Coded Two-Bucket Cameras for Computer Vision: Supplementary Materials

Mian Wei[1], Navid Sarhangnejad[2], Zhengfan Xia[2], Nikita Gusev[2], Nikola Katic[2], Roman Genov[2], and Kiriakos N. Kutulakos[1]

[1] Department of Computer Science, University of Toronto, Canada
{mianwei,kyros}@cs.toronto.edu
[2] Department of Electrical Engineering, University of Toronto, Canada
{sarhangn,xia,nikita,roman}@ece.toronto.edu, katic.nik@gmail.com

A C2B Pixel Design Considerations

Two- and four-bucket pixel layouts from the literature.   The design of our sensor's pixels builds upon several earlier pixel designs that include more than one bucket. Fig. A.1 shows four such designs, including an earlier C2B design from our group [3]. These layouts are shown here to provide some basic intuition about the two key design choices we made in our prototype's new sensor—and in C2B sensing more generally.

Design choice #1: per-pixel vs. per-bucket coding.   As explained in the main paper and illustrated in Figs. 1 and 3, the two buckets inside a C2B pixel are never in the same state at the same time (*i.e.*, both active or both inactive). This can be viewed as a form of coding that assigns a single bit to each pixel, so that bucket activities can be toggled. A natural question is whether it would be possible and/or desirable to control these buckets individually, so that they could be activated or de-activated independently of each other.

Unfortunately, although more flexible, per-bucket coding would cause poor CMOS performance for two fundamental reasons. First, it would require doubling the pixel's SRAM block in order to store the required two bits per pixel (*i.e.*, one per bucket). Such a block would be very large relative to the pixel's overall dimensions because even a one-bit memory uses lots of gates (Fig A.1, lower right). A two-bit in-pixel memory would therefore reduce the photo-sensitive region's relative size severely— and thus the pixel's overall quantum efficiency. Second, even ignoring the size of the photo-sensitive region, allowing both buckets to be active simultaneously would cause sluggish/inefficient pixels. This is because pixels operate by inducing a gradient field in the photo-sensitive region that "pulls" the light-generated charges toward the active bucket (see dashed lines in top row of Fig. A.1). Activating both buckets would induce competing gradient fields, causing charges to move slowly toward the individual buckets.

Design choice #2: C2B pixels versus memory-less multi-bucket pixels.   We believe that C2B sensors have several advantages: **(1)** they are scalable because they can pack S views into one C2B shot via coding, without hard-wiring S into the pixel's design. This is a particularly important feature because both structured-light and photometric-stereo systems typically use more than four illumination conditions for better accuracy and
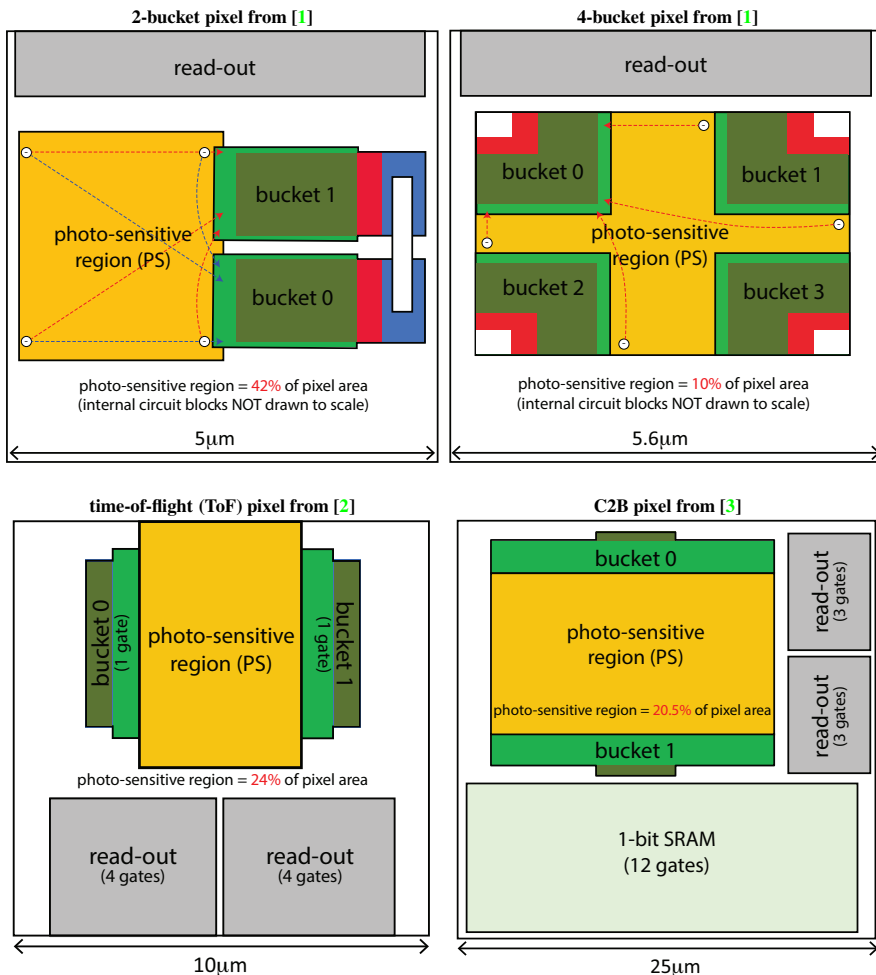
Fig. A.1: *CMOS layout of two- and four-bucket pixels from the literature.* Colors indicate circuit blocks of similar functionality across these layouts. Orange blocks denote the pixel's photosensitive region. White regions are occupied by wires, *etc.*. Except for the relative size of the photo-sensitive region, the dimensions of internal blocks were not included in [1]. The number of gates in each block is listed where known. Dashed curves in the top row indicate the approximate path traced by photo-generated charges when bucket 1 (red) or bucket 0 (blue), is active, respectively.

robustness. C2B sensors can achieve this—possibly at the expense of some spatial resolution compared to the results shown in the main paper—whereas a four-bucket sensor could not. **(2)** C2B pixels are much more light-efficient because each extra bucket takes up lots of space within a pixel. For example, the photo-sensitive region of the four-bucket pixel in the top for Fig. A.1 is four times smaller than its two-bucket counterpart (10% vs. 42% of pixel area, respectively). Adding even more buckets to a 2D sensor

could therefore lead to significant inefficiencies. **(3)** Moreover, it is unclear if 2D sensors with more than four buckets could offer acceptable imaging performance. This is because the charge-collection gradient fields of the S buckets must be identical up to rotation/reflection in order to ensure similar performance (*e.g.*, see charge-collection paths in Fig. A.1, top row), but the rectilinear CMOS design rules generally prevent achieving anything beyond four-fold gradient field symmetries. **(4)** S-bucket sensors cannot be used for applications that require sensor-plane coding because they lack the in-pixel memory required for this capability.

B   Proofs and Derivations

B.1   Sec. 2: Gaussian approximation of bucket-ratio random variables

Proposition B.1. The random variables r and $\hat{r}$ can be approximated by Gaussian random variable whose means and variances are given by:

$$\text{mean}(r) = \frac{j}{j + \hat{j}} \quad , \quad \text{var}(\hat{r}) = \frac{\sigma^2}{2(j + \hat{j})^2} \quad , \tag{B.1}$$

and

$$\text{mean}(\hat{r}) = \frac{\hat{j}}{j + \hat{j}} \quad , \quad \text{var}(\hat{r}) = \frac{\sigma^2}{2(j + \hat{j})^2} \quad , \tag{B.2}$$

where j and $\hat{j}$ are the noiseless bucket images and σ^2 is the variance of the noise.

Proof. Consider the bucket ratios r and $\hat{r}$ at frame F and pixel p, which are defined by combining bucket-1 and bucket-0 image intensities $\mathbf{i}^p[f]$ and $\hat{\mathbf{i}}^p[f]$ according to Eq. (5) from Sec. 2. For notational simplicity, we denote $\mathbf{i}^p[f]$ and $\hat{\mathbf{i}}^p[f]$ with i and $\hat{i}$ respectively. We model the two-bucket intensities i and $\hat{i}$ as having zero-mean Gaussian image noise with variance σ^2:

$$i = j + e \quad , \quad \hat{i} = \hat{j} + \hat{e} \quad , \tag{B.3}$$

where j, $\hat{j}$ are the noiseless bucket-1 and bucket-0 intensities and e, $\hat{e}$ are zero-mean Gaussian random variables with variance σ^2. In order to apply Lemma B.1 to r and $\hat{r}$, we require $\frac{\text{std}(i + \hat{i})}{j + \hat{j}} < 0.22$. This effectively amounts to requiring the SNR to be at least 5 which is an appropriate assumption for most imaging systems. Now we can apply Lemma B.1 with $k = 1$ to derive the approximate means and variances for r and $\hat{r}$. We first show the approximations for r:

$$\text{mean}(r) = \text{mean}\left(\frac{i}{i + \hat{i}}\right) = \frac{j}{j + \hat{j}} \quad , \quad \text{var}(r) = \text{var}\left(\frac{i}{i + \hat{i}}\right) = \frac{\sigma^2}{2(j + \hat{j})^2} \quad . \tag{B.4}$$

We now show the approximations for $\hat{r}$:

$$\text{mean}(\hat{r}) = \text{mean}\left(\frac{\hat{i}}{i + \hat{i}}\right) = \frac{\hat{j}}{j + \hat{j}} \quad , \quad \text{var}(\hat{r}) = \text{var}\left(\frac{\hat{i}}{i + \hat{i}}\right) = \frac{\sigma^2}{2(j + \hat{j})^2} \quad . \tag{B.5}$$

From Eq. (B.4) and Eq. (B.5), we can see that the approximate variance for both r and $\hat{r}$ are the same. Furthermore they are proportional to $\frac{1}{(j + \hat{j})^2}$. This means that the

approximate standard deviation for both of them is proportional to $\frac{1}{j+\hat{j}}$. Since j and j are linear function of the albedo, the approximate standard deviation for both r and $\hat{r}$ is weakly dependance on albedo. □

Lemma B.1. If X and Y are normally distributed with means μ_X, μ_Y respectively and variances σ^2, $k\sigma^2$ and $\frac{\text{std}(X+Y)}{\text{mean}(X+Y)} < 0.22$, then $\frac{X}{X+Y}$ can be approximated as a Gaussian random variable with the following mean and variance:

$$\text{mean}\left(\frac{X}{X+Y}\right) = \frac{\mu_X}{\mu_X + \mu_Y} \quad , \quad \text{var}\left(\frac{X}{X+Y}\right) = \frac{k\sigma^2}{(k+1)(\mu_X + \mu_Y)^2} \ . \tag{B.6}$$

Proof. Let X and Y be Gaussian random variances with means μ_X, μ_Y respectively and variances σ^2, $k\sigma^2$ respectively where k is the ratio between the variances of X and Y. Furthermore, let us assume that $\frac{\text{std}(X+Y)}{\text{mean}(X+Y)} < 0.22$. First, we use the observation from [4] that $kX - Y$ and $X + Y$ have Pearson correlation coefficient ρ equal to 0:

$$\rho = \frac{\text{cov}(kX - Y, X + Y)}{\text{std}(kX - Y)\text{std}(X + Y)} = \frac{\text{cov}(kX, X) - \text{cov}(Y, Y)}{\text{std}(kX - Y)\text{std}(X + Y)} = \frac{k\text{var}(X) - \text{var}(Y)}{\text{std}(kX - Y)\text{std}(X + Y)}. \tag{B.7}$$

Since $\text{var}(X) = \sigma^2$ and $\text{var}(Y) = k\sigma^2$, the numerator on the right side of Eq. (B.7), and therefore ρ, is equal to 0. Since we assume $\frac{\text{std}(X+Y)}{\text{mean}(X+Y)} < 0.22$, Table 1 from [5] tells us that we can reasonably approximate the ratio of two uncorrelated Gaussian random variables as a normal random variable with the following parameters:

$$\text{mean}\left(\frac{kX - Y}{X + Y}\right) = \frac{\text{mean}(kX - Y)}{\text{mean}(X + Y)} \quad , \quad \text{var}\left(\frac{kX - Y}{X + Y}\right) = \frac{\text{var}(kX - Y)}{\text{mean}(X + Y)^2} \ . \tag{B.8}$$

Now we expand Eq. (B.8) to simplify it:

$$\text{mean}\left(\frac{kX - Y}{X + Y}\right) = \frac{k\mu_X - \mu_Y}{\mu_X + \mu_Y} \quad , \quad \text{var}\left(\frac{kX - Y}{X + Y}\right) = \frac{k^2\sigma^2 + k\sigma^2}{(\mu_X + \mu_Y)^2} \ . \tag{B.9}$$

We can use Eq. (B.9) to now approximate the noise of $\frac{X}{X+Y}$. To do this, we first show that $\frac{X}{X+Y}$ is a linear function of $\frac{kX-Y}{X+Y}$:

$$\frac{X}{X + Y} = \frac{kX + X}{(k + 1)(X + Y)} = \frac{kX + X - Y + Y}{(k + 1)(X + Y)} = \frac{1}{k + 1} + \left(\frac{1}{k + 1}\right)\frac{kX - Y}{X + Y} \ . \tag{B.10}$$

Combining Eq. (B.9) and Eq. (B.10), we can approximate the mean of $\frac{X}{X+Y}$:

$$\text{mean}\left(\frac{X}{X + Y}\right) = \frac{1}{k + 1} + \left(\frac{1}{k + 1}\right)\frac{k\mu_X - \mu_Y}{\mu_X + \mu_Y} = \frac{\mu_X + \mu_Y + k\mu_X - \mu_Y}{(k + 1)(\mu_X + \mu_Y)} = \frac{\mu_X}{\mu_X + \mu_Y} \ , \tag{B.11}$$

which is the noiseless ratio. We can also approximate the variance of $\frac{X}{X+Y}$ using Eq. (B.9) and Eq. (B.10):

$$\text{var}\left(\frac{X}{X + Y}\right) = \frac{1}{(k + 1)^2}\text{var}\left(\frac{kX - Y}{X + Y}\right) = \frac{1}{(k + 1)^2}\frac{k^2\sigma^2 + k\sigma^2}{(\mu_X + \mu_Y)^2} = \frac{k\sigma^2}{(k + 1)(\mu_X + \mu_Y)^2} \ . \tag{B.12}$$

□

B.2 Sec. 3 derivations

B.2.1 Derivation of Eq. (11).

Consider the two image intensities, i_l and $\sum_k i_k$. We can model their values as a function of the unknown $\mathbf{x}$ (Eq. (10) from Sec. 3):

$$i_l = a\mathbf{d}_l\mathbf{x} + e_l \quad , \quad \sum_k i_k = a\sum_k \mathbf{d}_k\mathbf{x} + e_k \quad . \tag{B.13}$$

As $\sum_k \mathbf{d}_k$ is formed by summing the rows $\mathbf{d}_k$ of $\mathbf{D}$, we can express Eq. (B.13) as:

$$\sum_k i_k = a\mathbf{1}\mathbf{D}\mathbf{x} + \mathbf{1}\mathbf{e} \quad , \tag{B.14}$$

where $\mathbf{1}$ is a 1 by S vector of 1's and $\mathbf{1}\mathbf{e}$ aggregates the image noises. We can derive a constraint from the ratio of i_l and $\sum_k i_k$ from the noiseless Eq. (B.13) and Eq. (B.14):

$$\frac{i_l}{\sum_k i_k} = \frac{a\mathbf{d}_l\mathbf{x}}{a\mathbf{1}\mathbf{D}\mathbf{x}} = \frac{\mathbf{d}_l\mathbf{x}}{\mathbf{1}\mathbf{D}\mathbf{x}} \quad . \tag{B.15}$$

We denote $\frac{i_l}{\sum_k i_k}$ with r_l and multiply both sides with $\mathbf{1}\mathbf{D}\mathbf{x}$. $\qquad\square$

B.2.2 Derivation of Eq. (12).

Consider two image intensities, i_l and i_k. We have the following relation (Eq. (10) from Sec. 3):

$$i_l = a\mathbf{d}_l\mathbf{x} + e_l \quad , \quad i_k = a\mathbf{d}_k\mathbf{x} + e_k \quad , \tag{B.16}$$

where e_l and e_k are the noise for i_l and i_k respectively. Solving for $\mathbf{x}$ normally involves using noiseless Eq. (B.16) directly as constraints on $\mathbf{x}$:

$$i_l = a\mathbf{d}_l\mathbf{x} \quad , \quad i_k = a\mathbf{d}_k\mathbf{x} \tag{B.17}$$

However, instead of treating these as separate constraints, we can take their ratio:

$$\frac{i_l}{i_k} = \frac{a\mathbf{d}_l\mathbf{x}}{a\mathbf{d}_k\mathbf{x}} = \frac{\mathbf{d}_l\mathbf{x}}{\mathbf{d}_k\mathbf{x}} \quad , \tag{B.18}$$

and cross multiply the terms in Eq. (B.18) to derive Eq. (12). $\qquad\square$

B.3 Sec. 4 proofs

B.3.1 Proof of Observation 1.

Since $\mathbf{W}$ has S columns, we trivially know that $\operatorname{rank}\mathbf{W} \leq S$. It remains to show that $\operatorname{rank}\mathbf{W} \leq F + 1$. Since the rank of a matrix is equal to its row rank and the row rank of a matrix is determined by the dimensions spanned by its rows, we simply need to show that the rows of $\mathbf{W}$ which are $\mathbf{c}_1^p, \ldots \mathbf{c}_F^p, \overline{\mathbf{c}}_1^p, \ldots, \overline{\mathbf{c}}_F^p$ lie in a subspace V of dimension at most $F + 1$.

First, consider $\mathbf{c}_1^p$ and $\overline{\mathbf{c}}_1^p$ which are the first row vector of $\mathbf{W}$ and its compliment. Since they are complimentary, we know their sum, trivially:

$$\mathbf{c}_1^p + \overline{\mathbf{c}}_1^p = \mathbf{1} \; , \tag{B.19}$$

where $\mathbf{1}$ is a 1 by S vector of all 1's. Now let V be the subspace spanned by the vectors $\mathbf{c}_1^p, \ldots \mathbf{c}_F^p$, and $\overline{\mathbf{c}}_1^p$. Using Eq. (B.19), we can see that for all k, $\overline{\mathbf{c}}_k^p$ must lie in V since it can be expressed as the linear combination of vectors in V:

$$\overline{\mathbf{c}}_k^p = \mathbf{1} - \mathbf{c}_k^p = \mathbf{c}_1^p + \overline{\mathbf{c}}_1^p - \mathbf{c}_k^p \; . \tag{B.20}$$

Therefore given V, we know that $\mathbf{c}_1^p, \ldots \mathbf{c}_F^p$, and $\overline{\mathbf{c}}_1^p$ lie in V by construction and also all $\overline{\mathbf{c}}_k^p$'s lie in V. Since V is spanned by $F + 1$ vectors, we know that $\dim(V)$ is at most $F + 1$. Therefore the rows of $\mathbf{W}$ span a subspace with dimension at most $F + 1$, implying that $\operatorname{rank} \mathbf{W} \leq F + 1$. $\qquad\square$

B.3.2 Proof of Lemma 1.

We derive the lower bound for the MSE in 4 steps:

Step 1: Verify that the definition of the MSE used in [6] is still equal to $\frac{\sigma^2}{S}\operatorname{trace}\left[\left(\mathbf{W}'\mathbf{W}\right)^{-1}\right]$ for two-bucket cameras.

Step 2: Derive a lower bound on $\frac{\sigma^2}{S}\operatorname{trace}\left[\left(\mathbf{W}'\mathbf{W}\right)^{-1}\right]$ as a function of $\mathbf{W}'\mathbf{W}$'s largest eigenvalue λ_1.

Step 3: Bound the value of λ_1.

Step 4: Combine Step 2 and Step 3 results to derive the explicit lower bound.

Step 1: verification of MSE definition. Using Eq. (9) from Sec. 3, we know that the demultiplexed intensities $i_1^p \ldots i_S^p$ can be computed by applying the inverse of $\mathbf{W}$ to the $2F$ bucket intensities $\mathbf{i}^p$ and $\hat{\mathbf{i}}^p$. As in [6], we define the MSE of $\mathbf{W}$ as the mean variance of the demultiplexed images:

$$\mathrm{MSE} = \frac{1}{S}\sum_{k=1}^{S}\operatorname{var}(i_k^p) \; . \tag{B.21}$$

Let us denote $(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$ with $\mathbf{A}$. Let $\mathbf{A}_i$ denote the ith row of $\mathbf{A}$ and $\mathbf{A}_i[j]$ denote the element of $\mathbf{A}$ at row i and column j. Then we can express $\operatorname{var}(i_k^p)$ as:

$$\operatorname{var}(i_k^p) = \operatorname{var}\left(\mathbf{A}_k\begin{bmatrix}\mathbf{i}^p \\ \hat{\mathbf{i}}^p\end{bmatrix}\right) \tag{B.22}$$

$$= \operatorname{var}\left(\sum_{j=1}^{F}\mathbf{A}_k[j]\mathbf{i}_j^p + \mathbf{A}_k[j+F]\hat{\mathbf{i}}_j^p\right) \tag{B.23}$$

$$= \sum_{j=1}^{F}\mathbf{A}_k[j]^2\operatorname{var}(\mathbf{i}_j^p) + \mathbf{A}_k[j+F]^2\operatorname{var}(\hat{\mathbf{i}}_j^p) \; . \tag{B.24}$$

We will assume that $\mathbf{i}^p$ and $\hat{\mathbf{i}}^p$ have zero-mean Gaussian noise with variance σ^2. We can now simplify Eq. (B.24):

$$\text{var}(i_k^p) = \sum_{j=1}^{F} \mathbf{A}_k[j]^2 \sigma^2 + \mathbf{A}_k[j+F]^2 \sigma^2 = \sigma^2 \sum_{j=1}^{2F} \mathbf{A}_k[j]^2 \ , \tag{B.25}$$

and use Eq. (B.25) in Eq. (B.21):

$$\text{MSE} = \frac{1}{S} \sum_{k=1}^{S} \sigma^2 \sum_{j=1}^{2F} \mathbf{A}_k[j]^2 \tag{B.26}$$

$$= \frac{\sigma^2}{S} \sum_{k=1}^{S} \sum_{j=1}^{2F} \mathbf{A}_k[j]^2 \ . \tag{B.27}$$

Since the sum of squares of elements of $\mathbf{A}$ is equal to the trace of $\mathbf{A}'\mathbf{A}$, we can substitute this back into Eq. (B.27):

$$\text{MSE} = \frac{\sigma^2}{S} \text{trace}(\mathbf{A}'\mathbf{A}) \tag{B.28}$$

$$= \frac{\sigma^2}{S} \text{trace}\left[\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}' \right] \tag{B.29}$$

$$= \frac{\sigma^2}{S} \text{trace}\left[(\mathbf{W}'\mathbf{W})^{-1}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{W} \right] \tag{B.30}$$

$$= \frac{\sigma^2}{S} \text{trace}\left[(\mathbf{W}'\mathbf{W})^{-1} \right] \tag{B.31}$$

Step 2: bounding the MSE in terms of λ_1:   Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_S \geq 0$ be the eigenvalues of $\mathbf{W}'\mathbf{W}$. Then we can express $\text{trace}\left[(\mathbf{W}'\mathbf{W})^{-1} \right]$ as the sum of its eigenvalues:

$$\frac{\sigma^2}{S} \text{trace}\left[(\mathbf{W}'\mathbf{W})^{-1} \right] = \frac{\sigma^2}{S} \sum_{i=1}^{S} \frac{1}{\lambda_i} \tag{B.32}$$

$$= \frac{\sigma^2}{S\lambda_1} + \frac{\sigma^2}{S} \sum_{i=2}^{S} \frac{1}{\lambda_i} \tag{B.33}$$

$$= \frac{\sigma^2}{S\lambda_1} + \left(\frac{\sigma^2(S-1)}{S} \right)\left(\frac{1}{S-1} \right) \sum_{i=2}^{S} \frac{1}{\lambda_i} \ . \tag{B.34}$$

The right term in the addition in Eq. (B.34) is the arithmetic mean of $\frac{1}{\lambda_i}$'s. Since the arithmetic mean is always greater or equal than the harmonic mean, we have the following inequality:

$$\left(\frac{1}{S-1} \right) \sum_{i=2}^{S} \frac{1}{\lambda_i} \geq \left(\frac{\sum_{i=2}^{S}(\frac{1}{\lambda_i})^{-1}}{S-1} \right)^{-1} = \frac{S-1}{\sum_{i=2}^{S} \lambda_i} \ . \tag{B.35}$$

Since trace($\mathbf{W}'\mathbf{W}$) sums over all the eigenvalues of $\mathbf{W}'\mathbf{W}$, we know that:

$$\sum_{i=2}^{S} \lambda_i = \text{trace}(\mathbf{W}'\mathbf{W}) - \lambda_1 \ . \tag{B.36}$$

We can now combine Eq. (B.35) and Eq. (B.36) to turn Eq. (B.34) into an inequality:

$$\frac{\sigma^2}{S}\text{trace}\left[(\mathbf{W}'\mathbf{W})^{-1}\right] \geq \frac{\sigma^2}{S\lambda_1} + \sigma^2(\frac{S-1}{S})\frac{S-1}{\sum_{i=2}^{S}\lambda_i} \tag{B.37}$$

$$= \frac{\sigma^2}{S\lambda_1} + \sigma^2(\frac{S-1}{S})\frac{S-1}{\text{trace}(\mathbf{W}'\mathbf{W}) - \lambda_1}. \tag{B.38}$$

We compute trace($\mathbf{W}'\mathbf{W}$) before proceeding. Let $\mathbf{J}$ be the F by S matrix of 1's. Then we have the following:

$$\text{trace}(\mathbf{W}'\mathbf{W}) = \text{trace}[(\mathbf{C}^p)'\mathbf{C}^p + (\overline{\mathbf{C}}^p)'\overline{\mathbf{C}}^p] \tag{B.39}$$

$$= \text{trace}[(\mathbf{C}^p)'\mathbf{C}^p + (\mathbf{J} - \mathbf{C}^p)'(\mathbf{J} - \mathbf{C}^p)] \tag{B.40}$$

$$= \text{trace}[(\mathbf{C}^p)'\mathbf{C}^p + \mathbf{J}'\mathbf{J} - \mathbf{J}'\mathbf{C}^p - (\mathbf{C}^p)'\mathbf{J} + (\mathbf{C}^p)'\mathbf{C}^p] \tag{B.41}$$

$$= 2\text{trace}[(\mathbf{C}^p)'\mathbf{C}^p] + \text{trace}(\mathbf{J}'\mathbf{J}) - \text{trace}(\mathbf{J}'\mathbf{C}^p) - \text{trace}[(\mathbf{C}^p)'\mathbf{J}] \tag{B.42}$$

$$= 2\sum_{k=1}^{F}\sum_{j=1}^{S}\mathbf{C}_k^p[j]^2 + \sum_{k=1}^{F}\sum_{j=1}^{S}1 - \sum_{k=1}^{F}\sum_{j=1}^{S}\mathbf{C}_k^p[j] - \sum_{k=1}^{F}\sum_{j=1}^{S}\mathbf{C}_k^p[j] \ . \tag{B.43}$$

Since $\mathbf{C}^p$ is a binary matrix, we know that $\mathbf{C}_k^p[j]^2 = \mathbf{C}_k^p[j]$:

$$\text{trace}(\mathbf{W}'\mathbf{W}) = 2\sum_{k=1}^{F}\sum_{j=1}^{S}\mathbf{C}_k^p[j] + FS - \sum_{k=1}^{F}\sum_{j=1}^{S}\mathbf{C}_k^p[j] - \sum_{k=1}^{F}\sum_{j=1}^{S}\mathbf{C}_k^p[j] \tag{B.44}$$

$$= SF \ . \tag{B.45}$$

We can now substitute Eq. (B.45) into Eq. (B.38) yielding the following:

$$\frac{\sigma^2}{S}\text{trace}\left[(\mathbf{W}'\mathbf{W})^{-1}\right] \geq \frac{\sigma^2}{S\lambda_1} + \sigma^2(\frac{S-1}{S})\frac{S-1}{\text{trace}(\mathbf{W}'\mathbf{W}) - \lambda_1} \tag{B.46}$$

$$= \frac{\sigma^2}{S\lambda_1} + \sigma^2(\frac{S-1}{S})\frac{S-1}{SF - \lambda_1} \tag{B.47}$$

$$= \frac{\sigma^2}{S}(\frac{1}{\lambda_1} + \frac{(S-1)^2}{SF - \lambda_1}) \tag{B.48}$$

$$= \frac{\sigma^2}{S}\frac{SF - \lambda_1 + \lambda_1(S-1)^2}{(SF - \lambda_1)\lambda_1}) \tag{B.49}$$

$$= \frac{\sigma^2}{S}\frac{SF + \lambda_1 S^2 - 2S)}{(SF - \lambda_1)\lambda_1}) \tag{B.50}$$

$$= \sigma^2\frac{F + \lambda_1(S-2)}{(SF - \lambda_1)\lambda_1} \ . \tag{B.51}$$

Step 3: bounding the value of λ_1:   We know that the squared spectral norm of $\mathbf{W}$ is equal to λ_1:

$$\lambda_1 = \|\mathbf{W}\|_2^2 = \sup_{x \neq 0} \frac{\|\mathbf{W}x\|_2^2}{\|x\|_2^2} = \sup_{x \neq 0} \frac{x'(\mathbf{W}'\mathbf{W})x}{x'x} \quad . \tag{B.52}$$

We can use $x = \mathbf{1}$ where $\mathbf{1}$ is a S by 1 vector of all 1's on the right side of Eq. (B.52) as a lower bound of λ_1:

$$\lambda_1 \geq \frac{\mathbf{1}'(\mathbf{W}'\mathbf{W})\mathbf{1}}{\mathbf{1}'\mathbf{1}} = \frac{\sum_{k=1}^{2F}(\sum_{j=1}^{S}\mathbf{W}_k[j])^2}{S} \quad . \tag{B.53}$$

Since $\mathbf{W}$ is defined as the concatenation of $\mathbf{C}^p$ and $\overline{\mathbf{C}}^p$, we know that for $k \geq F$ and j:

$$\mathbf{W}_k[j] = 1 - \mathbf{W}_{k-F}[j] \quad . \tag{B.54}$$

We can now substitute Eq. (B.54) back into Eq. (B.53):

$$\lambda_1 \geq \frac{1}{S}\sum_{k=1}^{F}(\sum_{j=1}^{S}\mathbf{W}_k[j])^2 + \frac{1}{S}\sum_{k=F+1}^{2F}(\sum_{j=1}^{S}1 - \mathbf{W}_{k-F}[j])^2 \tag{B.55}$$

$$= \frac{1}{S}\sum_{k=1}^{F}(\sum_{j=1}^{S}\mathbf{W}_k[j])^2 + (\sum_{j=1}^{S}1 - \mathbf{W}_k[j])^2 \tag{B.56}$$

$$= \frac{1}{S}\sum_{k=1}^{F}(\sum_{j=1}^{S}\mathbf{W}_k[j])^2 + (S - \sum_{j=1}^{S}\mathbf{W}_k[j])^2 \tag{B.57}$$

$$= \frac{1}{2S}\sum_{k=1}^{F}2(\sum_{j=1}^{S}\mathbf{W}_k[j])^2 + 2(S - \sum_{j=1}^{S}\mathbf{W}_k[j])^2 \quad . \tag{B.58}$$

The inequality of arithmetic mean and geometric mean which states $a^2 + b^2 \geq 2ab$. This implies that $2a^2 + 2b^2 \geq a^2 + b^2 + 2ab \geq (a + b)^2$. We can use this on the summand of Eq. (B.58):

$$\lambda_1 \geq \frac{1}{2S}\sum_{k=1}^{F}(\sum_{j=1}^{S}\mathbf{W}_k[j] + S - \sum_{j=1}^{S}\mathbf{W}_k[j])^2 \tag{B.59}$$

$$= \frac{1}{2S}\sum_{k=1}^{F}S^2 \tag{B.60}$$

$$= \frac{SF}{2} \quad . \tag{B.61}$$

Additionally, we will prove an upper bound on λ_1. Let v be an eigenvector with eigenvalue λ_1 and at least one positive element and let $v[k]$ be the maximum element in v. Since $\mathbf{W}'\mathbf{W}$ only has non-negative values, we can bound λ_1 in the following way:

$$\lambda_1 v[k] = \sum_{j=1}^{S}(\mathbf{W}'\mathbf{W})_k[j]v[j] \leq \sum_{j=1}^{S}(\mathbf{W}'\mathbf{W})_k[j]v[k] \quad . \tag{B.62}$$

As we chose v to have at least one positive element, we know that $v[k] > 0$ so we can divide both sides of Eq. (B.62) with $v[k]$ and preserve the inequality:

$$\lambda_1 \leq \sum_{j=1}^{S} (\mathbf{W}'\mathbf{W})_k[j] \; . \tag{B.63}$$

Now, we simply bound the kth row sum for $\mathbf{W}'\mathbf{W}$ to bound λ_1:

$$\sum_{j=1}^{S} (\mathbf{W}'\mathbf{W})_k[j] = \sum_{j=1}^{S} \sum_{i=1}^{2F} \mathbf{W}_i[k]\mathbf{W}_i[j] \leq \sum_{j=1}^{S} \sum_{i=1}^{2F} \mathbf{W}_i[j] \; . \tag{B.64}$$

Since $\mathbf{W}$ is defined as the concatenation of $\mathbf{C}^p$ and $\overline{\mathbf{C}}^p$, it has exactly SF non-zero elements all equal to 1. The right side of Eq. (B.64) sums over all the elements of $\mathbf{W}$ so we can replace it with SF:

$$\sum_{j=1}^{S} (\mathbf{W}'\mathbf{W})_k[j] \leq \sum_{j=1}^{S} \sum_{i=1}^{2F} \mathbf{W}_i[j] = SF \; . \tag{B.65}$$

So the kth row sum of $\mathbf{W}'\mathbf{W}$ is bounded above by SF. Combining Eq. (B.61) and Eq. (B.65) gives us the interval in which λ_1 must lie:

$$\lambda_1 \in [\frac{SF}{2}, SF] \; . \tag{B.66}$$

Step 4: compute the explicit bound. Eq. (B.66) tells us that λ_1 has to be between $\frac{SF}{2}$ and SF. Eq. (B.51) expresses the lower bound of the MSE as the ratio between two functions of λ_1. The denominator, $(SF - \lambda_1)\lambda_1$, is a simple quadratic function of λ_1 which is maximized at $\lambda_1 = \frac{SF}{2}$ in the range $[\frac{SF}{2}, SF]$. The numerator, $F + \lambda_1(S - 2)$, is a simple linear function of λ_1 that is minimized also at $\lambda_1 = \frac{SF}{2}$ in the range $[\frac{SF}{2}, SF]$. Combining this information in Eq. (B.51) yields:

$$\frac{\sigma^2}{S}\text{trace}\left[(\mathbf{W}'\mathbf{W})^{-1}\right] \geq \sigma^2 \frac{F + \lambda_1(S - 2)}{(SF - \lambda_1)\lambda_1} \tag{B.67}$$

$$\geq \sigma^2 \frac{F + \frac{SF}{2}(S - 2)}{(SF - \frac{SF}{2})\frac{SF}{2}} \tag{B.68}$$

$$= 2\sigma^2 \frac{2F + SF(S - 2)}{(SF)^2} \tag{B.69}$$

$$= 2\sigma^2 \frac{1 + (S - 1)^2}{S^2 F} \; . \tag{B.70}$$

In the minimal configuration, $F = S - 1$. We can substitute this into Eq. (B.70):

$$\frac{\sigma^2}{S}\text{trace}\left[(\mathbf{W}'\mathbf{W})^{-1}\right] \geq 2\sigma^2 \frac{(S - 1)^2 + 1}{(S - 1)S^2} \; . \tag{B.71}$$

$\square$

B.3.3   Proof of Proposition 1.

The proof proceeds in 3 steps:

Step 1: We derive a simple expression of $\mathbf{W}'\mathbf{W}$.

Step 2: We compute an explicit formula for $(\mathbf{W}'\mathbf{W})^{-1}$.

Step 3: We compute the MSE of $(\mathbf{W}'\mathbf{W})^{-1}$ using the explicit formula and show that it is equal to the lower bound in Eq. (B.70).

First, we clarify some notation. Let $\mathbf{1}$ be a S by 1 vector of 1's, $\mathbf{J}$ be the $S - 1$ by S matrix of 1's, $\mathbf{H}$ be the S by S Hadamard matrix, and $\mathbf{I}_S$ be the S by S identity matrix. By construction, $\begin{bmatrix} \mathbf{1} & \widetilde{\mathbf{H}}' \end{bmatrix} = \mathbf{H}'$.

Step 1: simple expression for $\mathbf{W}'\mathbf{W}$.   We begin by expanding $\mathbf{W}'\mathbf{W}$:

$$\mathbf{W}'\mathbf{W} = \begin{bmatrix} (\mathbf{C}^p)' & (\overline{\mathbf{C}}^p)' \end{bmatrix} \begin{bmatrix} \mathbf{C}^p \\ \overline{\mathbf{C}}^p \end{bmatrix} \tag{B.72}$$

$$= (\mathbf{C}^p)'\mathbf{C}^p + (\overline{\mathbf{C}}^p)'\overline{\mathbf{C}}^p \tag{B.73}$$

$$= (\mathbf{C}^p)'\mathbf{C}^p + (\mathbf{J} - \mathbf{C}^p)'(\mathbf{J} - \mathbf{C}^p) \tag{B.74}$$

$$= \frac{1}{4}(\widetilde{\mathbf{H}} + 1)'(\widetilde{\mathbf{H}} + 1) + (\mathbf{J} - \frac{1}{2}(\widetilde{\mathbf{H}} + 1))'(\mathbf{J} - \frac{1}{2}(\widetilde{\mathbf{H}} + 1)) \tag{B.75}$$

$$= \frac{1}{4}(\widetilde{\mathbf{H}} + \mathbf{J})'(\widetilde{\mathbf{H}} + \mathbf{J}) + \frac{1}{4}(\mathbf{J} - \widetilde{\mathbf{H}})'(\mathbf{J} - \widetilde{\mathbf{H}}) \tag{B.76}$$

$$= \frac{1}{4}(\widetilde{\mathbf{H}}'\widetilde{\mathbf{H}} + \mathbf{J}'\mathbf{J} + \widetilde{\mathbf{H}}'\mathbf{J} + \mathbf{J}'\widetilde{\mathbf{H}}) + \frac{1}{4}(\widetilde{\mathbf{H}}'\widetilde{\mathbf{H}} + \mathbf{J}'\mathbf{J} - \widetilde{\mathbf{H}}'\mathbf{J} - \mathbf{J}'\widetilde{\mathbf{H}}) \tag{B.77}$$

$$= \frac{1}{2}(\widetilde{\mathbf{H}}'\widetilde{\mathbf{H}} + \mathbf{J}'\mathbf{J}) \ . \tag{B.78}$$

Since $\mathbf{J}$ is a $S - 1$ by S matrix of 1's, $\mathbf{J}'\mathbf{J}$ is a S by S matrix of $S - 1$'s which can be expressed as $(S - 1)\mathbf{1}\mathbf{1}'$:

$$\mathbf{W}'\mathbf{W} = \frac{1}{2}(\widetilde{\mathbf{H}}'\widetilde{\mathbf{H}} + (S - 1)\mathbf{1}\mathbf{1}') \tag{B.79}$$

$$= \frac{1}{2}(\widetilde{\mathbf{H}}'\widetilde{\mathbf{H}} + \mathbf{1}\mathbf{1}' + (S - 2)\mathbf{1}\mathbf{1}') \tag{B.80}$$

$$= \frac{1}{2}(\begin{bmatrix} \mathbf{1} & \widetilde{\mathbf{H}}' \end{bmatrix} \begin{bmatrix} \mathbf{1}' \\ \widetilde{\mathbf{H}} \end{bmatrix} + (S - 2)\mathbf{1}\mathbf{1}') \tag{B.81}$$

$$= \frac{1}{2}(\mathbf{H}'\mathbf{H} + (S - 2)\mathbf{1}\mathbf{1}') \ . \tag{B.82}$$

$\mathbf{H}'\mathbf{H}$ is easy to compute as it is orthogonal. In particular, $\mathbf{H}$ satisfies the condition:

$$\mathbf{H}'\mathbf{H} = S\mathbf{I}_S \ . \tag{B.83}$$

We substitute Eq. (B.83) into Eq. (B.82):

$$\mathbf{W}'\mathbf{W} = \frac{S}{2}\mathbf{I}_S + \frac{S - 2}{2}\mathbf{1}\mathbf{1}'. \tag{B.84}$$

Step 2: explicit formula for $(\mathbf{W}'\mathbf{W})^{-1}$.   The form $\mathbf{W}'\mathbf{W}$ takes in Eq. (B.82) is very convenient as we can compute its inverse using the Sherman-Morrison formula which

states:

$$(A + uv')^{-1} = A^{-1} - \frac{A^{-1}uv'A^{-1}}{1 + v'A^{-1}u} , \tag{B.85}$$

where A is a matrix and u, v are vectors. Since we will be using a special case of the formula where A is the identity matrix and u, v are both equal to $\sqrt{\frac{S-2}{S}}\mathbf{1}$, the formula becomes even simpler:

$$(A + uv')^{-1} = (\mathbf{I}_S + \frac{S-2}{S}\mathbf{11}')^{-1} \tag{B.86}$$

$$= \mathbf{I}_S - \frac{\frac{S-2}{S}\mathbf{11}'}{1 + \frac{S-2}{S}\mathbf{1}'\mathbf{1}} \tag{B.87}$$

$$= \mathbf{I}_S - \frac{\frac{S-2}{S}\mathbf{11}'}{1 + S - 2} \tag{B.88}$$

$$= \mathbf{I}_S - \frac{S-2}{(S-1)S}\mathbf{11}' . \tag{B.89}$$

We can use Eq. (B.89) to compute $(\mathbf{W}'\mathbf{W})^{-1}$:

$$(\mathbf{W}'\mathbf{W})^{-1} = (\frac{S}{2}\mathbf{I}_S + \frac{S-2}{2}\mathbf{11}')^{-1} \tag{B.90}$$

$$= \frac{2}{S}(\mathbf{I}_S + \frac{S-2}{S}\mathbf{11}')^{-1} \tag{B.91}$$

$$= \frac{2}{S}(\mathbf{I}_S - \frac{S-2}{(S-1)S}\mathbf{11}') . \tag{B.92}$$

Step 3: compute MSE for W. Using Eq. (B.92), we have:

$$\frac{\sigma^2}{S}\mathrm{trace}\left[(\mathbf{W}'\mathbf{W})^{-1}\right] = \frac{\sigma^2}{S}\mathrm{trace}\left[\frac{2}{S}\left(\mathbf{I}_S - \frac{S-2}{(S-1)S}\mathbf{11}'\right)\right] \tag{B.93}$$

$$= \frac{2\sigma^2}{S^2}\left[\mathrm{trace}(\mathbf{I}_S) - \mathrm{trace}\left(\frac{S-2}{(S-1)S}\mathbf{11}'\right)\right] \tag{B.94}$$

$$= \frac{2\sigma^2}{S^2}\left[S - \frac{S-2}{(S-1)S}S\right] \tag{B.95}$$

$$= \frac{2\sigma^2}{S^2}\frac{S^2 - 2S + 2}{S-1} \tag{B.96}$$

$$= 2\sigma^2\frac{(S-1)^2 + 1}{(S-1)S^2}. \tag{B.97}$$

$\square$

B.4 Generalization of Eq. (9) and Lemma 1 to bucket ratios

The results of Eq. (9) and Lemma 1 in the paper apply to raw intensities captured by C2B frames. Since the BRD method relies on demosaicing and then demultiplexing ratio mosaics, it is not obvious that we can use the same $\mathbf{W}$ to demultiplexing the ratio mosaics. Furthermore, $\mathbf{W}$ might not even be optimal for demultiplexing ratio mosaics.

This is because the MSE is intensity-based and defined using the noise model of the bucket images which differs from the noise model of the bucket ratios.

Proposition B.2 below generalizes Eq. (9) from Sec. 2 and shows that we can use the same $\mathbf{W}$ to demultiplex the bucket ratios to recover the illumination ratios, defined in Eq. (11) from Sec. 3. Additionally, Proposition B.3 generalizes Lemma 1 and shows that we can define an approximate MSE using the normal approximation of the demultiplexed illumination ratios which can be used to find optimal multiplexing matrices $\mathbf{W}$. We exhaustively search for matrices that minimized the approximate MSE and verified that $\mathbf{W}$'s, constructed from optimal $\mathbf{C}^p$'s in Table 2 from Sec. 4, that minimize the intensity-based MSE also minimize the approximate MSE.

Proposition B.2. If $i_1, \ldots, i_S$ are demultiplexed image intensities obtained from bucket images $\mathbf{i}$ and $\hat{\mathbf{i}}$ and multiplexing matrix $\mathbf{W}$, then $r_1, \ldots, r_S$ are demultiplexed illumination ratios obtained from bucket ratios $\mathbf{r}$ and $\hat{\mathbf{r}}$ and multiplexing matrix $\mathbf{W}$:

$$\begin{bmatrix} r_1 \\ \vdots \\ r_S \end{bmatrix} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}' \begin{bmatrix} \mathbf{r} \\ \hat{\mathbf{r}} \end{bmatrix} \ . \tag{B.98}$$

Proof. Given F frames, we have $2F$ images, $\mathbf{i}$, the bucket-1 images, and $\hat{\mathbf{i}}$, the bucket-0 images. From $\mathbf{i}$ and $\hat{\mathbf{i}}$, we can compute ratios $\mathbf{r}$ and $\hat{\mathbf{r}}$ using Eq. (5) from Sec. 2:

$$\mathbf{r} = \begin{bmatrix} \frac{\mathbf{i}[1]}{\mathbf{i}[1]+\hat{\mathbf{i}}[1]} \\ \vdots \\ \frac{\mathbf{i}[F]}{\mathbf{i}[F]+\hat{\mathbf{i}}[F]} \end{bmatrix} \quad , \quad \hat{\mathbf{r}} = \begin{bmatrix} \frac{\hat{\mathbf{i}}[1]}{\mathbf{i}[1]+\hat{\mathbf{i}}[1]} \\ \vdots \\ \frac{\hat{\mathbf{i}}[F]}{\mathbf{i}[F]+\hat{\mathbf{i}}[F]} \end{bmatrix} \ . \tag{B.99}$$

For a given frame f, $\mathbf{i}[f] + \hat{\mathbf{i}}[f]$ is equal to $\sum_k i_k$ (Eq. (8) from Sec. 3):

$$\mathbf{i}[f] + \hat{\mathbf{i}}[f] = \mathbf{c}_f \begin{bmatrix} i_1 \\ \vdots \\ i_S \end{bmatrix} + \overline{\mathbf{c}}_f \begin{bmatrix} i_1 \\ \vdots \\ i_S \end{bmatrix} = \mathbf{1} \begin{bmatrix} i_1 \\ \vdots \\ i_S \end{bmatrix} = \sum_k i_k \ , \tag{B.100}$$

where $\mathbf{1}$ is a 1 by S vector of 1's. Equation (B.100) tells us $\mathbf{r}$ and $\hat{\mathbf{r}}$ are equal to $\mathbf{i}$ and $\hat{\mathbf{i}}$ scaled by $1/\sum_k i_k$. Therefore, demultiplexing the bucket ratios will allow us to recover S demultiplexed intensities $i_1, \ldots, i_S$ also scaled by $1/\sum_k i_k$:

$$\begin{bmatrix} \frac{i_1}{\sum_k i_k} \\ \vdots \\ \frac{i_S}{\sum_k i_k} \end{bmatrix} = \frac{1}{\sum_k i_k} \begin{bmatrix} i_1 \\ \vdots \\ i_S \end{bmatrix} = \frac{1}{\sum_k i_k}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}' \begin{bmatrix} \mathbf{i} \\ \hat{\mathbf{i}} \end{bmatrix} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}' \begin{bmatrix} \mathbf{r} \\ \hat{\mathbf{r}} \end{bmatrix} \ . \tag{B.101}$$

Since we define the illumination ratio r_l as $i_l/\sum_k i_k$, the left side of Eq. (B.101) is exactly the vector of the S illumination ratios. $\square$

Proposition B.3. The approximate MSE of $\mathbf{W}$ for demultiplexing bucket ratios can be defined as:

$$\text{MSE} = \frac{\lambda^2}{S}\text{trace}\left[(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\begin{bmatrix}\mathbf{I}_F & -\mathbf{I}_F \\ -\mathbf{I}_F & \mathbf{I}_F\end{bmatrix}\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\right], \tag{B.102}$$

where λ^2 is the approximate variance of the bucket ratios.

Proof. Similarly to what we did in the proof of Lemma 1, we define the MSE as the mean variance of the demultiplexed illumination ratios:

$$\text{MSE} = \frac{1}{S}\sum_{k=1}^{S}\text{var}(r_k) . \tag{B.103}$$

Let us denote $(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$ with $\mathbf{A}$. Let $\mathbf{A}_i$ denote the ith row of $\mathbf{A}$ and $\mathbf{A}_i[j]$ denote the element of $\mathbf{A}$ at row i and column j. Let $\mathbf{r}[j]$ and $\hat{\mathbf{r}}[j]$ denote the jth element in $\mathbf{r}$ and $\hat{\mathbf{r}}$, respectively. Using Proposition B.2, we can express $\text{var}(r_k)$ as:

$$\text{var}(r_k) = \text{var}\left(\mathbf{A}_k\begin{bmatrix}\mathbf{r} \\ \hat{\mathbf{r}}\end{bmatrix}\right) = \text{var}\left(\sum_{j=1}^{F}\mathbf{A}_k[j]\mathbf{r}[j] + \mathbf{A}_k[j+F]\hat{\mathbf{r}}[j]\right) . \tag{B.104}$$

Notice that $\mathbf{r}[j]$ and $\hat{\mathbf{r}}[j]$ sum up to 1:

$$\hat{\mathbf{r}}[j] = \frac{\hat{\mathbf{i}}[j]}{\mathbf{i}[j]+\hat{\mathbf{i}}[j]} = \frac{\hat{\mathbf{i}}[j]+\mathbf{i}[j]-\mathbf{i}[j]}{\mathbf{i}[j]+\hat{\mathbf{i}}[j]} = 1 - \frac{\mathbf{i}[j]}{\mathbf{i}[j]+\hat{\mathbf{i}}[j]} = 1 - \mathbf{r}[j] . \tag{B.105}$$

Now we can substitute Eq. (B.105) into Eq. (B.104):

$$\text{var}(r_k) = \text{var}\left(\sum_{j=1}^{F}\mathbf{A}_k[j]\mathbf{r}[j] + \mathbf{A}_k[j+F](1-\mathbf{r}[j])\right) \tag{B.106}$$

$$= \text{var}\left(\sum_{j=1}^{F}(\mathbf{A}_k[j]-\mathbf{A}_k[j+F])\mathbf{r}[j]\right) \tag{B.107}$$

$$= \sum_{j=1}^{F}(\mathbf{A}_k[j]-\mathbf{A}_k[j+F])^2\text{var}(\mathbf{r}[j]) . \tag{B.108}$$

From Eq. (B.108), we see that the definition of the MSE depends on $\text{var}(\mathbf{r}[j])$. Using Lemma B.2, we can denote the approximate variance of $\mathbf{r}[j]$ as λ^2:

$$\text{MSE} = \frac{1}{S}\sum_{k=1}^{S}\sum_{j=1}^{F}(\mathbf{A}_k[j]-\mathbf{A}_k[j+F])^2\lambda^2 \tag{B.109}$$

$$= \frac{\lambda^2}{S}\sum_{k=1}^{S}\sum_{j=1}^{F}(\mathbf{A}_k[j]-\mathbf{A}_k[j+F])^2 . \tag{B.110}$$

To simplify Eq. (B.110), we first define a matrix whose elements are equal to $\mathbf{A}_k[j] - \mathbf{A}_k[j+F]$. Since we are always subtracting elements exactly F columns apart and in the same row. We know that this matrix is difference between the submatrix formed by taking the first F columns of $\mathbf{A}$ and the submatrix formed by taking the last F columns

of $\mathbf{A}$. We will denote these submatrices with $\mathbf{A}^L$ and $\mathbf{A}^R$ respectively. We can compute them as follows:

$$\mathbf{A}^L = \mathbf{A} \begin{bmatrix} \mathbf{I}_F \\ 0 \end{bmatrix} \qquad \mathbf{A}^R = \mathbf{A} \begin{bmatrix} 0 \\ -\mathbf{I}_F \end{bmatrix} \ , \tag{B.111}$$

where $\mathbf{I}_F$ is the F by F identity matrix. Taking their difference gives us the following expression:

$$\mathbf{A}^L - \mathbf{A}^R = \mathbf{A} \begin{bmatrix} \mathbf{I}_F \\ 0 \end{bmatrix} - \mathbf{A} \begin{bmatrix} 0 \\ -\mathbf{I}_F \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{I}_F \\ -\mathbf{I}_F \end{bmatrix} \ , \tag{B.112}$$

where $\mathbf{I}_F$ is the F by F identity matrix. Now summing all the squared elements of $\mathbf{A}^L - \mathbf{A}^R$ equates to computing the trace of $(\mathbf{A}^L - \mathbf{A}^R)'(\mathbf{A}^L - \mathbf{A}^R)$. Using this fact and substituting Eq. (B.112) into Eq. (B.110) allows us to simplify the MSE:

$$\text{MSE} = \frac{\lambda^2}{S} \sum_{k=1}^{S} \sum_{j=1}^{F} (\mathbf{A}_k[j] - \mathbf{A}_k[j+F])^2 \tag{B.113}$$

$$= \frac{\lambda^2}{S} \text{trace}[(\mathbf{A}^L - \mathbf{A}^R)(\mathbf{A}^L - \mathbf{A}^R)'] \tag{B.114}$$

$$= \frac{\lambda^2}{S} \text{trace}(\mathbf{A} \begin{bmatrix} \mathbf{I}_F \\ -\mathbf{I}_F \end{bmatrix} \begin{bmatrix} \mathbf{I}_F & -\mathbf{I}_F \end{bmatrix} \mathbf{A}') \tag{B.115}$$

$$= \frac{\lambda^2}{S} \text{trace}(\mathbf{A} \begin{bmatrix} \mathbf{I}_F & -\mathbf{I}_F \\ -\mathbf{I}_F & \mathbf{I}_F \end{bmatrix} \mathbf{A}') \ . \tag{B.116}$$

We can now substitute $\mathbf{A} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$ into Eq. (B.116) to derive Eq. (B.102). □

Lemma B.2. If the $2F$ bucket images $\mathbf{i}[1], \dots, \mathbf{i}[F]$ and $\hat{\mathbf{i}}[1], \dots \hat{\mathbf{i}}[F]$ all have zero-mean Gaussian noise with variance σ^2, then for any frame f, the bucket ratios r and $\hat{r}$ for $\mathbf{i}[f]$ and $\hat{\mathbf{i}}[f]$ both have approximate variance λ^2 which is independent of f.

Proof. From Proposition B.1, we know the approximate variances for r and $\hat{r}$ are equal:

$$\text{var}(r) = \text{var}(\hat{r}) = \frac{\sigma^2}{2(j+\hat{j})^2} \ , \tag{B.117}$$

where j and $\hat{j}$ are the means of $\mathbf{i}[f]$ and $\hat{\mathbf{i}}[f]$, respectively. From Eq. (B.100), we already know that $\mathbf{i}[f] + \hat{\mathbf{i}}[f]$ is equal to the sum of the S noiseless demultiplexed image intensities, $\sum_k i_k$. Therefore there are no terms on the right side of Eq. (B.117) that depends on f and we set $\lambda^2 = \frac{\sigma^2}{2(\sum_k i_k)^2}$. □

C   Details of Empirical SNR Analysis Shown in Fig.6(a),(b)

We considered the cases where $S = \{4, 5, 6, 7\}$ and $F = S - 1$. The case $S = 3$ was not evaluated because all non-degenerate 2×3 matrices $\mathbf{C}^p$ have the same expected SNR. We present the sub-optimal matrices we used as shown in Table C.1. We evaluated the SNR gain for both image intensities and illumination ratios which is defined in Eq. (11) from Sec. 3.

# Frames	$S=4$	$S=5$	$S=6$	$S=7$
$\mathrm{MSE_{sub}}$	0.9167	1	1.0667	1.1190
SNR Gain, Eq. (C.118)	1.4833	1.6269	1.7541	1.8553
Sup-optimal $\mathbf{C}^p$	1 0 0 0	1 0 0 0 0	1 0 0 0 0 0	1 0 0 0 0 0 0
	0 1 0 0	0 1 0 0 0	0 1 0 0 0 0	0 1 0 0 0 0 0
	0 0 1 0	0 0 1 0 0	0 0 1 0 0 0	0 0 1 0 0 0 0
		0 0 0 1 0	0 0 0 1 0 0	0 0 0 1 0 0 0
			0 0 0 0 1 0	0 0 0 0 1 0 0
				0 0 0 0 0 1 0

Table C.1: *Sup-optimal $\mathbf{C}^p$ used for comparison.* MSE and SNR gain for image intensities are computed for $\sigma = 1$ in the minimal configurations ($F = S - 1$).

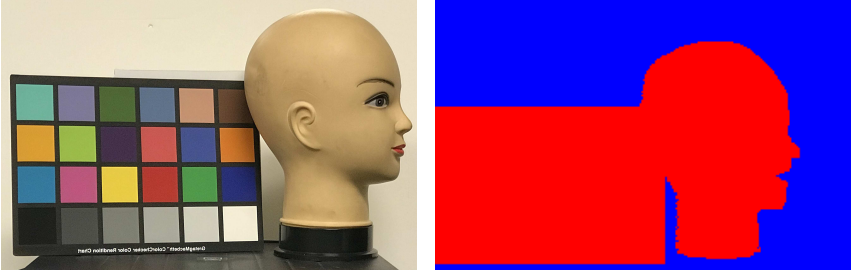**Fig. C.2:** *Objects used for performance evaluation.* **(a)** Color view of the objects, view not from C2B camera. **(b)** The binary foreground mask (red is foreground). The mask excludes pixels not belonging to the objects as well as shadows.

Empirical SNR gain for image intensities. To measure average SNR of a given $\mathbf{C}^p$, we **(1)** sequentially capture F C2B frames of a static scene using $S = F + 1$ sub-frames per frame, S structured-light patterns $\mathbf{l}_1, \ldots, \mathbf{l}_S$, and the same matrix $\mathbf{C}^p$ for all pixels—see Eq. (8) from Sec. 3; **(2)** demultiplex the $2F$ intensities collected in each pixel's buckets by multiplying with the $S \times 2F$ pseudoinverse of $\left[\frac{\mathbf{C}^p}{\overline{\mathbf{C}}^p}\right]$ to get intensities $i_1^p, \ldots, i_S^p$ for $\mathbf{l}_1, \ldots, \mathbf{l}_S$, respectively; **(3)** repeat steps 1&2 300 times; **(4)** compute $\frac{\text{mean}}{\text{standard deviation}}$ at each pixel across trials; and **(5)** average this ratio over all valid pixels in the S demultiplexed images using a binary foreground mask (Fig. C.2(b)).

Empirical SNR gain for illumination ratios. We compute the empirical SNR gain for illumination ratios. We use the same steps as those used to compute the empirical SNR gain for image intensities. The only difference here is that instead of demultiplexing bucket intensities, we demultiplex the bucket ratios.

Theoretical SNR gain. We define the theoretical SNR gain [7] as:

$$\text{SNR Gain} = \sqrt{\frac{\mathrm{MSE_{triv}}}{\mathrm{MSE_{opt}}}} \, , \qquad\qquad (C.118)$$

where $\mathrm{MSE_{triv}}$ and $\mathrm{MSE_{opt}}$ are computed using Eq. (13) from Sec. 4 for the sub-optimal matrix and optimal matrix, respectively. We list the theoretical SNR gains in Table C.1.
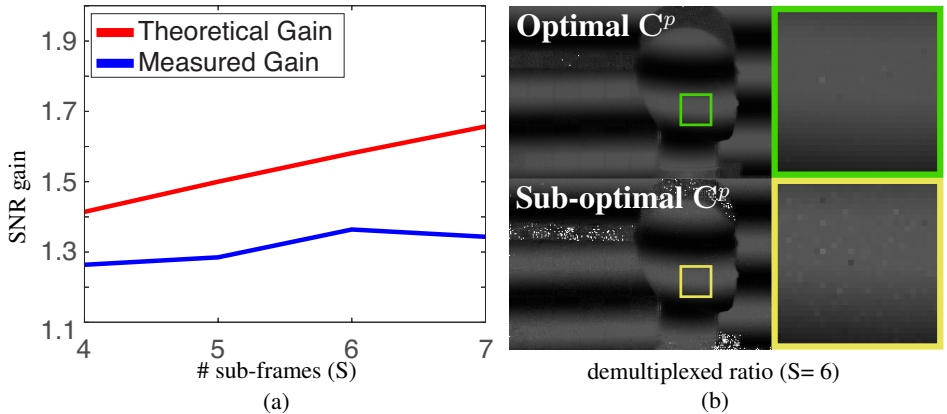
Fig. C.3: *SNR gains of illumination ratios.* **(a)** Plot of empirical SNR gains (in blue) and theoretical SNR gains (in red) for illumination ratios. **(b)** One out of S illumination ratios obtained with each $\mathbf{C}^p$. The optimal $\mathbf{C}^p$ yielded visibly less noisy images (please zoom in to the electronic copy).

Empirical evaluation of SNR gain for illumination ratios. We visualize the SNR gain for the illumination ratios in Fig. C.3(a). The same multiplexing matrices as those used for evaluating the SNR gain of image intensities are used here. We also show examples of demultiplexed illumination ratios with optimal and sub-optimal multiplexing matrices in Fig. C.3(b). The gap between the theoretical SNR gain and the empirical SNR gain could be explained by the noise model used. In particular, the theoretical SNR gain does not take into account Poisson noise which might cause the SNR gain to decrease in practise. We can see that similar trends appear for both the empirical SNR gains in illumination ratios and the empirical SNR gains in image intensities (Fig. 6(a) from Sec. 5). However, the theoretical and empirical performance gains suggest that while bucket ratios do provide albedo-invariant images, demultiplexing image intensities provides a greater SNR gain than demultiplexing bucket ratios.

D Details of Sec. 5 Simulations

Scene details. We set the image resolution to 300 x 300. We use orthographic projection for the camera and set the camera centre as the origin. We use a Lambertian model (Eq. (10) and Table 1 from Sec. 3) to render the scene so there are no interreflections and no occlusions in the scene. Each pixel is assumed to be in the line of sight of the light source. We set the mean reflectance intensity to be 80 and the variance to be 400. We use OpenCV's demosaicing function with *BayerBG2RGB_EA* as the flag. We sample the RGB color space with Gaussians centred at [0.5,0.5,0.5] and variance 0.1. color images have uniform color. Each frequency uses 20 samples to compute the RMSE. We run two set of experiments. One experiment keeps the the reflectance frequency fixed to 0.3 of the Nyquist limit and varies the geometric frequency (normal for photometric stereo and depth for structured light). The other experiment keeps the geometric frequency fixed to 0.3 of the Nyquist limit while varying the reflectance frequency.
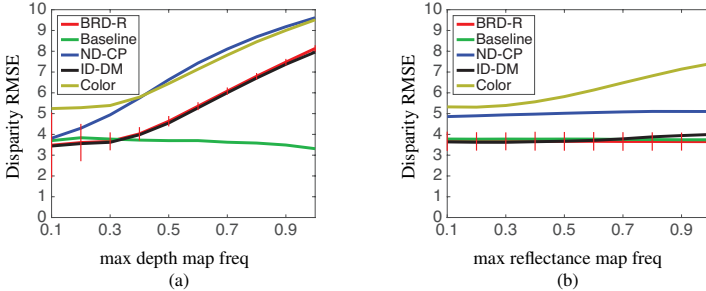
Fig. D.4: *Performance comparisons for structured light triangulation on synthetic scenes.* **(a)** Disparity root-mean-squared error (RMSE) as a function of the depth map's highest spatial frequency. Reflectance spatial frequency is set to 0.3 of Nyquist. **(b)** Disparity root-mean-squared error of disparity estimation as a function of the reflectance map's highest spatial frequency. Depth spatial frequency is set to 0.3 of Nyquist.

Structured light triangulation. We multiplex 4 phase-shifted sinusoidal patterns with spatial frequency of 8 using the 3 by 4 optimal $\mathbf{C}^p$ (Table 2 from Sec. 4). We set the centre of the projector to (20,0,0) and rotate it $10°$ in the XZ plane towards the camera. The camera image plane is a rectangle of length 50 and width 50 centred at the origin. For the depth map, we generate a 300 x 300 image of uniformly-sampled depths. We set the maximum depth frequency by band-pass filtering this depth map. We then rescale and translate the image to a depth map with a mean depth 100 (unitless in MATLAB) and a depth variance of 8.333. Phase unwrapping was done by finding the maximum and minimum disparity with depth bounds of 92.5 and 107.

Photometric stereo, additional details. We multiplex the 4 light sources into 3 pixels using the 3 by 4 optimal $\mathbf{C}^p$ (Table 2 from Sec. 4) for the C2B frame and Bayer color mosaic. We uniformly sample 4 light source directions until their condition number is below 3.

Discussion. We show the results of our simulations for structured light triangulation in Fig. D.4 with 95% confidence intervals on BRD-R. The experiments obtained by fixing the reflectance frequency show that when the reflectance frequency is low, ID-DM outperforms other techniques. The experiments fixing the depth frequency show that BRD-R outperforms ID-DM when there is high frequency reflectance variation. In general, these experiments show that the choice of shape estimation method has trade-offs. Reflectance-invariant methods like BRD-R are robust to arbitrary reflectance frequencies, but may not always help improve performance.

E   Details of Sec. 6 Experiments

E.1   Experimental conditions: additional details

Configuration for photometric stereo. Light source calibration consisted of two steps. First, we computed the direction of incidence of each source onto the scene using a

chrome sphere. Second, to account for the non-uniform illumination spot produced by the LED lenses, we placed a fronto-parallel, uniformly-white and diffuse poster board in front of the camera and computed the light source intensity at each pixel. This intensity was subsequently used for all shape computations at that pixel.

Configuration for structured-light triangulation. Our projector can store up to 96 binary patterns in its memory, and projects gray-scale patterns by dithering those binary patterns at rates up to 96 patterns per 20ms video frame. This limited projection rate, along with limitations of our camera's triggering firmware, made it impossible to project more than 24 binary patterns onto the scene within each sub-frame. As a result, our structured-light projection patterns were quantized significantly, taking on integer values in the range $[0, 23]$. Phase unwrapping was done by using depth bounds.

E.2   Quantitative experiments: one-shot versus multi-frame acquisition

Definition of "ground truth." The goal of our quantitative experiments was to evaluate how one-shot 3D estimates from a C2B camera differ from those computed by applying the identical shape estimation algorithm to full-resolution images acquired sequentially. Any differences between these estimates would necessarily be due to the demosaicing process or the coded two-bucket imaging procedure itself—not the accuracy of the underlying algorithms employed. We therefore treat the 3D estimates computed by multi-frame acquisition to be our reference, "ground-truth" shape. Because several pixels in the field of view do not produce valid "ground-truth" 3D data (due to shadows, isolated regions with specular reflectance, *etc.*) we define a foreground mask to exclude them (see Figs. E.5(c) and E.5(e)).

Evaluation procedure. For structured light we use one metric: (1) the percentage of bad-matching pixels [8]. We define bad-matching pixels to be pixels whose disparity is more than one pixel away from the ground-truth disparity. For photometric stereo, we use root-mean-squared angular error and median angular error, both measured in degrees. We compare the impact of thee different RGB demosaicing procedures: MATLAB's built-in function; OPENCV's function with edge-aware demosaicing; and the deep demosaicing system of Gharbi *et al.* [9] which represents the current state of the art on the problem. We used the same test objects for both photometric stereo and structured-light triangulation. The objects, shown in Fig E.5(a), consist of a book and a colorful hat whose high-frequency albedo variations make its reconstruction challenging to algorithms—such as color photometric stereo—as it breaks the constant-albedo assumption.

Photometric stereo results. Table E.2 compares the RMSE obtained by the various methods, for both one-shot and simulated one-shot images. These results suggest that the ID-DM method has the best performance when coupled with the deep-learning-based demosaicing algorithm [9]. As shown in Table E.3, which considers median error, the BRD-R method performs the best with Matlab's demosaic function. While there is a
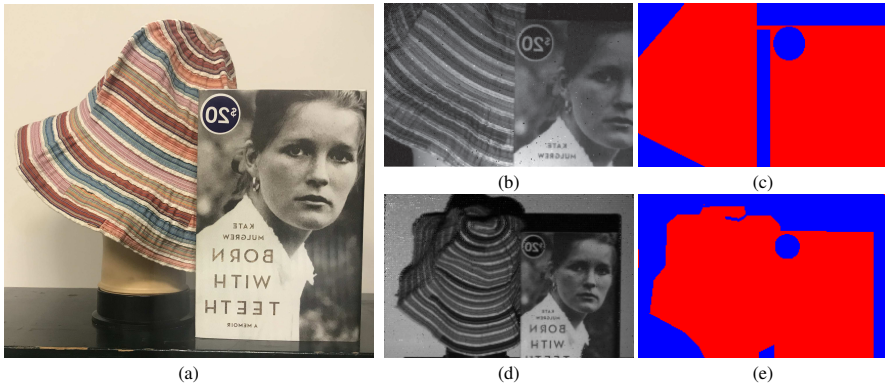
Fig. E.5: *Objects used for ground-truth evaluation.* **(a)** Color view of the objects. **(b)** C2B view of the scene for photometric stereo. **(c)** The binary foreground mask (red is foreground). The mask excludes pixels in shadow or pixels that exhibit glossy reflection. **(d)** C2B view of the scene for structured-light triangulation. **(e)** Binary foreground mask used for structured-light error calculations (red is foreground), non-object pixels are masked.

demosaicing algorithm	input images	ID-DM (degrees)	BRD-R (degrees)
MATLAB demosaic()	one-shot	*12.079*	*11.883*
	simulated one-shot	10.539	10.140
OPENCV demosaicing (BayerBG2RGB_EA)	one-shot	*12.224*	*12.322*
	simulated one-shot	10.057	9.703
Demosaic Net [9]	one-shot	***11.675***	*11.936*
	simulated one-shot	10.215	9.758

Table E.2: *Angular RMSE for photometric stereo.*

discrepancy in the best reconstruction method, we note that the errors are close between ID-DM and BRD-R, which is consistent with our simulations (Fig. 6(c) and 6(d) from Sec. 5). This suggests that even ID-DM is able to handle the albedo variations present in the scene. One thing to note is that the demosaicing algorithms are designed for image intensities and not illumination ratios so it is possible for a demosaicing algorithm more optimized for illumination ratios to improve the performance of the BRD-R method. In terms of overall magnitude, the differences between one-shot and multi-shot are within the range of previously-reported ground-truth evaluations of previous one-shot photometric stereo algorithms [10,11]. Comparing to the results obtained from simulated one-shot images it is clear that most of the reconstruction error is indeed due to demosaicing rather than our sensor's performance under C2B imaging conditions.

Structured-light triangulation results. Tables E.4 compares the percentage of bad-matching pixels for structured-light triangulation. Here the BRD-R method yields the best performance with OpenCV's demosaic function giving the lowest bad-pixel percentage. There is a 10-20% gap between the bad-pixel percentage of simulated one-shot and actual one-shot mosaics. This discrepancy is likely due to the fact that disparity errors in the actual one-shot results generally decrease for the simulated one-shot results. Since all disparity errors greater than 1 are treated equally in the bad-pixel percentage metric, this gap can be explained by disparity errors close to, but greater than 1 becom-

demosaicing algorithm	input images	ID-DM (degrees)	BRD-R (degrees)
MATLAB demosaic()	one-shot	*5.917*	***4.901***
	simulated one-shot	4.520	4.096
OPENCV demosaicing (BayerBG2RGB_EA)	one-shot	*5.453*	*4.980*
	simulated one-shot	3.947	3.745
Demosaic Net [9]	one-shot	*5.357*	*5.375*
	simulated one-shot	3.993	3.990

Table E.3: *Median angular error for photometric stereo.*

demosaicing algorithm	input images	ID-DM (bad%)	BRD-R (bad%)
MATLAB demosaic()	one-shot	*61.30*	*48.23*
	simulated one-shot	50.65	30.27
OPENCV demosaicing (BayerBG2RGB_EA)	one-shot	*56.62*	***45.29***
	simulated one-shot	48.34	30.84
Demosaic Net [9]	one-shot	*55.6*	*47.16*
	simulated one-shot	46.51	33.94

Table E.4: *Disparity bad-pixel percentage for structured-light triangulation.*

ing lower for simulated one-shots. Since BRD-R methods has lower RMSE (Fig. 7 from Sec. 6) than ID-DM methods, this also explains the gap there.

F Live 3D Acquisition Experiments

F.1 Photometric stereo experiments

Person-PS. Errors in the hair result from the appearance of hair violating the Lambertian image formation model. Sectioned colors in the background result from shadows. Since we are using 4 subframes and 3 pixels, we are not robust to shadows and occlusions in the scene. The different expressions made by the subject show the ability of the system to capture surface normals at different orientations. Notice that the estimated normals in areas in the line of sight of all LEDs are not severely affected by neighbouring shadowed pixels. This is because our technique does not rely on the integrability of the surface in order to estimate normals. Our normal estimates are stable over time and do not exhibit any clear temporal noise. We try a variety of hand gestures to test the robustness of our system.

Hat-PS. Since we use ID-DM, the normal estimates should not be robust to high reflectance frequencies. We observe this in the form of temporal jitter in the video, where we can see various artefacts that appear on the hat as it is being rotated and moved. However, even though there are artefacts, the estimated normals overall look reasonable. We also place a hand in the scene for comparison. Since the hand is textureless, there are little, if any, artefacts in the estimated normals of the hand. The hand crushes the hat and shows that our system is able to handle non-rigid deformations of objects in the scene. It is important to point out that colorful and textured objects are not apt for traditional one-shot methods so this is a hard object from which to estimate normals.

Dress-PS. The black edges in the normal map are caused by black edges in the original image and these are not artefacts from our algorithm. We see that the normal estimates are consistent during each period of the swing.

F.2   Structured-light triangulation experiments

Person-SL. There are pixels with poor performance, and they appear as temporal noise in the disparity map. These erroneous disparity is strictly due to sensor imperfections and not the algorithms proposed. Because we are using a single projector, we are not able to estimate the disparity of the pixels in the areas that are shadowed by the hand. We show the subject making different hand gestures. These hand gestures introduce discontinuities in the disparity map between the face and the hand as well as between different fingers. Even under these different deformations and discontinuities, the disparity map is quite stable.

Handkerchief-SL. We can see that while the object has texture, the resulting disparity map are largely invariant to the albedo. Due to motion in the scene however, they are not completely invariant and this is apparent as the edges of the textures can be traced out in the scene. We also place the handkerchief under different deformations, showing that albedo-invariance holds in many cases.

Toy Cloud-SL. Since the cloud is colorful, this is an object that is traditionally difficult for color-based one shot techniques. We can see that the disparity in our case is still smooth and the quality of the disparity map is not affected by the different colors in the toy cloud. We do note that right side of the cloud has a color that is very dark so the disparity map is naturally more noisy for that region; this is not a problem with our algorithm.

References

1. Wan, G., Li, X., Agranov, G., Levoy, M., Horowitz, M.: CMOS Image Sensors With Multi-Bucket Pixels for Computational Photography. IEEE J. Solid-State Circuits **47**(4) (2012) 1031–1042 2

2. Stoppa, D., Massari, N., Pancheri, L., Malfatti, M., Perenzoni, M., Gonzo, L.: A range image sensor based on 10-$mu hbox m$ lock-in pixels in 0.18-mum cmos imaging technology. IEEE Journal of Solid-State Circuits **46**(1) (2011) 248–258 2

3. Sarhangnejad, N., Lee, H., Katic, N., O'Toole, M., Kutulakos, K.N., Genov, R.: CMOS Image Sensor Architecture for Primal-Dual Coding. In: Int. Image Sensor Workshop, Hiroshima, International Image Sensor Society (2017) 1, 2

4. Geary, R.C.: The frequency distribution of the quotient of two normal variates. Journal of the Royal Statistical Society **93**(3) (1930) 442–446 4

5. Oliveira, A., Oliveira, T., Macías, S., Antonio: Distribution function for the ratio of two normal random variables. In: AIP Conference Proceedings. Volume 1648. (2015) 4

6. Ratner, N., Schechner, Y.Y., Goldberg, F.: Optimal multiplexed sensing: bounds, conditions and a graph theory link. Opt Express **15**(25) (2007) 17072–17092 6

7. Cossairt, O., Gupta, M., Nayar, S.K.: When does computational imaging improve performance? IEEE Transactions on Image Processing (2013) 16

8. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Computer Vision **47**(1) (2002) 7–42 19

9. Gharbi, M., Chaurasia, G., Paris, S., Durand, F.: Deep joint demosaicking and denoising. In: Proc. ACM SIGGRAPH Asia. (2016) 19, 20, 21

10. Fyffe, G., Yu, X., Debevec, P.E.: Single-shot photometric stereo by spectral multiplexing. In: ICCP. (2011) 20

11. Chakrabarti, A., Sunkavalli, K.: Single-image rgb photometric stereo with spatially-varying albedo. In: International Conference on 3D Vision (3DV). (2016) 20