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.

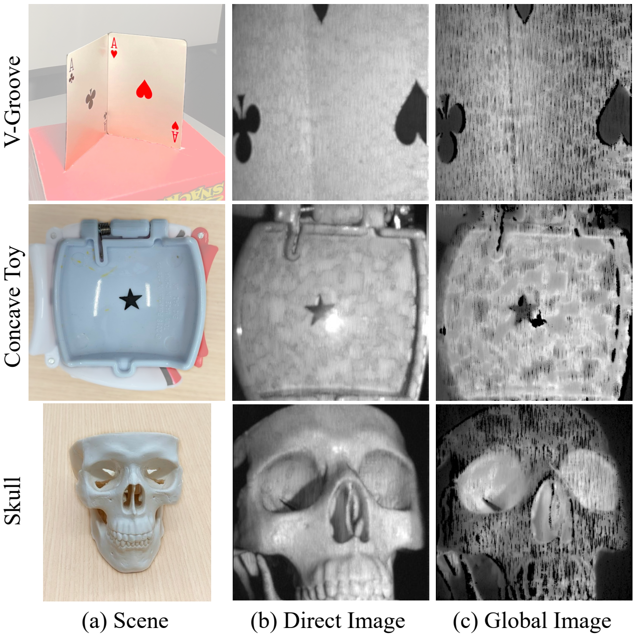
1.1. Direct and Global Images

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



(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.

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

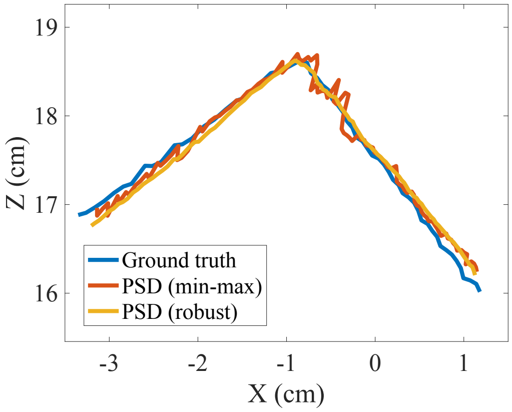


Figure 2: Depth slice of the V-groove comparing depth estimates using two methods to estimate the direct-only centroid on the PSD. Our proposed robust separation method (yellow) produces a smoother depth slice than the min-max method (red).

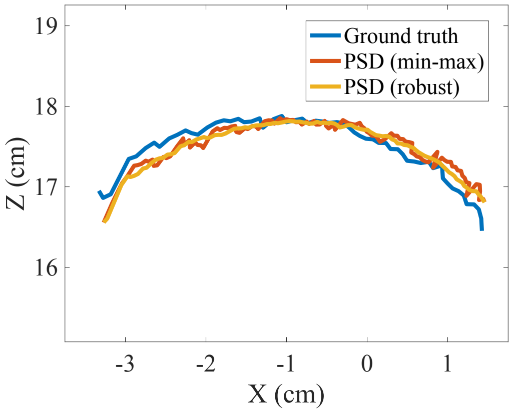
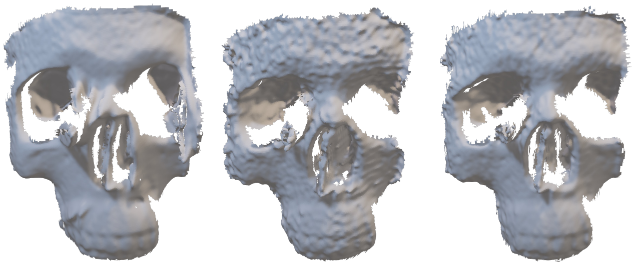


Figure 3: Depth slice of the concave toy comparing the depth estimates. Similar to the V-groove depth slice above, our global illumination suppression method improves the depth estimates.

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.

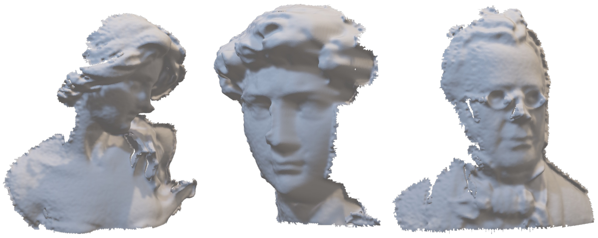


Figure 5: In simple scenes without global illumination, a single raster without global illumination suppression is sufficient. Here we present various reconstructions of opaque busts using a single 2-minute raster scan.

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:

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

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

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:

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

$$\mathrm{Var}\left[\widetilde{C_x}'\right] = \mathrm{Var}\left[\widetilde{C_x}\right] \tag{11}$$

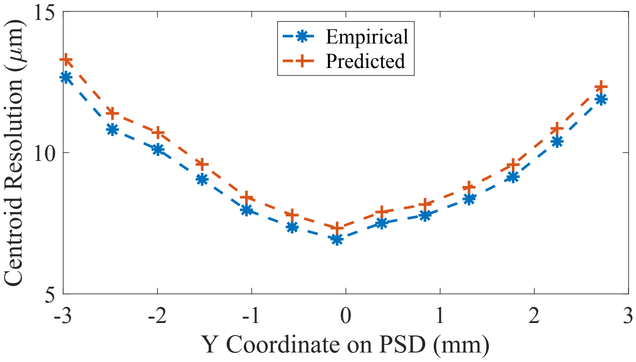


Figure 6: Empirical and predicted centroid resolution along a vertical line on the PSD. Over all points in a 13×13 point array, the predicted centroid resolution RMSE is 0.55μm. In this example, the centroid RMSE attains 7μm at the center of the diode.

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:

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

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

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
    $C_x(t), C_y(t) \leftarrow$
        $\textsc{RobustCentroid}(\{V_x^k(t), V_y^k(t), V_s^k(t)\}_{k=1}^M)$
end for

function $\textsc{RobustCentroid}(\{V_x^k, V_y^k, V_s^k\}_{k=1}^M)$
    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