Supplementary Document: Multiplexed Illumination for Scene Recovery in the Presence of Global Illumination

Jinwei Gu[1], Toshihiro Kobayashi[2], Mohit Gupta[3], Shree K. Nayar[3]
[1]Rochester Institute of Technology, [2]Canon Inc., [3]Columbia University
jwgu@cis.rit.edu, kobayashi.toshihiro558@canon.co.jp, {mohitg, nayar}@cs.columbia.edu

This supplementary document includes the following sections. Section 1 describes another multiplexed illumination method using a multiplexing scheme similar to that in Lemma 4.1 in the paper. The method also extracts N direct illuminations from $2N + 1$ images. Section 2 shows detailed derivation of the SNR analysis results. Section 3 discusses about the pros and cons of direct/global separation using checkerboard and sinusoid patterns. Please refer to the project website (www.cis.rit.edu/jwgu/research/demux) for more direct/global separation results.

1. Sinusoid Sequential Multiplexing

In addition to the proposed frequency modulated multiplexing method (Section 4.2 in the paper), we also designed another multiplexed method for direct/global separation for N light sources. This method uses a multiplexing scheme similar to the idea of the theoretical lower bound in Section 4.1 in the paper, except that it uses high frequency sinusoidal patterns instead of checkerboard patterns since read light sources cannot produce perfect step edges.

Again, at first, we turn on all the N light sources at half brightness and capture an image I_0:

$$I_0(x) = \sum_{i=1}^{N} \left(L_d^{(i)(x)} + L_g^{(i)(x)} \right) /2,$$

Next, instead of setting the i-th light source to be a checkerboard pattern, we modulate the i-th light source with a high frequency sinusoid pattern, while keeping all the other $N-1$ light sources at half brightness. Suppose the modulation sinusoidal pattern is $(1 + \sin(\phi(x)))/2$, where $\phi(x)$ is the phase of a scene point with respect to the light source. The captured image is

$$\begin{aligned} I_i^{(1)}(x) &= \left(\frac{L_d^{(i)}(x)}{2} \sin(\phi(x)) + \frac{L_d^{(i)}(x) + L_g^{(i)}(x)}{2} \right) \\ &\quad + \sum_{j=1, j\neq i}^{N} \frac{L_d^{(j)}(x) + L_g^{(j)}(x)}{2} \\ &= \frac{L_d^{(i)}(x)}{2} \sin(\phi(x)) + I_0 = \alpha_i(x) + I_0(x), \end{aligned}$$

where $\alpha_i(x) = L_d^{(i)}(x)/2 \sin(\phi(x))$. If we shift the sinusoidal pattern by $3\pi/2$ in phase and modulate the i-th light, and capture a second image, we have

$$I_i^{(2)}(x) = \frac{L_d^{(i)}(x)}{2} \sin(\phi(x) + \frac{3\pi}{2}) + I_0 = \beta_i(x) + I_0(x), \tag{1}$$

where $\beta_i(x) = L_d^{(i)}(x)/2 \cos(\phi(x))$. Since we modulate each light twice, together with I_0, we need to capture $2N + 1$ images. To make the analysis consistent with the frequency modulated multiplexing, we add an unknown $\gamma = I_0$, and thus these $2N + 1$ equations can be written in the matrix form $\mathbf{S} \cdot \mathbf{x} = \mathbf{b}$, where the unknowns are $\mathbf{x} := [\alpha_1, \beta_1, \cdots, \alpha_N, \beta_N, \gamma]^T$, the captured images are $\mathbf{b} = [I_0, I_1^{(1)}, I_1^{(2)}, \cdots, I_N^{(1)}, I_N^{(2)}]^T$, and the matrix $\mathbf{S}$ of size $(2N + 1) \times (2N + 1)$ is

$$\mathbf{S} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 0 & 1 \\ & & \cdots & & \\ 0 & 0 & \cdots & 1 & 1 \end{pmatrix}. \tag{2}$$

We note that $\det \mathbf{S} = (-1)^{N+1}$, and thus it is nonsingular. In theory, this sequential multiplex method can be used to separate the direct components. We call this method *sinusoid sequential multiplexing*. It is easy to implement — only a single light source is modulated with sinusoidal patterns at any time. However, both theoretical analysis and experiments show this method will amplify noise (see below).

The condition number of matrix $\mathbf{S}$ is close to $N+2$ that grows linearly with respect to the number of light sources. This indicates its robustness decreases as more and more lights are involved.

2. Signal-to-Noise Ratio (SNR) Analysis

We perform SNR analysis in terms of camera noise. In our problem, we are interested in the N direct illumination components, $L_d^{(i)}, i = 1, \cdots, N$. For simplicity, we assume the N light sources are of the same brightness and the scene is a flat surface with a uniform albedo. The MSE (Mean Square Error) of the N direct components is (see appendix for derivation)

$$\overline{\Delta L_d^2} = \frac{1}{N}\sum_{i=1}^{N}\overline{\left(\Delta L_d^{(i)}\right)^2} = \frac{2}{N}\sum_{i=1}^{N}\left(\Delta\alpha_i^2 + \Delta\beta_i^2\right)$$
$$= \frac{2}{N}\left(\text{Trace}\left(\left(\mathbf{M}^T\mathbf{M}\right)^{-1}\right) - \left(\mathbf{M}^T\mathbf{M}\right)^{-1}_{2N+1,2N+1}\right)\sigma^2. \quad (3)$$

where σ is the noise level in each of the captured images, $\mathbf{M}$ is the mixing matrix (for sinusoid sequential multiplexing, $\mathbf{M} = \mathbf{S}$; for frequency modulated multiplexing, $\mathbf{M} = \mathbf{F}$.) Compared to the sequential separation method (*i.e.*, non-multiplexing, $N = 1$), the SNR gain, G, is

$$G = \frac{\text{SNR}_{\text{multiplexed}}}{\text{SNR}_{\text{sequential}}} = \frac{\overline{\Delta L_d}_{\text{sequential}}}{\overline{\Delta L_d}_{\text{multiplexed}}}$$
$$= \sqrt{N \cdot \frac{\sigma_0^2}{\sigma^2} \cdot \frac{\text{Trace}\left(\left(\mathbf{M}_0^T\mathbf{M}_0\right)^{-1}\right) - \left(\mathbf{M}_0^T\mathbf{M}_0\right)^{-1}_{3,3}}{\text{Trace}\left(\left(\mathbf{M}^T\mathbf{M}\right)^{-1}\right) - \left(\mathbf{M}^T\mathbf{M}\right)^{-1}_{2N+1,2N+1}}} \quad (4)$$

where σ_0 is the noise level for the captured image under a single light source, and $\mathbf{M}_0$ is a 3×3 mixing matrix for a single light source, corresponding the matrix $\mathbf{F}$ when $N = 1$. For the matrix $\mathbf{F}$, we recall that $\mathbf{F} = \left[\mathbf{c}_1, \mathbf{s}_1, \cdots, \mathbf{c}_N, \mathbf{s}_N, \frac{1}{\sqrt{2}}\mathbf{1}\right]$, and we have

$$\mathbf{F}^T\mathbf{F} = \text{diag}\left(\frac{2N+1}{2}, \cdots, \frac{2N+1}{2}, \frac{2N+1}{2}\right), \quad (5)$$

and thus

$$\text{Trace}\left(\left(\mathbf{F}^T\mathbf{F}\right)^{-1}\right) - \left(\mathbf{F}^T\mathbf{F}\right)^{-1}_{2N+1,2N+1} = \frac{4N}{2N+1}. \quad (6)$$

For the matrix $\mathbf{S}$ (defined in Equation (2)), we have

$$\mathbf{S}^T\mathbf{S} = \begin{pmatrix} \mathbf{I}_{2N\times 2N} & \mathbf{1}_{2N\times 1} \\ \mathbf{1}^T_{2N\times 1} & 2N+1 \end{pmatrix} \quad (7)$$

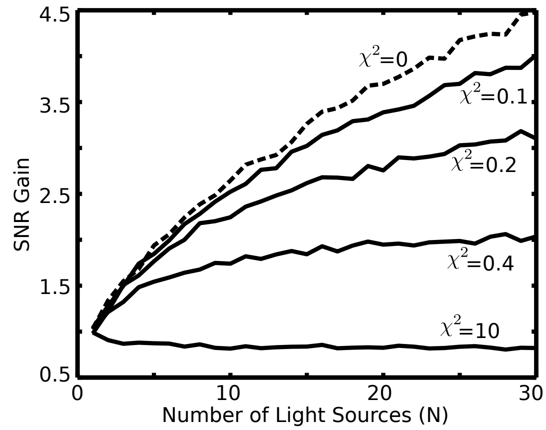

Figure 1. Simulation results of the SNR gain for the frequency modulated multiplexing methods under different noise characteristics. $\chi^2 = \sigma_p/\sigma_r$ describe the relative weight between photon noise and read noise.

and

$$(\mathbf{S}^T\mathbf{S})^{-1} = \begin{pmatrix} \mathbf{I}_{2N\times 2N} + \mathbf{1}_{2N\times 2N} & -\mathbf{1}_{2N\times 1} \\ -\mathbf{1}^T_{2N\times 1} & 1 \end{pmatrix}, \quad (8)$$

and thus

$$\text{Trace}\left((\mathbf{S}^T\mathbf{S})^{-1}\right) - (\mathbf{S}^T\mathbf{S})^{-1}_{2N+1,2N+1} = 4N. \quad (9)$$

Therefore, Equation (4) can be simplified as

$$G_{\mathbf{F}} = \frac{\sigma_0}{\sigma} \cdot \sqrt{\frac{2N+1}{3}}, \qquad G_{\mathbf{S}} = \frac{\sigma_0}{\sigma} \cdot \sqrt{\frac{1}{3}}, \quad (10)$$

where $G_{\mathbf{F}}$ and $G_{\mathbf{S}}$ are the SNR gain for the frequency modulated methods and the sinusoid sequential method, respectively. As shown, for the sequential multiplexing method, the SNR gain $G_{\mathbf{S}} < 1$ (since $\sigma_0 \leq \sigma$), and thus it is not recommended in practice. Below we analyze $G_{\mathbf{F}}$ in details.

A typical imaging system contains three noise sources: photon noise, read noise, and dark noise. For a single light source, suppose the photon noise is σ_p^2, the read noise is σ_r^2, and the dark noise is σ_d^2. Thus, the total noise should be $\sigma_0^2 = \sigma_p^2 + \sigma_r^2 + \sigma_d^2$. For the multiplexed illumination case, since there are N lights on, the photon noise should be $N\sigma_p^2$. Assuming the exposure time and ISO setting remain the same, the dark noise and read noise keep unchanged. Thus, the total noise is $\sigma^2 = N \cdot \sigma_p^2 + \sigma_r^2 + \sigma_d^2$. For typical DSLR cameras, the dark noise is relatively small compared to photon noise and read noise, and can be effectively reduced via cooling.

For the frequency modulated multiplexing, if the imaging system is read noise limited (*e.g.*, in low light), we have $\sigma^2 \approx \sigma_0^2$, and thus $G_{\mathbf{F}} = \sqrt{(2N+1)/3}$. If it is photon noise limited (*e.g.*, for long exposures), we have $\sigma^2 \approx N\sigma_0^2$, and thus $G_{\mathbf{F}} = \sqrt{\frac{2N+1}{3N}} \approx \sqrt{2/3}$, which means this is

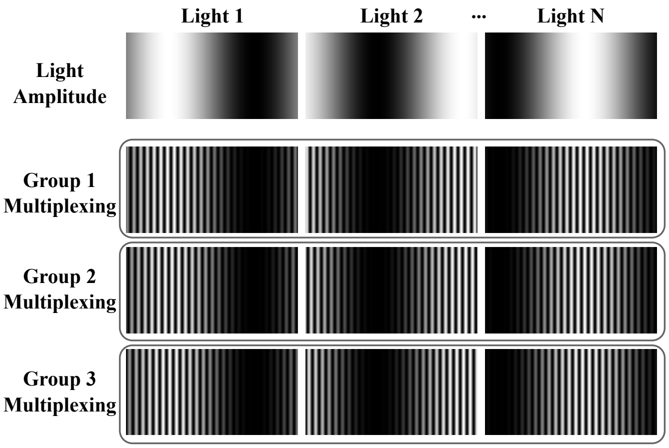

Figure 2. **Using conventional light multiplexing [2] for direct/global separation for multiple light sources.** For N light sources, the sequential separation method [1] will need $3N$ images. For each light sources, we modulate it with three shifted sinusoidal patterns (shown as columns in the above figure). One can perform conventional light multiplexing using Hadamard codes by first grouping these $3N$ images into three groups (shown as rows) and perform multiplexing for each group. This method will still need $3N$ images, but it will have SNR benefits for dim scenes.

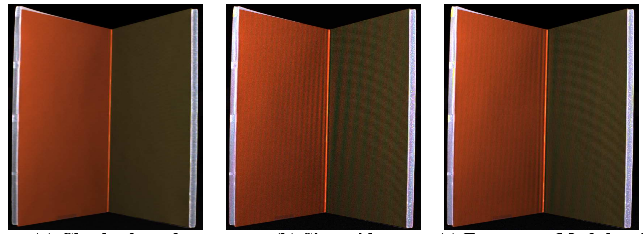

Figure 3. One of the direct components of a V-groove scene under $N = 3$ light sources. (a) is the result of the sequential separation using a checkerboard pattern [1], (b) is the result of the sequential separation using sinusoid patterns, and (c) is the result of the proposed frequency modulated multiplexing. As shown, the checkerboard method has the highest quality but requires much more images. The proposed method has better quality than the sequential separation using sinusoid patterns, while requiring fewer images ($2N + 1$ versus $3N$).

no benefit for multiplexing in terms of SNR. $G_{\mathbf{F}}$ depends on the noise characteristics (*i.e.*, the relative weight of read noise and photon noise) of the imaging system. This conclusion is similar to that of conventional light multiplexing without direct/global separation [3]. We define $\chi^2 = \sigma_p/\sigma_r$ as a measure for noise characteristics. Figure 1 shows the SNR gain versus the number of light sources for different χ^2 values. The curves are computed via simulation. As shown, when read noise dominates (*i.e.*, small χ^2 values), the SNR gain is proportional to $\sqrt{N}$. As photon noise increases (*i.e.*, large χ^2 values), the SNR gain reduces and finally fix on a constant value $\sqrt{2/3}$. The simulated results match well with our theoretical analysis.

In addition to the sequential separation method which turns on a single light source at a time, one can employ the conventional Hadamard code based multiplexing scheme [2], as explained in Fig. 2. While it still needs $3N$ images, for dim objects and read noise limited situation, it increases SNR by $\sqrt{\frac{(N+1)^2}{4N}}$. Compared to this Hadamard based method, the proposed frequency modulated multiplexing method has a SNR gain of $\sqrt{\frac{4N(2N+1)}{3(N+1)^2}} \approx \sqrt{8/3} \approx 1.6$. Overall, our proposed method has higher SNR, while requiring fewer images ($2N + 1$ vs. $3N$).

Bright Objects: So far, we assume the exposure time and ISO setting are common. For bright objects, to avoid saturation, we need to vary these settings. We assume the ISO setting is fixed and discuss the effect of the exposure time to the SNR gain (similar conclusion can be derived for ISO settings). Suppose L_d is the typical readout for exposure time t_0 for single source case. Suppose for multiplexed illumination, the exposure time is t, and thus the direct component will be $L_d t/t_0$, and thus the SNR gain $G_{\mathbf{F}}$ is

$$G_{\mathbf{F}} = \frac{\sigma_0}{\sigma(t)} \cdot \frac{t}{t_0} \cdot \sqrt{\frac{2N+1}{3}}, \tag{11}$$

where $\sigma(t)^2 = Nt\sigma_p^2/t_0 + \sigma_r^2 + t\sigma_d^2/t_0$. Note that both photon noise and dark noise increase with respect to exposure time. Suppose the saturation threshold for the imaging system is L_{max}, we have

$$N \cdot L_d t/t_0 \leq L_{max} \tag{12}$$

In this case, the SNR gain $G_{\mathbf{F}}$ (in either the photon noise limited or read noise limited case) is given by

$$G_{\mathbf{F}} = \frac{L_{max}}{L_d}\sqrt{\frac{2N+1}{3N^2}} \tag{13}$$

Thus, multiplexing will actually reduce SNR (since we have to shorten exposure time to prevent saturation, which undermines SNR).

3. Checkerboard vs. Sinusoids vs. Frequency Modulated Multiplexing

Nayar et al. [1] proposed two practical methods for direct/global separation for a single light source. The first method is to use a checkerboard pattern, which is shifted multiple times in both x and y directions. To overcome color bleeding, optical defocus, screen door, and other problems in projectors, usually we need to shift the pattern 25 times or more [1].

The second method is to use three sinusoid patterns with different phases and solve a 3×3 linear system to compute the direct and global components, as reviewed in Section 3 in the our paper. While it requires only three images, because of the quantization error and image noise for the projected sinusoid patterns, it is much easier to have artifacts presented in the separated components. Figure 3 shows an example of a V-groove under three light sources. Figure 3(a) shows one of the direct components using the checkerboard methods (with 25 images per light source). Figure 3(b) shows the direct component using three sinusoid patterns, which clearly shows the vertical stripe artifacts. Therefore, in our experiments, to fairly evaluate the performance of the proposed method, we used the checkerboard method (with 25 images per light) to obtain high-quality direct/global components as ground truth for comparison. The minimum number of images required for a satisfactory direct/global separation will vary with both the imaging system settings and the scene.

Figure 3(c) shows another interesting fact. It shows the separated direct component using the proposed method. In total, our method requires $2 \times 3 + 1 = 7$ images. The result is comparable to that of the checkerboard (Fig. 3(b)), and is better than that of the sequential sinusoid method (Fig. 3(b)). In other words, we can obtain better results with few images. This is expected, since as shown in Section 2, the SNR gain is $\sqrt{(2 \times 3 + 1)/3} = 1.63$.

References

[1] 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. *ACM Transactions on Graphics (SIGGRAPH)*, 25(3):935–944, 2006. 3

[2] Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur. A theory of multiplexed illumination. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, Nice, France, October 2003. 3

[3] Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur. Multiplexing for optimal lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:1339–1354, 2007. ISSN 0162-8828. doi: http://doi.ieeecomputersociety.org/10.1109/TPAMI.2007.1151. 3