measurements for centroid estimation

Figure 7: **Robust centroid estimation.** Our method displays masks on the SLM (a), captures the corresponding PSD measurements (V_s dual image shown in (b)), and performs a linear regression technique to estimate the centroid of the directly illuminated spot. In (c), we plot the pairwise differences of real data using Eq. (15) to illustrate the linear relationship (see text for details).

dual image pixel independently, we drop the dependence on $[m,n]$ in the following derivation.

Since the direct component corresponds to a tightly focused spot on the PSD—ideally, a delta function at a location (C_x, C_y)—we can express the measurements at the dual pixel $[m,n]$ as

$$\begin{bmatrix} V_x^k \\ V_y^k \\ V_s^k \end{bmatrix} = \begin{bmatrix} C_x \\ C_y \\ 1 \end{bmatrix} \alpha^k d + \begin{bmatrix} g_x \\ g_y \\ g \end{bmatrix}. \tag{13}$$

Here, g_x is the x-weighted global intensity, given as

$$\iint_{x,y} x\, g(x,y)\, m^k(x,y)\, dx\, dy. \tag{14}$$

The smoothness arguments made earlier for global illumination allows us to approximate this integral with an intensity g_x that is independent of the mask. Now consider a pair of difference measurements, $V_x^k - V_x^j$ and $V_s^k - V_s^j$; from (13), they are related as

$$V_x^k - V_x^j = C_x(V_s^k - V_s^j). \tag{15}$$

This suggests that the we can directly estimate the centroid of the direct component by applying linear regression techniques to these pairwise differences. Critically, this allows us to use *all* the measurements made by the PSD. Applying this across all projector pixels, we can estimate the centroids in the absence of global illumination. We visualize the associated measurements in Fig. 7.

Effectiveness of global suppression. We implemented a physically-accurate two-bounce renderer and simulated the setup in Fig. 6 for two scenes: a V-Groove and a concave bowl. Figure 8 shows relative performance of various methods on both scenes, and Fig. 9 characterizes performance as

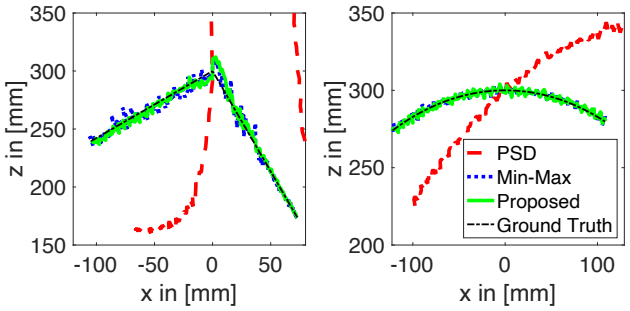


Figure 8: **Simulation results demonstrating global illumination suppression.** (left) a V-Groove, and (right) a concave bowl. We compare reconstuction without global suppression (denoted as "PSD"), Nayar *et al*. global suppression (denoted as "Min-Max"), and the proposed work that estimates centroids robustly. The plots show a 2D cross section of 3D reconstuctions. There are significant biases in the reconstructed shapes in the absence of global suppression. Both the Min-Max approach and the proposed approach suppress these biases; however, as seen in Fig. 9, their performance under noise is significantly different.

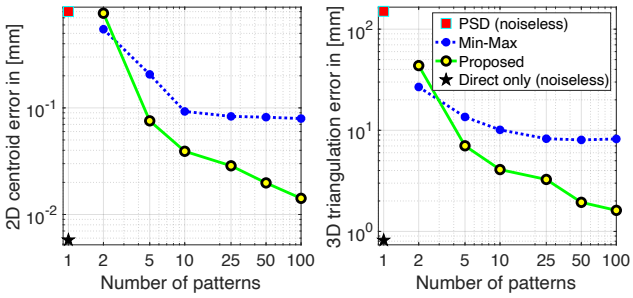
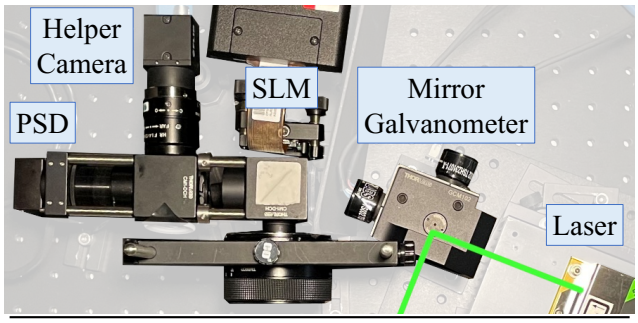


Figure 9: **Error as a function as number of SLM patterns.** Simulation-based performance evaluation as a function of the number of SLM mask patterns used, characterizing 2D centroid localization of the laser spot (left) and the 3D triangulation error (right). The min-max global suppression approach, adapted from Nayar et al., which relies only on information on maximum and minimum intensity measurements does not significantly benefit from an increase in the number of patterns. In contrast, the robust centroid estimation significantly improves with increasing number of patterns. For comparison, we provide the reconstruction error associated with a direct-only reconstruction.

a function of the number of patterns used for global suppression. For Nayar *et al*. and the proposed technique, we used random binary patterns for the mask. Since the laser spot on the scene had a finite radius corresponding to ~ 8 pixels on the SLM, we also performed a parameter sweep over the size of the binary patches and picked the best performing patch size for each technique in each trial.



Component	Part Number
PSD	Thorlabs PDP90a
Helper camera	Point Grey FL3-U3-13E4C-C
Spatial Light Modulator	Holoeye HES6001
2D Galvonometer	Thorlabs GVS002
Laser	Coherent Sapphire LP 532-150
ADC	NI USB-6343

Figure 10: **Component layout of our prototype.**

These simulation results provide two clear conclusions. First, we observe that in the absence of global suppression, the recovered depth is grossly inaccurate. Second, Nayar's method does not benefit as much as the proposed approach from increasing the number of measurements; this can be attributed to their reliance on just minimum and maximum statistics, which ignores a large majority of captured data.

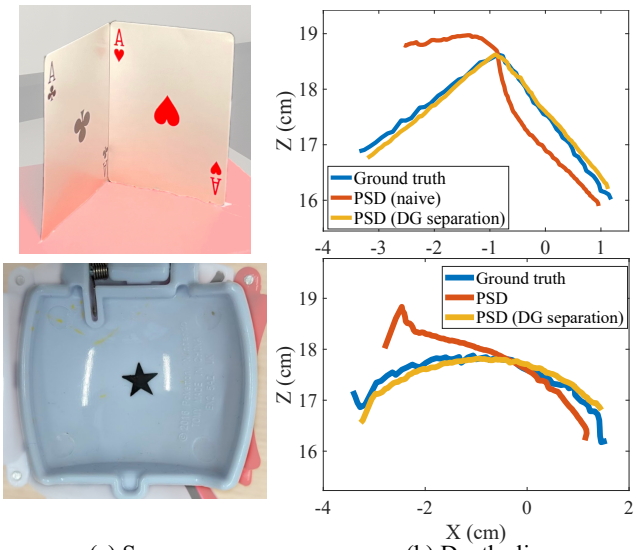
5. Experimental Results

We now validate the performance of our technique on a lab prototype.

5.1. Lab Prototype

We describe the hardware prototype and calibration procedure. Our system depicted in Figs. 6 and 10 uses a 10×10mm PSD. The PSD observes the scene through a 24mm lens. The helper camera observes the scene from the PSD's viewpoint and is only used for initial calibration and obtaining ground truth depth. The projector is a 165mW 532nm laser into a 2D mirror galvanometer. The ADC measures the system's five measurements in Eq. (10) and controls the galvo mirrors at 15KHz.

To calibrate the system, we obtain 3D-2D point correspondences by imaging a 10×10 point array on a planar checkerboard at 4 different depths. The calibrated helper camera provides the point array locations in world coordinates. Following the camera calibration procedure described in [14], we initially estimate the intrinsic and extrinsic parameters via RANSAC with the direct linear transform. We refine this estimate and include a radial distortion model via Levenberg-Maraquardt optimization. The final PSD calibrations attains 14.1μm 2D reprojection RMSE



(a) Scene (b) Depth slice

Figure 11: **Effectiveness of global illumination suppression.** A line scan of two scenes with strong interreflections demonstrates our method's robustness to global illumination. Each object's concave shape creates large, systemic centroid errors, resulting in erroneous depth estimates (red). In both scenes, our proposed method (yellow) closely tracks the ground-truth depth obtained using the helper camera.

and 98.1μm 3D RMSE over the calibration inliers ($n = 258$ point correspondences).

To obtain robustness to ambient light, we capture a single measurement without any illuminated laser spot and subtract it from subsequent measurements. We could further suppress ambient light using a narrow band-pass filter to only pass the laser illumination.

Effectiveness of global illumination suppression on experimental data. Figure 11 showcases the effectiveness of the proposed direct-global separation scheme on two objects with strong interreflections using the lab prototype described in Sec. 5.1. Here we scan a line across two objects: a V-groove and concave object. We compare the resulting depth slices with the ground truth depth obtained using a helper camera. Our proposed method (yellow) closely tracks the ground truth depth, while the naive centroid estimates (red) are severely biased due to interreflections.

To further quantify the method's effectiveness, we fit two planes to the V-groove surface in Fig. 1. The reconstruction using our direct-global separation achieves a 320μm plane fit root-mean-square error (RMSE) with an angle $\theta = 89°$ between the two cards (ground truth: $92°$). By contrast, the surface reconstruction without global illumination suppression attains a 1598μm plane fit RMSE, with an angle $\theta = 113°$.

Surface reconstructions. Figure 12 presents results using our proposed direct-global separation method on scenes with strong long-range global illumination. Each scene is captured under different masks on the SLM, and we compare the reconstructions using both the min-max and regression technique to estimate the centroid of the directly illuminated light spot. Our regression technique is less noisy than the min-max technique.

Scan time. Our method to suppress global illumination necessitates multiple captures. Each surface reconstruction using the direct-global separation procedure uses 32 captures: 31 masks with checkerboards and random tiling patterns, along with a single capture using an all-white mask. With this approach, our direct-global separation captures take approximately 10 minutes. For single raster scans without direct-global separation, we perform a slower, 2 minute raster scan to improve robustness to noise.

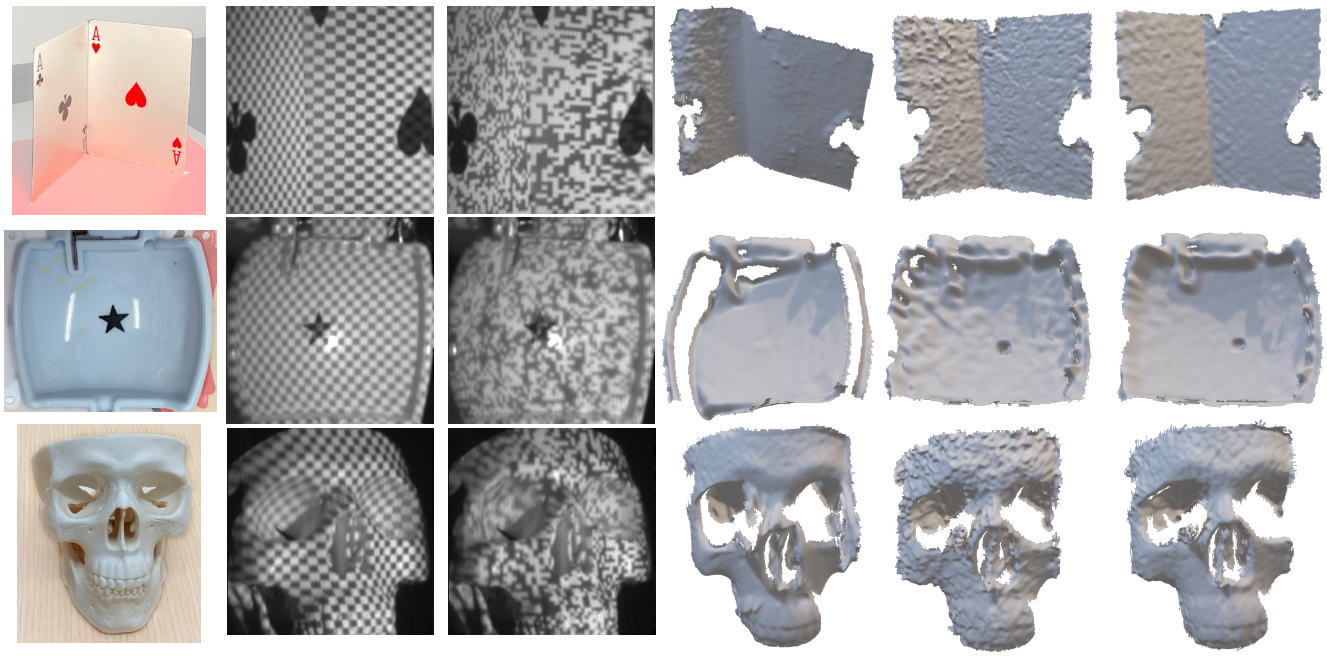
The prototype's imaging speed is constrained by engineering limitations. While our system operates at 15KHz, significantly higher bandwidth PSDs, transimpedance amplifiers, and ADCs would speed up the scanning. Alongside this, replacing the 2D mirror galvanometer with a MEMS laser projector is a natural extension for faster scans.

6. Conclusion and Discussion

This paper presents a new approach for SL 3D where we use the position sensing capabilities of an unconventional diode to obtain 3D scans. We present a theoretical treatment of PSDs for SL applications along with a practical system design that robustly scans scenes in the presence of global illumination. The results in this work suggest that, for many vision tasks, coarse spatial statistics suffice and, hence, the use of unconventional sensors can provide a different set of tradeoffs. Given this, we hope that this work renews interest in PSDs and other non-traditional sensors in the computational imaging community.

Comparison with CMOS SL. It is expected that well-established CMOS SL systems outperform our prototype as they have enjoyed decades of sustained research and development. However, we claim PSDs hold potential in imaging regimes in which there is an inherent difficulty fabricating high-resolution sensors. Two immediate examples are 3D scanning in shortwave infrared, where high-resolution InGaAs sensors are extremely expensive, and SPAD arrays for LIDAR, which require per-pixel timing circuits. In both cases, a pixelated sensor does not scale well at higher resolutions. Solutions built on PSDs could, in principle, provide better tradeoffs in both instances.

The PSD's other advantage over CMOS lies in its minimal readout. We postulate that the PSD's minimal sen-



(a) Scene | (b) Dual Image (Checkerboard) | (c) Dual Image (Random) | (d) Surface (Naïve raster scan) | (e) Surface (Min-Max) | (f) Surface (Robust)

Figure 12: **Gallery of results.** To reduce the effect of global illumination, we repeatedly raster scan the scene using different high-frequency patterns on the SLM. This provides a corresponding set of dual images in (b) and (c). Observe how the dual image is from the projector's view as if the object were illuminated using a masked light source at the PSD. Surface reconstructions without any global illumination suppression are shown in (d); note the significant errors in the object concavities. The final two columns illustrate the improvements using min-max (e) and our more robust regression technique (f) to suppress global illumination.

Table 1: **Comparison with existing SL techniques.** Readout cost for different SL systems capturing an $n \times n$ depth map using an $n \times n$ pixel array. For our PSD system, M denotes the number of SLM patterns. C_0, C_1, C_2 are application-specific constants, and note that C_0 and M are for robustness to global illumination.

SL System	Readout Cost
CMOS point scanning	n^4
CMOS line scanning	n^3
CMOS gray codes	$C_0 \cdot n^2 \log n$
Event-based line scanning	$C_1 \cdot n^2$
PSD	$C_2 \cdot Mn^2$

sor readout (see Table 1) in combination with our proposed global illumination suppression technique could render it a more practical solution for ultra-high resolution scanning.

Acknowledgement

This work was supported by the National Science Foundation under awards 1652569, 1943149, and 1730147.

References

[1] S. Achar, J. Bartels, W. Whittaker, K. Kutulakos, and S. Narasimhan. Epipolar time-of-flight imaging. *ACM Transactions on Graphics*, 36(4):37:1–37:8, July 2017. 2

[2] E. Antonsson and R. Mann. Automatic 6-D.O.F. Kinematic Trajectory Acquisition and Analysis. *Journal of Dynamic Systems, Measurement, and Control*, 111(1):31–39, 03 1989. 2

[3] W. Chen, P. Mirdehghan, S. Fidler, and K. Kutulakos. Autotuning structured light by optical stochastic gradient descent. In *CVPR*, 2020. 2

[4] V. Couture, N. Martin, and S. Roy. Unstructured light scanning robust to indirect illumination and depth discontinuities. *International Journal of Computer Vision*, 108:204–221, 2014. 2

[5] K. J. Dana and S. K. Nayar. Histogram Model for 3D Textures. In *CVPR*, 1998. 1

[6] D. Freedman, E. Krupka, Y. Smolin, I. Leichter, and M. Schmidt. SRA: Fast removal of general multipath for ToF sensors. In *ECCV*, 2014. 2

[7] J. Gu, T. Kobayashi, M. Gupta, and S.K. Nayar. Multiplexed illumination for scene recovery in the presence of global illumination. In *ICCV*, 2011. 2

[8] M. Gupta, A. Agrawal, A. Veeraraghavan, and S. Narasimhan. A practical approach to 3D scanning in the presence of interreflections, subsurface scattering and defocus. *International Journal of Computer Vision*, 102:33–55, 2013. 2

[9] M. Gupta and N. Nakhate. A geometric perspective on structured light coding. In *ECCV*, 2018. 2

[10] M. Gupta, S.K. Nayar, M. Hullin, and J. Martin. Phasor Imaging: A Generalization of Correlation-Based Time-of-Flight Imaging. *ACM Trans. Graphics*, 34:156:1–18, 2015. 2

[11] M. Gupta and S. K. Nayar. Micro phase shifting. In *CVPR*, 2012. 2

[12] M. Gupta, Y. Tian, S. Narasimhan, and L. Zhang. (De) Focusing on global light transport for active scene recovery. In *CVPR*, 2009. 2

[13] E. Hadjidemetriou, M.D. Grossberg, and S.K. Nayar. Spatial Information in Multi-resolution Histograms. In *CVPR*, 2001. 1

[14] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, 2004. 7

[15] M. Idesawa. High-precision image position sensing methods suitable for 3-D measurement. *Optics and Lasers in Engineering*, 10(3):191–204, 1989. 2

[16] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar. Coded time of flight cameras: Sparse deconvolution to address multipath interference and recover time profiles. *ACM Transactions on Graphics*, 32:167:1–10, 2013. 2

[17] S. Kalbitzer and W. Melzer. On the charge dividing mechanism in position sensitive detectors. *Nuclear Instruments & Methods*, 56(2):301–304, 1967. 3

[18] S. Kalbitzer and W. Stumpfi. A nomogram for the design of position sensitive silicon detectors. *Nuclear Instruments & Methods*, 77(2):300–302, 1970. 3

[19] M. Kazhdan and H. Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics*, 32(3):1–13, 2013. 5

[20] Y. Kim, A. Kapoor, T. Mansi, and A. Kamen. A wide-area, low-latency, and power-efficient 6-DoF pose tracking system for rigid objects. *IEEE Sensors Journal*, 22(5):4558–4568, 2022. 2

[21] J. Klotz, M. Gupta, and A. C. Sankaranarayanan. Project website. https://github.com/Image-Science-Lab-cmu/PSD_3D/. 2

[22] P. Lichtsteiner, C. Posch, and T. Delbruck. A 128 × 128 120 db 15 μs latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits*, 43(2):566–576, 2008. 2

[23] G. Lucovsky. Photoeffects in nonuniformly irradiated p-n junctions. *Journal of Applied Physics*, 31(6):1088–1095, 1960. 2

[24] N. Matsuda, O. Cossairt, and M. Gupta. MC3D: Motion contrast 3D scanning. In *ICCP*, 2015. 2

[25] P. Mirdehghan, W. Chen, and K. Kutulakos. Optimal structured light a la carte. In *CVPR*, 2018. 2

[26] M. Muglikar, G. Gallego, and D. Scaramuzza. ESL: Event-based structured light. In *International Conference on 3D Vision (3DV)*, pages 1165–1174, 2021. 2

[27] C. Narayanan, A. Buckman, and I. Busch-Vishniac. Noise analysis for position-sensitive detectors. *IEEE Transactions on Instrumentation and Measurement*, 46(5):1137–1144, 1997. 3

[28] S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar. Fast separation of direct and global components of a scene using high frequency illumination. In *ACM Transactions on Graphics*, volume 25, pages 935–944. 2006. 2, 5, 6

[29] M. O'Toole, S. Achar, S. Narasimhan, and K. Kutulakos. Homogeneous codes for energy-efficient illumination and imaging. *ACM Transactions on Graphics*, 34(4):35:1–35:13, July 2015. 2

[30] M. O'Toole, F. Heide, L. Xiao, M. Hullin, W. Heidrich, and K. Kutulakos. Temporal frequency probing for 5D transient analysis of global light transport. *ACM Transactions on Graphics*, 33:87:1–11, 2014. 2

[31] M. O'Toole, J. Mather, and K. Kutulakos. 3D shape and indirect appearance by structured light transport. In *Proc. IEEE CVPR*, 2014. 2

[32] J. Park, D. Won, K. Park, S. Baeg, and M. Baeg. Development of a real time locating system using PSD under indoor environments. In *SICE-ICASE International Joint Conference*, 2006. 2

[33] M. Rahimi and Y. Shen. PSD microscopy: a new technique for adaptive local scanning of microscale objects. *Robotics and Biomimetics*, 4(1):1–19, 10 2017. 2

[34] S. Richter, M. Stutz, A. Gratzke, Y. Schleitzer, G. Krampert, F. Hoeller, U. Wolf, L. Riedel, and D. Doering. Position sensing and tracking with quasistatic mems mirrors. In *Proceedings of SPIE - The International Society for Optical Engineering*, 2013. 2

[35] D. Rodríguez-Navarro, J. Lázaro-Galilea, Á. De-La-Llana-Calvo, I. Bravo-Muñoz, A. Gardel-Vicente, G. Tsirigotis, and J. Iglesias-Miguel. Indoor positioning system based on a PSD detector, precise positioning of agents in motion using AoA techniques. *Sensors*, 17(9), 2017. 2

[36] P. Sen, B. Chen, G. Garg, S. Marschner, M. Horowitz, M. Levoy, and H. Lensch. Dual photography. *ACM Transactions on Graphics*, 24(3):745–755, 2005. 4

[37] J. Wallmark. A new semiconductor photocell using lateral photoeffect. *Proceedings of the IRE*, 45(4):474–483, 1957. 2

[38] H. Woltring. Single- and dual-axis lateral photodetectors of rectangular shape. *IEEE Transactions on Electron Devices*, 22(8):581–590, 1975. 2, 3

Computational 3D Imaging with Position Sensors
Supplementary Material

Jeremy Klotz*, Mohit Gupta†, Aswin C. Sankaranarayanan‡
Columbia University*, University of Wisconsin–Madison†, Carnegie Mellon University‡

Abstract

We present supplementary analyses of our global illumination suppression technique in Sec. 1. We then detail the centroid variance and present proofs of system invariants in Sec. 2 and conclude with an implementation notes in Sec. 3.

1. Surface Reconstructions

We detail the method and results using the proposed global illumination suppression technique.

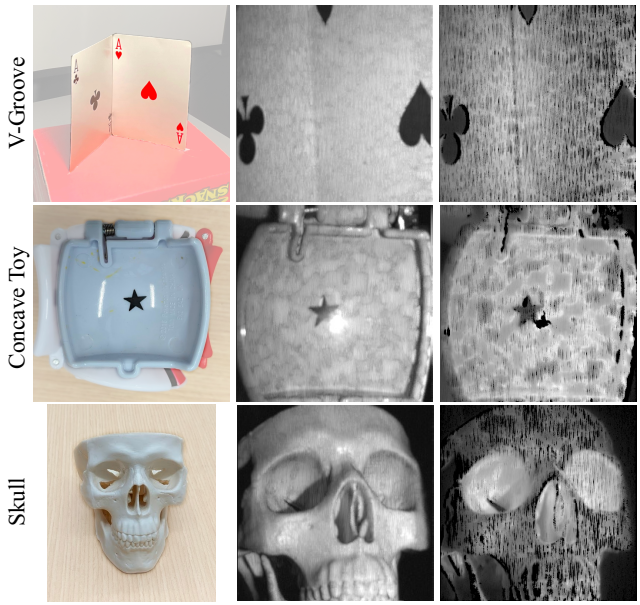
1.1. Direct and Global Images

Figure 1 shows the direct and global images of three different scenes presented in the main paper. For each scene, we follow the min-max procedure in [3] to decompose the dual image stack into direct and global images. Specifically, for each dual image pixel, we identify the SLM mask corresponding to the largest intensity value on the PSD. Along with the capture under that mask's complement, we compute the direct and global images following the method described in [3]. Despite the use of multiple SLM patterns, this separation technique produces vertical artifacts in Fig. 1 (b) and (c). This observation motivates our robust separation technique detailed in the main paper that uses the entire stack of measurements rather than a single pair to estimate the image centroids.

Each global image highlights strong interreflections in the object concavities that would otherwise bias the centroid measurements if not properly handled. Specifically, the V-groove interreflections are strongest at the center, the plastic toy's interreflections are strongest within the concavity, and the skull shows strong global illumination within the eye sockets.

1.2. Min-Max v. Regression Technique

Alongside these reconstructions, we also present line scans of the V-groove and concave toy to compare the min-max processing with our proposed regression technique for suppressing global illumination. The line scans in Fig. 2 and



(a) Scene (b) Direct Image (c) Global Image

Figure 1: Direct and global images of three different scenes. Since our imaging setup corresponds to the optical dual of Nayar *et al.* [3], the direct and global images in (b) and (c) are seen from the projector's view. Each direct and global image is computed using the min-max technique. Note that the images are scaled and gamma corrected for visualization, and the global image is further scaled by 2× to increase contrast.

Fig. 3 demonstrate the effectiveness of our regression-based global illumination suppression method.

1.3. Scan Time Comparison

The results presented in the main paper using our global illumination suppression method takes 10 minutes to scan. By contrast, a single raster scan without global illumination suppression takes 2 minutes. To evaluate the difference in surface reconstructions, Fig. 4 shows three different reconstructions of the skull using various scanning times. In Fig. 4 (a) and (b), we compare a single raster scan with the

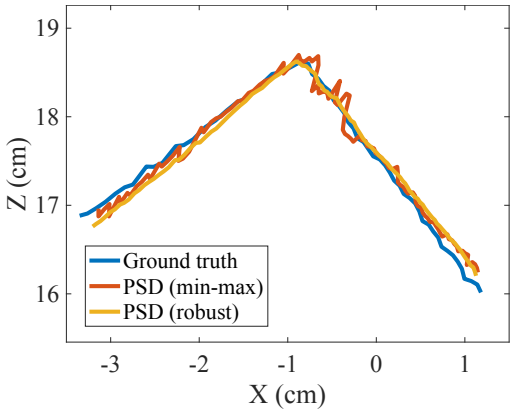


Figure 2: Depth slice of the V-groove comparing depth estimates using two methods to estimate the direct-only centroid on the PSD. Our proposed robust separation method (yellow) produces a smoother depth slice than the min-max method (red).

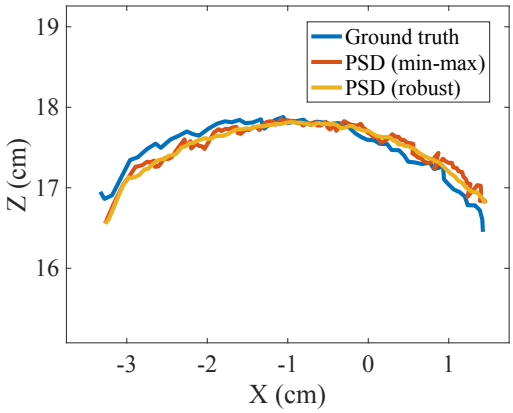
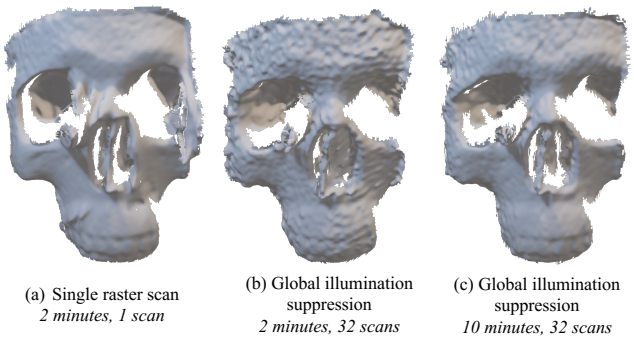


Figure 3: Depth slice of the concave toy comparing the depth estimates. Similar to the V-groove depth slice above, our global illumination suppression method improves the depth estimates.

global illumination suppression method under same time budget (2 minutes). In this case, the scan with global illumination suppression is significantly noisier. In Fig. 4 (c), we show the same object using global illumination suppression using a slower scan, which provides a higher quality reconstruction.

In simpler scenes, a single raster scan is sufficient. Figure 5 shows surface reconstructions on opaque busts using a single scan.



(a) Single raster scan
2 minutes, 1 scan

(b) Global illumination suppression
2 minutes, 32 scans

(c) Global illumination suppression
10 minutes, 32 scans

Figure 4: We compare a single, slow scan (a) with our global illumination suppression method under the same time budget (b) and an extended time budget (c). Note that suppressing global illumination is necessary in this scene since interreflections in the eye cavities bias the depth estimates. Since our global illumination suppression method uses 32 raster scans, the faster raster scans produce a noisier surface in (b). In (c), we scan the same object using global illumination suppression with slower raster scans and observe a qualitative improvement in the surface.

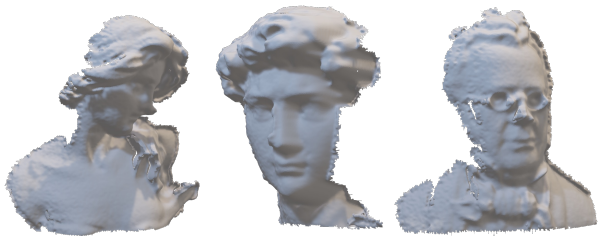


Figure 5: In simple scenes without global illumination, a single raster without global illumination suppression is sufficient. Here we present various reconstructions of opaque busts using a single 2-minute raster scan.

2. Details of PSD Analysis and Invariants

2.1. Centroid Variance Approximation

We consider the variance of the centroid measurement in two separate cases. In case 1, $V_x = 0$ when the light spot is at the center of the diode. This corresponds to the configuration of our lab prototype, thus our empirical results follow this form. In case 2, $V_x = 0$ when the light spot strikes the edge of the diode. This configuration follows the simple form used in section 2 of the main paper to explain the principles of a PSD.

The centroid variance for each case differs slightly. For a fixed light level, the centroid variance is minimized when $V_x = 0$, corresponding to different physical locations on the diode. In practice, the engineer may choose the configuration by adding an appropriate offset voltage to V_x.

Case 1: $V_x = 0$ at the center of the diode. Consider a PSD with length L_x such that $V_x = 0$ when the light spot strikes the center of the diode. We will denote the centroid as C_{1x} using the subscript to differentiate it from the centroid C_{2x} in case 2. Given the following voltages

$$\begin{bmatrix} V_x \\ V_s \end{bmatrix} = G \iint_{x,y} \begin{bmatrix} x \\ 1 \end{bmatrix} i(x,y)\, dx\, dy, \tag{1}$$

the centroid, with $C_{1x} = 0$ at the diode's center, is:

$$C_{1x} = \frac{L_x}{2}\frac{V_x}{V_s}. \tag{2}$$

Let $\widetilde{V_x}, \widetilde{V_s}$ be measurements with uncorrelated signal-independent noise:

$$\widetilde{V_x} = V_x + \mathcal{N}(0, \sigma^2) \tag{3}$$

$$\widetilde{V_s} = V_s + \mathcal{N}(0, \sigma^2) \tag{4}$$

Then, the estimated centroid $\widetilde{C_{1x}}$ is:

$$\widetilde{C_{1x}} = \frac{L_x}{2}\frac{\widetilde{V_x}}{\widetilde{V_s}} \tag{5}$$

There is no analytical random variable describing $\widetilde{C_{1x}}$ when $\widetilde{V_x}$ and $\widetilde{V_s}$ have non-zero means. However, $\widetilde{C_{1x}}$ is approximately normal when $\sigma \ll V_x, V_s$ [1]. Following the first-order Taylor series expansion of $\widetilde{C_{1x}}$ described in [1], we approximate the measured centroid's variance as:

$$\text{Var}\left[\widetilde{C_{1x}}\right] \approx \frac{L_x^2}{4}\frac{\sigma^2}{V_s^2}\left(\frac{V_x^2}{V_s^2} + 1\right) \tag{6}$$

$$\approx \frac{\sigma^2}{V_s^2}\left(C_{1x}^2 + \frac{L_x^2}{4}\right) \tag{7}$$

Note that the least variance occurs when V_x is small and V_s is large, corresponding to a bright light spot at the center of the diode.

The normal approximation above requires V_x and V_s to be strictly positive. When the light spot appears on the other side of the diode, V_x will be negative. We can show by substitution that the normal approximation still holds in this case:

Suppose $\mathbb{E}[\widetilde{V_x}] < 0$. Let $\widetilde{V_x}' = -\widetilde{V_x}$. Then:

$$\widetilde{C_x}' = \frac{\widetilde{V_x}'}{\widetilde{V_s}} = -\widetilde{C_x} \tag{8}$$

$$\tag{9}$$

The approximation holds for $\widetilde{C_x}'$ since the $\widetilde{V_x}'$ and $\widetilde{V_s}$ have strictly positive means, and the moments are related as:

$$\mathbb{E}[\widetilde{C_x}'] = -\mathbb{E}[\widetilde{C_x}], \tag{10}$$

$$\text{Var}\,[\widetilde{C_x}'] = \text{Var}\,[\widetilde{C_x}] \tag{11}$$

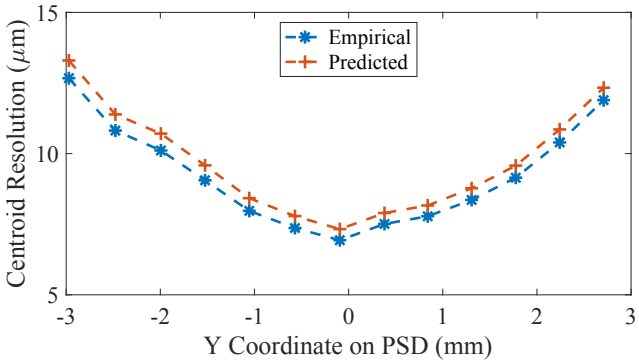


Figure 6: Empirical and predicted centroid resolution along a vertical line on the PSD. Over all points in a 13×13 point array, the predicted centroid resolution RMSE is 0.55μm. In this example, the centroid RMSE attains 7μm at the center of the diode.

We validate this noise model by capturing a 13×13 point array on a plane and comparing the empirical and predicted centroid resolutions. At each point, the empirical centroid resolution is the root-mean-square error (RMSE) of the centroid over 59k samples; the predicted centroid resolution is the square root of Eq. (7), replacing V_x, V_s with their respective sample means $E[\widetilde{V_x}], E[\widetilde{V_s}]$ and C_{1x} with $\frac{L_x}{2}\frac{E[\widetilde{V_x}]}{E[\widetilde{V_s}]}$.

Fig. 6 shows agreement between the predicted and empirical centroid resolution of points along a vertical line near the center of the PSD. Over all points ($n = 169$), the predicted centroid resolution from Eq. (7) attains an RMSE of 0.55μm from the empirical resolution.

Case 2: $V_x = 0$ at the edge of the diode. In section 2 of the main paper, we defined V_x such that $V_x = 0$ at the edge of the diode. In this case, we denote the centroid as C_{2x} to differentiate it from the centroid C_{1x} in case 1:

$$C_{2x} = L_x\frac{V_x}{V_s}$$

In the presence of signal-independent measurement noise on each channel as described above, the measured centroid is:

$$\widetilde{C_{2x}} = L_x\frac{\widetilde{V_x}}{\widetilde{V_s}}$$

Applying the same normal approximation to $\widetilde{C_{2x}}$ as in case 1, we have:

$$\text{Var}\,[\widetilde{C_{2x}}] \approx \frac{L_x^2\sigma^2}{V_s^2}\left(\frac{V_x^2}{V_s^2} + 1\right) \tag{12}$$

$$\approx \frac{\sigma^2}{V_s^2}\left(C_{2x}^2 + L_x^2\right), \tag{13}$$

corresponding to eq. 5 in the main paper.

Normal approximation limitations. The normal approximation requires strictly positive samples from the numerator ($\widetilde{V_x}$) and denominator ($\widetilde{V_s}$). This breaks down at low light levels when $V_s \to 0$ or when near the location on the diode corresponding to $C_x = 0$ (configuration-dependent).

2.2. Proof of Invariance to Uniform Image Scaling

Let α be a uniform scaling term on the image $I(x, y)$. The measured centroid is:

$$\begin{bmatrix} C'_x \\ C'_x \end{bmatrix} = \frac{\iint_{x,y} \begin{bmatrix} x \\ y \end{bmatrix} \alpha I(x,y)\, dx\, dy}{\iint_{x,y} \alpha I(x,y)\, du\, dv} \tag{14}$$

$$= \frac{\iint_{x,y} \begin{bmatrix} x \\ y \end{bmatrix} I(x,y)\, dx\, dy}{\iint_{x,y} I(x,y)\, dx\, dy} \tag{15}$$

$$= \begin{bmatrix} C_x \\ C_y \end{bmatrix} \tag{16}$$

2.3. Proof of Defocus Invariance

Let $I(x, y)$ be the in-focus image and $b(x, y)$ be the kernel of a spatially-invariant point spread function. The centroids of $I(x, y)$ and $b(x, y)$ are given by:

$$\begin{bmatrix} C_x \\ C_y \end{bmatrix} = \frac{\iint_{x,y} \begin{bmatrix} x \\ y \end{bmatrix} I(x,y)\, dx\, dy}{\iint_{x,y} I(x,y)\, dx\, dy} \tag{17}$$

$$\begin{bmatrix} C^b_x \\ C^b_y \end{bmatrix} = \frac{\iint_{x,y} \begin{bmatrix} x \\ y \end{bmatrix} b(x,y)\, dx\, dy}{\iint_{x,y} b(x,y)\, dx\, dy} \tag{18}$$

The centroid of the measured image on an infinitely large sensor is given by:

$$\begin{bmatrix} C'_x \\ C'_y \end{bmatrix} = \frac{\iint_{x,y} \begin{bmatrix} x \\ y \end{bmatrix} (I * b)(x,y)\, dx\, dy}{\iint_{x,y} (I * b)(x,y)\, dx\, dy} \tag{19}$$

$$= \frac{\iint_{x,y} \begin{bmatrix} x \\ y \end{bmatrix} \iint_{u,v} b(u,v) I(x-u, y-v)\, du\, dv\, dx\, dy}{\iint_{x,y} \iint_{u,v} b(u,v) I(x-u, y-v)\, du\, dv\, dx\, dy} \tag{20}$$

$$= \frac{\iint_{u,v} b(u,v) \iint_{x,y} \begin{bmatrix} x \\ y \end{bmatrix} I(x-u, y-v)\, dx\, dy\, du\, dv}{\iint_{u,v} b(u,v) \iint_{x,y} I(x-u, y-v)\, dx\, dy\, du\, dv} \tag{21}$$

Since we are integrating over an infinitely large sensor, any image translation by a finite (u, v) will not change the final integral. Thus, we can remove the dependence on (u, v) in the image in the denominator.

$$= \frac{\iint_{u,v} b(u,v) \iint_{x,y} \begin{bmatrix} x \\ y \end{bmatrix} I(x-u, y-v)\, dx\, dy\, du\, dv}{\iint_{u,v} b(u,v)\, du\, dv \iint_{x,y} I(x,y)\, dx\, dy} \tag{22}$$

$$= \frac{\iint_{u,v} b(u,v) \begin{bmatrix} C_x + u \\ C_y + v \end{bmatrix} du\, dv}{\iint_{u,v} b(u,v)\, du\, dv} \tag{23}$$

By separating the integral, this expression simplifies to:

$$= \begin{bmatrix} C_x \\ C_y \end{bmatrix} \frac{\iint_{u,v} b(u,v)\, du\, dv}{\iint_{u,v} b(u,v)\, du\, dv} + \frac{\iint_{u,v} \begin{bmatrix} u \\ v \end{bmatrix} b(u,v)\, du\, dv}{\iint_{u,v} b(u,v)\, du\, dv} \tag{24}$$

$$= \begin{bmatrix} C_x + C^b_x \\ C_y + C^b_y \end{bmatrix} \tag{25}$$

Ignoring boundary conditions where the point spread kernel extends beyond the finite sensor area, the measured centroid is uniformly biased by the point spread kernel's centroid.

3. Implementation Notes

Pseudocode for global illumination suppression. Algorithm 1 shows pseudocode to estimate direct-only image centroids using the regression technique.

Calibration procedure. Both the PSD and projector are calibrated in the helper camera's coordinate frame. To obtain 2D-3D point correspondences for the PSD calibration, we project a 10×10 point array on a planar checkerboard of known geometry at varying depths. At each point in the

Algorithm 1 Robust centroid estimation algorithm.

M: Number of masks
T: Number of points on the dual image grid
for $k = 1 \dots M$ **do**
  Display mask m^k on SLM and raster scan
  $\{V_x^k(t), V_y^k(t), V_s^k(t)\} \leftarrow$ PSD readout
end for
for $t = 1 \dots T$ **do**
  $C_x(t), C_y(t) \leftarrow$
    RobustCentroid($\{V_x^k(t), V_y^k(t), V_s^k(t)\}_{k=1}^M$)
end for

function RobustCentroid($\{V_x^k, V_y^k, V_s^k\}_{k=1}^M$)
  Compute pairwise differences
  $\mathcal{P} \leftarrow \{(i,j) \in \{1,\dots,M\} \times \{1,\dots,M\} \ : \ i \neq j\}$
  $D_x \leftarrow [(V_x^i - V_x^j) \ \forall \ (i,j) \in \mathcal{P}]$
  $D_y \leftarrow [(V_y^i - V_y^j) \ \forall \ (i,j) \in \mathcal{P}]$
  $D_s \leftarrow [(V_s^i - V_s^j) \ \forall \ (i,j) \in \mathcal{P}]$

  Estimate centroids using linear regression
  $C_x \leftarrow \sum (D_s \cdot D_x) \,/\, \left(\sum (D_s \cdot D_s) + 10^{-8}\right)$
  $C_y \leftarrow \sum (D_s \cdot D_y) \,/\, \left(\sum (D_s \cdot D_s) + 10^{-8}\right)$
  return C_x, C_y
end function

point array, we capture a high-dynamic range image using the helper camera, measure the image centroid on the PSD, and record the galvo mirror angles. We know the 2D location of the laser spot in the helper camera's image, and the view of the checkerboard gives us the plane orientation used to compute the 3D spot location.

As described in the main paper, we follow the calibration procedure in [2] to estimate the intrinsics, extrinsics, and lens distortion coefficients of the PSD. Since the 3D locations are in the helper camera's coordinate frame, the PSD's extrinsics are also in that coordinate frame.

The distance between the galvo mirrors in our lab prototype causes the projector to violate the pinhole projection model, thereby creating distortion and triangulation errors. To account for this, we fit a polynomial model to map the mirror angles $(\tan\theta_p, \tan\psi_p)$ from the calibration point array to rays in 3D space. At each mirror angle in the calibration point array, the helper camera gives us the 3D location of the projector ray's intersection with the calibration plane at different depths. The polynomial model attains $66.1\mu\text{m}$ RMSE in estimating the 3D points from measured mirror angles on the calibration data.

References

[1] E. Díaz-Francés and F. Rubio. On the existence of a normal approximation to the distribution of the ratio of two independent normal random variables. *Statistical Papers*, 54(2):309–323, 2012. 3

[2] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, 2004. 5

[3] S.K. Nayar, G. Krishnan, M.D. Grossberg, and R. Raskar. Fast separation of direct and global components of a scene using high frequency illumination. In *ACM SIGGRAPH*, pages 935–944. 2006. 1