Frequency-Space Decomposition and Acquisition of Light Transport under Spatially Varying Illumination

Dikpal Reddy[1], Ravi Ramamoorthi[1], and Brian Curless[2]

[1] University of California, Berkeley
[2] University of Washington

Abstract. We show that, under spatially varying illumination, the light transport of diffuse scenes can be decomposed into direct, near-range (subsurface scattering and local inter-reflections) and far-range transports (diffuse inter-reflections). We show that these three component transports are redundant either in the spatial or the frequency domain and can be separated using appropriate illumination patterns. We propose a novel, efficient method to sequentially separate and acquire the component transports. First, we acquire the direct transport by extending the direct-global separation technique from floodlit images to full transport matrices. Next, we separate and acquire the near-range transport by illuminating patterns sampled uniformly in the frequency domain. Finally, we acquire the far-range transport by illuminating low-frequency patterns. We show that theoretically, our acquisition method achieves the lower bound our model places on the required number of patterns. We quantify the savings in number of patterns over the brute force approach. We validate our observations and acquisition method with rendered and real examples throughout.

1 Introduction

Image-based acquisition of light transport in real scenes poses significant challenges despite the parallelism afforded by cameras and projectors. Efficient acquisition of the light transport in a projector-camera system is particularly difficult since the illumination is spatially varying at the scene points and the transport does not lend itself to simple mathematical models (Figure 1). Furthermore, the acquired light transport simply gives a relationship between the light source and camera and does not provide insight into different physical phenomena such as inter-reflections and subsurface scattering.

In this paper, we propose a technique to decompose and efficiently acquire the light transport under spatially varying illumination. We show that the light transport can be decomposed into three physically meaningful component transports: direct, near-range (sub-surface scattering and local inter-reflections) and far-range (diffuse inter-reflections). Since the three component transports differ in their response to different frequencies of sinusoidal illumination patterns,

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part VI, LNCS 7577, pp. 596–610, 2012.

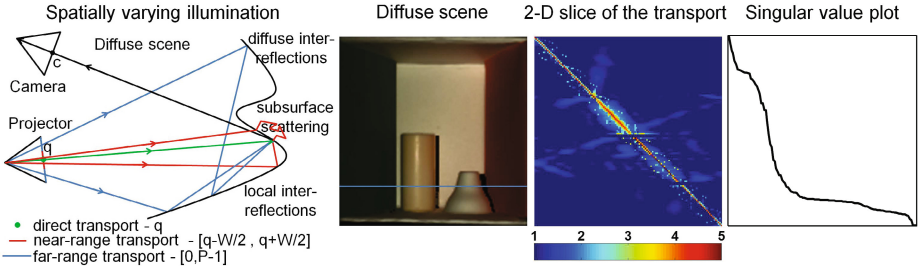

Fig. 1. Spatially varying illumination at the scene using a projector-camera system. Camera pixel c is lit directly by projector pixel q. The near-range effects (subsurface scattering and local inter-reflections) at c are caused by a neighborhood of size W around q. The far-range effects (diffuse inter-reflections) at c are potentially due to all projector pixels $[0, P-1]$. We implement a non-coaxial projector-camera system and acquire the transport of a diffuse scene. A 2-D slice of the transport matrix in log-scale at the blue row shows a diagonally dominant, dense matrix with slowly decaying singular values.

we show that they can be separated through appropriate choice of these patterns. We also show that compared to the full transport, each of the component transports can be modeled parsimoniously, enabling their efficient acquisition. In Figure 2 we show an example decomposition and acquisition of the full transport with a projector-camera system.

We build on the observation that at a camera pixel, the three components of the light transport are influenced by an increasing range of projector pixels as illustrated in Figure 1. The direct component at a camera pixel is due to a single projector pixel which is localized in the projector's spatial domain. The light received at the same camera pixel due to subsurface scattering and local inter-reflections is due to projector pixels in the small neighborhood of the direct projector pixel. Finally, the diffuse inter-reflection component at a camera pixel is due to projector pixels spread over a wider area. Hence, we term the component effects direct, near-range and far-range.

Since the component effects at a camera pixel have increasing support in the projector's spatial domain, the opposite holds true in the projector's frequency domain. The far-range transport has been observed to be a low-frequency phenomenon in previous work [1],[2]. Near-range transport responds to a wider range of frequencies but does not respond to the highest frequency illumination [3]. The direct transport responds to the entire frequency spectrum of illuminations. But these observations have not been used for light transport acquisition. We validate these observations in Figure 3(4) through frequency analysis of the transport matrix of a rendered scene. We also observe that at a camera pixel, the component transports are redundant either in the projector's spatial or frequency domain. The direct and near-range transports have small support in the spatial domain and the far-range transport has small bandwidth in the frequency domain.

Based on the above observations, we choose illumination frequencies which allow us to decompose the transport into its components and acquire them

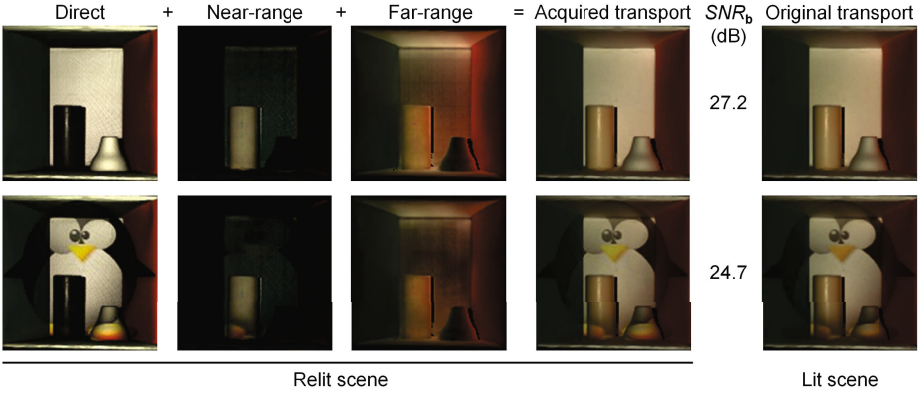

Fig. 2. 16384 × 16384 transport matrix of a real Cornell box like scene is acquired using 1060 illumination patterns, a 15.4× saving. We acquire the full transport by decomposing it and acquiring the component transport matrices: direct, near-range and far-range. The acquired component and full transports are visualized by relighting with all-white floodlight (top row) and a penguin image (bottom row) and compare well to the images of the scene lit with the same illumination (last column). Note that the direct lighting and inter-reflections (far-range) are well separated and the near-range transport captures the subsurface effects of the candle.

individually with only a few patterns as illustrated in Figure 3(5). We propose a method where using highest frequency sinusoidal illumination, we first separate the direct transport and estimate it in the spatial domain. Then we use sinusoidal patterns sampled uniformly in the frequency domain to separate the near-range transport and estimate it in the spatial domain as well. Finally, we use a small number of low-frequency sinusoidal patterns and estimate the far-range transport in the frequency domain.

Our key insight is that sampling the near-range transport in frequency domain as opposed to spatial domain enables almost perfect separation of near and far-range effects. Hence, theoretically our method can acquire the light transport and its components using as few patterns as permitted by our model.

Contributions:

1. We show that the light transport under spatially varying illumination can be decomposed into physically meaningful component transports.
2. We propose a method to separate and efficiently acquire the component transports based on our observation that each component has varying bandwidth in the projector's frequency domain.
3. We show that by sampling the near-range component in frequency domain, our method acquires the full transport using minimal number of patterns.

2 Related Work

Direct-Global Separation: Nayar et al. [4] proposed a method using high-frequency illumination to separate the direct from the indirect component under flood light. But [4] does not apply to other light patterns or the full transport. Our work is inspired by [4] but it differs in two significant ways. First, we acquire the full transport via decomposition, enabling separation under any light. This implies selective component transports in a scene can be edited and enhanced under any lighting. Second, we decompose the global transport further into mathematically parsimonious near and far-range, making the acquisition efficient. Gu et al. [5] extended the direct-global separation in [4] to multiple light sources achieving savings in the number of necessary illumination patterns.

Light Transport Acquisition: Previous methods on efficient light transport acquisition focused on distant lighting case whereas we focus on spatially varying illumination. Different mathematical models of redundancy in the transport matrix were proposed to acquire it efficiently. Fuchs et al.[6] adaptively chose the lighting direction based on the properties of the reflectance fields. Both Wang et al. [7] and O'Toole and Kutulakos [8] assumed that the transport matrix is low rank. In compressive light transport (CLT) acquisition [9], the rows of the matrix are assumed to be sparse in the wavelet basis.

Spatially Varying Illumination: Previous methods on light transport acquisition under spatially varying illumination either used brute force or used simple models which ignored diffuse inter-reflections. Masselus et al. [10], acquired the 6-D reflectance field through brute force by capturing 2-D images with a 4-D incident light field. Sen et al. [11] exploited Helmholtz reciprocity and parallelized the acquisition by assuming only near-range effects. Garg et al. [12] captured the entire 8-D transport by adaptively projecting hierarchical patterns designed to estimate diffuse components modeled as rank-1 off-diagonal submatrices. Our approach treats the near-range effects as mathematically different from the far-range effects and achieves efficiency in acquisition.

Separation of Component Transports: Previous work on light transport separation focused on extracting the multiple bounces of light [13] and formulated the light transport inversion as canceling of these bounces [14]. Recent methods have focused on separating components of the global transport. Wu et al. [15] use expensive time-of-flight imaging to decompose the transport into direct, subsurface scattering and inter-reflections. O'Toole et al. [16] proposed a novel matrix probing technique to highlight selective components of the transport in a single image. We focus not just on separation of the component transports but also their efficient acquisition. Also, our method uses a regular non-coaxial projector-camera system with sinusoidal illumination patterns.

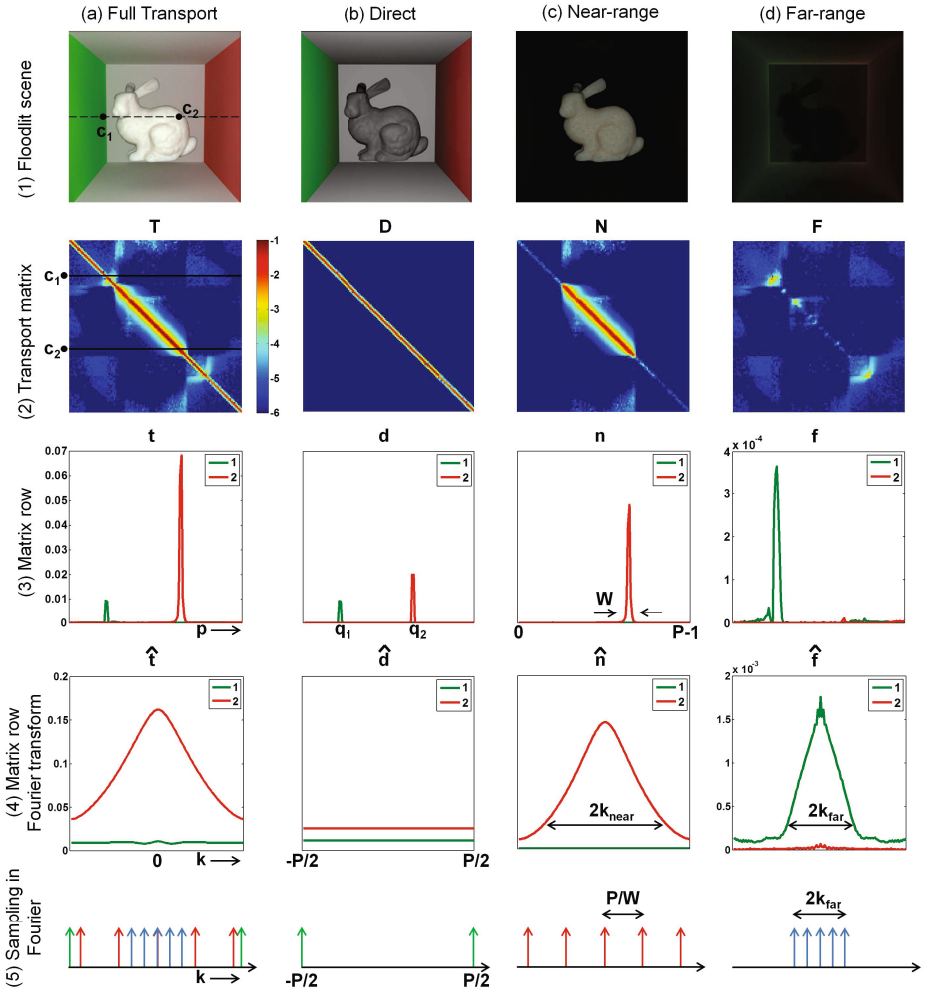

Fig. 3. (1) We render the full, direct, near-range and far-range transports with a coaxial projector-camera setup and include only subsurface scattering in the near-range. (2) 128×128 slice of the transport matrix is generated along the dashed line. The component matrices have diagonal, banded-diagonal and small magnitude off-diagonal structure. (3) Compare the rows of the transport matrices at camera pixels c_1 and c_2 lit directly by projector pixels q_1 and q_2. At all camera pixels c, a window of size W around q describes the near-range. (4) Compare Fourier transforms of the two rows of the transport matrix. The bandwidth of direct, near and far-range obeys $2\frac{P}{2} > 2k_{near} > 2k_{far}$. (5) We acquire the matrix with sinusoidal illumination patterns. Direct is separated and acquired with illumination frequencies $|k| = \frac{P}{2}$, near-range with frequencies spaced uniformly at $\frac{P}{W}$ and far-range with frequencies from its entire bandwidth.

3 Light Transport Modeling

Let the $C \times P$ matrix $\mathbf{T}$ denote the light transport. $\mathbf{T}(c, p)$ is the transport between the projector pixel p and the camera pixel c. In practice, $C > P$ but for analysis we assume square $\mathbf{T}$ with one-one correspondence between the C camera pixels and P projector pixels i.e. $C = P$. We present our idea on a 1-D projector and camera and extension to 2-D is straightforward. Under projector lighting $\mathbf{l}_k$, the camera image $\mathbf{b}_k$ is given by

$$\mathbf{b}_k(c) = \sum_p \mathbf{T}(c, p)\mathbf{l}_k(p). \tag{1}$$

We use discrete Fourier transform (DFT) basis illumination $\mathbf{l}_k(p) = e^{-j2\pi\frac{k}{P}p}$ of frequency $\frac{k}{P}$ cycles/pixel where $k \in \{-\frac{P}{2}, \ldots, -1, 0, 1, \ldots, \frac{P}{2}\}$. Under novel illumination $\mathbf{l} = \sum_k \alpha_k \mathbf{l}_k$, the image is given by $\mathbf{b} = \sum_k \alpha_k \mathbf{b}_k$. In this paper, we treat each camera pixel independently. Hence, we drop index c and write (1) as

$$\begin{aligned} b_k &= \textstyle\sum_p \mathbf{t}(p)\mathbf{l}_k(p) \\ &= \mathbf{t} \cdot \mathbf{l}_k \end{aligned} \tag{2}$$

where $b_k = \mathbf{b}_k(c)$ and $\mathbf{t}(p) = \mathbf{T}(c, p)$.

Transport matrix $\mathbf{T}$ can be reconstructed row-wise by reconstructing $\mathbf{t}$ at each camera pixel from measurements b_k as

$$\mathbf{t}(p) = \frac{1}{P} \sum_k b_k \mathbf{l}_k^*(p) \tag{3}$$

where $\mathbf{l}_k^*(p) = e^{j2\pi\frac{k}{P}p}$ is the inverse DFT basis. In general, we need all P measurements $\{b_k : k = -\frac{P}{2}, \ldots, -1, 0, 1, \ldots, \frac{P}{2}\}$ to reconstruct $\mathbf{t}$. The same P patterns are sufficient to acquire the matrix $\mathbf{T}$ as well since the camera pixels are acquired in parallel and processed independently. We show that, under spatially varying illumination, $\mathbf{T}$ can be decomposed into physically meaningful components, each of which can be acquired row-wise using many fewer illumination patterns.

3.1 Light Transport Decomposition

Under spatially varying illumination, the transport matrix $\mathbf{T}$ exhibits full-rank. The matrix has a dominant diagonal corresponding to direct illumination and significant energy in off-diagonal elements corresponding to global illumination as shown in Figure 3. The matrix is neither sparse nor low rank making its analysis and acquisition difficult. But $\mathbf{T}$ can be decomposed into component transport matrices $\mathbf{D}$, $\mathbf{N}$ and $\mathbf{F}$ corresponding to direct, near-range and far-range effects. We validate this observation by rendering a synthetic scene with subsurface scattering and diffuse inter-reflections as shown in Figure 3.

$$\begin{aligned} \mathbf{T} &= \mathbf{D} + \mathbf{N} + \mathbf{F}. \\ \mathbf{t} &= \mathbf{d} + \mathbf{n} + \mathbf{f}. \end{aligned} \tag{4}$$

Accordingly, we decompose the camera pixel intensity b_k into component intensities b_k^d, b_k^n and b_k^f respectively. The direct transport $\mathbf{D}$ is diagonal and full-rank. The near-range transport $\mathbf{N}$ is sparse, banded diagonal and high rank. Conventionally this is interpreted column-wise as the response in a camera region to a projector pixel's illumination. We interpret this row-wise as the influence of a projector region on a single camera pixel. The far-range transport matrix $\mathbf{F}$ exhibits a dense structure but also has rapidly decaying singular values consistent with the previous observations that $\mathbf{F}$ is low rank [1],[2]. In comparison to other components, $\mathbf{F}$ has small magnitude entries which cumulatively contribute significant energy to the final image.

3.2 Spatial Domain Modeling

The support of row $\mathbf{t}$ indicates the projector pixels illuminating the camera pixel either directly or globally (Figure 3(3)). In the direct component, a single projector pixel q influences a camera pixel in correspondence, resulting in a diagonal transport matrix $\mathbf{D}$. This implies $\mathbf{d}(p) = 0, \forall\ p \neq q$ and only a single value $\mathbf{d}(q)$ needs to be acquired in each row. In the absence of other components, turning on q and measuring intensity at c gives us $\mathbf{d}(q)$. Since the camera pixels are acquired in parallel, in the absence of near and far-range effects, a single illumination pattern, such as floodlight, is sufficient to acquire all rows $\mathbf{d}$ simultaneously i.e. acquire $\mathbf{D}$.

In the near-range component, a small neighborhood of projector pixels around q influence the camera pixel resulting in a banded diagonal transport matrix $\mathbf{N}$. The support set of these projector pixels is given by $\mathcal{P}_W = [q - \frac{W}{2}, q + \frac{W}{2}]$[1] i.e. a window of size $W \ll P$ centered at q. W is user defined and we use a single global value at each camera pixel. In practice, W is set sufficiently large to account for the near-range effects at all camera pixels. Since $\mathbf{n}(p) = 0, \forall\ p \notin \mathcal{P}_W$, it is sufficient to acquire W non-zero entries of $\mathbf{n}$. Though W is the same at every camera pixel, note that the set $\mathcal{P}_W$ is distinct.

Although the support $\mathcal{P}_W$ changes at each camera pixel, in the absence of other components, only W illumination patterns in all are sufficient to acquire $\mathbf{n}$ at every camera pixel i.e. acquire $\mathbf{N}$. We show in Section 4.2 that projecting W sinusoidal patterns with frequencies spaced uniformly in the spectrum as shown in Figure 3(5c) enables better separation of near and far-range components.

The far-range effect is exhibited over a much wider range of projector pixels as illustrated in Figure 1 and validated in Figure 3. We assume that the support set includes all the projector pixels $[0, P-1]$. Note that $\mathbf{f}$ has a large number of low magnitude entries compared to $\mathbf{d}$ and $\mathbf{n}$. While the exact nature of influence of projector pixels is dictated by the form factor and BRDF of scene points, we observe that this effect is primarily low frequency for diffuse scenes as shown in Figure 3 (4). Hence only a few low-frequency illumination patterns are sufficient to acquire $\mathbf{f}$ in the absence of other components.

[1] For simplicity we consider even W and show a symmetric $\mathcal{P}_W$ with $W+1$ elements.

3.3 Frequency Domain Modeling

We denote the DFT of $\mathbf{t}$, $\mathbf{d}$, $\mathbf{n}$ and $\mathbf{f}$ by $\hat{\mathbf{t}}$, $\hat{\mathbf{d}}$, $\hat{\mathbf{n}}$ and $\hat{\mathbf{f}}$ respectively and make the following observations. Since $\mathbf{d}(p) = \mathbf{d}(p)\delta(p-q)$, its DFT is given by $\hat{\mathbf{d}}_k = \mathbf{d}(q)\mathbf{l}_k(q)$ where $\mathbf{l}_k(q)$ is the phase. The magnitude $|\hat{\mathbf{d}}_k| = \mathbf{d}(q)$ is a constant independent of k implying the bandwidth of the direct transport is $k_{dir} = \frac{P}{2}$ as shown in Figure 3(4b). We acquire $\mathbf{d}$ by illuminating the scene with the highest frequency $|k| = \frac{P}{2}$ as shown in Figure 3(5b) and discussed in Section 4.1.

The near-range $\mathbf{n}$ has finite support $\mathcal{P}_W$ but exhibits smoothness over the support as shown in Figure 3(3c). As a result, its frequency response is bandlimited approximately to k_{near} i.e. $\hat{\mathbf{n}}_k = 0, \forall\ |k| > k_{near}$. Typically, $\mathbf{n}$ is parsimonious in the spatial domain compared to the frequency domain i.e. $2k_{near} > W$. Hence, we reconstruct $\mathbf{n}$ by exploiting its sparsity in the spatial domain but obtain the measurements by illuminating W sinusoidal patterns whose frequencies are uniformly spaced as shown in Figure 3(5c) and discussed in Section 4.2.

The far-range $\mathbf{f}$ is a low-frequency effect [1],[2] and we validate this observation in Figure 3(4d). $\mathbf{f}$ is bandlimited approximately to k_{far} i.e. $\hat{\mathbf{f}}_k \approx 0, \forall\ |k| > k_{far}$. In practice, setting $k_{far} = \frac{P}{20}$ (10% of the spectrum) is sufficient.

The far-range transport $\mathbf{f}$ has wider spatial support compared to near-range transport $\mathbf{n}$ and varies smoothly implying $k_{far} \ll k_{near}$. Similarly, $\mathbf{n}$ has wider support compared to $\mathbf{d}$ and hence $k_{near} < \frac{P}{2}$. This fact enables us to choose illumination $\mathbf{l}_k$ which separates one component from the other. Since the components are redundant in either projector's spatial or frequency domain, only a few of these discriminant illumination patterns also suffice to acquire them.

4 Light Transport Acquisition

Based on the spatial and frequency domain models of Sections 3.2 and 3.3, we make the following observations critical to our method.

1. We can pick an illumination $\mathbf{l}_k$ of frequency k such that if $k_{near} < |k| \leq \frac{P}{2}$ then $b^f_k = 0$, $b^n_k = 0$ and if $k_{far} < |k| \leq k_{near}$ then $b^f_k = 0$. This allows us to selectively suppress component transports with our choice of illumination.
2. Each component transport is parsimonious either in spatial or frequency domain. $\mathbf{d}$ has 1 unknown in spatial domain, $\mathbf{n}$ has W unknowns in spatial domain and $\mathbf{f}$ has k_{far} unknowns in frequency domain.
3. For acquiring the W entries $\{\mathbf{n}(p) : p \in \mathcal{P}_W\}$ at every camera pixel, the illumination patterns must satisfy the following conditions
   a) be well-conditioned over the support $\mathcal{P}_W$ at all camera pixels
   b) have minimal response from the far-range component.
   A key contribution of our paper is that we pick W basis illumination patterns $\{\mathbf{l}_k : k = \frac{P}{W} * [-\frac{W}{2}, \ldots, -1, 0, 1, \ldots, \frac{W}{2}]$. We show that these patterns are well conditioned over $\mathcal{P}_W$ at all camera pixels and only a small number of them have far-range response i.e. satisfy $|k| < k_{far}$.

4.1 Direct Transport Acquisition

We choose the highest frequency basis illumination $\mathbf{l}_k$, $k = \frac{P}{2}$ shown as two impulses in Figure 3(5b). Since, $b_k^n = b_k^f = 0$, we have $b_k = b_k^d = \mathbf{d}(q)\mathbf{l}_k(q)$. Since q is known at each camera pixel after correspondence, $\mathbf{d}$ is simply given as $\mathbf{d}(q) = b_k^d \mathbf{l}_k^*(q)$. We denote the acquired direct transport as $\tilde{\mathbf{d}}$. Like Nayar et al. [4], we pick $k = \frac{P}{2}$ but we acquire the direct transport whereas [4] estimates the direct image under flood light.

4.2 Near-Range Transport Acquisition

Uniform Sampling in the Fourier Domain: We pick W illumination patterns $\mathbf{l}_k$, $k \in \mathcal{K}_W = \frac{P}{W} * [-\frac{W}{2}, \ldots, -1, 0, 1, \ldots, \frac{W}{2}]$. The sampling pattern is shown in Figure 3(5c). The frequencies of the patterns $\mathbf{l}_k$ are $\{\frac{k}{P} = \frac{1}{W} * [-\frac{W}{2}, \frac{W}{2}]\}$. Note that this is the same set of frequencies as those in the DFT basis $e^{-j2\pi \frac{k'}{W} p'}$ of a length W signal. The only difference is that $\mathbf{l}_k$ is length P and hence periodic with period $\frac{P}{W}$. It is easy to show that over any window $\mathcal{P}_W$ of size W, the patterns in $\mathcal{K}_W$ are orthonormal i.e.

$$\sum_{p \in \mathcal{P}_W} \mathbf{l}_{k_1}^*(p)\mathbf{l}_{k_2}(p) = \begin{cases} W : k_1, k_2 \in \mathcal{K}_W, k_1 = k_2 \\ 0 \quad : k_1, k_2 \in \mathcal{K}_W, k_1 \neq k_2 \end{cases} \tag{5}$$

This implies that the W patterns in $\mathcal{K}_W$ are well-conditioned over any window $\mathcal{P}_W$. *They are sufficient to estimate* $\boldsymbol{n}$ *robustly at any camera pixel.*

Removing the Direct Component: Since $b_k^d \neq 0$, we remove the contribution $\tilde{\mathbf{d}}$ from the pixel intensity.

$$\begin{aligned} \tilde{b}_k = b_k - \tilde{\mathbf{d}} \cdot \mathbf{l}_k &= (\mathbf{t} - \tilde{\mathbf{d}}) \cdot \mathbf{l}_k \\ \tilde{b}_k &\approx \mathbf{n} \cdot \mathbf{l}_k + \mathbf{f} \cdot \mathbf{l}_k \\ &= \textstyle\sum_{p \in \mathcal{P}_W} \mathbf{n}(p)\mathbf{l}_k(p) + b_k^f. \end{aligned} \tag{6}$$

Reconstructing n: Since $\{\mathbf{l}_k : k \in \mathcal{K}_W\}$ form an orthonormal basis over $\mathcal{P}_W$, $\mathbf{n}$ can be reconstructed from $\{\tilde{b}_k : k \in \mathcal{K}_W\}$ through simple inversion of the basis. Equation (5) implies that the inversion is nothing but a linear combination of basis $\{\mathbf{l}_k^* : k \in \mathcal{K}_W\}$ with $\{\tilde{b}_k : k \in \mathcal{K}_W\}$ as the coefficients.

$$\frac{1}{W} \sum_{k \in \mathcal{K}_W} \tilde{b}_k \mathbf{l}_k^*(p') = \frac{1}{W} \sum_{k \in \mathcal{K}_W} \sum_{p \in \mathcal{P}_W} \mathbf{n}(p)\mathbf{l}_k(p)\mathbf{l}_k^*(p') + \frac{1}{W} \sum_{k \in \mathcal{K}_W} b_k^f \mathbf{l}_k^*(p') \tag{7}$$

$$\tilde{\mathbf{n}}(p') = \mathbf{n}(p') + \mathbf{n}_{err}(p'). \tag{8}$$

Removing Far-Range Effect: Note that $b_k^f \neq 0$ for $|k| \leq k_{far}$ and the estimate $\tilde{\mathbf{n}}$ is corrupted by far-range effect at these frequencies. Since, the set $\mathcal{K}_W \cap \{k : |k| \leq k_{far}\}$ is small, the error $\mathbf{n}_{err}$ is small as well. In Figure 3(5), this set is the intersection of the red and the blue samples. Since $k_{far} = 0.05P$, if $W < 20$

then only the DC ($k = 0$) is common to both sets. Even when $W \geq 20$, only a small number of frequencies are in common. More importantly, the fact that only a small number of patterns in $\mathcal{K}_W$ have a response from far-range component enables almost perfect separation between near and far-range effects. It also provides the flexibility to drop the frequencies $\mathcal{K}_W \cap \{k : |k| \leq k_{far}\}$ from $\mathcal{K}_W$. In practice, we eliminate the error $\mathbf{n}_{err}$ by only considering frequencies $\mathcal{K}_W \backslash \{k : |k| \leq k_{far}\}$. Then we have $b_k^f = 0$, $\forall k$ and

$$\tilde{b}_k = \sum_{p \in \mathcal{P}_W} \mathbf{n}(p)\mathbf{l}_k(p). \tag{9}$$

Reconstructing n with a Prior: Since the equations (9) are now underdetermined by a small number, we impose a prior on $\mathbf{n}$. We found the sparsity prior on $\mathbf{n}$ to be robust on real data and we solve

$$\tilde{\mathbf{n}} = \arg\min_{\mathbf{n}} \sum_{p \in \mathcal{P}_W} |\mathbf{n}(p)| \quad \text{s.t.} \quad \sum_{\mathcal{K}_W \backslash \{k:|k| \leq k_{far}\}} |\tilde{b}_k - \sum_{p \in \mathcal{P}_W} \mathbf{n}(p)\mathbf{l}_k(p)|^2 \leq \epsilon. \tag{10}$$

where ϵ is the noise variance noise across all measurements at a camera pixel.

The main contribution of our work is the choice of illumination frequencies which allow near-range to be acquired with little interference from the far-range effects. *Our approach is significant since it allows the separation of near and far-range effects through a simple choice of illumination frequencies.*

4.3 Far-Range Transport Acquisition

We pick illumination patterns $\{\mathbf{l}_k : k = [-k_{far}, k_{far}]\}$. Since $b_k^n \neq 0$ and $b_k^d \neq 0$, we remove their contribution $\tilde{\mathbf{d}}$, $\tilde{\mathbf{n}}$ from the pixel intensities.

$$\tilde{b}_k = b_k - \tilde{\mathbf{d}} \cdot \mathbf{l}_k - \tilde{\mathbf{n}} \cdot \mathbf{l}_k = (\mathbf{t} - \tilde{\mathbf{d}} - \tilde{\mathbf{n}}) \cdot \mathbf{l}_k \tag{11}$$

$$\tilde{b}_k \approx \mathbf{f} \cdot \mathbf{l}_k \tag{12}$$

Since $b_k^f = \mathbf{f} \cdot \mathbf{l}_k = 0$, $\forall |k| > k_{far}$, we have acquired the entire frequency spectrum of $\mathbf{f}$ and $\mathbf{f}$ is reconstructed through a simple inverse DFT.

4.4 Theoretical Analysis of Savings

The transports $\mathbf{d}$ and $\mathbf{n}$ are parameterized by 1 and W values in projector's spatial domain and $\mathbf{f}$ is parameterized by $2k_{far}$ values in projector's frequency domain. In total, the number of parameters in 1-D are $1 + W + 2k_{far}$. In 2-D, assuming a $P \times P$ projector, the corresponding number of parameters are 1, W^2 and $4k_{far}^2$ with sum total $1 + W^2 + 4k_{far}^2$ parameters. We have shown in Sections 4.1-4.3 that the components $\mathbf{d}$, $\mathbf{n}$ and $\mathbf{f}$ can be acquired using exactly $1 + W + 2k_{far}$ illumination patterns which is just the number of parameters (equivalently, $1 + W^2 + 4k_{far}^2$ in 2-D). This shows that theoretically our method

achieves the lower bound placed by our model on the number of measurements needed to acquire the transport matrix $\mathbf{T}$. Assuming $W = \alpha P$ and $2k_{far} = \beta P$, the savings in number of measurements in 2-D over the brute force is

$$\tau = \frac{P^2}{1 + \alpha^2 P^2 + \beta^2 P^2} \approx \frac{1}{\alpha^2 + \beta^2}. \tag{13}$$

Typically, we pick $\alpha = 0.15$ and $\beta = 0.1$ giving us a theoretical saving of $\tau \approx 31$.

5 Results

5.1 Implementation Details

Hardware Setup: We acquire the transport of a real, diffuse scene with a Canon 5D Mark II and a 50mm f1.4 lens at exposure 1/10sec. We use an Optoma TX779P operating at 120Hz for projecting sinusoidal illumination. We use real-valued, DC offset illumination patterns (in 1-D) $0.5 + 0.5\cos(2\pi\frac{k}{P}p)$ and $0.5 + 0.5\sin(2\pi\frac{k}{P}p)$ in place of complex $\mathbf{l}_k$ and $\mathbf{l}_{-k}$. To accurately account for the DC offset of the projector, we project a pattern $0.5 + 0.5\cos(2\pi\frac{k}{P}p)$ and its inverse $0.5 - 0.5\cos(2\pi\frac{k}{P}p)$, resulting in half the savings suggested by theory.

Projector-Camera Correspondence: For correspondence, at each camera pixel c we determine the projector pixel q directly lighting it. We create a square transport matrix T, by averaging the intensity of all camera pixels lit by the same projector pixel q. Under global illumination, we establish correspondence by scanning the scene with horizontal and vertical black stripes [4].

Savings in Measurements: Equation (13) suggests theoretical savings of $\tau \approx 31\times$ but practically the savings are halved to $\tau \approx 15.5\times$. In our acquisition, we achieve savings of $\tau = 15.4\times$ in Figure 2 and $\tau = 20.8\times$ in Figure 4. Note that as a pre-capture process, the projector-camera correspondence requires $2P$ projections which is linear in P and a small fraction of P^2.

Reconstruction: We set W sufficiently large, typically $W \sim 0.15P$, to model the near-range accurately at all camera pixels. This implies the entries of $\mathbf{n}$ in $\mathcal{P}_W$ are sparse. If $W \sim 0.15P$ and $k_{far} \sim 0.05P$, the set $\mathcal{K}_W \cap \{k : |k| \leq k_{far}\}$ has only a small fraction of patterns compared to W^2 and Equation (9) is slightly under-determined. We reconstruct $\mathbf{n}$ robustly in practice by solving (10) for sparse $\mathbf{n}$ at each camera pixel independently using a Bayesian algorithm [17].

We quantify the reconstruction accuracy of relighting by the signal-to-noise ratio $SNR_{\mathbf{b}} = 20\log_{10}\frac{\|\mathbf{Tl}\|}{\|(\mathbf{T}-\tilde{\mathbf{T}})\mathbf{l}\|}$ where $\mathbf{b} = \mathbf{Tl}$ is the image of the scene under lighting $\mathbf{l}$ and $\tilde{\mathbf{T}}\mathbf{l}$ is the image under relighting. Additionally, the reconstruction accuracy of row $\mathbf{t}$ of a camera pixel is quantified by $SNR_{\mathbf{t}} = 20\log_{10}\frac{\|\mathbf{t}\|}{\|(\mathbf{t}-\tilde{\mathbf{t}})\|}$.

5.2 Experimental Results

Acquisition of Light Transport: We acquire the transport matrix of the scenes shown in Figure 2 and Figure 4. Both the matrices are large for brute

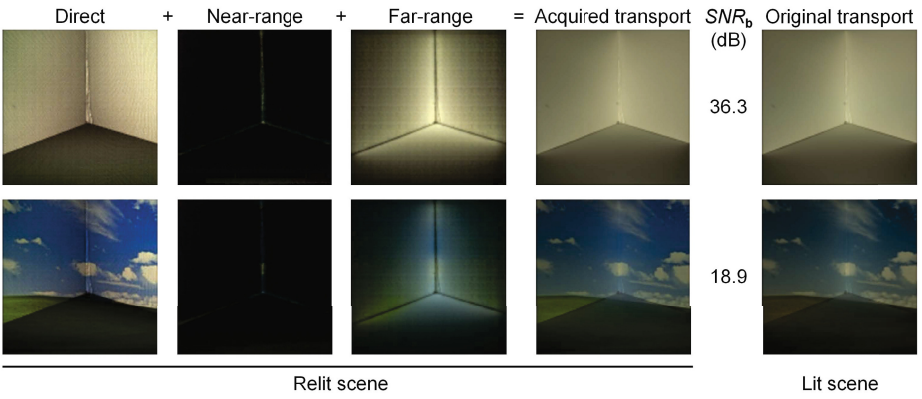

Fig. 4. We acquire the 16384 × 16384 transport matrix of a 128 × 128 large scene of a corner with 788 illumination patterns. We set $W = 15$. Under relighting with the landscape image (second row), notice that the blue sky casts blue inter-reflections on the floor and the green ground casts green inter-reflections on the walls.

force acquisition but we accurately acquire with savings of 15.4× and 20.8× respectively. Additionally, we decompose the transport into direct, near-range and far-range components. In Figure 2, a scene with significant inter-reflections and subsurface scattering, the acquired transport under relighting is close to the original lit scene. The candle's specularities appear in the direct component. The candle exhibits significant subsurface scattering which appears in the near-range. In the far-range, the candle has a red and green tinge due to the red and green wall on either side. Likewise the far-range has red diffuse inter-reflection from the wall on the cup and it displays none of the high frequencies in the penguin image. In Figure 4, we accurately acquire the transport at a corner. The near-range has the local inter-reflections at the edges of the wedge and the diffuse inter-reflections appear in the far-range.

Simulation on Acquired Light Transport: We simulate our method on the transport of a real scene (Figure 5) acquired through brute force and compare with the ground truth for validation. We decompose and acquire the transport matrix with just the minimal required $1+W^2+4k_{far}^2$ patterns. We pick $W = 19$ and set $2k_{far} = 0.1P$ in both the dimensions i.e. $(7, 10)$ respectively. We drop 9 near-range illumination patterns which fall within (or closest to) $2k_{far}$. Notice that $SNR_{\mathbf{b}}$ of relighting is high under floodlight and a natural image.

$SNR_{\mathbf{b}}$ computed at different illumination frequencies k and $SNR_{\mathbf{t}}$ computed at different camera pixels c are shown as a 2-D plot in Figure 5(Right). Across all frequencies and pixels, the SNR is high and atleast 25dB. The lowest $SNR_{\mathbf{b}}$ is at k_{far}, the transition frequencies between the far and near-range. Likewise, the lowest $SNR_{\mathbf{t}}$ is at the camera pixels in the transition zone between the near and far-range close to the edge.

Comparison to CLT [9]: CLT was proposed for acquiring transport matrices under distant lighting and exploits sparsity of each row (reflectance function) in

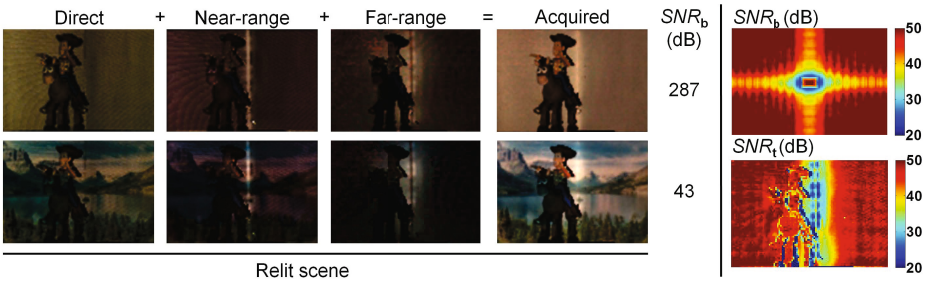

Fig. 5. We simulate our method on a real light transport acquired through brute force. (Left) We use 423 illumination patterns to acquire and decompose the 6800×6800 matrix of a 68×100 wedge-with-toy scene. The near-range captures the local inter-reflections at the edge and the far-range captures the diffuse inter-reflections. (Right) We plot $SNR_{\mathbf{b}}$ as a function of the sinusoidal illumination frequency $k = (k_x, k_y)$ and $SNR_{\mathbf{t}}$ as a function of the camera pixel $c = (c_x, c_y)$.

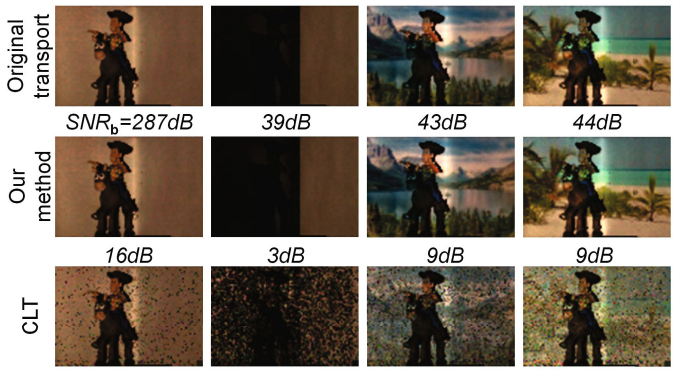

Fig. 6. We compare our method with CLT [9] under four relighting examples. Both methods acquire 6800×6800 transport matrix using 423 patterns. Our method significantly outperforms CLT since CLT is not designed for spatially varying illumination.

Haar wavelet basis. It also imposes spatial consistency by jointly reconstructing the reflectance of neighboring camera pixels. Under spatially varying illumination the transport matrix is diagonally dominant and full rank, making the joint modeling and reconstruction of rows harder. Hence, for comparison with CLT, we treat each row independently and do not impose spatial consistency in its implementation. For acquisition we use pseudo-random illumination patterns $\mathbf{l}_k$ with entries from a Bernoulli distribution. We reconstruct row $\mathbf{t}$ from the measurements $\{b_k : k = 1, 2, \ldots, K\}$ by imposing sparsity of $\mathbf{t}$ in Haar wavelet basis. The experiments in Figure 6 show that our method outperforms CLT.

6 Conclusions

We have presented a theoretical model of the light transport and its components under spatially varying illumination. Our paper separates the near-range effects

from the far-range allowing us to acquire the transport using minimal number of patterns. We show practical results by acquiring the transport of real scenes efficiently and decomposing them into component transports.

Practically, we treat the projector-camera correspondence and light transport acquisition separately. An acquisition technique which treats them in the same framework is an interesting avenue for future work. Theoretically, modeling of the near-range and far-range in a single framework is desirable for graceful decomposition and efficient acquisition.

In summary, light transport acquisition is an important and challenging problem due to the size of the matrix. However, the matrix exhibits structure and is decomposable and we believe there is significant future work in exploiting this structure for more efficient acquisition.

Acknowledgements. This work was supported by ONR PECASE grant N00014-09-1-0741. We thank Intel Science and Technology Center for Visual Computing for the equipment. We thank Jiamin Bai, Milos Hasan and Rahul Narain for their kind support and the anonymous reviewers for their comments.

References

1. Nayar, S.K., Ikeuchi, K., Kanade, T.: Shape from interreflections. International Journal of Computer Vision 6, 173–195 (1991)
2. Koenderink, J.J., van Doorn, A.J.: Geometrical modes as a general method to treat diffuse interreflections in radiometry. J. Opt. Soc. Am. 73, 843–850 (1983)
3. Jensen, H.W., Marschner, S.R., Levoy, M., Hanrahan, P.: A practical model for subsurface light transport. In: SIGGRAPH 2001, pp. 511–518. ACM (2001)
4. Nayar, S.K., Krishnan, G., Grossberg, M.D., Raskar, R.: Fast separation of direct and global components of a scene using high frequency illumination. ACM Trans. Graph. 25, 935–944 (2006)
5. Gu, J., Kobayashi, T., Gupta, M., Nayar, S.: Multiplexed illumination for scene recovery in the presence of global illumination. In: ICCV 2011, pp. 691–698 (2011)
6. Fuchs, M., Blanz, V., Lensch, H.P., Seidel, H.P.: Adaptive sampling of reflectance fields. ACM Trans. Graph. 26 (2007)
7. Wang, J., Dong, Y., Tong, X., Lin, Z., Guo, B.: Kernel nyström method for light transport. ACM Trans. Graph. 28, 29:1–29:10 (2009)
8. O'Toole, M., Kutulakos, K.N.: Optical computing for fast light transport analysis. In: SIGGRAPH ASIA 2010, pp. 164:1–164:12. ACM (2010)
9. Peers, P., Mahajan, D.K., Lamond, B., Ghosh, A., Matusik, W., Ramamoorthi, R., Debevec, P.: Compressive light transport sensing. ACM Trans. Graph. 28, 3:1–3:18 (2009)
10. Masselus, V., Peers, P., Dutré, P., Willems, Y.D.: Relighting with 4D incident light fields. In: SIGGRAPH 2003, pp. 613–620. ACM (2003)
11. Sen, P., Chen, B., Garg, G., Marschner, S.R., Horowitz, M., Levoy, M., Lensch, H.P.A.: Dual photography. In: SIGGRAPH 2005, pp. 745–755. ACM (2005)
12. Garg, G., Talvala, E.V., Levoy, M., Lensch, H.P.A.: Symmetric photography: Exploiting data-sparseness in reflectance fields. In: Eurographics Symposium, pp. 251–262. Eurographics Association (2006)
13. Seitz, S.M., Matsushita, Y., Kutulakos, K.N.: A theory of inverse light transport. In: ICCV 2005, pp. 1440–1447 (2005)

14. Chandraker, M., Bai, J., Ng, T.T., Ramamoorthi, R.: On the duality of forward and inverse light transport. IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 2122–2128 (2011)
15. Wu, D., O'Toole, M., Velten, A., Agrawal, A., Raskar, R.: Decomposing global light transport using time of flight imaging. In: CVPR 2012, pp. 366–373 (2012)
16. O'Toole, M., Raskar, R., Kutulakos, K.N.: Primal-dual coding to probe light transport. ACM Trans. Graph. 31, 39:1–39:11 (2012)
17. Babacan, S., Molina, R., Katsaggelos, A.: Bayesian compressive sensing using laplace priors. IEEE Transactions on Image Processing 19, 53–64 (2010)