

A Combined Theory of Defocused Illumination and Global Light Transport

Mohit Gupta · Yuandong Tian ·
Srinivasa G. Narasimhan · Li Zhang

Received: 27 September 2010 / Accepted: 3 October 2011 / Published online: 27 October 2011
© Springer Science+Business Media, LLC 2011

Abstract Projectors are increasingly being used as light-sources in computer vision applications. In several applications, they are modeled as point light sources, thus ignoring the effects of illumination defocus. In addition, most active vision techniques assume that a scene point is illuminated only directly by the light source, thus ignoring global light transport effects. Since both defocus and global illumination co-occur in virtually all scenes illuminated by projectors, ignoring them can result in strong, systematic biases in the recovered scene properties. To make computer vision techniques work for general real world scenes, it is thus important to account for both these effects.

In this paper, we study the interplay between defocused illumination and global light transport. We show that both these seemingly disparate effects can be expressed as low pass filters on the incident illumination. Using this observation, we derive an invariant between the two effects, which can be used to separate the two. This is directly useful in scenarios where limited depth-of-field devices (such as projectors) are used to illuminate scenes with global light trans-

port and significant depth variations. We show applications in two scenarios: (a) accurate depth recovery in the presence of global light transport, and (b) factoring out the effects of illumination defocus for correct direct-global component separation. We demonstrate our approach using scenes with complex shapes, reflectance properties, textures and translucencies.

Keywords Defocused illumination · Global light transport · Depth recovery · Direct-global component separation · Projectors · Physics-based vision

1 Introduction

Active vision techniques use illumination as a probe for recovering a variety of scene properties. Recovery involves building models of the interaction of light with the scene (light transport) and *inverting* the models. In general, light transport can be complex and inverting the models can become intractable. Consequently, most such techniques have historically used simplified models of light transport; it is assumed that each scene point is illuminated only directly by the light source. Additionally, it is also assumed that the light source is a point or a distant light source or that the scene is roughly planar, so that illumination defocus effects are not modeled.

In most real world scenes, both these effects co-occur, making these assumptions excessively severe. Global light transport effects, such as sub-surface scattering, inter reflections and volumetric scattering are ubiquitous in any real world scene. Additionally, if the light source is an area light source or a limited depth of field device (such as projectors), scene points will receive defocused illumination. This

<http://graphics.cs.cmu.edu/projects/DefocusGlobal/>

M. Gupta (✉) · Y. Tian · S.G. Narasimhan
Robotics Institute, Carnegie Mellon University, Pittsburgh,
PA, 15213, USA
e-mail: mohitg@cs.cmu.edu

Y. Tian
e-mail: tydsh@cs.cmu.edu

S.G. Narasimhan
e-mail: srinivas@cs.cmu.edu

L. Zhang
Computer Science Department, University of Wisconsin,
Madison, WI 53706, USA
e-mail: lizhang@cs.wisc.edu

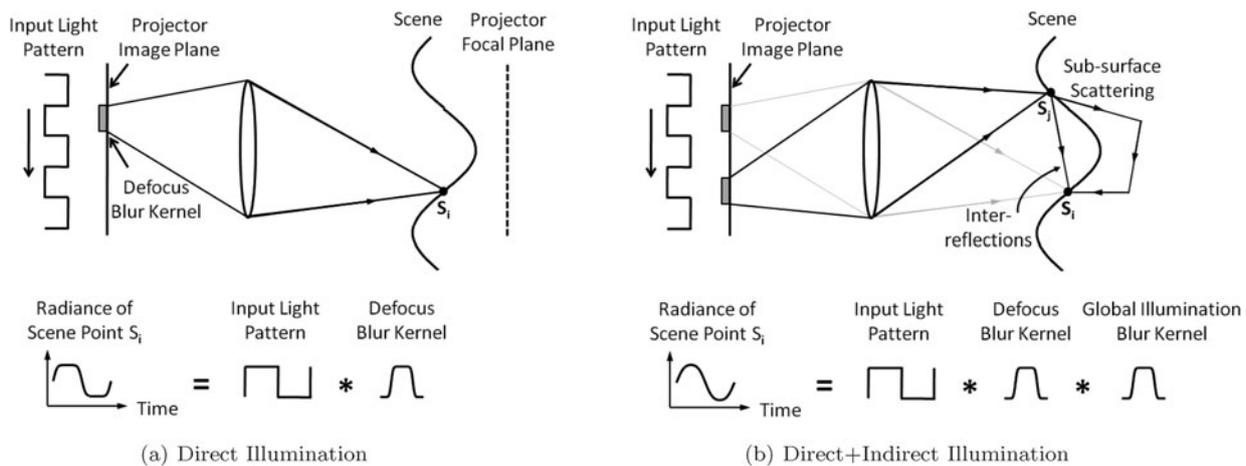


Fig. 1 Image formation model. **(a)** A periodic illumination pattern is projected on the scene using a projector. The temporal radiance profiles of scene points which are not in focus are blurred. The amount of defocus blur is a function of the scene depths. **(b)** The presence of

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

assumption is specially pertinent as projectors are being increasingly used as programmable illumination in computer vision applications. Consequently, defocused illumination and global light transport effects introduce strong, systematic biases in the recovered scene properties. To make active vision techniques work for general real world scenes, it is thus important to account for both these effects.

The goal of this paper is to study the interplay of both these effects. It may seem that defocused illumination and global light transport are completely different physical phenomena. Defocus is a result of the optics of the source, and encodes scene depths. On the other hand, global light transport encodes the intrinsic properties of the scene, such as 3D geometry and material properties. Our key observation is that both these effects manifest as low pass filters on the incident illumination during image formation. This observation allows analyzing both the effects using signal processing tools, without having to explicitly model either of them. 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 Fig. 1). We then derive an approximate invariant between the global light transport blur and defocus, 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 applications in two scenarios which require separation of the two effects. *First*, accurate depth recovery in the presence of global light transport (sub-surface scattering and inter-reflections). In the presence of global illumination, techniques such as photometric stereo (Woodham 1980),

shape from shading (Horn 1975), structured light scanning, shape from projector defocus (Zhang and Nayar 2006) produce erroneous results. For the depth cue of illumination defocus (Zhang and Nayar 2006), we show that global light transport can be separated from the depth cue without explicitly modeling or measuring light transport. We follow the frequency domain approach of Zhang and Nayar (2006) 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.

The *second* scenario that we consider is the separation of the direct and global components of light transport for scenes with depth variations larger than the narrow depth of field of projectors (<0.3 m). We follow the spatial domain approach of Nayar et al. (2006) and derive defocus-invariant measures of global light transport. 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. It is interesting to note the duality between the two applications in terms of their respective *noise* and *signal*: in the first application, global illumination is *noise* and defocus is the *signal*, while in the second application, defocus is the *noise* and global illumination is the *signal*.

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 our analysis is done independently at each pixel, we do not impose any smoothness constraints, or require presence of scene texture.

2 Related Work

Modeling Global Light Transport and Defocused Illumination A lot of work has been done in the computer graphics literature on modeling and simulating forward light transport. However, most of these models are too complex for the purpose of recovering scene properties. For volumetric scattering, the single scattering assumption is used to simplify light transport and thus, to recover scene properties (Treibitz and Schechner 2006; Narasimhan et al. 2005, 2006). However, the single scattering assumption is not valid in general for other modes of global light transport such as sub-surface scattering and inter-reflections. There has been extensive work on modeling camera defocus (Subbarao and Lu 1992) and using it to recover scene depths (Nayar and Nakagawa 1994; Watanabe and Nayar 1998). However, there has been limited work on modeling illumination defocus due to area light sources or due to limited depth of field devices such as projectors. Most active vision techniques either assume a point light source or the scene to be planar to avoid defocused illumination. Examples are shape recovery from structured light 3D scanning, shape from inverse light transport (Liu et al. 2010) and analysis of light transport (Bai et al. 2010; O’Toole and Kutulakos 2010; Mukaigawa et al. 2010).

Shape Recovery Under Global Light Transport Most existing shape-from-intensity techniques (Woodham 1980; Horn 1975; Zhang and Nayar 2006) 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. (2006). 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. (1991) 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 (Sen et al. 2005; Garg et al. 2006) can be used to remove inter-reflection from images (Seitz et al. 2005). 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 (Godin et al. 2001). Researchers have used polarization (Chen et al. 2007), modulation with a high-frequency illumination

pattern (Chen et al. 2008) and fluorescence (Hullin et al. 2008) 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) (Nayar and Nakagawa 1994; Hasinoff and Kutulakos 2006) and depth from camera defocus (DFD) (Watanabe and Nayar 1998) 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 Fig. 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.

Another class of techniques measure density distribution of volumetric media using active lighting (Atcheson et al. 2008; Hawkins et al. 2005; Gu et al. 2008). Confocal imaging techniques recover partially transparent volumes by focusing the illumination and sensor simultaneously on slices of the volume (Fuchs et al. 2008; Levoy et al. 2004). Morris and Kutulakos (2007) and Kutulakos and Steger (2008) reconstruct specular and transparent scenes by capturing multiple images under varying illumination and varying imaging geometry. 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.

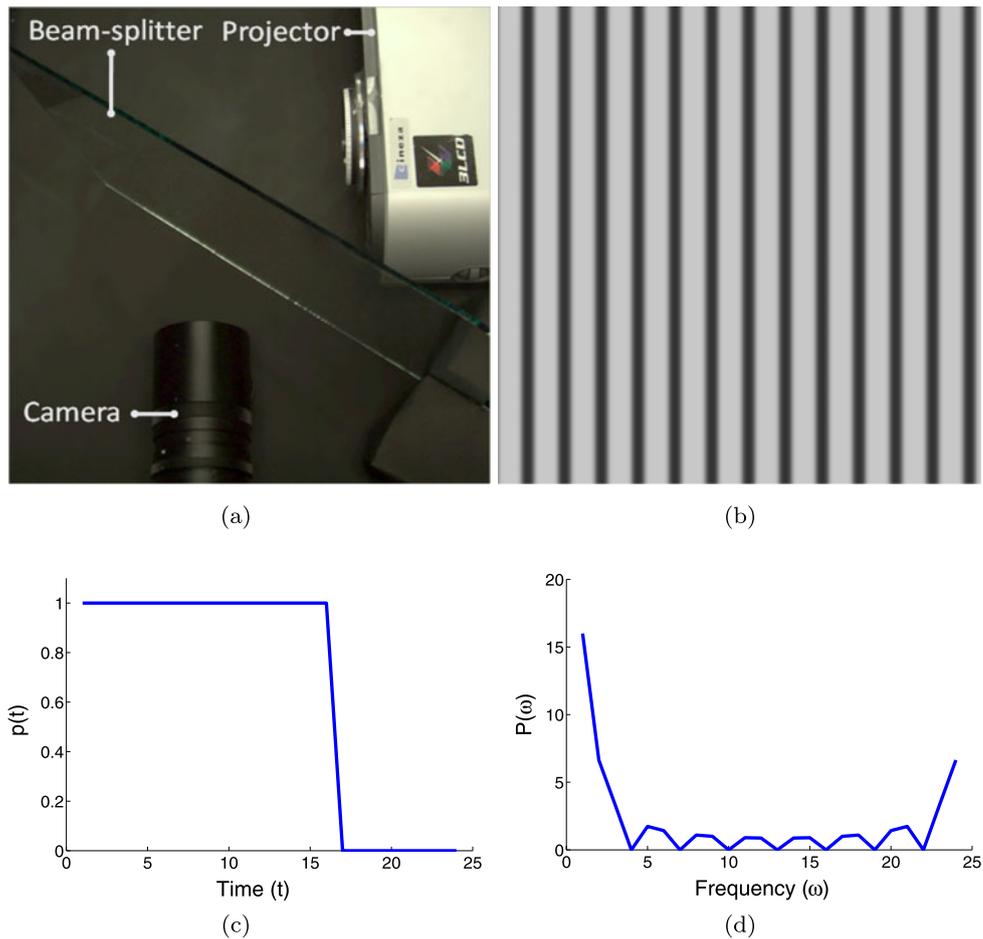
3 Image Formation Model

Consider a scene being illuminated by a projector with a periodic high frequency pattern. An example pattern is shown in Fig. 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 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 Fig. 1(b)).

Direct Illumination Consider the illustration in Fig. 1(a). The direct component of the radiance $e_i^d(t, f)$ at the scene

¹Although DFD and DFF also suffer from occlusion, the effects are not as severe due to a much smaller base-line (Schechner and Kiryati 2000).

Fig. 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. The pattern is shifted horizontally one pixel at a time and an image is captured for each shift. (c) Temporal profile $p(t)$ of the incident illumination. (d) Discrete Fourier transform $P(w)$ of $p(t)$



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 :^{2,3}

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

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 Global illumination at a scene point S_i is due to radiance received from other scene points, as shown in Fig. 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:

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

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

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

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

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

We have implicitly included the α_i term with m_{ii} . Since the projector is the only light source illuminating the scene, there is no ambient illumination. The effects of ambient illumination can be easily accounted for by adding a constant DC offset to the temporal intensity profile $e_i(t, f)$. Taking the Fourier transform of (4):

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

²We 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.

³We assume that there is no camera defocus. Experimentally, this is achieved by using a small camera aperture.

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 $\exp(-Iw\phi_j)$. Rearranging the terms:

$$E_i(w, f) = P(w)B_i(w, f)G_i(w, f), \quad (6)$$

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

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} . Equation (6) states that the observed radiance profile $E_i(w, f)$ at S_i is the convolution of the input pattern with the defocus blur kernel $B_i(w, f)$ and the global illumination blur kernel $G_i(w, f)$. This is illustrated in Fig. 1. Note that the above analysis does not make any assumption on the particular form of the blur kernels.

Modeling the Blur Kernels as Temporal Functions In the above derivation, both the blur kernels are modeled as temporal functions. This might appear counter-intuitive as both phenomena are essentially forms of spatial blurring. But the effect of defocus and global illumination can also be modeled temporally. For each scene point, both phenomena operate on the temporal incident illumination profile (Fig. 2(c)) and result in a blurred temporal radiance profile $e(t, f)$. This is shown in Fig. 1(b). The blur kernels $B(w, f)$ and $G(w, f)$ model this temporal blurring effect. The temporal and the spatial blurring effects are strongly related. If the support of the spatial blur kernel at a scene point is large, the corresponding temporal blur kernel has a proportionately large support as well. The advantage of doing a temporal analysis over spatial analysis is that it can be done independently for each pixel, without making strong assumptions on the scene structure, such as local smoothness of depth.

4 Relationship Between Global Illumination Blur and Projector Defocus

As shown in the previous section, the observed blur kernel is the convolution of the blur kernels due to both global illumination and illumination defocus. Since the two kernels encode different scene properties, it is useful to separate them. How can the two kernels be separated? Fortunately, we know that the defocus kernel $B(w, f)$ can be modulated by changing the projector focus settings. In this section, we study the dependence of the global illumination blur kernel $G(w, f)$ on the projector focus setting f .

Global Illumination Blur Is Approximately Invariant to Projector Focus Setting Consider the expression for the global illumination blur at a scene point S_i , as given in (7). Notice that the form factors m_{ij} are a function of the scene geometry and material properties, and do not depend on the projector focus setting. Similarly, the phase terms depend on the scene and illumination geometry and illumination frequency, and are independent of the focus setting f .

The only term dependent on f is the ratio $\frac{B_j(w, f)}{B_i(w, f)}$. Consider points S_j which are in a local neighborhood around S_i . Since the corresponding defocus kernels B_j and B_i are depth dependent, they vary in a similar manner as the focus setting is changed. Thus, the ratio $\frac{B_j(w, f)}{B_i(w, f)}$ remains nearly constant with changing f . Now consider points S_j which are in a neighborhood N_{dist} that is distant from S_i . For these points, the defocus kernels B_j and B_i vary differently as f is changed. But, in the summation (7), the contribution from these points is low because the form factors m_{ij} fall rapidly with distance. Moreover, for points S_j within N_{dist} , the phase terms vary much more rapidly (between -1 and 1) as compared to the variation in the form factors m_{ij} . Thus, the contributions from these points *cancel out* among themselves. Consequently, the global illumination blur kernel $G_i(w, f)$ is *nearly invariant* to the projector focus setting.

Dependence of the Invariant on Scene Content and the Illumination Frequency In order to understand the dependence of the invariant on the illumination patterns and the scene content, it is important to highlight the assumptions made while deriving the invariant:

- The form factors m_{ij} fall rapidly with distance between points S_j and S_i .
- For points S_j within neighborhoods distant from S_i , the phase terms vary much more rapidly as compared to the variation in the form factors m_{ij} .

The first assumption is satisfied by most scenes since the effect of global light transport phenomena (sub-surface scattering, diffusion and inter-reflections) decreases with distance due to light attenuation and fall-off. For the second assumption to hold, the frequency of the incident illumination should be higher than the frequency with which the form factors vary. For general scenes, it is hard to derive a closed form expression for the required illumination frequencies as the form factors are complex functions of scene geometry and material properties. Intuitively, the smaller the period of the illumination pattern, the better is the approximation. A similar argument was used in Nayar et al. (2006) for separating the direct and global components of light transport. One scenario that violates these assumptions is the presence of a distant mirror in the scene. This would result in high-frequency inter-reflections which are also strong despite be-

ing distant. In this scenario, the invariance will not hold. An example scene is shown in Fig. 12.

In the following, we provide empirical validation for the above observation using scenes exhibiting strong inter reflections and sub-surface scattering. In Appendix, we provide validation using simulations for different distributions of scene points.

4.1 Validation Using Real Experiments

For the purpose of validation, we measure $G(w, f)$ for a wide range of projector focus settings f . For a scene point S_i , we can compute $G_i(w, f)$ 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 (6) and noting that $B_i(f) = B_j(f)$:

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

Experimental Setup We use a co-located camera-projector system as shown in Fig. 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. As mentioned in the previous section, for the invariant to hold, it is important to use an illumination pattern with a small period (high-frequency). We use the pattern shown in Fig. 2(b) to illuminate the scene. This pattern has a period of 24 pixels in the horizontal direction (Zhang and Nayar 2006). For each focus setting, we acquire 24 images as the pattern is translated horizontally, one pixel at a time. From these images, we measure the temporal radiance profiles $e_i(t, f)$. 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. We compute $E_i(w, f)$ by taking the Discrete Fourier Transform of the observed radiance profiles $e_i(t, f)$.

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 Fig. 3(a). Figures 3(b–d) show sample input images for three out of six focus settings. We compute $E(w, f)$ at different focus settings for the scene point B , which receives global illumination due to inter-reflections. We repeat the experiment for the same set of focus settings by removing the right red colored plane (Fig. 3(e)). In this case, the scene point A does not receive any global illumination. Figures 3(i–j) show temporal intensity profiles at Point A and B respectively. The profiles at B are more blurred than the profiles at A due to the additional global illumination blur. Figures 3(k–l) show discrete-time Fourier transform $E(\omega, f)$ of (i) and (j). Figure 3(m)

shows the plot of $E(3, f)$ for points A (no global illumination) and B (with global illumination). The global illumination blur $G(w, f)$ is computed by taking the point-wise ratio of the two curves, according to (8). Figure 3(n) shows the plot of scaled $G(w, f)$ for $w = 1, 2, 3, 4$ at point B. Figures 3(o–p) show global illumination blur kernels and defocus blur kernels computed at different focus settings.

Two observations can be made from the plots. *First*, as shown in Fig. 3(n), the total variation in $G(w, f)$ for $w = 1, 2, 3$ is less than 5% over the entire range of focal plane positions (0.3 m–2.5 m).⁴ Figures 3(o–p) illustrate the global illumination blur kernels and the defocus blur kernels at different focus settings, computed using the first three DFT coefficients. The global illumination kernels remain nearly constant, while the defocus kernels show significant variation. This validates that the global illumination blur resulting from inter-reflections is insensitive to the projector focus setting. *Second*, we observe that the plots for $E(w, f)$, with and without global illumination, achieve maxima at the same focal plane position, as shown in Fig. 3(m).

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 Fig. 4(a). The bottom (green) part of the candle also receives inter-reflections from the base on which the candle is kept. 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 $E(w, f)$ for A and B in Fig. 4(i). As before, $G(w, f)$ at B is computed by taking the point-wise ratio of the two curves. We observe similar results as in the case of inter-reflections: the low frequency components of the global illumination blur kernel remain nearly constant as the projector focus setting is changed significantly.

From a practical point of view, since the first three coefficients remain nearly constant with changing focus setting, any function of these can be used as a defocus invariant measure of the global illumination blur. For our techniques, we use the third coefficient of the DFT ($w = 3$). 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.

⁴For coefficients corresponding to higher frequencies ($w \geq 4$), the variation in $G(w, f)$ is large. This can be explained from the fact that the global illumination blur kernel is defined in terms of ratios of defocus blur kernels (7). Since defocus blur kernels are low pass-filters, they suppress large frequencies. This can be observed in the DFT plots of the intensity profiles (Fig. 3(k–l)). The amplitude is nearly zero for $w \geq 4$. Consequently, the estimation of $G(w, f)$ is unstable and unreliable for $w \geq 4$. While it is possible to estimate $G(w, f)$ for higher frequencies by using coded apertures for projectors and designing broadband defocus blur kernels, currently it is beyond the scope of this paper.

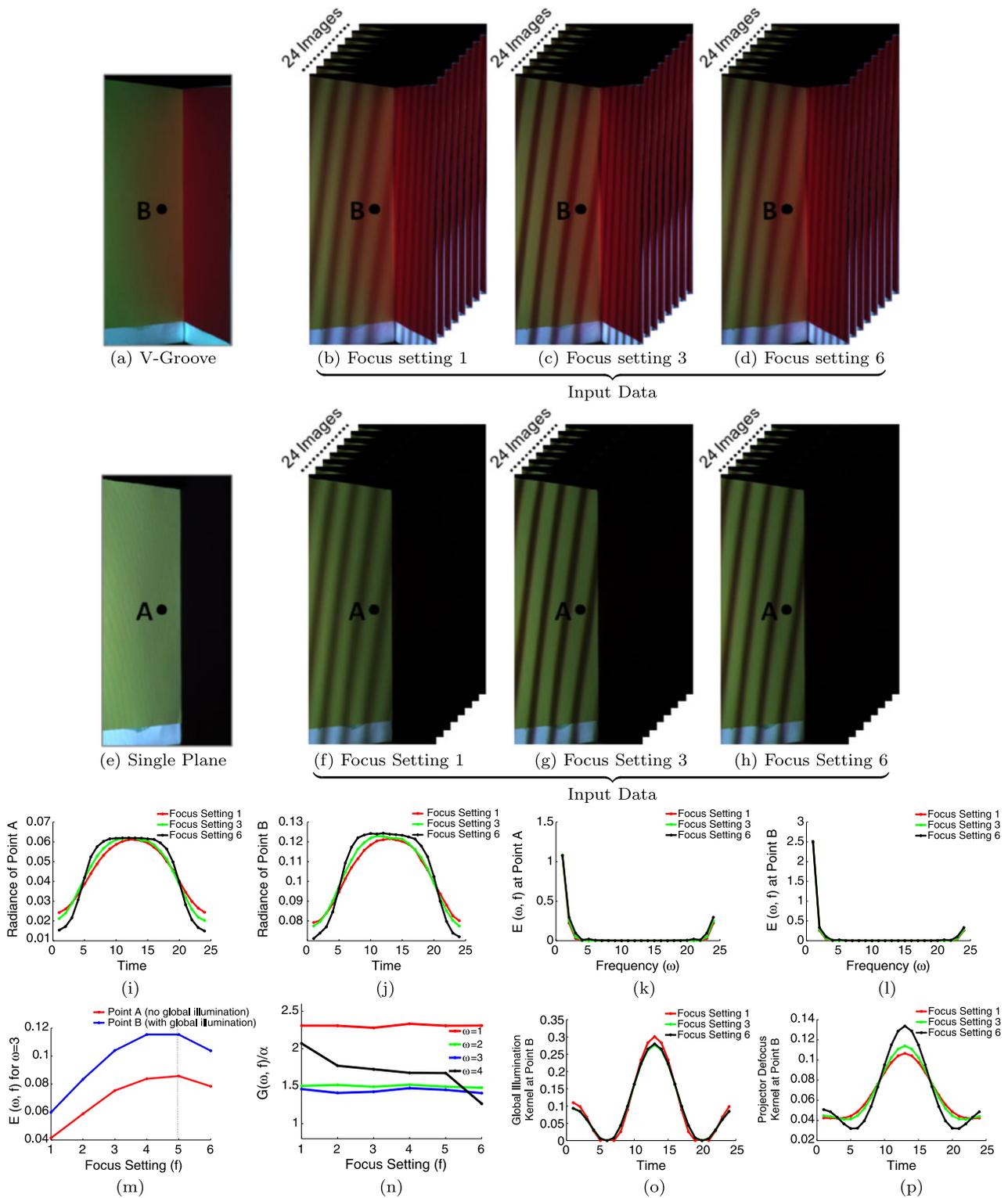


Fig. 3 Experiments to show the invariance of global illumination blur (due to inter-reflections) to the projector focus setting. (a) A V-groove constructed by placing two planes. (b–d) Sample input images for three out of six focus settings. (e) A single plane after removing the right, red colored plane from the V-Groove. (f–h) Sample input images for three out of six focus settings. (i–j) Temporal intensity profiles at Point A and B respectively for one out of six focus settings. (k–l) Discrete-time Fourier transform $E(\omega, f)$ of (i) and (j). (m) Plot of $E(\omega, f)$ for $\omega=3$

Points A (no global illumination) and B (with global illumination). (n) Plot of scaled $G(w, f)$ for $w = 1, 2, 3, 4$ at point B. The relative variation in $G(w, f)$ is less than 5% for $w = 1, 2, 3$. (o) Global illumination blur kernels computed at different focus settings using the first three frequency components. (p) Defocus blur kernels at different focus settings. The global illumination kernels remain nearly constant, while the defocus kernels show significant variation

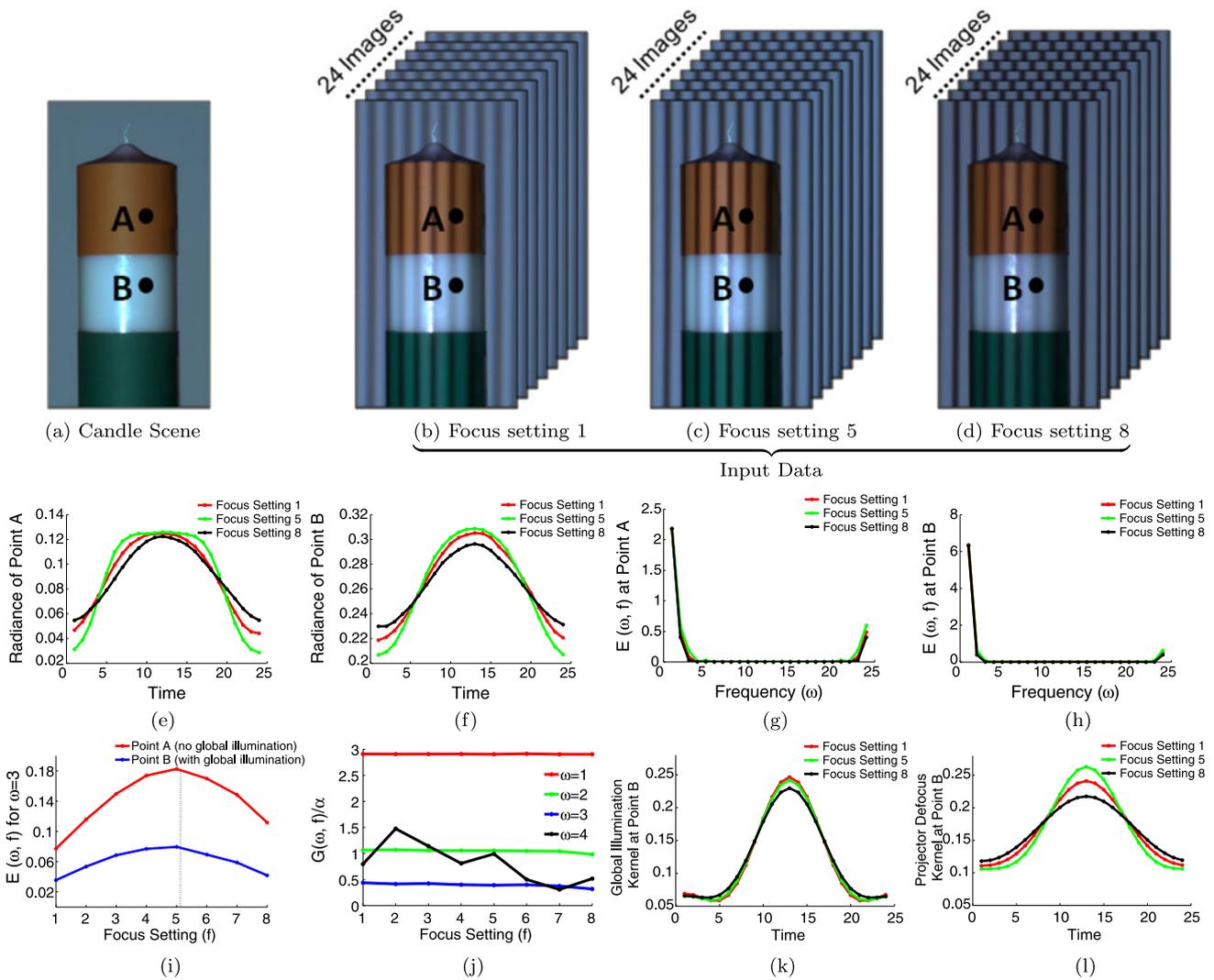


Fig. 4 Experiments to show the invariance of global illumination blur (due to sub-surface scattering) to the projector focus setting. (a) Candle scene. (b–d) Sample input images for three out of eight focus settings. (e–f) Temporal intensity profiles at Point A and B respectively for one out of six focus settings. (g–h) Discrete-time Fourier transform $E(\omega, f)$ of (e) and (f). (i) Plot of $E(3, f)$ for Points A (no global illumination) and B (with global illumination). (j) Plot of scaled $G(w, f)$

at point B for $w = 1, 2, 3, 4$. This is computed by taking the ratio of the two curves in (i) as in (8). The relative variation in $G(w, f)$ is less than 7% for $w = 1, 2, 3$. (k) Global illumination blur kernels computed at different focus settings using the first three frequency components. (l) Defocus blur kernels at different focus settings. The global illumination kernels remain nearly constant, while the defocus kernels show significant variation

5 Depth Recovery Under Global Illumination

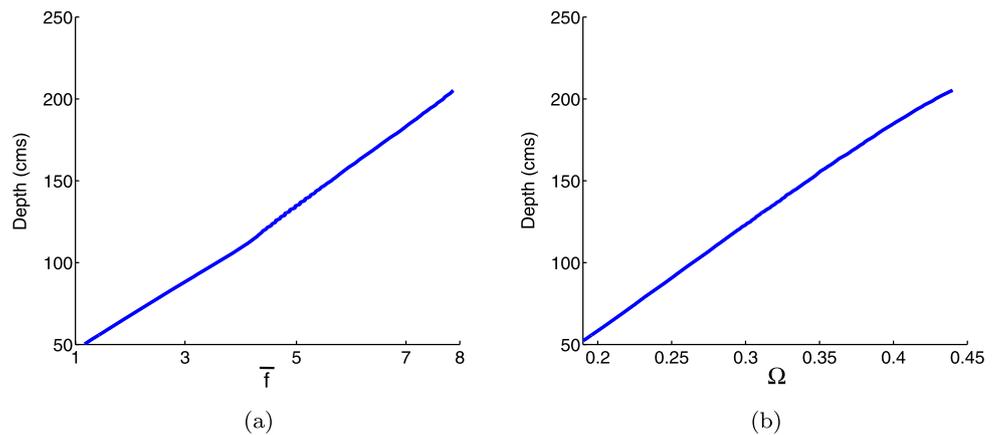
Based on the invariant derived in the previous section, we present two algorithms for recovering depths in the presence of global light transport. The first algorithm (Sect. 5.1) requires a sweep of the focal plane across the scene, acquiring images at multiple focus settings. Recall that the blur in the intensity profile measured at a single focal plane setting is a convolution of the defocus blur and the global illumination blur, both of which are unknown. Thus, we need intensity profiles for *at least two* focal settings in order to separate the two blur kernels. The second algorithm

(Sect. 5.2) requires capturing images at only two focus settings.

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. Assuming that the defocus blur kernel $B(f)$ is *unimodal*, the plot of $B(f)$ attains a unique maximum at the focal plane position f , when the corre-

Fig. 5 Mappings between (a) scene depths and the focus measure \bar{f} , (b) scene depths and the defocus measure Ω



sponding scene point is the best in focus.⁵ Since the global illumination blur $G(f)$ is invariant to f , the plot of $E(f)$ reflects the behavior of the defocus blur $B(f)$ (see (6)). Consequently, $E(f)$ and $B(f)$ attain a unique maximum at the same focal plane location f . It follows that the plots of $E(f)$ for two different scene points at the same depth but receiving different amounts of global illumination share the same maximum location. Two examples are shown in Figs. 3(m) and 4(i). This suggests the maximum location \bar{f}_i as a global-illumination invariant depth measure:

$$\bar{f}_i = \arg \max_f E_i(f). \quad (9)$$

The resolution of the above depth measure is limited by the number of focal settings used. The resolution can be improved by assuming that the blur kernel is smooth and that it can be well approximated with a Gaussian. Then, the discrete focus measure values $E_i(f)$ can be interpolated by fitting a Gaussian (Nayar and Nakagawa 1994). As a one time calibration step, we compute a one-to-one mapping between scene depths and \bar{f}_i using a planar, diffuse reflective board, whose depths are known a priori (see Fig. 5(a)). This mapping, along with the estimates of \bar{f} , is used to compute the actual depths. 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 estimate 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 (6), we define the following ratio measure which is invariant to global illumination:

$$\Omega_i = \frac{E_i(f_2)}{E_i(f_1)} = \frac{B_i(f_2)}{B_i(f_1)}. \quad (10)$$

⁵If the projector has a coded aperture (Levin et al. 2007), the defocus blur kernel might be multi-modal and $B(f)$ might attain multiple maxima. In this paper, we consider unimodal defocus blur kernels.

We compute a mapping between scene depths and Ω_i using a planar calibration board, as shown in Fig. 5(b). This mapping, along with the estimates of Ω is used to estimate the actual depths for a given scene.

5.3 Results

Figure 6 shows results of our techniques for the V-groove and the candle scenes. The single focal plane algorithm (Zhang and Nayar 2006) over-estimates the defocus blur resulting in incorrect depth estimates near the concavity of the V-groove (inter-reflections), for the middle (exposed) part of the candle (sub-surface scattering) and for the bottom (green) part of the candle (inter-reflections from the base). Our depth from two planes (Sect. 5.2) and multiple planes (Sect. 5.1) algorithms significantly mitigate the errors. The residual errors on the middle (exposed) part and the bottom (green) part of the candle are due to the approximate nature of the invariant. The relative RMS error for both our algorithms is $\sim 1\%$ for the V-groove and $\sim 5\%$ for the candle.

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 Fig. 6 were acquired using a calibration plane with pre-measured depths. In the following, we demonstrate our algorithms on a variety of scenes with complex shapes and material properties, and significant global illumination.

Candle and Marble Scene (Fig. 7) This scene consists of 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 single focal plane algorithm (Zhang and Nayar 2006) does not account for global light transport effects. Notice the incorrect depths, most noticeably on the candle and inside the V-groove concavity. The apparent details on the two statues are also spurious; they appear due

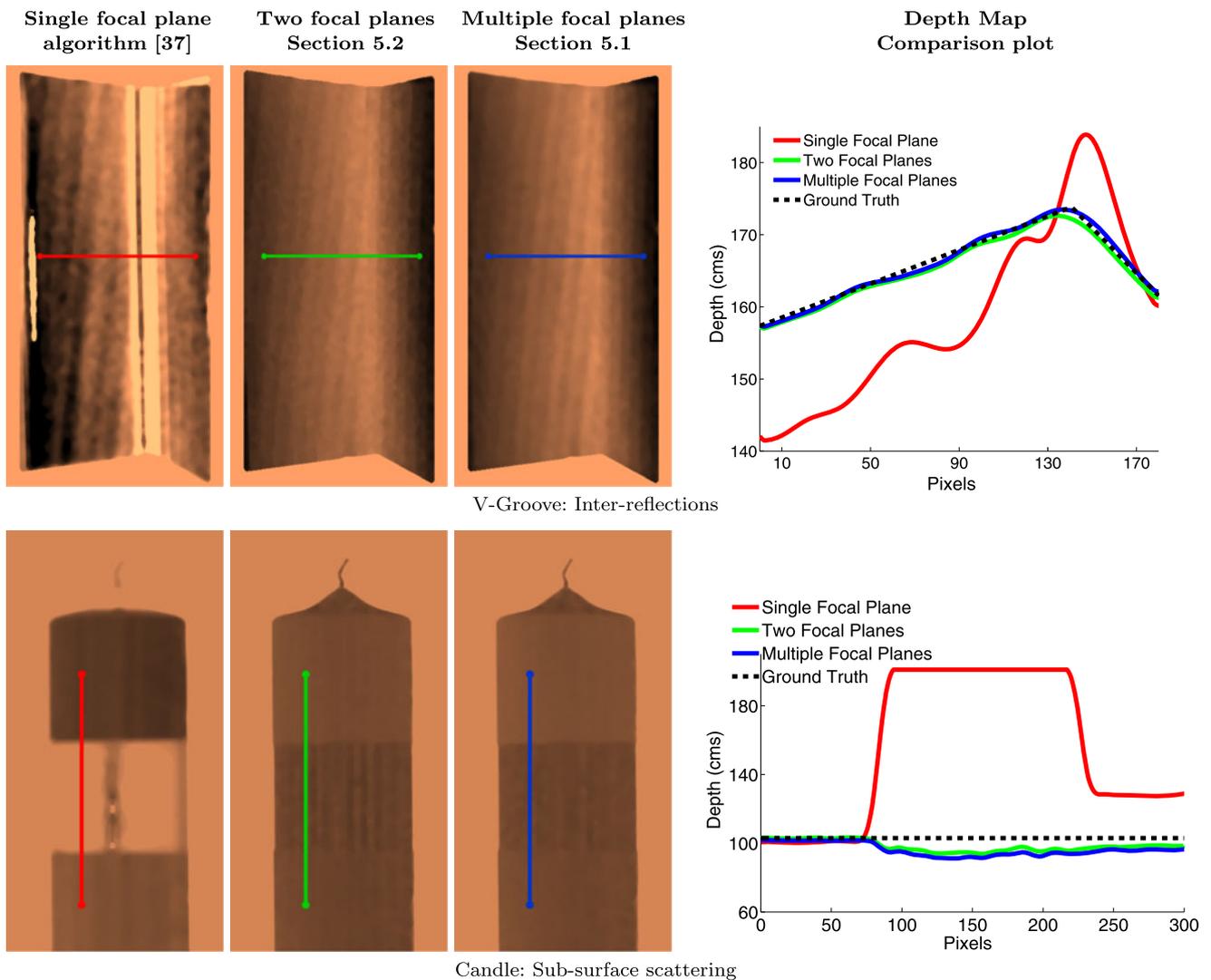


Fig. 6 (Color online) Comparison of the three depth recovery techniques for the V-groove (Fig. 3) and the candle (Fig. 4) scenes. The single focal plane algorithm (Zhang and Nayar 2006) over-estimates the defocus blur in the presence of inter-reflections and sub-surface scattering. This results in incorrect depth estimates near the concavity of the V-groove (inter-reflections), for the middle (exposed) part of the

candle (sub-surface scattering) and for the bottom (green) part of the candle (inter-reflections from the base). 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

to inter-reflections between the folds on the statues. In contrast, on the depth maps computed using our techniques, the errors due to global illumination are significantly reduced.⁶

Synthetic Materials Scene (Fig. 8) This scene consists of objects with complex and anisotropic BRDF’s (metal, velvet and fur) and intricate shapes. The single focal plane algorithm computes incorrect depths at the base of the objects

due to inter-reflections. Notice the sharp variation in depth at the base of the red-cylinder. The correct depth map should have a smooth depth transition, as can be noticed on depth maps computed using our techniques. Similarly, in the scene consisting of various industrial parts (Fig. 11), the depth map computed using the single focal plane algorithm has errors due to sharp inter-reflections and different material properties. In the depth map computed using our technique, the errors are significantly reduced.

⁶The 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.

Real and Fake Materials Scene (Fig. 9) This scene consists of real and fake flowers, real and fake fruits and milk with different fat content. Objects present in this scene exhibit

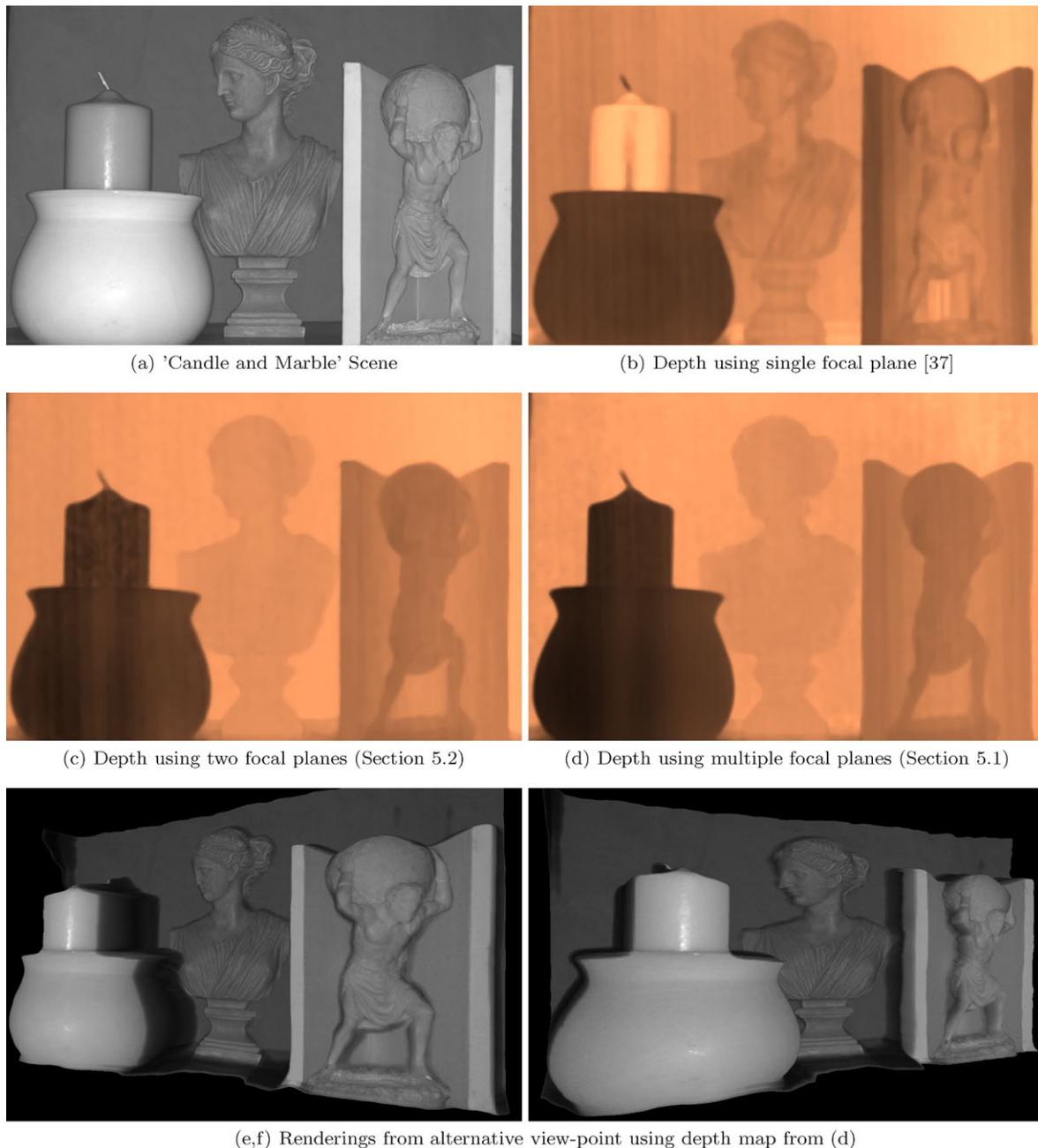


Fig. 7 Depth computation for the 'Candle and Marble' scene. (a) This scene consists of a wax candle inside a white pot closest to the projector, a marble statue, a V-groove and a polyresin bust, in this depth order. There is significant global light transport in form of sub-surface scattering (candle and the marble statue) and inter-reflections (inside the v-groove and between the folds on the statues). (b) Depth map using the single plane algorithm (Zhang and Nayar 2006). Notice the incor-

rect depths, most noticeably on the candle and inside the V-groove concavity. The apparent details on the two statues are also spurious; they appear due to inter-reflections between the folds on the statues. (c, d) Depth maps using our two focal planes and multiple focal planes algorithms respectively. The errors due to global illumination are significantly reduced. (e–f) Texture-mapped 3D model of the scene computed using (d)

varying degrees of sub-surface scattering due to different material properties. The single focal plane algorithm does not account for different material properties, thus computing incorrect depths. For instance, in the correct depth map, the two milk glasses should have the same depths, which is the

case with our results.⁷ Similar effect can be noticed in the

⁷Points on the boundary of the lemons are in attached shadow as the normals face away from the illumination direction. This results in incorrect depth estimates and the bright 'halos'.

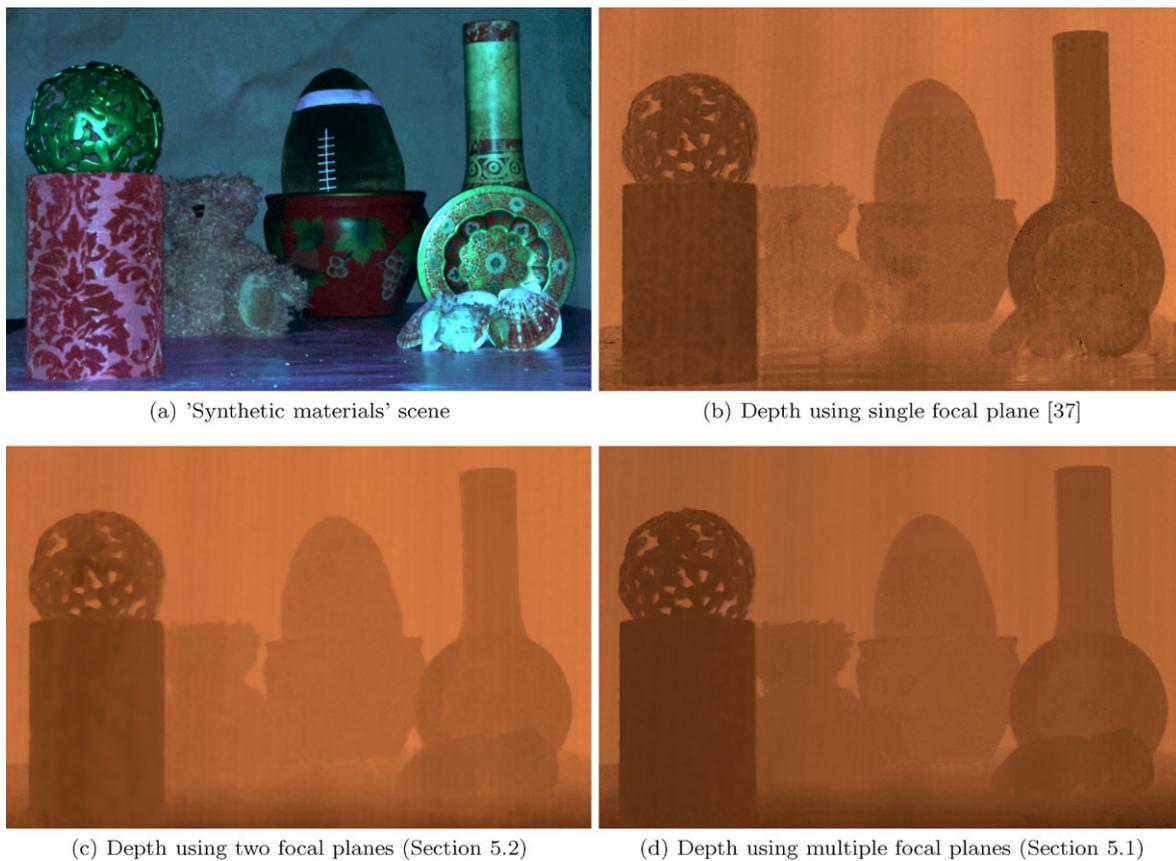


Fig. 8 (Color online) *Depth computation for the 'Synthetic materials' scene.* (a) This scene consists of objects with complex and anisotropic BRDF's (metal, velvet and fur) and intricate shapes. (b) The single focal plane algorithm computes incorrect depths at the base of the ob-

jects due to inter-reflections. Notice the sharp variation in depth at the base of the *red-cylinder*. The correct depth map should have a smooth depth transition. (c, d) Depth maps computed using our techniques

candles and soaps scene (Fig. 10), where all the objects are placed at roughly the same depth. However, the single focal plane algorithm computed significantly different depths due to different material properties. In comparison, the depth variation in the results computed by our algorithms is significantly smaller.

Failure Case As mentioned in Sect. 4, the presence of a distant mirror in a scene violates the assumptions that are made to derive the invariance of the global illumination blur to projector defocus. An example is illustrated in Fig. 12. The pot receives strong, specular inter-reflections from a distant mirror (not visible in the image). Consider a point S_i on the pot which receives inter-reflections from a point S_j on the mirror. In the following, we analyze the global illumination blur kernel at S_i using (7).

For S_i and S_j , the defocus kernels B_i and B_j vary differently as the projector focus setting f is changed. Consequently, the ratio $\frac{B_j(w, f)}{B_i(w, f)}$ varies as the projector focus setting f is changed. Since the inter-reflections are from a mirror surface, the form factor m_{ij} is not negligible despite the distance between S_i and S_j being large. Thus, in the summation

(7), the contribution from S_j is significant. Finally, the frequency with which the form factors m_{ij} vary in a neighborhood around S_j is much more than the frequency of the incident illumination. Thus, the contributions of points within a neighborhood around S_j do not cancel out among themselves. As a result, the global illumination blur kernel at S_i is not invariant to f .

In this case, our techniques fail to completely account for the errors due to global illumination. Depth computed using the single focal plane method is incorrect for points on the pot which receive specular inter-reflections. Although the errors in the depth maps computed using our techniques are mitigated, they are not completely removed. Note that the projector and camera were not co-located in this experiment and depth computation was not performed in the shadow regions.

6 Direct-Global Separation Under Defocus

The algorithm proposed in Nayar et al. (2006) separates the direct and global components of light transport with a sin-



Fig. 9 Depth computation for the 'Real and fake materials' scene. (a) This scene consists of real and fake flowers, real and fake fruits and milk with different fat content. Materials present in this scene exhibit varying degrees of translucency. (b) The single focal plane algorithm

does not account for different material properties, thus computing incorrect depths. In the correct depth map, the two milk glasses should have the same depths. (c, d) Depth maps computed using our techniques

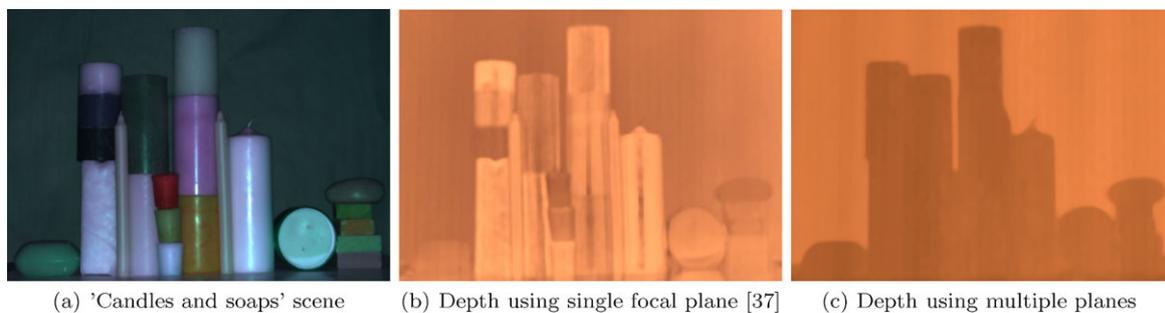


Fig. 10 Depth computation for the 'Candles and soaps' scene. (a) This scene consists of soaps and candles with different material properties. (b) Depth map computed using the single focal plane algo-

rithm has significant errors; in the correct depth map, all the soaps and candles should have the same depths. (c) Depth map computed using our technique

gle projector focal plane position. This technique, however, does not take into account the effects of defocused illumination. In the presence of defocus blur, a single focal plane is not sufficient to recover the correct separation. 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 in addition to 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 Nayar et al. (2006), the

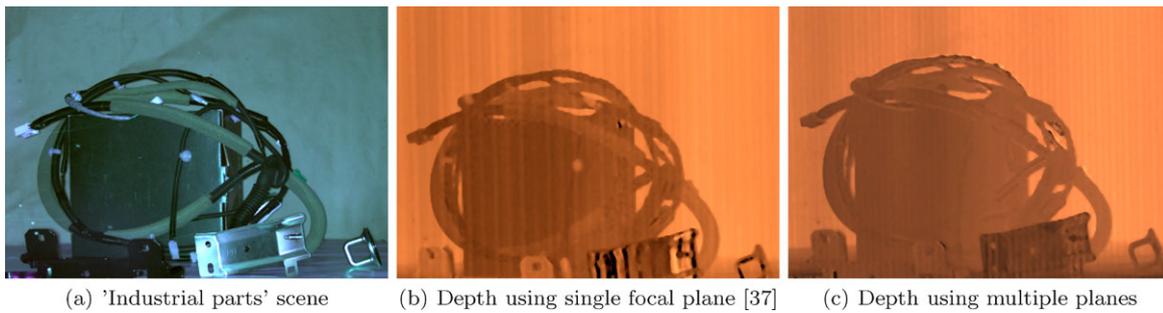


Fig. 11 Depth computation for the ‘Industrial parts’ scene. (a) This scene consists of different industrial parts. (b) Depth map computed using the single focal plane algorithm has errors due to sharp inter-

reflections and different material properties. (c) In the depth map computed using our technique, the errors are significantly reduced

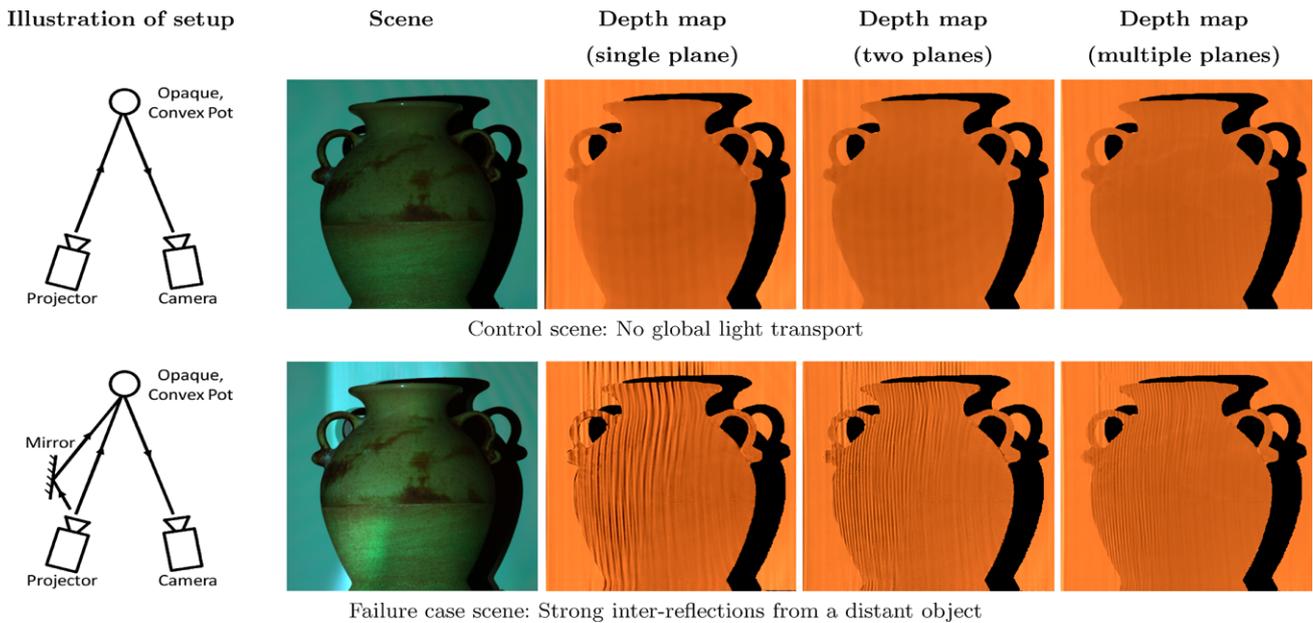


Fig. 12 Illustration of failure case: *Top row:* Control scene. In the absence of global light transport, all three techniques compute accurate depth maps. *Bottom row:* Failure case scene. The pot receives strong, specular inter-reflections from a distant mirror (not visible in the image). Depth map computed using the single focal plane method has

errors on scene points which receive specular inter-reflections. In this case, because of significant light transport among distant scene points, the global illumination blur is not invariant to the projector focus setting. Consequently, although the errors in the depth maps computed using our techniques are mitigated, they are not completely removed

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 (1) and (3), we write the expression for $e^+(f)$ in the presence of defocus:

$$e_i^+(f) = \beta_i^+(f)e_i^d + 0.5e_i^g, \tag{11}$$

$$\beta_i^+(f) = \max_t \{p_i(t) * b_i(t, f)\}, \tag{12}$$

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, $e^-(f)$:

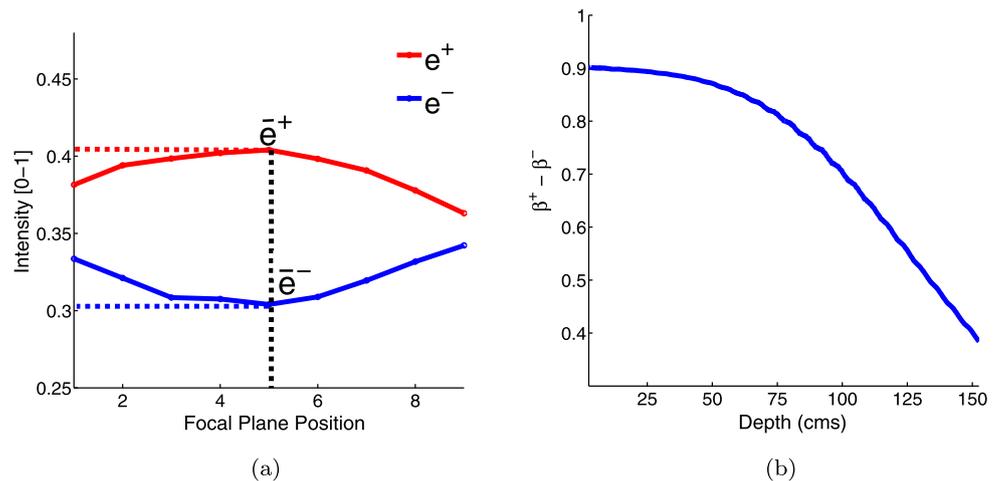
$$e_i^-(f) = \beta_i^-(f)e_i^d + 0.5e_i^g, \quad \text{where} \tag{13}$$

$$\beta_i^-(f) = \min_t \{p_i(t) * b_i(t, f)\}. \tag{14}$$

These equations are generalizations of the separation equations given in Nayar et al. (2006), 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 $\beta_i^-(f) = 0$.

Removing the Effects of Illumination Defocus The effects of defocus in the illumination space are different from the effects of defocus in the camera space. In the camera space, the support of the blur kernel for a camera pixel lies on a set of scene points, which can potentially be non-planar. Since

Fig. 13 (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 (15) and (16). (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 (17) and (18)



the scene's in-focus intensities are unknown, the blur kernel can not be computed a priori. In contrast, in the illumination space, the support of the blur kernel for a scene point lies on the planar image plane of the projector (Fig. 1). In the context of direct-global separation, the effect of illumination defocus can be expressed by only two scalars, $\beta_i^+(f)$ and $\beta_i^-(f)$. Additionally, these scalars depend only on the depth of S_i , and not on other scene points. Thus, $\beta_i^+(f)$ and $\beta_i^-(f)$ can be computed a priori as a function of scene depths. This makes it significantly simpler to remove the effects of defocus in the illumination space as compared to the camera space.

The two separation techniques that we present are different in the following important way. In the first approach, we compute an image of the scene as if the entire scene is receiving focused illumination. This requires sweeping the projector focal plane across the depth of the scene. This is dual to computing an *all-in-focus* image from a focal stack in the camera space. Once the all-in-focus images are computed, we simply apply the standard separation techniques as in Nayar et al. (2006). In the second approach, we exploit the fact that the effects of illumination defocus depend only on the scene depths, and thus can be pre-computed. Then, given a depth-map of the scene, the effects of defocus are calculated for each scene point individually, and then removed.

6.1 Separation Using Multiple Focal Planes

In this section, we present a separation technique using multiple focal plane positions. We use a checker-board illumination pattern as in Nayar et al. (2006). Input images (about 25)⁸ are acquired at different focus settings. Figure 15 shows sample input images for the V-groove scene for 3 out of 6

focus settings. For a point S_i , we compute $e_i^+(f)$ and $e_i^-(f)$ at each focus setting. As in the previous section, we assume that the defocus blur kernel is smooth and unimodal. Then the curve for $e_i^+(f)$ attains a unique maximum, while the curve for $e_i^-(f)$ attains a unique minimum. An example plot for a point on the candle is shown in Fig. 13(a). Also, we can use Gaussian interpolation to compute \bar{e}_i^+ and \bar{e}_i^- , the extrema values of $e_i^+(f)$ and $e_i^-(f)$ respectively. The computed images \bar{e}_i^+ and \bar{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:

$$\bar{e}_i^+ = e_i^d + 0.5e_i^g, \quad (15)$$

$$\bar{e}_i^- = 0.5e_i^g. \quad (16)$$

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

Accuracy of the Multiple Focal Plane Algorithm vs. the Sampling of the Focal Planes The accuracy of the multiple focal plane algorithm is a function of the illumination frequency (period) and the sampling of the focal plane space. We consider two cases. First, when the sampling of the focal planes is sufficiently fine so that for each scene point, the minimum defocus kernel support⁹ is smaller than half the illumination period (i.e., the defocus kernel lies completely within one square of the checkerboard pattern). In this case, $\beta_i^+(f) = 1$ and $\beta_i^-(f) = 0$ ((11) and (13)), and \bar{e}_i^+ and \bar{e}_i^- are the max and min intensities respectively as if S_i is in focus. Thus, the separation computed using the multiple focal planes algorithm is the same as the *ground-truth* separation (the ground-truth direct-global separation is when the scene is in focus).

⁸Theoretically, only 2 images are required. For robustness, we acquire multiple images.

⁹For a scene point, the size of the blur kernel support is the smallest for the focal plane position closest to the scene point.

Fig. 14 Accuracy of the multiple focal planes algorithm for direct-global separation vs. the number of focal planes. The average RMS error for the direct-global components computed using different number of focal plane positions. Direct-global separation computed using 7 focal plane positions is considered the ground truth

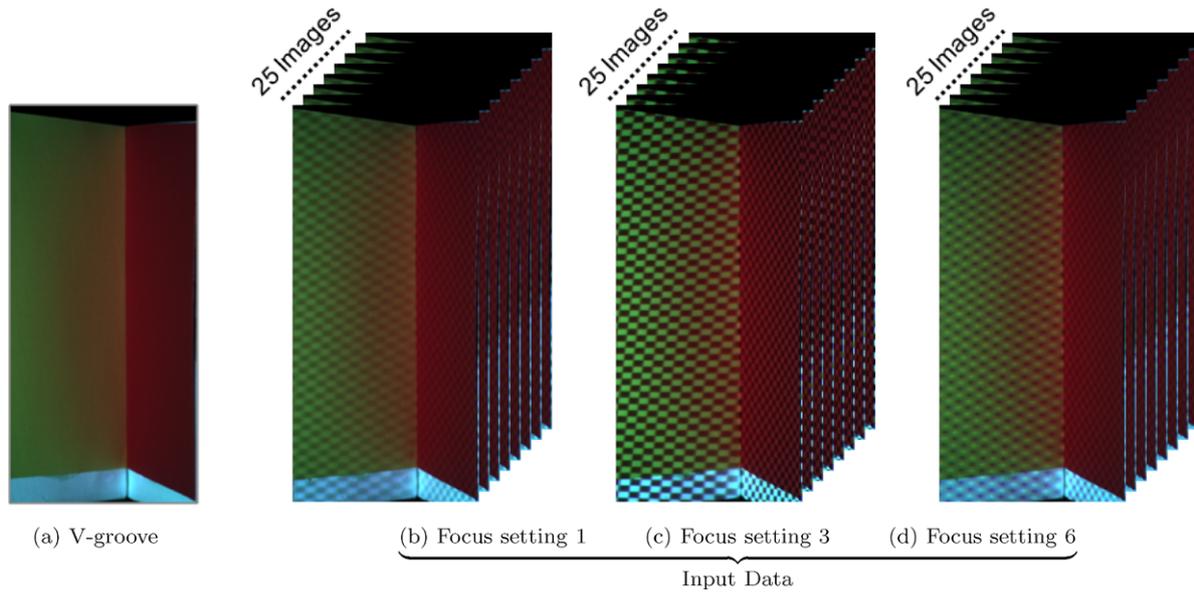
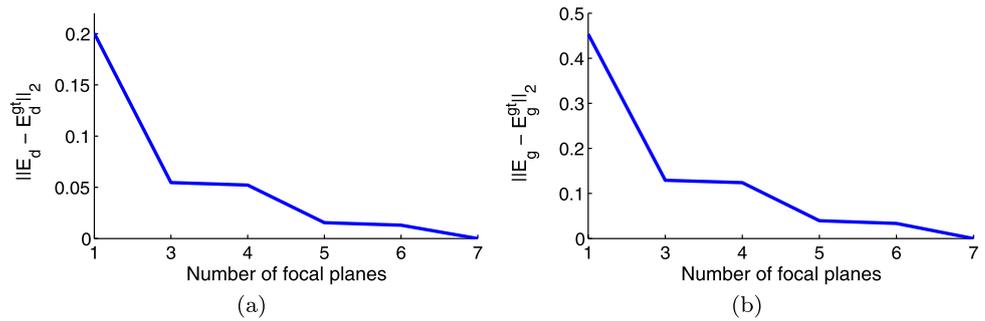


Fig. 15 Input images for direct-global separation. The multiple focal planes algorithm (Sect. 6.1) requires images captured at multiple (at least 3) focus settings. Shown above are sample input images at 3 out

of 6 different focus settings. The single focal plane + depth map algorithm (Sect. 6.2) requires images captured only at a single focal plane, and a depth map of the scene

Let N_b^{focal} be the number of evenly placed focal plane positions required to ensure that for each point, the minimum blur kernel support lies within half the illumination period. Let N_f^{focal} be the number of focal plane positions required to ensure that each point is perfectly in focus for some focal position. Then, $N_b^{focal} < N_f^{focal}$. Suppose D_{scene} is the depth of the scene and D_{dof} is the depth of field of the projector, then $N_b^{focal} < N_f^{focal} = \frac{D_{scene}}{D_{dof}}$. For example, if the depth of the scene is 150 cms and the depth of field of the projector is 30 cms, 5 focal plane positions are sufficient to achieve ground-truth separation.

The second case is when the focal plane sampling is coarse, i.e., the minimum support of the blur kernel is larger than half the illumination period. In this case, direct-global components are approximated by interpolation. The approximation error depends on the sampling frequency and how well the blur kernel fits the interpolation model. Since defocus blur kernels are frequently approximated with a Gaussian model, we used Gaussian interpolation.

Figure 18 shows the separation results computed using different number of focal plane samples. Qualitatively, all the results look similar, and are different from the result obtained using only a single focal plane (Fig. 17). Figure 14 shows a quantitative comparison—the average RMS error for the direct-global components computed using different number of focal plane positions. The depth of the scene is approximately 150 cms and the depth of field of the projector is approximately 30 cms. Direct-global separation computed using 7 focal plane positions is taken as the ground truth. As expected, the error falls sharply from 1 to 3 planes, and plateaus after 5 planes.

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 Sect. 5. For a scene point S_i , the direct and the global component are given using (11) and (13):

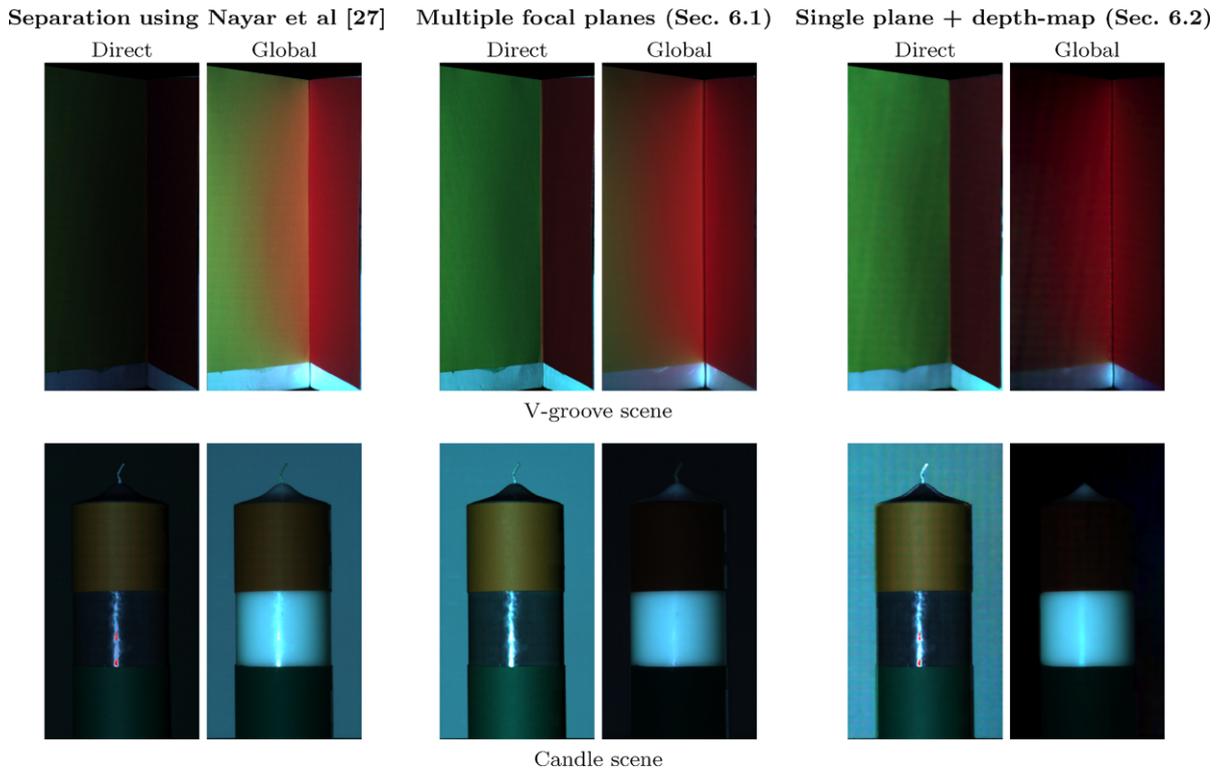
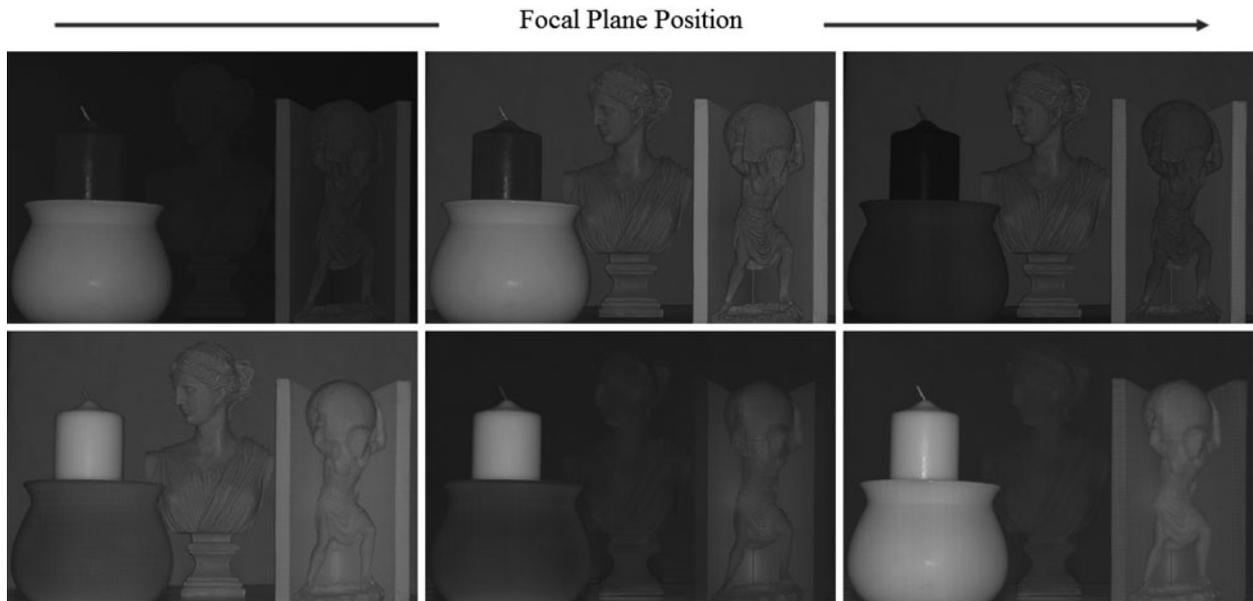


Fig. 16 Comparison of the three direct-global separation techniques. The technique in Nayar et al. (2006) does not account for illumination defocus and incorrectly estimates the direct and global components. The direct component is underestimated and the global component is over-estimated on the planes of the V-groove and on the background

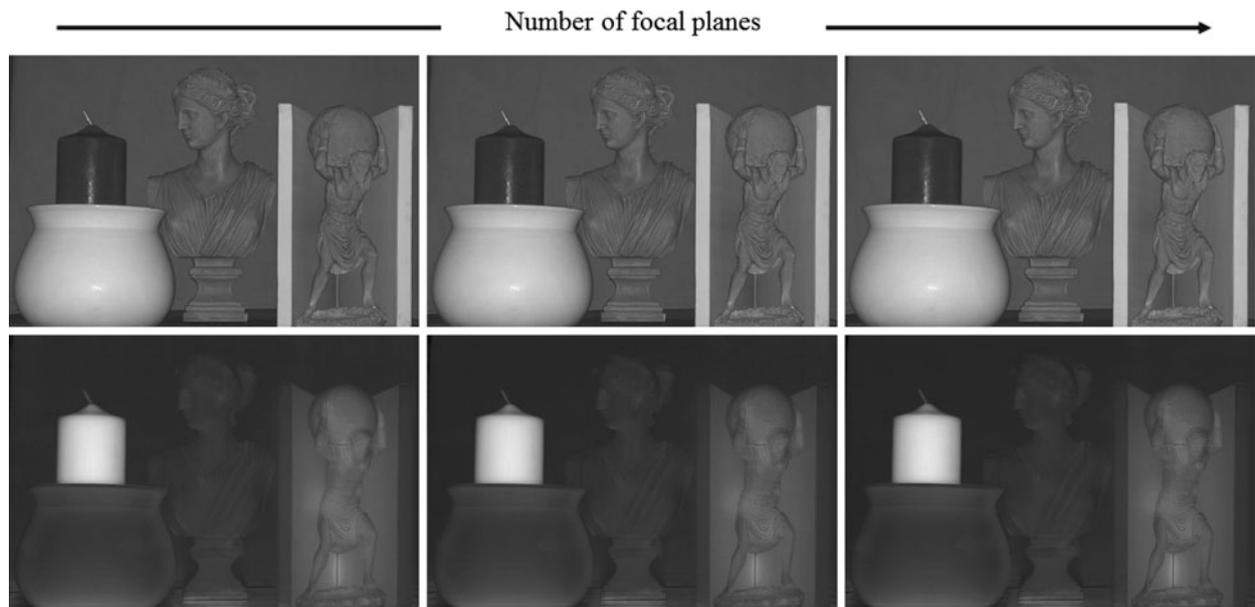
plane in the candle scene. Our techniques account for defocus while computing the direct-global separation. Notice the color-bleeding due to inter-reflections inside the V-groove and large global component on the exposed parts of the candle due to sub-surface scattering



(a) Direct (top) and Global (bottom) components computed at different focal plane positions using [27]. From left to right: projector focal plane in front of the scene, in the middle of the scene and behind the scene.

Fig. 17 Direct-Global separation using the technique in Nayar et al. (2006) for the marbles and candle scene. The depth of the scene is approximately 150 cms, larger than the depth of field of projectors (~30 cms). Thus, illuminating this scene with a projector results in defo-

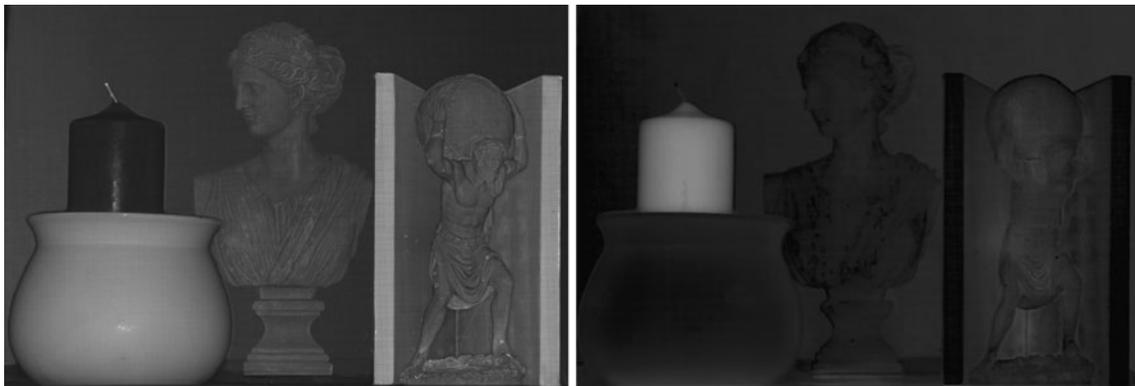
cused illumination. Because of defocus, the technique in Nayar et al. (2006) incorrectly produces different direct-global separation for different focus settings. At each focus setting, correct separation is produced only for the scene points which are in focus



(a) Direct (top) and Global (bottom) components computed using the multiple focal planes method (Sec. 6.1) From left to right: 3 focal planes, 5 focal planes and 7 focal planes.

Fig. 18 Direct-Global separation using our multiple focal plane algorithm (Sect. 6.1) for the marbles and candle scene. The multiple focal plane algorithm mitigates the errors due to defocus by computing all-in-focus images. The accuracy of the algorithm increases with the number of focal planes used. In this example, direct-global separation

computed using 7 focal plane positions is taken as the ground truth. Notice the large global component on the candle due to sub-surface scattering and inside the V-groove due to inter-reflections. Figure 14 shows a quantitative comparison



(c) Direct (left) and Global (right) components computed using single focal plane + depth map method (Sec. 6.2)

Fig. 19 Direct-Global separation using the single focal plane + depth map algorithm (Sect. 6.2) for the marbles and candle scene

$$e_i^d = \frac{e_i^+(f) - e_i^-(f)}{\beta_i^+(f) - \beta_i^-(f)}, \tag{17}$$

$$e_i^g = e_i - e_i^d, \tag{18}$$

where e_i is the observed intensity when the scene is fully lit. The denominator in (17) 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 $(\beta_i^+(f) - \beta_i^-(f))$ and scene depths using a flat diffuse inclined plane with known depths and no global illumination, as shown in Fig. 13(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:

$$\beta_r^+(f) - \beta_r^-(f) = \frac{e_r^+(f) - e_r^-(f)}{e_r}, \tag{19}$$

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 (17) with (19), to recover the direct and global components.

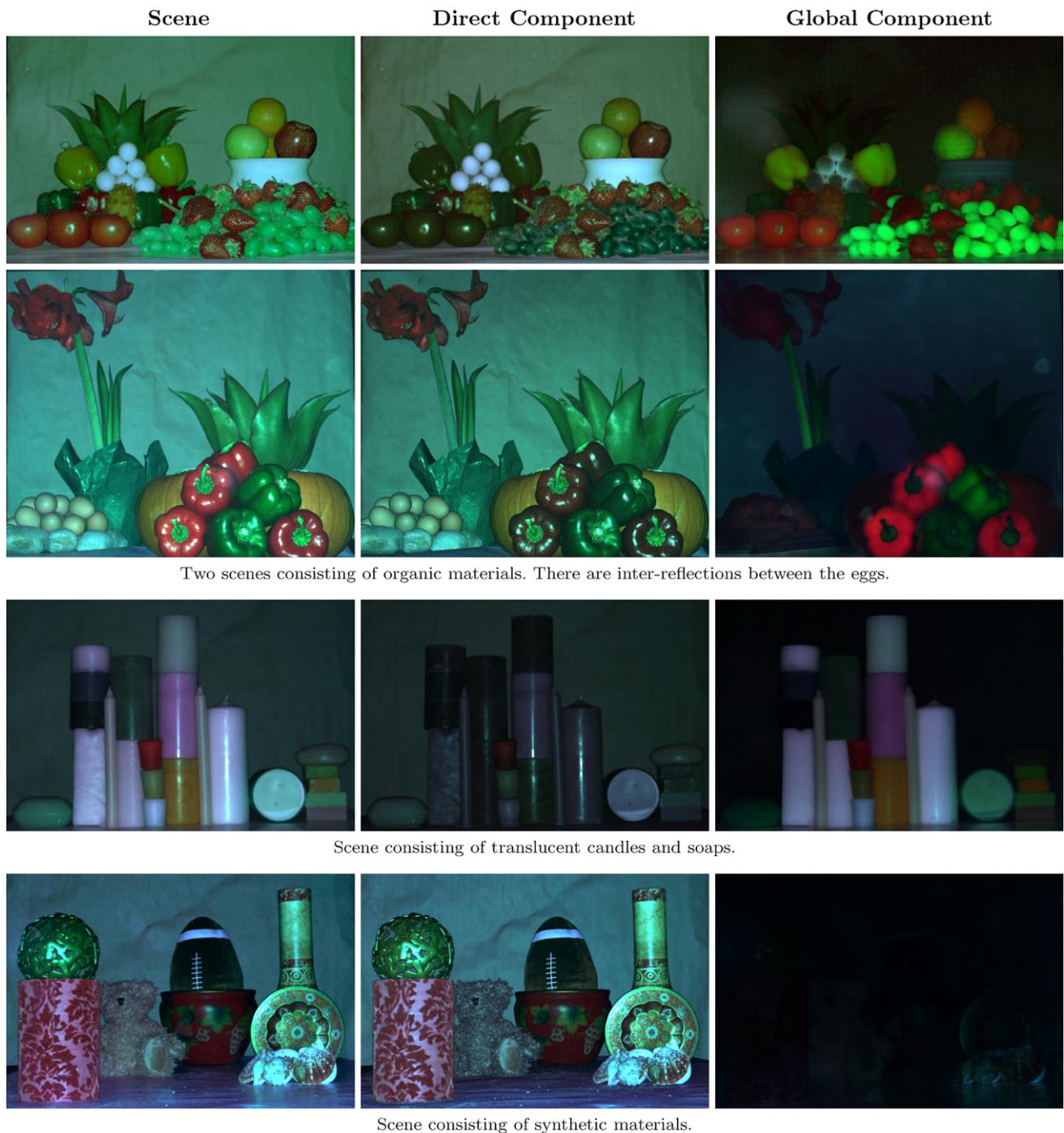


Fig. 20 Direct-Global separation for different scenes computed using the multiple focal planes algorithm (Sect. 6.1). The depth range of all the scenes is larger than 150 cms, more than the depth of pro-

jectors (~ 30 cms). For large depth scenes, it is critical to account for illumination defocus to compute the correct direct-global separation. See the project web-page (see footnote 10) for more results

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 checkerboard pattern with checkers of size 8×8 pixels. The pattern 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 16 shows the direct-global separation results for the candle and the V-groove scene. The focal plane was placed in front of the scene so that the objects are not in focus. The technique in Nayar et al. (2006) does not account

for illumination defocus and incorrectly estimates the direct and global components. The direct component is underestimated and the global component is over-estimated on the planes of the V-groove and on the background plane in the candle scene. In contrast, our techniques account for defocus while computing the direct-global separation. Notice the color-bleeding due to inter-reflections inside the V-groove and large global component on the exposed parts of the candle due to sub-surface scattering.

We also consider scenes with large depth variations (0.3 m–2 m), significantly more than the depth of field of the projector, as shown in Figs. 17–19. The technique in Nayar et al. (2006) produces different direct-global separation for different projector focus settings. This is incorrect since the direct-global separation is inherent to the scene, and should not depend on the projector focus setting. Our separation algorithms account for the defocus blur, and recover the correct direct and global components. Notice the large global component on the candle due to sub-surface scattering and inside the V-groove due to inter-reflections.

Figure 20 shows results for more scenes containing objects with a variety of material properties and different geometries. The depth range of all the scenes is larger than 150 cms, more than the depth of field of projectors (~30 cms). For more results and comparison, see the project web-page.¹⁰

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. Based on the invariant, we have shown two applications: First, accurate depth recovery in the presence of global illumination (sub-surface scattering and inter-reflections). Second, factoring out the effects of defocus for correct direct-global separation in large depth scenes. We now discuss some limitations of our approaches.

We have discussed defocused illumination in the context of projectors. However, illumination defocus is a more general effect which can be observed wherever area light sources are used. For example, in outdoor settings where sun is the illumination source, a vertical pole casts a shadow on the ground which is sharp near the base and gets blurred as we move away. This effect is similar to defocus observed with projectors. This suggests that sun can be used as an area light source for recovering outdoor scene properties. Sun has previously been used for outdoor scene recovery (Bouguet and Perona 1998; Nayar et al. 2006). However, illumination

defocus effects have not been considered, thus requiring the occluder to be very close to the scene.

We now discuss some limitations of our approaches. Our approaches do not handle perfectly mirrored objects due to high frequency global illumination. In the presence of specular reflections from mirrored objects, a scene point may receive global illumination from distant scene points. In this case, the global illumination blur is not invariant to the projector focus setting. As a result, our techniques do not fully account for the effects of global illumination, as shown in Fig. 12.

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.

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.

Acknowledgements This research is supported in parts by an ONR grant N00014-08-1-0330 and NSF awards CCF-0541307, IIS-0643628 and IIS-0845916. Parts of the results presented in this paper have previously appeared in Gupta et al. (2009).

Appendix: Validation Using Simulations

In this section, we verify the invariance of global illumination blur to projector focus settings using simulations in MATLAB. We compute $G(f)$ according to (7). To account for intensity fall-off (Hardy 1967), 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:

$$m_{ij} \propto \frac{1}{D_{ij}^2}. \quad (20)$$

For sub-surface scattering, the term m_{ij} encodes the additional exponential decay due to attenuation (Narasimhan and Nayar 2002):

$$m_{ij} \propto \frac{1}{D_{ij}^2} \exp(-D_{ij}). \quad (21)$$

For diffusion, we use the following expression for the form factor:

$$m_{ij} \propto \frac{1}{D_{ij}} \exp(-D_{ij}). \quad (22)$$

¹⁰Webpage: <http://graphics.cs.cmu.edu/projects/DefocusGlobal/>.

Fig. 21 (Color online) Simulation result for symmetric case. The *first row* is for the first DFT coefficient ($w = \pi/12$ in our setup). The *second row* is for the second DFT coefficient ($w = \pi/6$). For each row, the *first column* is the defocus kernel at that frequency, the *second column* is average global illumination blur over 100 simulations. *Blue curve* is for inter-reflection, *black* for sub-surface scattering and *red* for diffusion

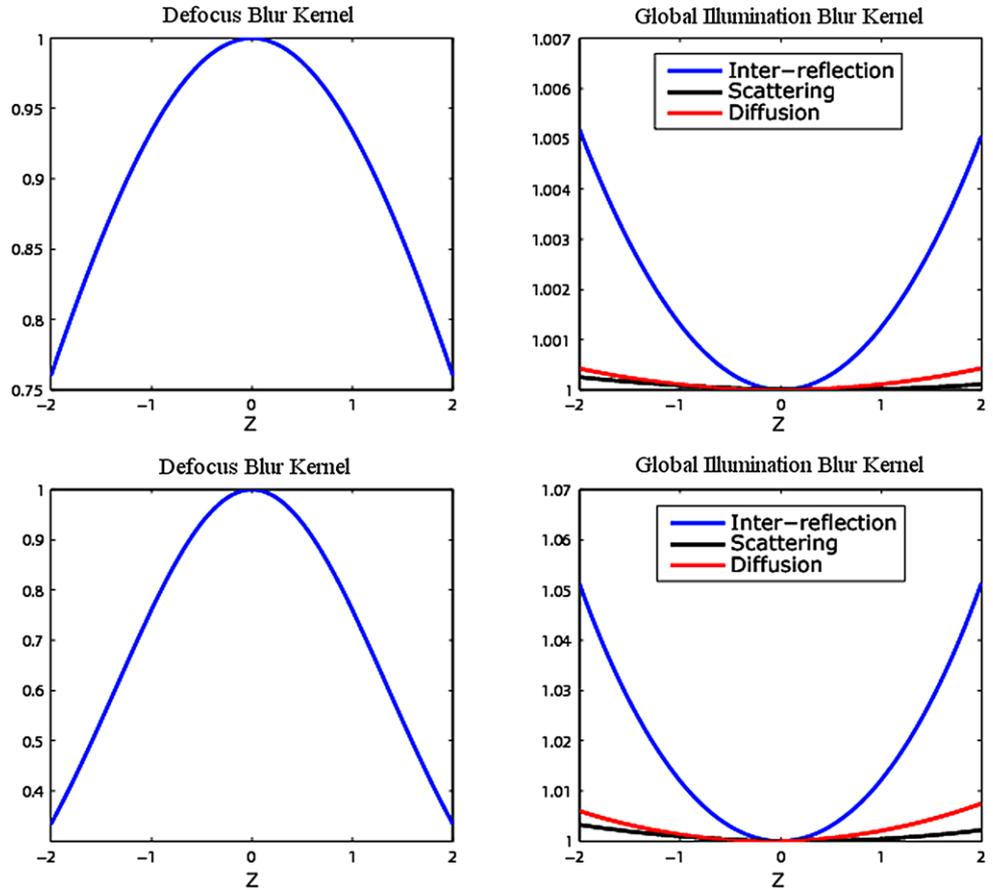
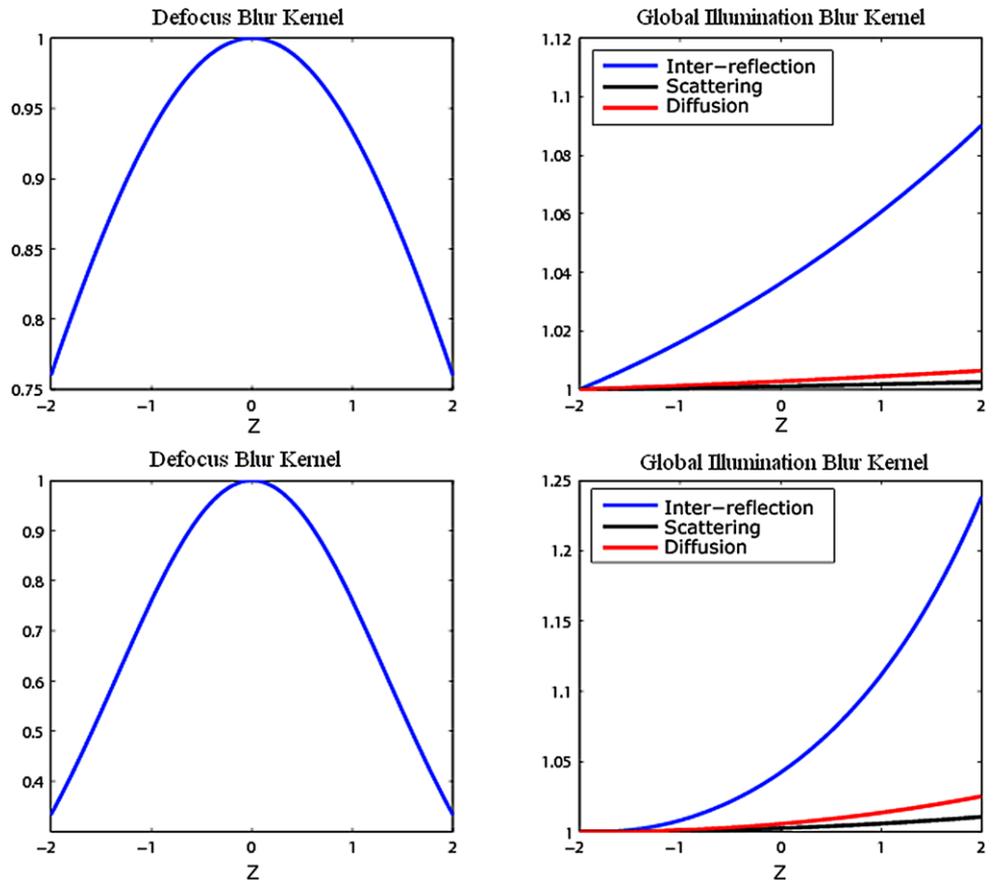


Fig. 22 Asymmetric case. See the caption of Fig. 21 for labels. The variation of geometry blur is higher than symmetric case, yet it still remains nearly constant over a large range of f



The scene is modeled as a 2D symmetric uniform distribution of points around point S_i , which is assumed to be at the origin. 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.

Results Figure 21 shows our simulation result. The global illumination blur has far less variation ($\sim 0.5\%$ for the inter-reflection case for the second DFT component) over f as compared to the defocus kernel (25–40 % variation). The variation is even lesser for sub-surface scattering and diffusion. This is because the form-factors for sub-surface scattering and diffusion fall-off much more rapidly with distance as compared to inter-reflections. Figure 22 shows results for a non-symmetric scene distribution. In this case, the plots are not symmetric around the origin. However, $G_i(w, f)$ still remains nearly constant over a large range of f .

References

- Atcheson, B., Ihrke, I., Heidrich, W., Tevs, A., Bradley, D., Magnor, M., & Seidel, H. (2008). Time-resolved 3d capture of non-stationary gas flows. In *SIGGRAPH*.
- Bai, J., Chandraker, M., Ng, T.-T., & Ramamoorthi, R. (2010). A dual theory of inverse and forward light transport. In *ECCV*.
- Bouguet, J.-Y., & Perona, P. (1998). 3d photography on your desk. In *Proc. IEEE ICCV*.
- Chen, T., Lensch, H. P. A., Fuchs, C., & Seidel, H.-P. (2007). Polarization and phase-shifting for 3d scanning of translucent objects. In *CVPR*.
- Chen, T., Seidel, H.-P., & Lensch, H. P. A. (2008). Modulated phase-shifting for 3d scanning. In *CVPR*.
- Fuchs, C., Heinz, M., Levoy, M., Seidel, H.-P., & Lensch, H. P. A. (2008). Combining confocal imaging and descattering. *Computer Graphics Forum*, 27(4), 1245–1253.
- Garg, G., Talvala, E.-V., Levoy, M., & Lensch, H. P. A. (2006). Symmetric photography: Exploiting data-sparseness in reflectance fields. In *EGSR*.
- Godin, G., Beraldin, J.-A., Rioux, M., Levoy, M., Cournoyer, L., & Blais, F. (2001). An assessment of laser range measurement of marble surfaces. In *Fifth conference on optical 3-D measurement techniques*.
- Gu, J., Nayar, S. K., Grinspun, E., Belhumeur, P. N., & Ramamoorthi, R. (2008). Compressive structured light for recovering inhomogeneous participating media. In *ECCV*.
- Gupta, M., Tian, Y., Narasimhan, S. G., & Zhang, L. (2009). (De) focusing on global light transport for active scene recovery. In *Proc. IEEE CVPR*.
- Hardy, A. (1967). How large is a point source? *Journal of Optical Society of America*, 57(1), 44–47.
- Hasinoff, S. W., & Kutulakos, K. N. (2006). Confocal stereo. In *ECCV (1)*.
- Hawkins, T., Einarsson, P., & Debevec, P. (2005). Acquisition of time-varying participating media. In *SIGGRAPH*.
- Horn, B. (1975). Obtaining shape from shading information. *The Psychology of Computer Vision* 19(1), 115–155.
- Hullin, M. B., Fuchs, M., Ihrke, I., Seidel, H.-P., & Lensch, H. P. A. (2008). Fluorescent immersion range scanning. In *SIGGRAPH*.
- Kutulakos, K. N., & Steger, E. (2008). A theory of refractive and specular 3d shape by light-path triangulation. *IJCV*, 76(1), 13–29.
- Levin, A., Fergus, R., Durand, F., & Freeman, B. (2007). Image and depth from a conventional camera with a coded aperture. In *SIGGRAPH*.
- Levoy, M., Chen, B., Vaish, V., Horowitz, M., McDowall, I., & Bolas, M. (2004). Synthetic aperture confocal imaging. In *SIGGRAPH*.
- Liu, S., Ng, T.-T., & Matsushita, Y. (2010). Shape from second-bounce of light transport. In *ECCV*.
- Morris, N. J. W., & Kutulakos, K. N. (2007). Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. In *ICCV*.
- Mukaigawa, Y., Yagi, Y., & Raskar, R. (2010). Analysis of light transport in scattering media. In *CVPR*.
- Narasimhan, S. G., Gupta, M., Donner, C., Ramamoorthi, R., Nayar, S. K., & Jensen, H. W. (2006). Acquiring scattering properties of participating media by dilution. *ACM Transactions on Graphics*, 25(3), 1003–1012.
- Narasimhan, S. G., & Nayar, S. K. (2002). Vision and the atmosphere. *IJCV*, 48(3), 233–254.
- Narasimhan, S. G., Nayar, S. K., Sun, B., & Koppal, S. J. (2005). Structured light in scattering media. In *Proc. IEEE ICCV* (pp. 420–427).
- Nayar, S., Ikeuchi, K., & Kanade, T. (1991). Shape from Interreflections. *IJCV*, 6(3), 173–195.
- Nayar, S. K., Krishnan, G., Grossberg, M. D., & Raskar, R. (2006). Fast separation of direct and global components of a scene using high frequency illumination. In *SIGGRAPH*.
- Nayar, S. K., & Nakagawa, Y. (1994). Shape from focus. *PAMI*, 16(8), 824–831.
- O’Toole, M., & Kutulakos, K. N. (2010). Optical computing for fast light transport analysis. In *Proc. SIGGRAPH Asia*.
- Schechner, Y. Y., & Kiryati, N. (2000). Depth from defocus vs. stereo: How different really are they? *IJCV*, 39(2), 141–162.
- Seitz, S. M., Matsushita, Y., & Kutulakos, K. N. (2005). A theory of inverse light transport. In *ICCV*.
- Sen, P., Chen, B., Garg, G., Marschner, S. R., Horowitz, M., Levoy, M., & Lensch, H. P. A. (2005). Dual photography. In *SIGGRAPH*.
- Subbarao, M., & Lu, M.-C. (1992). Computer modeling and simulation of camera defocus. In *Machine vision and applications* (pp. 277–289).
- Treibitz, T., & Schechner, Y. Y. (2006). Instant 3Descatter. In *Proc. IEEE CVPR* (Vol. 2, pp. 1861–1868).
- Watanabe, M., & Nayar, S. (1998). Rational filters for passive depth from defocus. *IJCV*, 27(3), 203–225.
- Woodham, R. (1980). Photometric method for determining surface orientation from multiple images. *OptEng*, 19(1), 139–144.
- Zhang, L., & Nayar, S. K. (2006). Projection defocus analysis for scene capture and image display. In *SIGGRAPH*.