(De) Focusing on Global Light Transport for Active Scene Recovery *

Mohit Gupta†, Yuandong Tian†, Srinivasa G. Narasimhan† and Li Zhang‡
† Robotics Institute, Carnegie Mellon University, USA
‡ Computer Science Department, University of Wisconsin-Madison, USA

Abstract

Most active scene recovery techniques assume that a scene point is illuminated only directly by the illumination source. Consequently, global illumination effects due to inter-reflections, sub-surface scattering and volumetric scattering introduce strong biases in the recovered scene shape. Our goal is to recover scene properties in the presence of global illumination. To this end, we study the interplay between global illumination and the depth cue of illumination defocus. By expressing both these effects as low pass filters, we derive an approximate invariant that can be used to separate them without explicitly modeling the light transport. This is directly useful in any scenario where limited depth-of-field devices (such as projectors) are used to illuminate scenes with global light transport and significant depth variations. We show two applications: (a) accurate depth recovery in the presence of global illumination, and (b) factoring out the effects of defocus for correct direct-global separation in large depth scenes. We demonstrate our approach using scenes with complex shapes, reflectances, textures and translucencies.

1. Introduction

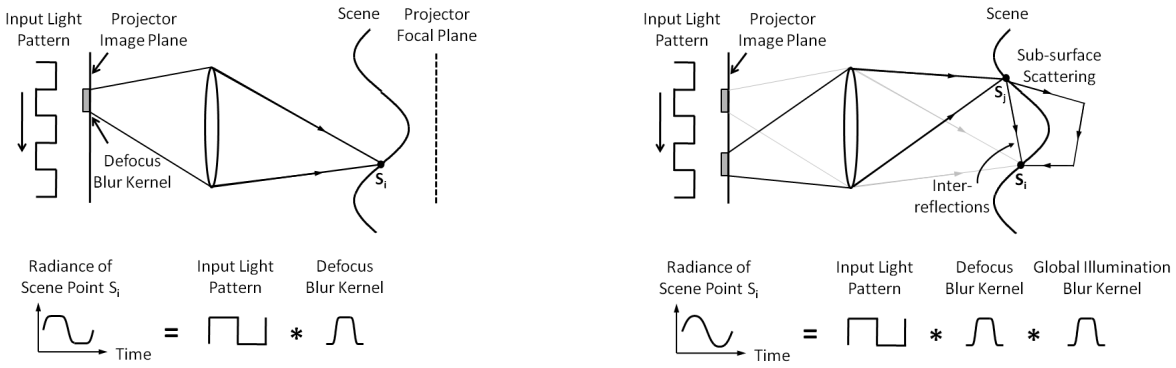
Light interacts with the world around us in complex ways, resulting in phenomena such as inter-reflections, sub-surface scattering and volumetric scattering. These phenomena are collectively termed as global light transport or global illumination. Historically, the effects of global illumination have largely been ignored in the computer vision literature. Most active shape recovery techniques make the simplifying assumption that a scene point is illuminated only directly by the illumination source. In the presence of global illumination, techniques such as photometric stereo [22], shape from shading [11], structured light scanning, shape from projector defocus [23] produce erroneous results. For instance, inter-reflections make concave objects appear shallower [14]. Sub-surface scattering in translucent objects can confound structured light based methods, leading to incorrect depths [7]. It is fair to say that most active scene recovery techniques have limited applicability in real life settings where global illumination is ubiquitous.

The goal of this work is to recover scene properties in the presence of global illumination. In general, separating the effects of global illumination from a shape cue requires explicit modeling of global light transport. Given the complexity of global illumination, this can be intractable. We consider illumination defocus, the depth cue for the shape-from-projector-defocus approach [23]. In this case, we show that global illumination can be separated from the depth cue without explicitly modeling or measuring light transport. The key observation is that both illumination defocus and global illumination manifest as low pass filters during image formation. If the scene is illuminated with a periodic illumination pattern, we show that the observed radiance at each pixel over time can be modeled as a convolution of the input pattern with the two blur kernels associated with defocus and global illumination (see Figure 1).

Expressing both the shape cue (defocus) and global illumination as blur kernels can appear to be counter-productive, as it may make it harder to separate the two. However, illumination defocus and global illumination are different physical phenomena. Illumination defocus is a result of the optics of the source, and encodes scene depths. On the other hand, global illumination encodes the intrinsic properties of the scene, such as 3D geometry and material properties. Thus, although changing the projector focus setting changes the defocus blur, the global illumination blur remains approximately constant. Based on this observation, we derive an invariant between global illumination blur and defocused illumination which can be used to separate the two effects. This invariant is directly useful in scenarios where limited depth-of-field devices (such as projectors) are used to illuminate scenes with global light transport and large depth variations.

We show two applications which require separation of defocus and global illumination: **First**, accurate depth recovery in the presence of global illumination (sub-surface scattering and inter-reflections). We follow the frequency domain approach of Zhang et al [23] and derive two depth estimation algorithms. The first algorithm requires a sweep of the projector focal plane across the scene and is dual to shape-from-camera-focus techniques. The second algorithm requires only two focal plane settings and is similar in spirit to shape-from-camera-defocus methods. **Second**, separation of the direct and global components of light transport for scenes with depth variations larger than the narrow depth of field of projectors ($< 0.3m$). We follow the spatial domain approach of Nayar et al [15] and derive defocus-invariant measures of global illumination. Again, we present two algorithms for separation based on (a) multiple focal plane positions and (b) single focal plane position and a depth map estimated in the first application.

*This research was supported in parts by an ONR grant N00014-08-1-0330 and NSF awards CCF-0541307, IIS-0643628 and IIS-0845916.



(a) Direct Illumination (b) Direct+Indirect Illumination

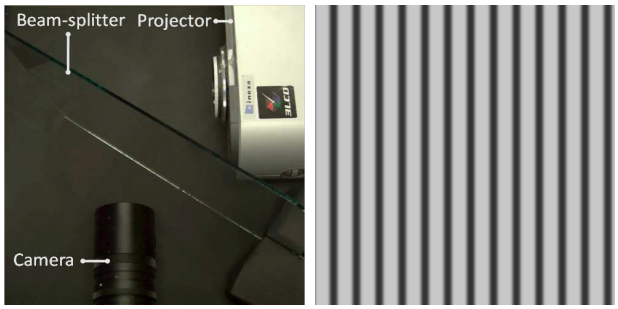
Figure 1. Image formation model. (a) A periodic illumination pattern is projected on the scene using a projector. The temporal radiance profiles of scene points which are not in focus are blurred. The amount of defocus blur is a function of the scene depths. (b) The presence of global light transport due to sub-surface scattering and inter-reflections introduces an additional blur. We show that the blur due to global illumination is independent of the projector focal plane position. This enables depth recovery even in the presence of global light transport.

We demonstrate our approaches using scenes with complex shapes and material properties including (a) marble, wax and natural objects such as fruits, milk and plants that show strong subsurface scattering, (b) objects with complex reflectance properties and textures such as fur, velvet, metal, wood and (c) objects with occlusions and concavities with strong inter-reflections. Since we do not impose any smoothness constraints, we recover depths independently at every pixel. Our techniques do not require complex calibration and are simple to implement.

2. Related Work

Most existing shape-from-intensity techniques [22, 11, 23] account for only the direct component of light transport. One possibility is to remove the global component a priori using the approach of Nayar et al [15]. However, this approach requires the projector's illumination to be focused on the entire 3D scene, making it unamenable for depth recovery using projector defocus analysis. Nayar et al [14] recovered depths in the presence of inter-reflections for scenes made of a few Lambertian planar facets. Approaches based on explicitly measuring the light transport matrix [19, 6] can be used to remove inter-reflection from images [18]. Such approaches require measuring a large number of impulse responses of the scene. Our methods do not require explicit modeling or estimation of the light transport matrix.

For structured light based techniques, the presence of sub-surface scattering and inter-reflections hinders the detection of the light sheet intersection with the objects [7]. Researchers have used polarization [3], modulation with a high-frequency illumination pattern [4] and fluorescence [12] to mitigate the adverse effects of global illumination. However, polarization does not reduce the effects of inter-reflections, and the fluorescence based technique requires submerging the scene in a fluorescent dye. Moreover, as with any triangulation based technique, structured lighting suffers from the presence of occlusions in complex scenes. Depth from camera focus (DFF) [16, 9] and depth from camera defocus (DFD) [21] techniques can compute complete depth maps[1], but they rely on scene texture for accurate scene recovery. We use a co-located camera-projector setup for data acquisition, as shown in Figure 2 (a). Using this setup prevents shadows due to occlusions, enabling recovery of complete, hole-free depth-maps. Also, our techniques can handle scenes with or without textures.



(a) (b)

Figure 2. Data acquisition setup. (a) Co-located camera-projector setup enables recovery of hole-free depth maps. (b) The periodic pattern used to illuminate the scene.

Another class of techniques measure density distribution of volumetric media using active lighting [2, 10, 8]. Confocal imaging techniques recover partially transparent volumes by focusing the illumination and sensor simultaneously on slices of the volume [5, 13]. The focus of this work is reconstructing opaque and translucent surfaces. It will be interesting to analyze the effects of volumetric scattering and transparency on our techniques in the future.

3. Image Formation Model

Consider a scene being illuminated by a projector with a periodic high frequency pattern. An example pattern is shown in Figure 2 (b). The pattern is translated horizontally, one pixel at a time, and an image is acquired for each translation. In the following, we show that the temporal radiance

[1] Although DFD and DFF also suffer from occlusion, the effects are not as severe due to a much smaller base-line [17].

profile at each pixel can be modeled as a convolution of the input pattern with the two blur kernels associated with illumination defocus and global illumination (see Figure 1(b)).

Direct Illumination: Consider the illustration in Figure 1 (a). The direct component of the radiance $e_i^d(t,f)$ at scene point S_i is the convolution of the illumination pattern, $p_i(t)$, and the defocus blur kernel $b_i(t,f)$ at S_i [2]:

$$e_i^d(t,f) \quad = \quad \alpha_i \; p_i(t) * b_i(t,f)\,. \tag{1}$$

where t denotes time, and f is the location of the projector focal plane. The blur kernel $b_i(t,f)$ depends on the depth of S_i and the position of the projector focal plane, f. The scale factor α_i accounts for the BRDF of the scene point, orientation of the surface with respect to the illumination source and the sensor, and the intensity fall-off.

Global Illumination: The global illumination at a scene point S_i is due to radiance received from other scene points, as shown in Figure 1(b). Let m_{ij} be the fraction of the direct radiance at the scene point S_j that reaches S_i, possibly after multiple inter-reflections and sub-surface scattering. Then the global component $e_i^g(t,f)$ is obtained by adding the contributions from all other scene points:

$$e_i^g(t,f) = \sum_{S_j \in Scene, j \neq i} m_{ij}\, p_j(t) * b_j(t,f)\,. \tag{2}$$

The total radiance $e_i(t,f)$ at S_i is the sum of the direct and the global components:

$$e_i(t,f) \quad = \quad e_i^d(t,f) + e_i^g(t,f)\,. \tag{3}$$

We compactly write the expression for radiance at scene point S_i using Eqs. 1, 2 and 3:

$$e_i(t,f) \quad = \quad \sum_{S_j \in Scene} m_{ij}\, p_j(t) * b_j(t,f)\,. \tag{4}$$

We have implicitly included the α_i term with m_{ii}. Taking the Fourier transform of Eq. 4:

$$E_i(w,f) = P(w) \sum_{S_j} m_{ij}\, exp(-I\, w\, \phi_j)\, B_j(w,f)\,, \tag{5}$$

where, uppercase symbols denote the Fourier transforms of the corresponding lower-case symbols. The variable w represents the frequency. Since $p_j(t)$ is a shifted version of $p_i(t)$, their Fourier transforms have the same magnitude $P(w)$ and differ only in the phase term $exp(-I\, w\, \phi_j)$. Rearranging the terms:

$$E_i(w,f) \quad = \quad P(w)\;\; B_i(w,f)\;\; G_i(w,f)\,, \tag{6}$$

$$G_i(w,f) \quad = \quad \sum_{S_j} m_{ij}\, exp(-I\, w\, \phi_j)\, \frac{B_j(w,f)}{B_i(w,f)}\,. \tag{7}$$

[2] We assume that both incoming and outgoing radiance remain constant within the small solid angles ($< 1^\circ$) subtended by the projector and camera apertures respectively at the scene point.

The term $B_i(w,f)$ is the Fourier transform of the defocus blur kernel at S_i. This term encodes scene depths and is independent of global illumination. We define $G_i(w,f)$ as the Fourier transform of the *global illumination blur kernel* at S_i. The term $G_i(w,f)$ encodes the optical interactions between different scene points via the light transport coefficients m_{ij}. Thus, the observed blur $E_i(w,f)$ is a function of both the blur due to defocus $B_i(w,f)$ and the blur due to global illumination $G_i(w,f)$. Note that this analysis and the techniques presented in the paper do not make any assumption on the particular form of the blur kernels.

We compute $E_i(w,f)$ by taking the Discrete Fourier Transform of the observed radiance profile. We use the third coefficient of the DFT ($w = 3$) as a measure of the amount of blur, as we empirically found it to be the most informative coefficient. In the rest of the paper, for brevity, we drop the argument w, i.e. $E(w,f)$, $G(w,f)$ and $B(w,f)$ will be denoted as $E(f)$, $G(f)$ and $B(f)$ respectively.

4. Invariance of Global Illumination Blur to Illumination Defocus

In this section, we establish the invariant that the global illumination blur $G(f)$ is insensitive to the projector focus setting f. We show this using both real experiments and simulations. An analytical proof for a particular distribution of scene points and symmetric defocus kernels is given in the technical report [20].

4.1. Validation using Real Experiments

For the purpose of validation, we measure $G(f)$ for a wide range of projector focus settings f. For a scene point S_i, we can compute $G_i(f)$ up to a constant scale factor by identifying another scene point S_j which does not receive any global illumination, and has the same depth as S_i. Using Eq. 6 and noting that $B_i(f) = B_j(f)$:

$$\frac{G_i(f)}{\alpha_j} \quad = \quad \frac{E_i(f)}{E_j(f)}\,, \tag{8}$$

Experimental Setup: We use a co-located camera-projector system as shown in Figure 2 (a). Our system consists of a Sony Cineza 3-LCD video projector and a Lumenera Lu165C 12-bit camera. The projector focus setting is changed by rotating the focus ring manually. Markings were made on the focus ring to be able to replicate the focus settings. We use the pattern shown in Figure 2 (b) to illuminate the scene. This pattern has a period of 24 pixels in the horizontal direction [23]. For each focus setting, we acquire 24 images as the pattern is translated horizontally, one pixel at a time. The total number of images acquired is $24 \times F$, where F is the number of focus settings used. The acquisition time is approximately 1 minute per focus setting.

Validation Results: We design experiments to establish the invariant for both sub-surface scattering and inter-reflections. For inter-reflections, we construct a V-groove using two diffuse planes, as shown in Figure 3 (a). We compute $E(f)$ for the scene point A for different focus settings,

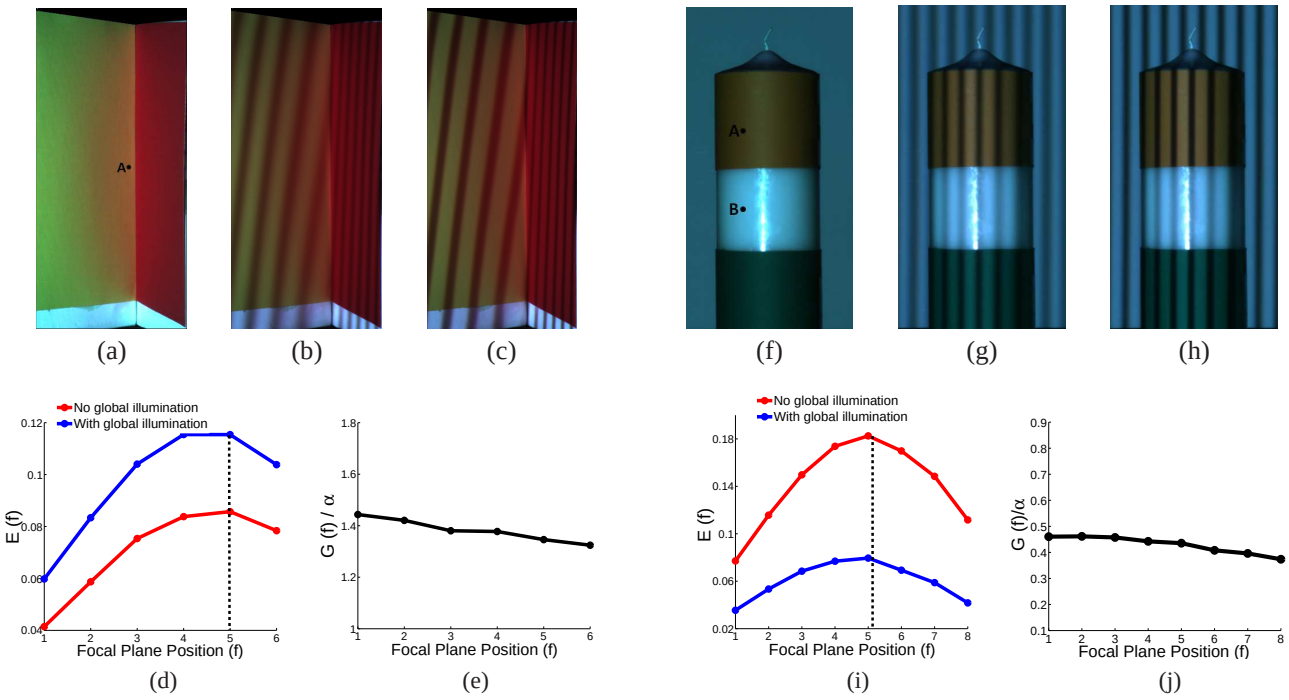


Figure 3. Experiments to show the invariance of the global illumination blur to the projector focus setting. (a) V-groove scene. (b-c) Sample input images for two out of six focus settings. (d) Blue curve is the plot of $E(f)$ vs. f at point A (with global illumination). $E(f)$ is computed by taking the DFT of the radiance profile. Red curve is the plot of $E(f)$ at point A with the right plane removed (no global illumination). (e) Plot of scaled $G(f)$ vs. the focal plane position. This is computed by taking the ratio of the two curves in (d) as in Eq. 8. The relative variation in $G(f)$ is less than 5% across the range of projector settings. (f) Candle Scene. (g-h) Sample input images for two out of eight focus settings. (i) Plot of $E(f)$ for points A (red curve - no global illumination) and B (blue curve - with global illumination). (j) Plot of $G(f)$ at point B. The variation in $G(f)$ is less than 7%.

which receives global illumination due to inter-reflections. We repeat the experiment for the same set of focus settings by removing the right plane (red colored). In this case, the scene point A does not receive any global illumination. In Figure 3(d), we plot $E(f)$ for A, both with and without global illumination. The global illumination blur $G(f)$ is computed by taking the point-wise ratio of the two curves, according to Eq. 8. For sub-surface scattering, we use a wax candle with the top and the bottom part covered with diffuse reflective paper, leaving the center exposed, as shown in Figure 3 (f). We choose a point B on the exposed part which receives global illumination in the form of sub-surface scattering. Point A, on the same vertical column and lying on the diffuse paper, is at the same depth as B but receives no global illumination. We plot $E(f)$ for A and B in Figure 3(i). As before, $G(f)$ at B is computed by taking the point-wise ratio of the two curves.

Two observations can be made from the plots. **First**, as shown in Figures 3(e) and 3(j), the total variation in $G(f)$ is less than 7% over the entire range of focal plane positions (0.3m-2.5m). This validates our claim that the global illumination blur resulting from sub-surface scattering and inter-reflections is insensitive to the projector focus setting. **Second**, we observe that the plots for $E(f)$, with and without global illumination, achieve maxima at the same focal plane position, as shown in Figures 3(d) and 3(i).

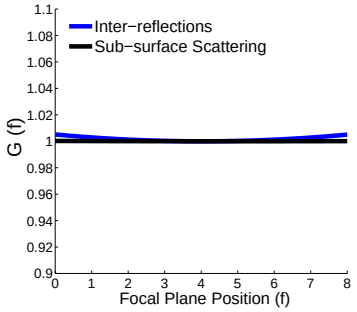
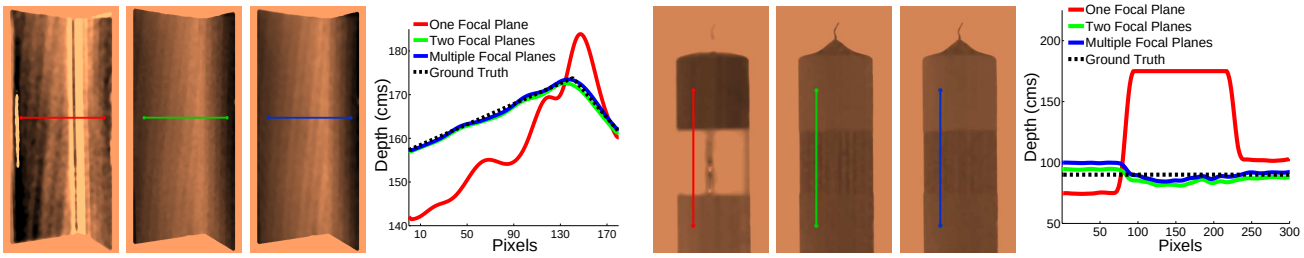


Figure 4. Invariance of $G(f)$ to the focal plane position f using simulations. The plot is normalized for scale and offset.

4.2. Validation using Simulations

We also validate the invariance of global illumination blur to projector focus settings using simulations. The scene is modeled as a 2D uniform distribution of points. We compute $G(f)$ according to Equation 7. We test our model for two cases of the light transport coefficients m_{ij} : inter-reflection and subsurface scattering. To account for intensity fall-off, occlusions and multiple bounces, we assume that the transfer coefficients m_{ij} between two points S_i and S_j to be inversely proportional to D_{ij}^2, the square-distance between them. Thus, for inter-reflections:

$$m_{ij} \propto \frac{1}{D_{ij}^2} \tag{9}$$



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 5. Comparison of the three depth recovery techniques for the V-groove and the candle scenes. (a, e) Single focal plane algorithm [23]. (b, f) Two Focal Planes (Section 5.2). (c, g) Multiple Focal Planes (Section 5.1). (d) Depth profiles for the V-groove along the indicated row. (h) Depth profiles for the candle along the indicated column. The single focal plane algorithm over-estimates the defocus blur in the presence of inter-reflections and sub-surface scattering resulting in incorrect depth estimates near the concavity and for the exposed parts of the candle. On the other hand, the relative RMS error for our algorithms is less than 1% for the V-groove and less than 5% for the candle. The ground truth depths were acquired using a calibration inclined plane with pre-measured depths.

For sub-surface scattering, the term m_{ij} encodes the additional exponential decay due to attenuation:

$$m_{ij} \quad \propto \quad \frac{1}{D_{ij}^2} \; exp(-D_{ij}) \tag{10}$$

For validation, we assume a Gaussian model for defocus blur. The spread of the gaussian is given by the distance between the scene point and the focal plane. We sample 100000 scene points from the distribution over 100 trials. We compute the average global illumination blur over all the scene points for different focal plane positions. Figure 4 shows the simulation results. The global illumination blur, both due to inter-reflections and sub-surface scattering remains nearly constant across different focus settings.

5. Depth Recovery under Global Illumination

Based on the invariant derived in the previous section, we present two algorithms for recovering depths in the presence of global light transport. The first algorithm requires a sweep of the focal plane across the scene, acquiring images at multiple focus settings. The second algorithm requires only two focus settings. Recall that the blur in the intensity profile measured at a single focal plane setting is a convolution of both the defocus blur and the global illumination blur. Thus, we need intensity profiles at *atleast* two focal settings in order to separate the two blur kernels.

5.1. Depth from multiple projector focal planes

In this algorithm, the DFT coefficients $E(f)$ are computed for multiple (≥ 3) focal plane positions f spanning the depth-range of the scene. Since the global illumination blur $G(f)$ is invariant to f, the plot of $E(f)$ against f reflects the behavior of the defocus blur $B(f)$. In other words, it attains a maxima when the corresponding scene point is the best in focus. It follows that scene points at the same depth but receiving different amounts of global illumination share the same maxima location. Two examples are shown in Figures 3(d,i). This suggests the maxima location $\overline{f}_i$ as a global-illumination invariant depth measure:

$$\overline{f}_i = \arg\max_f \; E_i(f) \tag{11}$$

The resolution of the depth measure, limited by the number of focal settings used, is improved by interpolating the focus measures $E_i(f)$ between the discrete focal plane settings [16]. As a one time calibration step, we compute a one-to-one mapping between scene depths and $\overline{f}_i$ using a planar, diffuse reflective board, whose depths are known a priori (see Figure 6 (a)). This mapping, along with the estimates of $\overline{f}$, is used to compute the actual depths for a given scene. This algorithm can be considered a dual to the shape-from-camera-focus technique, where depths are computed by sweeping the camera focal plane across the scene.

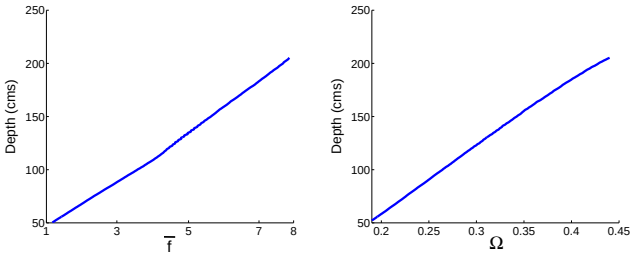
5.2. Depth from two projector focal planes

In this algorithm, we compute depths as a function of a defocus measure defined using only two focal positions f_1 and f_2. Since $G_i(f)$ is invariant to f, $G_i(f_1) = G_i(f_2)$. Using Eq. 6, we define the following ratio measure which is invariant to global illumination:

$$\Omega_i \quad = \quad \frac{E_i(f_2)}{E_i(f_1)} \quad = \quad \frac{B_i(f_2)}{B_i(f_1)} \, . \tag{12}$$

We compute a mapping (monotonic) between scene depths and Ω_i using a planar calibration board, as shown in Figure 6 (b). This mapping, along with the estimates of Ω is used to compute the actual depths for a given scene.

Results: We demonstrate our algorithms on scenes with complex shapes and material properties, and significant global illumination. Figure 5 shows results for the V-groove and the candle scenes. The single focal plane algorithm [23] over-estimates the defocus blur due to inter-reflections and sub-surface scattering resulting in incorrect depth estimates near the concavity and for the exposed parts of the candle. Our depth from two planes (Section 5.2) and multiple planes (Section 5.1) algorithms reconstruct both the shapes accurately. Theoretically, 3 focal planes are sufficient for the multiple focal planes algorithm. For robustness, we used 6 to 8 focal plane positions. Since we compute depths independently at every pixel, fine details such as the wick of the candle are reconstructed as well. The ground truth depths in Figure 5 were acquired using a calibration plane with pre-measured depths.



(a) (b)

Figure 6. Mappings between (a) scene depths and the focus measure $\overline{f}$, (b) scene depths and the defocus measure Ω.

Similar results can be observed in the more complex scenes of Figure 9 and 10. As before, the single plane algorithm over-estimates depths for the candle, and around the concavity of the V-groove. See the project web-page [1] for more results and comparisons. The striped artifacts visible in the depth maps are due to aliasing of the illumination pattern resulting from limited spatial resolution and non-ideal optics of the projector. The aliasing is mitigated by pre-filtering the pattern before projection.

6. Direct-Global Separation under Defocus

The algorithm proposed in [15] separates the direct and global components of light transport with a single projector focal plane position. However, in the presence of defocus blur, we need more information. Such a situation would arise if the depth range of the scene is larger than the depth of field of the projector. In this section, we present two algorithms for separating the direct and global components of radiance in the presence of defocus blur. The first algorithm uses multiple focal planes, and the second uses a single focal plane along with a depth map of the scene, which can be recovered using approaches of the previous section.

First, we derive the separation equations in the presence of defocus blur. Suppose we use a high-frequency pattern $p_i(t)$ with an equal number of on and off pixels to illuminate the scene. Then, following [15], the max-image, $e^+(f)$, computed by taking pixel-wise maximum, receives *approximately half* the global component. In the presence of defocus blur, the illumination pattern gets blurred. However, since the period of the pattern remains the same, this approximation still holds. Thus, using Eqs. 1 and 3, we write the expression for $e^+(f)$ in the presence of defocus:

$$e_i^+(f) \;=\; \beta_i^+(f)\; e_i^d \;+\; 0.5\; e_i^g\,, \qquad (13)$$

$$\beta_i^+(f) \;=\; \max_t \left\{ p_i(t) * b_i(t,f) \right\}. \qquad (14)$$

where $\alpha_i = e_i^d$. Note that e_i^d and e_i^g are the direct and global components respectively at S_i when the scene is fully illuminated. Similarly, we compute the min-image, $e^-(f)$:

$$e_i^-(f) \;=\; \beta_i^-(f)\; e_i^d \;+\; 0.5\; e_i^g\,, \text{ where} \qquad (15)$$

$$\beta_i^-(f) \;=\; \min_t \left\{ p_i(t) * b_i(t,f) \right\}. \qquad (16)$$

These equations are generalizations of the separation equations given in [15], as they account for defocus blur as well. The coefficients $\beta_i^+(f)$ and $\beta_i^-(f)$ depend on the defocus blur kernel $b_i(t,f)$ at S_i. If S_i is in perfect focus at the focus setting f, $\beta_i^+(f) = 1$ and $\beta_i^-(f) = 0$.

6.1. Separation using multiple focal planes

In this section, we present a separation technique by computing $e_i^+(f)$ and $e_i^-(f)$ for multiple focal planes. We use a checker-board illumination pattern as in [15]. Using a Gaussian interpolation scheme similar to previous section, we compute $\overline{e}_i^+$ and $\overline{e}_i^-$, the extrema values of $e_i^+(f)$ and $e_i^-(f)$ respectively. An example plot for a point on the candle is shown in Figure 8 (a). Note that the curve for $e_i^-(f)$ attains a maximum, while the curve for $e_i^-(f)$ attains a minimum. The computed images $\overline{e}_i^+$ and $\overline{e}_i^-$ are the max and min image respectively *as if* the scene is in perfect focus. Thus, we can write the separation equations as:

$$\overline{e}_i^+ \;=\; e_i^d \;+\; 0.5\; e_i^g \qquad (17)$$

$$\overline{e}_i^- \;=\; 0.5\; e_i^g \qquad (18)$$

The direct and global components can then be computed, respectively, as $e_i^d = \overline{e}_i^+ - \overline{e}_i^-$ and $e_i^g = 2\,\overline{e}_i^-$.

6.2. Separation using one plane and a depth map

Here, we present an algorithm to compute separation in the presence of defocus blur using a single focal plane and a depth map of the scene computed using the techniques presented in Section 5. For a scene point S_i, the direct and the global component are given using Eqs. 13 and 15:

$$e_i^d \;=\; \frac{e_i^+(f) \;-\; e_i^-(f)}{\beta_i^+(f) \;-\; \beta_i^-(f)}\,, \qquad (19)$$

$$e_i^g \;=\; e_i \;-\; e_i^d\,, \qquad (20)$$

where e_i is the observed intensity when the scene is fully lit. The denominator in Eq. 19 encodes the effects of defocus blur, and needs to be eliminated in order to recover the direct and global components. To this end, we build a mapping between $\left(\beta_i^+(f) - \beta_i^-(f)\right)$ and scene depths using a flat diffuse inclined plane with known depths and no global illumination, as shown in Figure 8 (b). For a point S_r on the inclined plane, we compute the max and the min images, $e_r^+(f)$ and $e_r^-(f)$ respectively. Then:

$$\beta_r^+(f) - \beta_r^-(f) \;=\; \frac{e_r^+(f) - e_r^-(f)}{e_r}\,, \qquad (21)$$

where e_r is the intensity at S_r when the plane is fully lit. If S_r and S_i are at the same depth, we can substitute for the denominator in Eq. 19 with Eq. 21, to recover the direct and global components.

Experiments and results for direct-global separation: For direct-global separation, we use the same setup as for depth estimation. We illuminate the scene with a checker-board pattern with checkers of size 8×8 pixels. The pattern

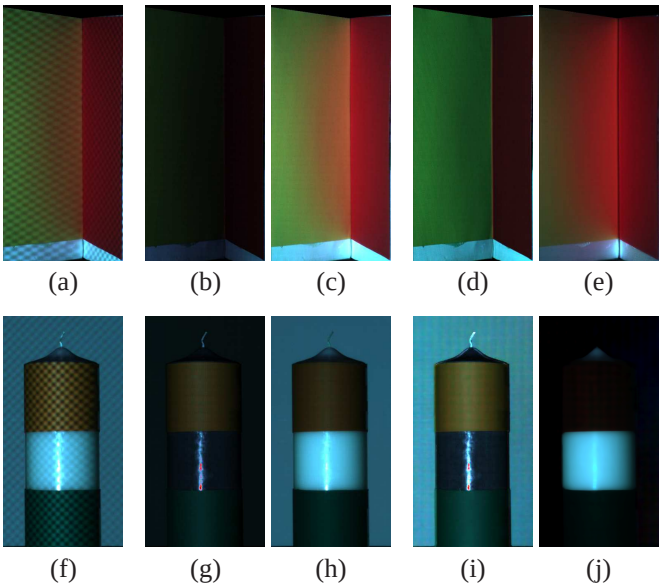


Figure 7. Comparison of the three separation techniques. (a, f) Input images. The technique in [15] incorrectly estimates the direct component (b, g) and the global component (c, h). The direct component is underestimated and the global component is overestimated on the planes of the V-groove and on the background plane in the candle scene. Correct separation using our multiple focal planes (d-e) and the single focal plane technique (i-j). **Zoom into the images for details**.

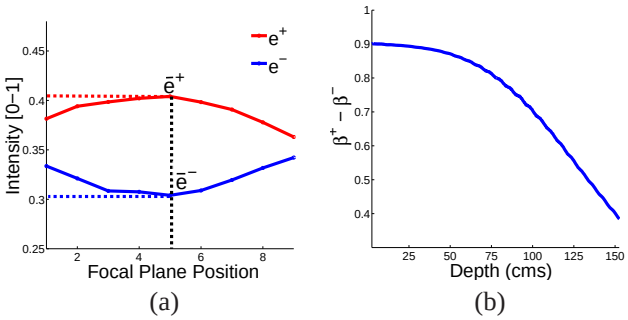


Figure 8. (a) Separation using multiple focal planes. We compute the extrema values of $e_i^+(f)$ and $e_i^-(f)$ and use them for separation in Eqs. 17 and 18. (b) Separation using one focal plane. Mapping between $\beta_i^+(f) - \beta_i^-(f)$ and scene depths. Given a depth map of the scene, this mapping is used to recover the correct separation using Eqs. 19 and 20.

is shifted 5 times by 3 pixels in both dimensions to acquire a total of 25 images per focal setting. The max-image and min-image are computed by simply taking the pixel-wise maximum and minimum respectively.

Figure 7 shows the separation results for the candle and the V-groove scene. First, the focal plane was placed in front of the scene so that the objects are not in focus. The technique in [15] under-estimates the direct component and over-estimates the global component on the planes of the V-groove and on the background plane in the candle scene. In contrast, our single focal plane algorithm (Section 6.2) recovers the correct separation, as shown for the candle. Notice the differences in the direct component on the background plane and the global component on the covered portions of the candle. Our multiple focal planes algorithm recovers the correct separation as well, as shown for the V-groove. Notice the *glow* due to inter-reflections around the concavity in the global component.

We also consider scenes with large depth variations (0.3m - 2m), significantly more than the depth of field of the projector, as shown in Figures 9 and 10. Here, for the technique in [15], the focal plane was placed in front of the scene, resulting in incorrect separation on the background and the polyresin bust in particular, as shown in Figure 9 (f-g). Our separation algorithms account for the defocus blur, and recover the correct direct and global components. For more results and comparisons, see the project web-page [1].

7. Discussion and Limitations

We have studied the interplay between defocused illumination and global illumination and derived an invariant which can be used to separate the two effects for scene recovery. We now discuss some limitations of our approaches.

If a scene point does not see the entire projector aperture, for example due to occlusions at depth discontinuities, the illumination defocus kernel at that scene point would be underestimated. As a result, the two/one plane algorithms are prone to errors close to depth discontinuities. However, the multiple focal plane algorithms rely on identifying the focus setting where there is no defocus. Thus, they are more robust at depth discontinuities (for instance, the wick of the candle) as compared to the two/one focal plane algorithms.

Our approaches do not handle perfectly mirrored objects due to high frequency global illumination. Another challenging problem is to analyze the effects of volumetric scattering and transparency on our techniques. Currently, the data acquisition process for our algorithms is not real-time. An avenue of future work is to extend our techniques for dynamic scenes. Finally, it will be interesting to account for camera defocus to combine the advantages of our techniques with those of shape from camera focus/defocus.

References

[1] Webpage. *http://graphics.cs.cmu.edu/projects/DefocusGlobal/*.

[2] B. Atcheson, I. Ihrke, W. Heidrich, A. Tevs, D. Bradley, M. Magnor, and H. Seidel. Time-resolved 3d capture of non-stationary gas flows. *SIGGRAPH*, 2008.

[3] T. Chen, H. P. A. Lensch, C. Fuchs, and H.-P. Seidel. Polarization and phase-shifting for 3d scanning of translucent objects. *CVPR*, 2007.

[4] T. Chen, H.-P. Seidel, and H. P. A. Lensch. Modulated phase-shifting for 3d scanning. *CVPR*, 2008.

[5] C. Fuchs, M. Heinz, M. Levoy, H.-P. Seidel, and H. P. A. Lensch. Combining confocal imaging and descattering. *Computer Graphics Forum*, 27(4), 2008.

[6] G. Garg, E.-V. Talvala, M. Levoy, and H. P. A. Lensch. Symmetric photography: Exploiting data-sparseness in reflectance fields. In *EGSR*, 2006.

[7] G. Godin, J.-A. Beraldin, M. Rioux, M. Levoy, L. Cournoyer, and F. Blais. An assessment of laser range measurement of marble surfaces. In *Fifth Conference on optical 3-D measurement techniques*, 2001.

(a) Scene (b) Depth using multiple planes (Section 5.1) (c) Depth using single plane [23]

(d) Direct (Section 6.1) (e) Global (Section 6.1) (f) Direct [15] (g) Global [15]

Figure 9. Results for the marbles and candle scene. (a) In this scene, we have a wax candle inside a white pot closest to the projector, a marble statue of *atlas*, a V-groove and a polyresin bust, in this depth order. The approximate depths of the clay-pot, the marble statue and the polyresin bust are $150cm$, $200cm$ and $270cm$ respectively. (b) Depth map using our multiple planes algorithm. **Zoom into the images for details**. (c) Incorrect depth map using the single plane algorithm. (d-e) Separation using our multiple planes technique. Since, there is a significant depth variation in the scene, the technique in [15] results in incorrect separation (f-g).

Scene Depth Map (Section 5.1) Direct Component (Section 6.1) Global Component (Section 6.1)

Figure 10. Scene recovery results using our techniques. This scene consists of organic materials: a few peppers, two pumpkins, the green plant, some marbles and eggs and a flower-pot, in this depth order. See the project web-page [1] for comparisons with previous techniques.

[8] J. Gu, S. K. Nayar, E. Grinspun, P. N. Belhumeur, and R. Ramamoorthi. Compressive Structured Light for Recovering Inhomogeneous Participating Media. In *ECCV*, 2008.

[9] S. W. Hasinoff and K. N. Kutulakos. Confocal stereo. In *ECCV (1)*, 2006.

[10] T. Hawkins, P. Einarsson, and P. Debevec. Acquisition of time-varying participating media. *SIGGRAPH*, 2005.

[11] B. Horn. Obtaining shape from shading information. *The Psychology of Computer Vision*, 19(1), 1975.

[12] M. B. Hullin, M. Fuchs, I. Ihrke, H.-P. Seidel, and H. P. A. Lensch. Fluorescent immersion range scanning. *SIGGRAPH*, 2008.

[13] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas. Synthetic aperture confocal imaging. *SIGGRAPH*, 2004.

[14] S. Nayar, K. Ikeuchi, and T. Kanade. Shape from Interreflections. *IJCV*, 6(3), 1991.

[15] 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. *SIGGRAPH*, 2006.

[16] S. K. Nayar and Y. Nakagawa. Shape from Focus. *PAMI*, 16(8):824–831, 1994.

[17] Y. Y. Schechner and N. Kiryati. Depth from defocus vs. stereo: How different really are they? *IJCV*, 39(2).

[18] S. M. Seitz, Y. Matsushita, and K. N. Kutulakos. A theory of inverse light transport. In *ICCV*, 2005.

[19] P. Sen, B. Chen, G. Garg, S. R. Marschner, M. Horowitz, M. Levoy, and H. P. A. Lensch. Dual photography. *SIGGRAPH*, 2005.

[20] Y. Tian, M. Gupta, S. G. Narasimhan, and L. Zhang. Relationship between projector defocus and global illumination for statistically-modeled scenes. Technical Report CMU-RI-TR-09-10, Carnegie Mellon University, March 2009.

[21] M. Watanabe and S. Nayar. Rational Filters for Passive Depth from Defocus. *IJCV*, 27(3).

[22] R. Woodham. Photometric method for determining surface orientation from multiple images. *OptEng*, 19(1), 1980.

[23] L. Zhang and S. K. Nayar. Projection defocus analysis for scene capture and image display. *SIGGRAPH*, 2006.