3D Shape and Indirect Appearance By Structured Light Transport: Supplemental Document

Matthew O'Toole John Mather Kiriakos N. Kutulakos
Department of Computer Science
University of Toronto
{motoole,jmather,kyros}@cs.toronto.edu

A. Proofs of Propositions 1 and 2

A.1. Proof of Proposition 1

Proposition 1. *If* $\mathbf{T}$ *is the discretized form of a transport function that is measurable and positive over the rectified projector and image planes, then*

$$\lim_{\epsilon \to 0} \frac{\mathbf{T}^{\mathrm{EI}}\,\mathbf{p}}{\mathbf{T}^{\mathrm{NE}}\,\mathbf{p}} \;=\; \mathbf{0} \tag{17}$$

where division is entry-wise and ϵ *is the pixel size for discretization.*

Proof sketch. We begin by identifying the rectified projector and image planes with the continuous domain $\mathcal{D} = [-1,1] \times [-1,1] \subset \Re^2$. Let $p = (p_x, p_y)$ be a point on the projector plane and let $\mathbf{I}_\epsilon(p)$ be an indicator function over $\mathcal{D}$ that specifies the spatial extent of the discrete epipolar line through the origin:

$$\mathbf{I}_\epsilon(p) = \begin{cases} 1 & \text{if } |p_x| \leq \frac{\epsilon}{2} \text{ and } |p_y| \leq 1 \\ 0 & \text{otherwise .} \end{cases} \tag{18}$$

In the continuous setting, light transport from the projector plane to the image plane is described by the *light transport equation* [2]. Given an image point $i \in \mathcal{D}$ on the epipolar line through the origin, this equation describes the total radiance transported to i from points on the projector plane:

$$\mathcal{I}(i) \;=\; \underbrace{\mathcal{T}(\hat{p},\, i)\, \mathcal{P}(\hat{p})}_{\text{direct}} \;+\; \underbrace{\int_{\mathcal{D}-\{\hat{p}\}} \mathcal{T}(p,\, i)\, \mathcal{P}(p)\, dp}_{\text{indirect}} \tag{19}$$

where $\hat{p}$ is the projector point in stereo correspondence with image point i; $\mathcal{P}(p)$ is the radiance along the ray through projector point p; and $\mathcal{T}(p, i)$ is the *transport function* describing the proportion of radiance from p that gets transported to i.

Without loss of generality, we prove the continuous form of the ratio in Eq. (1) for an image point i; this point is taken to be inside a discrete image pixel of dimension $\epsilon \times \epsilon$ on the epipolar line through the origin.

More specifically, we consider the epipolar indirect, total indirect, and non-epipolar indirect contributions at i:

$$\mathcal{I}^{\mathrm{EI}}(i) \;=\; \int_{\mathcal{D}-\{\hat{p}\}} \mathbf{I}_\epsilon(p)\, \mathcal{T}(p,i)\, \mathcal{P}(p)\, dp \tag{20}$$

$$\mathcal{I}^{\mathrm{I}}(i) \;=\; \int_{\mathcal{D}-\{\hat{p}\}} \mathcal{T}(p,i)\, \mathcal{P}(p)\, dp \tag{21}$$

$$\mathcal{I}^{\mathrm{NE}}(i) \;=\; \mathcal{I}^{\mathrm{I}}(i) \;-\; \mathcal{I}^{\mathrm{EI}}(i)\;. \tag{22}$$

We now show that for any $\delta > 0$, there is an $\epsilon > 0$ such that

$$\left| \frac{\mathcal{I}^{\mathrm{EI}}(i)}{\mathcal{I}^{\mathrm{NE}}(i)} \right| \;<\; \delta\;. \tag{23}$$

Since $\mathcal{T}()$ is measurable, we can apply the Cauchy-Schwarz inequality [4] to Eq. (4) to get an upper bound on the epipolar indirect contributions:

$$\begin{aligned} \mathcal{I}^{\mathrm{EI}}(i) \;&\leq\; \left\{ \int_{\mathcal{D}-\{\hat{p}\}} \mathbf{I}_\epsilon(p) \mathrm{d}p \right\}^{\frac{1}{2}} \left\{ \int_{\mathcal{D}-\{\hat{p}\}} \left[\mathcal{T}(p,\, i)\, \mathcal{P}(p)\right]^2 \mathrm{d}p \right\}^{\frac{1}{2}} \\ &=\; (2\epsilon)^{\frac{1}{2}} \left\{ \int_{\mathcal{D}-\{\hat{p}\}} \left[\mathcal{T}(p,\, i)\, \mathcal{P}(p)\right]^2 \mathrm{d}p \right\}^{\frac{1}{2}}\;. \end{aligned} \tag{24}$$

By combining Eqs. (5), (6) and (8) we also get a lower bound on the non-epipolar contributions:

$$\begin{aligned} \mathcal{I}^{\mathrm{NE}}(i) \;\geq\; &\int_{\mathcal{D}-\{\hat{p}\}} \mathcal{T}(p,i)\, \mathcal{P}(p)\, dp \;- \\ &(2\epsilon)^{\frac{1}{2}} \left\{ \int_{\mathcal{D}-\{\hat{p}\}} \left[\mathcal{T}(p,\, i)\, \mathcal{P}(p)\right]^2 \mathrm{d}p \right\}^{\frac{1}{2}}\;. \end{aligned} \tag{25}$$

Equation (7) now follows by choosing ϵ to be

$$\epsilon \;=\; \frac{1}{2}\left(\frac{\delta}{2 \,+\, \delta}\right)^2 \frac{\left\{ \int_{\mathcal{D}-\{\hat{p}\}} \mathcal{T}(p,i)\, \mathcal{P}(p)\, dp \right\}^2}{\int_{\mathcal{D}-\{\hat{p}\}} \left[\mathcal{T}(p,\, i)\, \mathcal{P}(p)\right]^2 \mathrm{d}p}\;. \tag{26}$$

Specifically, substituting Eq. (10) into Eqs. (8) and (9) we get

$$\left| \frac{\mathcal{I}^{\mathrm{EI}}(i)}{\mathcal{I}^{\mathrm{NE}}(i)} \right| \;\leq\; \frac{\frac{\delta}{2+\delta}}{1 - \frac{\delta}{2+\delta}} \;=\; \frac{\delta}{2} \;<\; \delta\;. \tag{27}$$

□

A.2. Proof of Proposition 2

We prove Proposition 2 for scenes consisting of a finite collection of objects, each of which is an open set in $\Re^3$ bounded by a smooth generic surface [1, 3].

Proposition 2. *Two generic n-bounce specular transport paths that originate from corresponding epipolar lines do not intersect for $n > 1$.*

Proof. For simplicity, we reverse the direction of light travel through image pixels, treating the camera as a second projector that also sends light onto the scene.

Let $\mathcal{L}, \mathcal{L}'$ be a pair of corresponding epipolar lines on the (continuous) projector and image planes, respectively, and let $p \in \mathcal{L}$ and $i \in \mathcal{L}'$ be points on them.

Suppose that the light originating at p and i undergoes $n \geq 1$ consecutive specular bounces upon entering the scene. Furthermore, suppose that the associated transport paths are generic, *i.e.*, they remain stable under infinitesimal perturbations of the scene's surfaces and of the points p and i. To prove the proposition, we show that the following cannot hold simultaneously:

1. the transport paths through p and i intersect at their $(n+1)$-th bounce, *i.e.*, their $(n+1)$-th bounce occurs at the same surface point in the scene; and

2. this intersection is generic, *i.e.*, it occurs for all points p, i in an open interval $\mathcal{Q} \subset \mathcal{L}$ and $\mathcal{Q}' \subset \mathcal{L}'$, respectively.

In particular, let $l_n(p)$ be the 3D ray that light follows after n specular bounces from projector point p. Similarly, let $l'_n(i)$ be the corresponding 3D ray for image point i. Since the transport paths through p and i are generic, the mappings $p \mapsto l_n(p)$ and $i \mapsto l'_n(i)$ are smooth functions for some open neighborhood $\mathcal{Q} \subset \mathcal{L}$ and $\mathcal{Q}' \subset \mathcal{L}'$ of p and i, respectively. These mappings define a pair of ruled surfaces in $\Re^3$: intuitively, as point p ranges over $\mathcal{Q}$, the 3D ray $l_n(p)$ twists and translates in space, tracing a ruled surface.

Now, for the transport paths through p and i to have their $(n+1)$-th bounce in common, three surfaces must meet at a point: ruled surface $l_n(\mathcal{Q})$, ruled surface $l'_n(\mathcal{Q}')$, and a surface in the scene. This, however, is *not* a generic condition because surfaces $l_n(\mathcal{Q})$ and $l'_n(\mathcal{Q}')$ transversally intersect along a curve and this curve will transversally intersect the scene's surfaces at isolated points [1]. □

B. Expanded derivations of selected equations

B.1. Derivation of Eq. (11)

Combining Eqs. (6) and (9) we have

$$\mathbf{i}_e \;=\; \frac{1}{T}\sum_{t=1}^{T}\sum_{f=1}^{E} \overline{\mathbf{q}_e(t)} \circ [\, \mathbf{T}_{ef}\; \mathbf{q}_f(t)\,] \tag{28}$$

where the $1/T$ factor captures the fact that each term in the sum is allocated $1/T$ of the total exposure time. We now split the sum into its epipolar and non-epipolar terms

$$\mathbf{i}_e \;=\; \frac{1}{T}\sum_{t=1}^{T}\Bigg\{ \overline{\mathbf{q}_e(t)} \circ [\, \mathbf{T}_{ee}\; \mathbf{q}_e(t)\,] \;+\; \sum_{\substack{f=1\\ f\neq e}}^{E} \overline{\mathbf{q}_e(t)} \circ [\, \mathbf{T}_{ef}\; \mathbf{q}_f(t)\,] \Bigg\} \tag{29}$$

and observe that the first term is always a vector of zeros. Therefore,

$$\mathbf{i}_e \;=\; \frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{f=1\\ f\neq e}}^{E} \overline{\mathbf{q}_e(t)} \circ [\, \mathbf{T}_{ef}\; \mathbf{q}_f(t)\,] \;. \tag{30}$$

Letting $T \to \infty$ and applying the Central Limit Theorem to Eq. (14) we get the expected image $\mathcal{E}[\mathbf{i}_e]$ for epipolar line e:

$$\mathcal{E}[\mathbf{i}_e] \;=\; \mathcal{E}\Bigg[\sum_{\substack{f=1\\ f\neq e}}^{E} \overline{\mathbf{q}_e} \circ [\, \mathbf{T}_{ef}\; \mathbf{q}_f\,] \Bigg] \tag{31}$$

$$\;=\; \mathcal{E}[\overline{\mathbf{q}_e}] \;\circ\; \mathcal{E}\Bigg[\sum_{\substack{f=1\\ f\neq e}}^{E} \mathbf{T}_{ef}\; \mathbf{q}_f \Bigg] \tag{32}$$

$$\;=\; \mathcal{E}[\overline{\mathbf{q}_e}] \;\circ\; \sum_{\substack{f=1\\ f\neq e}}^{E} \mathbf{T}_{ef}\; \mathcal{E}[\mathbf{q}_f] \tag{33}$$

$$\;=\; 0.25 \sum_{\substack{f=1\\ f\neq e}}^{E} \mathbf{T}_{ef}\; \mathbf{1} \;, \tag{34}$$

where Eq. (16) follows from the fact that epipolar lines e and f are distinct and thus their corresponding random vectors $\mathbf{q}_e$ and $\mathbf{q}_f$ are independent.

B.2. Derivation of Eq. (14)

Combining Eqs. (6) and (9) for the indirect-invariant mask and pattern we have:

$$\mathbf{i}_e = \frac{1}{T}\sum_{t=1}^{T}\sum_{f=1}^{E} \mathbf{m}_e(t) \circ \Big\{ \mathbf{T}_{ef}\; [\, \mathbf{m}_f(t) \circ \mathbf{r}_f(t) + \overline{\mathbf{m}_f(t)} \circ \overline{\mathbf{r}_f(t)}\,] \Big\} \tag{35}$$

We split the sum in Eq. (19) into its epipolar and non-epipolar terms,

$$\mathbf{i}_e \;=\; \frac{1}{T}\sum_{t=1}^{T}\Big\{\mathbf{m}_e(t)\circ\mathbf{T}_{ee}\,[\,\mathbf{m}_e(t)\circ\mathbf{r}_e(t)\,]\;+\;\mathbf{m}_e(t)\circ\mathbf{T}_{ee}\,[\,\overline{\mathbf{m}_e(t)}\circ\overline{\mathbf{r}_e(t)}\,]\;+\;\sum_{\substack{f=1\\f\neq e}}^{E}\mathbf{m}_e(t)\circ\mathbf{T}_{ef}\,[\,\mathbf{m}_f(t)\circ\mathbf{r}_f(t)\,]\;+\;\sum_{\substack{f=1\\f\neq e}}^{E}\mathbf{m}_e(t)\circ\mathbf{T}_{ef}\,[\,\overline{\mathbf{m}_f(t)}\circ\overline{\mathbf{r}_f(t)}\,]\Big\} \quad (36)$$

and note that the second term of Eq. (20) is always a vector of zeros. Letting $T\rightarrow\infty$ and applying the Central Limit Theorem to Eq. (20) we get the expected image for epipolar line e:

$$\mathcal{E}[\mathbf{i}_e] \;=\; \mathcal{E}\Big[\,\mathfrak{q}_e\circ[\,\mathbf{T}_{ee}\,(\mathfrak{q}_e\circ\mathfrak{r}_e)\,]\;+\;\sum_{\substack{f=1\\f\neq e}}^{E}\mathfrak{q}_e\circ[\,\mathbf{T}_{ef}\,(\mathfrak{q}_f\circ\mathfrak{r}_f+\overline{\mathfrak{q}_f}\circ\overline{\mathfrak{r}_f})\,]\,\Big]\;. \quad (37)$$

Now, $\mathfrak{q}_e$ is a random binary vector whose probability of being either $\mathbf{1}$ or $\mathbf{0}$ is 0.5. Using this fact as well as $\mathfrak{q}_e$'s independence from all other random vectors, the expectation in Eq. (21) becomes

$$\mathcal{E}[\mathbf{i}_e] \;=\; 0.5\,\mathbf{T}_{ee}\,\mathcal{E}[\mathfrak{r}_e]\;+\;0.5\sum_{\substack{f=1\\f\neq e}}^{E}\mathbf{T}_{ef}\,\mathcal{E}[\mathfrak{q}_f\circ\mathfrak{r}_f+\overline{\mathfrak{q}_f}\circ\overline{\mathfrak{r}_f}]\;. \quad (38)$$

Finally, using the definition of binary random vector $\mathfrak{r}_f$ in Eq. (13) the expectation becomes

$$\mathcal{E}[\mathbf{i}_e] \;=\; 0.5\,\mathbf{T}_{ee}\,\mathbf{p}_e\;+\;0.5\sum_{\substack{f=1\\f\neq e}}^{E}\mathbf{T}_{ef}\big\{\,\mathrm{Prob}[\,\mathfrak{q}_f=\mathbf{1}\,]\,\mathbf{p}_e+\mathrm{Prob}[\,\mathfrak{q}_f=\mathbf{0}\,]\,(\mathbf{1}-\mathbf{p}_e)\,\big\} \quad (39)$$

which is equal to

$$\mathcal{E}[\mathbf{i}_e] \;=\; 0.5\,\mathbf{T}_{ee}\,\mathbf{p}_e\;+\;0.25\sum_{\substack{f=1\\f\neq e}}^{E}\big\{\mathbf{T}_{ef}(\,\mathbf{p}_e\;+\;(\mathbf{1}-\mathbf{p}_e)\,\big\} \quad (40)$$

$$\;=\; 0.5\,\mathbf{T}_{ee}\,\mathbf{p}_e\;+\;0.25\sum_{\substack{f=1\\f\neq e}}^{E}\mathbf{T}_{ef}\mathbf{1}\;. \quad (41)$$

C. Experimental Prototypes

To encourage reproducibility, we include the complete parts list for our two experimental systems:

- a low-speed, low-cost system for video-rate indirect-only and epipolar-only imaging (Figure 8 of the paper) whose components are listed in Table 1; and
- a high-speed system for indirect-invariant shape acquisition and one-shot multi-pattern imaging (Figure 1), whose components are listed in Table 2.

Indirect-only and epipolar-only imaging We used a color AVT GT1920C camera for acquisition, a Texas Instruments LightCrafter for pixel masking and a 100 lumen Keynote Photonics LightCrafter kit for projection. The DMDs were synchronized at 2.7kHz, permitting $T = 96$ patterns and masks per video frame. The camera and DMD resolutions were quite different—1936x1456 versus 608x684—with each DMD pixel mapping to a 2×2 block of camera pixels. System calibration consists of computing the epipolar geometry between the two DMDs. We did this by first computing correspondences between the camera and each DMD separately. Patterns are uploaded to both DMDs once, at the beginning of an imaging session.

Indirect-invariant imaging For these experiments we used a monochrome AVT GT1920 camera and a pair of high-end DMDs from Texas Instruments (DLi 4130) that use a 2000 lumen light source. These operate at 22.2kHz, permitting $T = 800$ patterns per video frame. Although the DMD resolution was fairly high at 1024×768, its effective resolution was much lower, 484×364, because of the different physical dimensions and orientation of the camera sensor and DMD.

One-shot multi-pattern imaging Effective DMD resolution was even lower, 256×256, because of the scene's limited extent within the camera's field of view.

D. Generation of Masks & Projection Patterns

The mathematical definition of the patterns and masks we use in our SLT prototypes is discussed in Section 5 of the paper. Here, we show in Figure 2 examples of actual mask/pattern pairs uploaded on our DMDs and illustrate their construction according to Eqs. (10), (12), (13) and (16) in the paper.

Indirect-only imaging We use random mask/pattern pairs like those shown in Row 1 of Figure 2. To reduce the sensation of flicker by users who are physically present during video acquisition, we generate a random sequence of $T/2$ mask/pattern pairs ($T = 800$ or 96 depending on the prototype) and then generate a second mask/pattern sequence

Item #	Part Description	Quantity	Model Name	Company
1	color camera	1	GT1920C	Allied Vision Technologies
2	DMD projector	1	LC3000-Pro Pico Projector	Keynote Photonics
3	power supply	1	LC3000-Pro Power Supply	Keynote Photonics
4	DMD projector (mask)	1	DLP LightCrafter	Texas Instruments
5	power supply	1	T1228-Z12P-ND	Digi-Key Corporation
6	connector housing	2	WM1722-ND	Digi-Key Corporation
7	crimp	4	WM1142CT-ND	Digi-Key Corporation
8	Hirose contact plug	1	HR1623-ND	Digi-Key Corporation
9	12mm f/1.4 objective lens	1	Cinegon 1.4/12-0906	Schneider Optics
10	visible achromatic doublet pairs	2	MAP10100100-A	Thorlabs
11	300 grooves/mm transmission grating	1	GT25-03	Thorlabs
12	ring-activated threaded iris diaphragm	2	SM1D12D	Thorlabs
13	C-mount to SM1 adapter	1	SM1A9	Thorlabs
14	SM1 to C-mount adapter	1	SM1A10	Thorlabs
15	SM1 Coupler	1	SM1T10	Thorlabs
16	SM1 Lens Tube, 2 inch Thread Depth	1	SM1L20	Thorlabs
17	SM1 Lens Tube, 3 inch Thread Depth	1	SM1L30	Thorlabs
18	SM1-threaded cage plate	1	CP4S	Thorlabs
19	cage plate with 1.2 inch double bore	5	CP12	Thorlabs
20	cage plate with 35 mm aperture	4	CP03/M	Thorlabs
21	cylindrical lens mount	1	CH1A	Thorlabs
22	rod swivel coupler (set of four)	1	C2A	Thorlabs
23	rod end swivel connector (set of four)	1	C3A	Thorlabs
24	cage assembly rod, 2 inch long	2	ER2	Thorlabs
25	cage assembly rod, 3 inch long	4	ER3	Thorlabs
26	cage assembly rod, 4 inch long	4	ER4	Thorlabs
27	aluminum breadboard	1	MB3030/M	Thorlabs
28	12.7 mm x 40 mm optical post	1	TR40/M	Thorlabs
29	12.7 mm x 100 mm optical post	6	TR100/M	Thorlabs
30	post holder, 40 mm	1	PH40/M	Thorlabs
31	post holder, 100 mm	6	PH100/M	Thorlabs
32	studded pedestal base adapter	7	BE1/M	Thorlabs
33	small clamping fork	7	CF125	Thorlabs
34	30 mm single axis translation stage	1	#66-397	Edmund Optics
35	bottom adapter plate	1	#66-620	Edmund Optics
36	top adapter plate	1	#66-493	Edmund Optics
37	metric base plate	1	#54-975	Edmund Optics
38	thread-to-thread adapter	1	#56-323	Edmund Optics

Table 1: List of parts for the system shown in Figure 8 of the paper. The camera outputs live video at a rate of 28 frames per second. Each video frame requires 96 binary masks/projection patterns, *i.e.*, each mask/pattern is active for $375\mu sec$ of the frame's $36000\mu sec$ total exposure time.

Item #	Part Description	Quantity	Model Name	Company
1	monochrome camera	1	GT1920	Allied Vision Technologies
2	high-speed DMD	2	DLi4130VIS-7XGA	Digital Light Innovations
3	high-power LED	1	High Power S2+ w/ LED	Digital Light Innovations
4	connector housing	2	WM1728-ND	Digi-Key Corporation
5	fixed filter holder 40 mm Sq.	1	#54-997	Edmund Optics
6	45 degree mounting adapter	1	#59-001	Edmund Optics

Table 2: The high-speed system in Figure 1 uses identical optics to the low-speed one; the only differences between the two systems are (1) a faster DMD projector and mask, (2) their mounts, and (3) a monochrome camera that is identical to the low-speed system's color camera, but without the RGB filter mosaic. We only list the differing components in this table; the remaining parts are items 7-38 in Table 1. The system's camera outputs live video at a rate of 28 frames per second. Each video frame consists of 800 binary mask/projection patterns, *i.e.*, each mask/pattern is active for $45\mu sec$ of the frame's $36000\mu sec$ total exposure time.

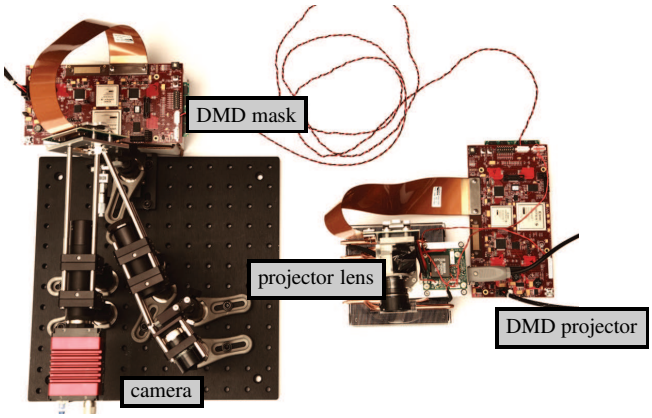

Figure 10: Our high-speed system. The key differences between this system and that shown in Figure 8 of the paper are a monochrome camera, the DMD mask, and the DMD projector.

whose projection patterns are the binary complement of the first $T/2$ projection patterns. This ensures a stable perception because the image integrated by the eye (or by a maskless camera) over the period of one video frame corresponds to a view of the scene under an all-white projection pattern.[1]

For indirect-only imaging, it is also important to ensure that no direct light "leaks" accidentally through the DMD mask. Such leaks can occur because of pixel misalignments between the DMD mask and the camera's sensor; because of the binary rasterization of epipolar lines; and because of projector/camera defocus. To make indirect-only acquisition robust to such effects, we slightly dilate the "off" regions on the generated masks. This reduces the occurrence of such leaks at the expense of a slight reduction in light efficiency. We found this approach to be very effective in practice; a similar idea was used in [5].

Epipolar-only imaging We generate epipolar-only video by operating the camera at 56fps and configuring the DMD of our low-speed prototype as follows:

- *odd video frames:* display 48 all-on mask/pattern pairs
- *even video frames:* display a sequence of 48 indirect-only mask/pattern pairs.

Epipolar-only video at 28fps is generated by (1) scaling the odd frames by 0.25 to account for the reduced intensity of indirect-only imaging (Eq. (11)) and (2) subtracting in real time the even frames from the scaled odd ones.

Indirect-invariant imaging and indirect-invariant 3D reconstruction We generate a sequence of 800 mask/pattern pairs for each of 9 grayscale structured-light patterns, as illustrated in rows 2-4 of Figure 2. We then capture one raw image of the scene for each of the 9 generated mask/pattern sequences. These 9 images are supplied, unaltered, to the 3D reconstruction algorithm.

One-shot, multi-pattern, indirect-invariant imaging We generate a sequence of 792 mask/pattern pairs, as outlined in row 5 of Figure 2, and upload them to the DMDs. We then apply the algorithm outlined in Section 5 independently to each frame of the raw live video stream.

E. Discussion of Videos in `SLT-supp.zip`

E.1. Videos in directory `figure5_videos/`

These videos are meant to be used in conjunction with Figure 5 in the paper.

- *Stripe scan:* The purpose of this video is to show that when we sweep a stripe along the scene, the indirect light received at an epipolar line (red line in the video) can be very significant.
- *Spot scan:* In this video we restrict projector illumination to the corresponding epipolar line. This enforces epipolar-only transport for pixels on the epipolar line shown in red. The video shows that the indirect light received at the red epipolar line is minimal, even for a highly-complex scene. Taken together, the stripe-scan and spot-scan videos demonstrate that the bulk of indirect light reaching the red epipolar line originates from projector pixels *outside* the corresponding epipolar line.
- *Acquiring* $\mathbf{T}_{ee}$*:* This video shows how an epipolar block $\mathbf{T}_{ee}$ can be acquired: for every frame in the spot-scan video, we collect all pixels along the red epipolar line and place them as a column of $\mathbf{T}_{ee}$, also shown in red. The video shows how the epipolar block is acquired, column by column, by sliding the spot along the epipolar line from left to right.
- *Acquiring* $\sum_f \mathbf{T}_{ef}$*:* An analogous visualization of the acquisition of $\sum_f \mathbf{T}_{ef}$, *i.e.*, by sweeping a vertical stripe from left to right on the projector plane.

E.2. Videos in directory `live_SLT_imaging/`

We show conventional, indirect-only and epipolar-only streams for a variety of scenes. These were recorded live with the prototype shown in Figure 8 of the paper. All streams were captured with the same camera and the same settings (exposure time, white balance, *etc.*) The three streams shown in each video were captured sequentially, with the only difference being the masks/projection patterns used. This is because we only have one color SLT prototype and it can operate in one of three modes at any given time (*i.e.*, conventional, indirect-only, epipolar-only).

[1] We emphasize that flicker is a purely subjective sensation that may be experienced by users who view the scene directly, without the benefit of the DMD mask. In particular, flicker *does not* occur in the videos captured by our prototypes.

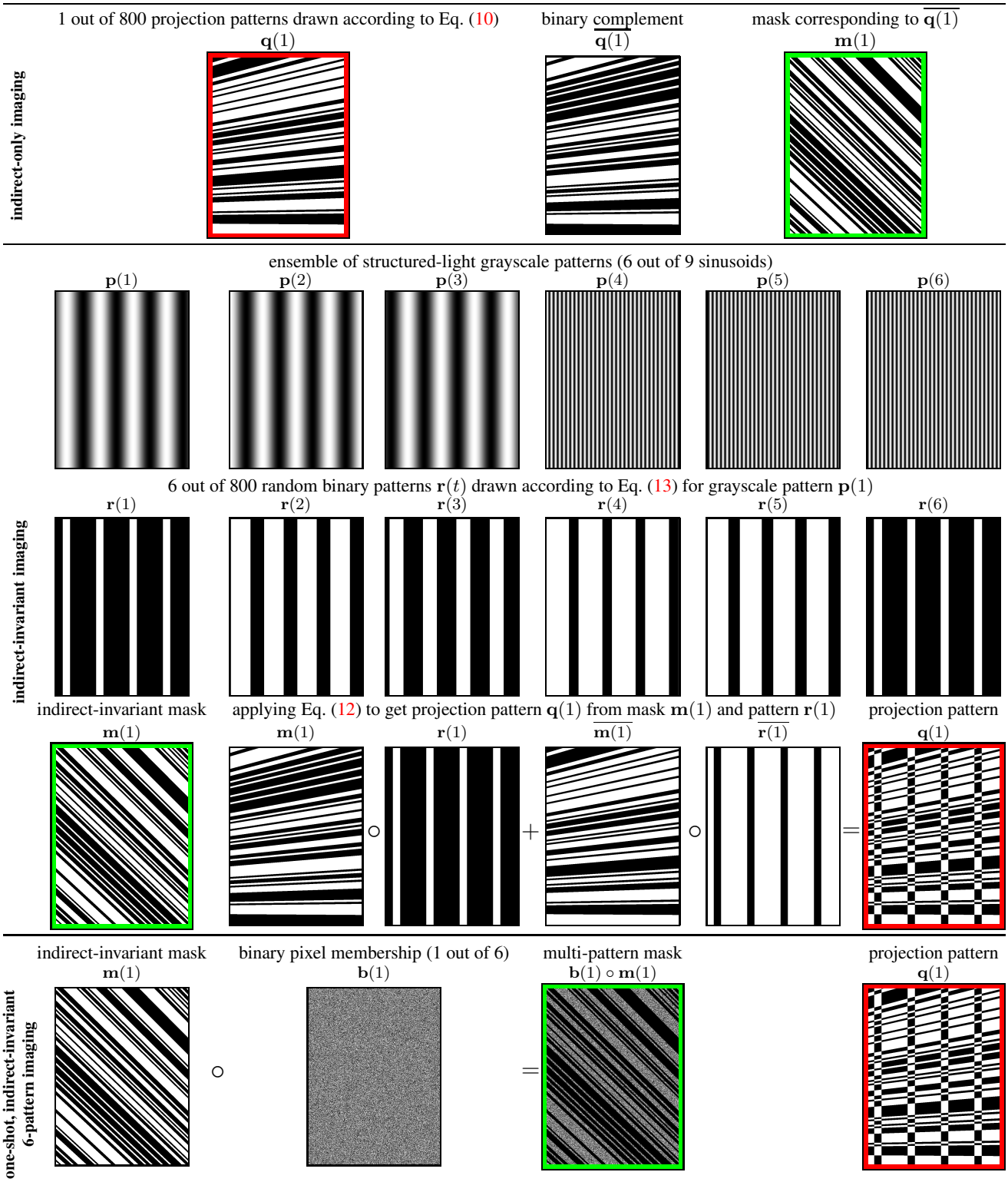

Figure 11: Deriving random pattern/mask pairs for three cases of SLT imaging. The derived patterns and masks are indicated with red and green borders, respectively. **Row 1:** For indirect-only imaging, the patterns and masks are constant along epipolar lines, with approximately half of them "on." **Row 2:** Six of the nine structured-light patterns we used. **Rows 3-4:** The masks for indirect-invariant imaging are identical to those for indirect-only imaging but the projection patterns differ. To generate them for a given grayscale structured-light pattern, we first generate a random sequence of binary patterns (Row 3) and then use that sequence, along with the sequence of masks, to compute the projection patterns. Row 4 shows one such example. **Row 5:** We generate pattern/mask pairs for 6-shot imaging as follows: (1) create 6 random binary images representing pixel membership for each pattern; (2) generate a sequence of 132 indirect-invariant binary pattern/mask pairs for each of 6 grayscale structured-light patterns, as outlined in Rows 2-4; (3) use the 792 projection patterns as is, and (4) multiply the masks element-wise with the associated pixel memberships. Row 5 shows one such calculation, for grayscale pattern $\mathbf{p}(1)$.

Conventional streams were captured under an all-on projection pattern with an all-on binary mask. Indirect-only streams were captured with 96 mask/pattern pairs, as explained in Section D. The conventional and indirect-only videos are recordings of the live, raw, video stream output by the camera.[2] Epipolar-only videos are created by pairwise differencing of adjacent video frames.

- *Candle:* A translucent candle.
- *Foam:* A piece of packing foam. The apparent speckles in the epipolar-only video are real: they correspond to momentary specular reflection from small, shiny membranes on the foam's surface.
- *Faux fur:* Note the marked difference between the epipolar-only component, which appears very shiny due direct near-specular reflection off the faux fur, versus the diffuse appearance of the indirect-only component, caused by sub-surface scattering. The occasional yellow tint in the epipolar-only component is due to saturation.
- *Glass:* This is the beer glass shown in Figures 5 and 1. Note that the glass appears essentially opaque in the epipolar-only component. This is because the light transmitted through the glass undergoes refraction, yielding non-linear paths that almost never lie on a single epipolar plane.
- *Hand:* Note the veins visible in the indirect-only component; also note the significant difference in apparent color of the hand in the indirect-only and epipolar-only components (due to sub-surface absorption and direct surface reflection, respectively).
- *Mug:* This is the mug shown in Figure 1. Note that artifacts appear occasionally on the white background behind the mug in the epipolar-only video. These occur because we do not compensate for motion when doing frame differencing. None of these artifacts appear in the indirect-only video, where no such differencing takes place.
- *Metal:* The indirect-only video clearly shows the very interesting caustics formed by the surface of this shiny metal plate. These caustics move very quickly; as a result, the frame-differencing we do for epipolar-only imaging causes ghosting on the white background. Again, none of these artifacts appear in the indirect-only video, which does not rely on frame differencing.
- *Water:* This example demonstrates our ability to successfully image a highly-complex, time-varying phenomenon, such as pouring water into a glass. As in the previous examples, the water appears mostly specular and opaque in the epipolar-only video whereas the caustics produced by light are clearly visible in the indirect-only video.
- *Wet hand:* The indirect-only video makes apparent the very dramatic changes in a hand's reflectance properties when water flows over it. We hypothesize that these changes are caused by scattering in the thin film of water flowing over the hand.
- *Small candle, paper:* More examples.

[2] To reduce file size for inclusion in the supplementary materials, it was necessary to compress these videos. As a result, some compression artifacts may be present.

E.3. Videos in directory `transport_robust_3D/`

This directory contains input images and 3D reconstruction results for the two scenes shown in rows 2-3 of Figure 9. Rows 2-4 of Figure 2 show the derivation of one of the mask/pattern pairs we used for this purpose.

- *Conventional phase shift input:* 9 input images acquired by conventional projection of phase-shifted patterns.
- *Conventional phase shift recontruction:* raw 3D points reconstructed from those input images.
- *Indirect-invariant phase shift input:* 9 input images acquired by indirect-invariant imaging with the same 9 patterns.
- *Indirect-invariant phase shift reconstruction:* raw 3D points reconstructed from those input images.
- *Conventional stripe-based recontruction:* as another example, we show reconstruction results obtained by sweeping a vertical stripe across the scene, as in conventional triangulation-based 3D laser scanning (768 images total). Despite the fact that stripe scanning relies on a much larger input dataset, our approach produces comparable results for the bowl scene and a far more complete model for the face scene.

E.4. Videos in directory `live_3D_capture/`

These videos provide details on the process of reconstructing a dynamic scene from video acquired by one-shot, 6-pattern indirect-invariant imaging. Row 5 of Figure 2 shows one of the mask/pattern pairs we constructed for this purpose.

- *Hand raw:* 169 frames of a moving hand, recorded live using one-shot, indirect-invariant, multi-pattern imaging.
- *Hand demosaic:* the 6 full-resolution videos resulting from demosaicing each video frame individually to obtain indirect-invariant views of the scene under 6 phase-shifted patterns.
- *Hand depth, albedo:* the raw, unprocessed depth and albedo maps reconstructed by applying conventional phase shifting to the demosaiced images.
- *Hand view albedo 1,2:* the reconstructed and texture-mapped geometry, shown from two different view-

points.

- *Hand view depth 1,2:* the reconstructed depth map, shown from two different viewpoints without texture mapping.

References

[1] V. Guillemin and A. Pollack. Differential Topology. *Prentice-Hall, 1974*.

[2] J. T. Kajiya. The Rendering Equation. *Proc. SIGGRAPH'86*.

[3] J. J. Koenderink. Solid Shape. *Cambridge University Press, 1990*.

[4] W. Rudin. Principles of Mathematical Analysis. *McGraw-Hill, 1976*.

[5] M. O'Toole, R. Raskar, K. N. Kutulakos. Primal-dual coding to probe light transport. *Proc. SIGGRAPH'12*.