

Fast Separation of Direct and Global Components of a Scene using High Frequency Illumination

Shree K. Nayar*
Columbia University

Gurunandan Krishnan†
Columbia University

Michael D. Grossberg‡
City University of New York

Ramesh Raskar§
MERL



(a) Scene



(b) Direct Component



(c) Global Component

Figure 1: (a) A scene lit by a single source of light. The scene includes a wide variety of physical phenomena that produce complex global illumination effects. We present several methods for separating the (b) direct and (c) global illumination components of the scene using high frequency illumination. In this example, the components were estimated by shifting a single checkerboard pattern 25 times to overcome the optical and resolution limits of the source (projector) and sensor (camera). The direct and global images have been brightness scaled by a factor of 1.25. In theory, the separation can be done using just 2 images. When the separation results are only needed at a resolution that is lower than those of the source and sensor, the separation can be done with a single image.

Abstract

We present fast methods for separating the direct and global illumination components of a scene measured by a camera and illuminated by a light source. In theory, the separation can be done with just two images taken with a high frequency binary illumination pattern and its complement. In practice, a larger number of images are used to overcome the optical and resolution limitations of the camera and the source. The approach does not require the material properties of objects and media in the scene to be known. However, we require that the illumination frequency is high enough to adequately sample the global components received by scene points. We present separation results for scenes that include complex interreflections, subsurface scattering and volumetric scattering. Several variants of the separation approach are also described. When a sinusoidal illumination pattern is used with different phase shifts, the separation can be done using just three images. When the computed images are of lower resolution than the source and the camera, smoothness constraints are used to perform the separation using a single image. Finally, in the case of a static scene that is lit by a simple point source, such as the sun, a moving occluder and a video camera can be used to do the separation. We also show several simple examples of how novel images of a scene can be computed from the separation results.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, texture; I.4.1 [Image Processing and Computer Vision]: Digitization and Image capture—Radiometry.

Keywords: direct illumination, global illumination, interreflections, subsurface scattering, volumetric scattering, translucency, coded illumination, image decomposition, image manipulation.

1 Introduction

When a scene is lit by a source of light, the radiance of each point in the scene can be viewed as having two components, namely, direct and global. The direct component is due to the direct illumination of the point by the source. The global component is due to the illumination of the point by other points in the scene. Consider the scene point P shown in Figure 2. The light ray A represents its direct illumination by the source and hence is the sole cause for the direct component of the radiance measured by the camera. The rays B, C, D and E are received by P from other points in the scene, and together they contribute to the global component of its radiance measured by the camera. These global illumination light rays are caused by different physical phenomena that are common in the real world. Ray B represents the interreflection of light between scene points; ray C results from subsurface scattering within the medium beneath the surface; ray D is due to volumetric scattering by a participating medium in the scene; and ray E is due to diffusion of light by a translucent surface.

It is highly desirable to have a method for measuring the direct and global components of a scene, as each component conveys important information about the scene that cannot be inferred from their sum. For instance, the direct component gives us the purest measurement of how the material properties of a scene point interact with the source and camera. Therefore, a method that can measure just the direct component can be immediately used to enhance a wide range of scene capture techniques that are used in computer vision and computer graphics. The global component conveys the complex optical interactions between different objects and media in the scene. We know that it is the global component that makes photo-

*e-mail: nayar@cs.columbia.edu

†e-mail: gkguru@cs.columbia.edu

‡e-mail: grossberg@cs.ccny.cuny.edu

§e-mail: raskar@merl.com

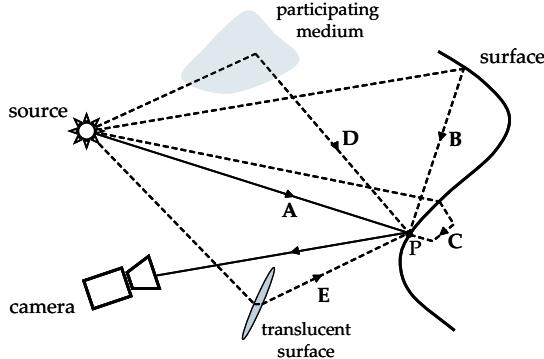


Figure 2: The radiance of scene point is due to direct illumination of the point by the source (A) and global illumination due to other points in the scene. The global illumination can arise from interreflections (B), subsurface scattering (C), volumetric scattering (D) and translucency (E). Separation of the direct and global components of measured radiance is useful as these components convey different properties of the scene.

torealistic rendering a hard and computationally intensive problem. A measurement of this component could provide new insights into these interactions that in turn could aid the development of more efficient rendering algorithms. Furthermore, measurement of the direct and global components can enable new types of image manipulations that are more faithful to the physical laws that govern scattering.

The goal of this paper is to develop efficient methods for separating the direct and global components of a scene lit by a single light source and viewed by a camera. One way to measure the global component is to illuminate each point of the scene independently and capture an image to determine the contribution of the point to all other points, as recently proposed in [Seitz et al. 2005]. While this approach is valid from a theoretical standpoint, it becomes prohibitively expensive for large and complex scenes. We show that the direct and global components at all scene points can be efficiently measured by using high frequency illumination patterns. This approach does not require the scattering properties of the scene points to be known. We only assume that the global contribution received by each scene point is a smooth function with respect to the frequency of the lighting. This assumption makes it possible, in theory, to do the separation by capturing just two images taken with a dense binary illumination pattern and its complement. In practice, due to the resolution limits imposed by the source and the camera, a larger set of images (25 in our setting) is used. We show separation results for several scenes that include diffuse and specular interreflections, subsurface scattering due to translucent surfaces and volumetric scattering due to dense media. We also show how the direct and global components can be used to generate new images that represent changes in the optical properties of the objects and the media in the scene.

We present several variants of our method that seek to minimize the number of images needed for separation. We show that by using a sinusoid-based illumination pattern, the separation can be done using just three images taken by changing the phase of the pattern. When the resolution of the camera and the source are greater than the desired resolution of the direct and global images, the separation can be done with a single image by assuming neighboring scene points to have similar direct and global components. In the case of just a simple point light source, such as the sun, the source cannot be controlled to generate the required high frequency illumination patterns. In such cases, the shadow of a line or mesh occluder can be swept over the scene while it is captured by a video camera. The captured video can then be used to compute the direct and global components. We conclude the paper with a discussion on several extensions of our approach that are planned for the future.

2 Related Work

There has been limited work on separating the direct and global illumination components of a scene from images. Classical shape-from-brightness algorithms, such as photometric stereo [Woodham 1980], do not account for global illumination due to interreflections and hence produce incorrect shape and reflectance estimates for scenes with concavities. For Lambertian surfaces, Nayar et al.[1991] analyzed the properties of the incorrect shape and reflectance produced by photometric stereo and showed that the actual shape and reflectance can be recovered from the incorrect ones. This recovery process implicitly separates the direct and global components of the scene. However, this approach is hard to extend to non-Lambertian scenes with complex geometries of the type we are interested in.

In the case of pure interreflections produced by any opaque surface, the direct and global components can be interpreted in a simple manner. The direct component is due to a single reflection at the surface, while the global component is due to multiple reflections. An interesting theoretical decomposition based on this interpretation was recently proposed by Seitz et al. [2005]. They also presented a method for estimating the interreflection contribution due to any given number of reflections. While the decomposition itself is applicable to surfaces with arbitrary BRDF, the method for estimating the decomposition is based on the Lambertian assumption. Moreover, this method requires a very large number of images to be acquired as it needs to know the photometric coupling between all pairs of scene points.

In order to separate the illumination components of arbitrary scenes, one needs to go beyond the realm of interreflections and be able to handle more complex phenomena such as subsurface scattering in translucent objects and volumetric scattering by participating media. A general approach to this problem is to estimate the dependence of the light field [Levoy and Hanrahan 1996; Gortler et al. 1996] of a scene on an arbitrary illumination field. This dependence is expressed as a linear transformation called a transport matrix. Due to its enormous dimensionality, estimation of the transport matrix requires a very large number of images and illuminations. Several techniques have been proposed to reduce the number of images by using coded illumination fields [Zongker et al. 1999; Debevec et al. 2000; Chuang et al. 2000; Lin et al. 2002; Peers and Dutré 2003; Zhu and Yang 2004; Shim and Chen 2005; Sen et al. 2005]. Even so, typically, several tens or even hundreds of images are needed to obtain acceptable estimates of the transport matrix. In our work, we do not aim to recover the entire transport matrix. Our goal is less ambitious – it is to separate the appearance of a scene lit by a single source into its direct and global components. In this setting, we show that the separation can be done with a very small number of images – under certain conditions, just a single image is sufficient.

It has been recently shown that by separately computing the direct and global components, the rendering of a synthetic scene can be sped up [Sloan et al. 2002; Arikan et al. 2005]. We have computed direct and global images for several scenes and these images include a variety of non-intuitive effects. We therefore believe that our separation methods could provide new insights that lead to more efficient rendering algorithms. Finally, we show several simple examples of how the separation results for a scene can be used to generate novel images that represent changes in the physical properties of the scene.

3 Theory of Fast Separation

3.1 Definitions for Direct and Global Illumination

Consider a surface viewed by a camera and illuminated by a point source, as shown in Figure 3(a). Let us assume that the source generates a set of illumination rays, each ray corresponding to a single source element, as in the case of a digital projector. We assume that each point of the surface could cause a significant scattering event in the direction of the camera if lit by the source. The radiance of a surface point measured by the camera due to such a scattering event is

referred to as the *direct component*, L_d . The exact value of the direct component is determined by the BRDF of the surface point, which can be arbitrary¹. For our separation method to work, we assume that each camera pixel can observe at most one significant scattering event, i.e. two different source elements cannot produce a direct component along a camera pixel's line of sight².

The remaining radiance measured by the camera pixel is referred to as the *global component*, L_g . In computer graphics, this term is typically used to denote interreflections – light received by a surface point after reflection by other scene points. Here, we are using a more general definition. In addition to interreflections, the global illumination received by the surface point may be due to volumetric scattering, subsurface scattering or even light diffusion by translucent surfaces (see Figure 2). The case of diffusion by a translucent surface works similarly to interreflections. In the case of volumetric scattering, the global component arises from the illumination of the surface point by light scattered from particles suspended in a participating medium. In the case of subsurface scattering, the surface point receives light from other points within the surface medium. Finally, the global component also includes volumetric and subsurface effects that may occur within the camera pixel's field of view but outside the volume of intersection between the fields of view of the pixel and the source element that produces a significant scattering event at the pixel. These are considered to be global effects as they are not significant scattering events caused by individual source elements³.

In all cases, the total radiance measured at a camera pixel is the sum of the direct and global components:

$$L = L_d + L_g. \quad (1)$$

3.2 The Nature of the Light Source

In our work, we restrict ourselves to the use of a single camera and a single source. While we will use a point source to describe our separation method, this is not a strict requirement. We only require that each point in the scene be directly illuminated by at most one source element. In other words, the light rays corresponding to the source elements should not intersect within the working volume of the setup used to perform the separation. Any source (point or extended) that satisfies this condition may be used.

3.3 Separation using High Frequency Illumination

Let us assume that the scene in Figure 3(a) includes a single opaque surface of arbitrary BRDF immersed in a non-scattering medium so that the global component arises solely from interreflections. As we will see, our analysis of this case is applicable to other phenomena such as subsurface and volumetric scattering.

Let us divide the surface into a total of N patches, M of which are directly visible to the source. Each of these M visible patches corresponds to a single pixel of the source. We denote the radiance of the patch i measured by the camera c as $L[c, i]$, and its two components as $L_d[c, i]$ and $L_g[c, i]$, so that $L[c, i] = L_d[c, i] + L_g[c, i]$. The global component of i due to interreflections from all other surface patches can be written as:

¹Since, in practice, cameras and sources have finite resolutions, the direct component is the aggregate of all the scattering that occurs within the volume of intersection between the fields of view of the camera pixel that observes the surface point and the source element that illuminates it.

²A scenario that violates this assumption is the case of a transparent (and yet reflective) surface in front of another surface, where a camera pixel's line of sight could produce two significant scatterings due to different source elements (pixels in the case of a projector).

³This claim does not hold true when the source and the camera are collocated. In this special case, the field of view of each camera pixel is illuminated by a single source element and hence the volumetric and subsurface scattering effects within the pixel's field of view will indeed be significant scattering events and hence appear in the direct component.

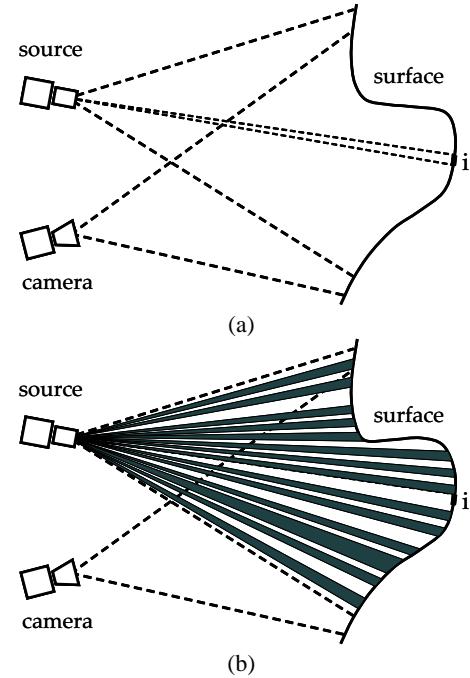


Figure 3: (a) A simple scenario where the radiance of each patch i includes a direct component due to scattering of light incident directly from the source and a global component due to light incident from other points in the scene. (b) When the source radiates a high frequency binary illumination pattern, the lit patches include both direct and global components while the unlit patches have only a global component. In theory, two images of the scene taken with such an illumination pattern and its complement are sufficient to estimate the direct and global components for all patches in the scene.

$$L_g[c, i] = \sum_{j \in P} A[i, j] L[i, j], \quad (2)$$

where, $P = \{j | 1 \leq j \leq N, j \neq i\}$. $L[i, j]$ is the radiance of patch j in the direction of patch i and $A[i, j]$ incorporates the BRDF of i as well as the relative geometric configuration of the two patches⁴. We can further decompose $L_g[c, i]$ into two components as $L_g[c, i] = L_{gd}[c, i] + L_{gg}[c, i]$, where $L_{gd}[c, i]$ is due to the direct component of radiance from all scene patches and $L_{gg}[c, i]$ is due to the global component of radiance from all scene patches:

$$L_{gd}[c, i] = \sum_{j \in P} A[i, j] L_d[i, j], \quad L_{gg}[c, i] = \sum_{j \in P} A[i, j] L_g[i, j]. \quad (3)$$

Now let us assume that only a fraction α of the source pixels are activated and that these activated pixels are well-distributed over the entire scene to produce a high frequency illumination pattern, as shown in Figure 3(b). The set of illuminated patches can be denoted as $Q = \{k | k \in N \text{ and } \text{lit}(k) = 1\}$, where the function lit indicates whether a patch is illuminated or not. Then, the above components become:

$$L_{gd}^+[c, i] = \sum_{j \in Q} A[i, j] L_d[i, j], \quad L_{gg}^+[c, i] = \sum_{j \in P} A[i, j] L_g^+[i, j]. \quad (4)$$

Note that $L_{gd}^+[c, i]$ differs from $L_{gd}[c, i]$ only in that the lit αM patches rather than all the M patches have a direct component and hence make a contribution. Therefore, if the geometry and reflectance term $A[i, j]$ and the direct component $L_d[i, j]$ are smooth with respect to the frequency of the illumination pattern, we have:

⁴Details on the form of $A[i, j]$ can be found in [Chandrasekhar 1950; Koenderink and van Doorn 1983; Kajiya 1986; Immel et al. 1986; Forsyth and Zisserman 1991; Nayar et al. 1991; Seitz et al. 2005]. In our work, the exact form of $A[i, j]$ is not relevant as it is not used explicitly during separation.

$$L_{gd}^+[c, i] = \alpha L_{gd}[c, i]. \quad (5)$$

A brief frequency domain analysis of the illumination frequency that makes the above relation valid is given in Appendix A.

Now, let us consider the second term, $L_{gg}^+[c, i]$. Since $L_g^+[i, j]$ in Equation (4) is the result of higher orders of interreflection than $L_{gd}^+[c, i]$, it is even smoother and hence less affected by the non-uniformity of the illumination. However, it is directly proportional to the average power of the illumination, which is reduced by α in the case of the high frequency pattern. Therefore, $L_g^+[i, j] = \alpha L_g[i, j]$ and we get:

$$L_{gg}^+[c, i] = \alpha L_{gg}[c, i]. \quad (6)$$

Consider two captured images of the scene, where, in the first image L^+ the scene is lit with high frequency illumination that has fraction α activated source pixels and in the second image L^- it is lit with the complementary illumination that has fraction $1 - \alpha$ activated source pixels. If the patch i is lit directly by the source in the first image then it is not lit by the source in the second image, and we get:

$$L^+[c, i] = L_d[c, i] + \alpha L_g[c, i], \quad L^-[c, i] = (1 - \alpha)L_g[c, i]. \quad (7)$$

Therefore, if we know α , we can compute the direct and global components at each camera pixel from just two images. Thus far, we have assumed that when a source pixel is not activated it does not generate any light. In the case of a projector, for example, this is seldom completely true. If we assume the brightness of a deactivated source element is a fraction b , where $0 \leq b \leq 1$, of an activated element, then the above expressions can be modified as:

$$\begin{aligned} L^+[c, i] &= L_d[c, i] + \alpha L_g[c, i] + b(1 - \alpha)L_g[c, i], \\ L^-[c, i] &= bL_d[c, i] + (1 - \alpha)L_g[c, i] + \alpha bL_g[c, i]. \end{aligned} \quad (8)$$

Again, if α and b are known, the separation can be done using just two images. Note that if α is either close to 1 or 0, the scene will be lit (sampled) very sparsely in one of the two images. Since we wish to maximize the sampling frequency of the illumination in both images, a good choice is $\alpha = \frac{1}{2}$. In this case, we get:

$$\begin{aligned} L^+[c, i] &= L_d[c, i] + (1 + b)\frac{L_g[c, i]}{2}, \\ L^-[c, i] &= bL_d[c, i] + (1 + b)\frac{L_g[c, i]}{2}. \end{aligned} \quad (9)$$

Based on the above results, we will develop a variety of separation methods. In each case, we will record a set of brightness values at each camera pixel and use L_{max} and L_{min} to denote the maximum and minimum of these values. In the above case of two images taken with $\alpha = \frac{1}{2}$, $L^+ \geq L^-$ and hence $L_{max} = L^+$ and $L_{min} = L^-$.

3.4 Generality and Limitations

While we have used a simple scene with just interreflections to describe the separation method, it is applicable to a wide range of scenarios. The direct component can include diffuse and specular reflections. The global component can arise from not just interreflections but also volumetric and subsurface scattering. In the presence of these scattering effects, the surface element j in the above equations represents voxels of intersection between the fields of view of camera and source pixels that are distributed in 3D space, rather than 2D surface patches.

In the case of volumetric scattering, as mentioned earlier, two effects are captured by the measured global component. The first is the illumination of each surface point by the participating medium. This effect works like interreflections and hence is included in the measured global component. The second effect is the brightness of the participating medium within the pixel's field of view. Consider the entire set of source rays that pass through the line of sight of a single camera pixel. In the first image, a fraction α of the rays will pass

through the line of sight and in the second image a fraction $1 - \alpha$ of the rays will pass through it. Therefore, if the illumination frequency is high enough, even if the medium is non-homogeneous, the second effect is also included in the measured global component.

In the case of subsurface scattering, the direct component is produced by the BRDF of the surface interface while the global component is produced by the BSSRDF (not including the BRDF of the surface interface) of the surface medium.

On the other hand, there are several extreme scenarios where the illumination frequency will not be sufficiently high for the separation method to produce accurate results. For example, if the scene includes highly specular or refractive surfaces, the surface point may be strongly illuminated from multiple directions and hence exhibit multiple significant scattering events. Furthermore, the global component may be underestimated due to undersampling by the illumination. In such cases, the separation results will include undesirable artifacts, as we will show in Section 4.

3.5 Verification

Before we describe our separation methods, we present several verification results on the use of high frequency illumination for estimating the global component. For these experiments, we have used the scene shown in Figure 4(a) (also shown in Figure 1) that includes a wide variety of physical phenomena. The scene was lit by a Sanyo PROExtraX digital projector (with 1024×768 pixels) and images were captured using a Point Grey Dragonfly camera (with 1024×768 pixels). Since the camera detector has a Bayer color mosaic, all separation calculations were done using the raw mosaiced images and then demosaicing was applied to obtain color information. Since this camera is noisy for our purposes, for each projected illumination we captured and averaged 32 images to reduce noise.

We present results for the points marked in Figure 4(a) which include diffuse interreflections on the white board (A), specular interreflections due to the nutshell (B), subsurface scattering in the marble object (C), subsurface scattering in the candle wax (D), translucency of the frosted glass (E), volumetric scattering by dilute milk in the plastic cup (F) and global illumination in a cast shadow (G).

In the first experiment, we studied the sensitivity of the global component to the size of the illumination patch used to construct the high frequency pattern. For this the scene was fully lit except for square patches of size $p \times p$ (p is in projector pixels) centered around the above 7 points. Each point therefore receives only global illumination⁵. In Figure 4(b), the global component (measured in the green channel and normalized by the value for a 3×3 patch) is plotted as a function of the unlit patch size p^6 . Note that between $p = 3$ and $p = 11$ the global estimate varies by less than 10% for all the samples except for the marble point. This demonstrates that if the illumination is sufficiently high in frequency a good separation can be obtained for a wide range of physical phenomena. In the case of the marble, we see a quick fall in the global component with p because the global component mainly arises from subsurface scattering which is very local due to the high density of the marble. Therefore, the marble represents a good example of a phenomenon that needs to be sampled with a higher illumination frequency than what our current system can produce or measure. For this reason, we do not include the marble point (C) in the subsequent experiments.

In the second experiment, we lit the scene with 100 illumination patterns. Each of these patterns was made of non-overlapping 6×6 patches, where exactly half the patches were activated but the placement of the activated patches was random. The above points of interest were kept unlit with 6×6 patches for all the illumination patterns

⁵The projector was calibrated to account for the effect of the inactive brightness factor b in Equation 8.

⁶We did not use a patch size of 1×1 as the value of b cannot be precisely measured in this case due to the limitations of the projector's optics.

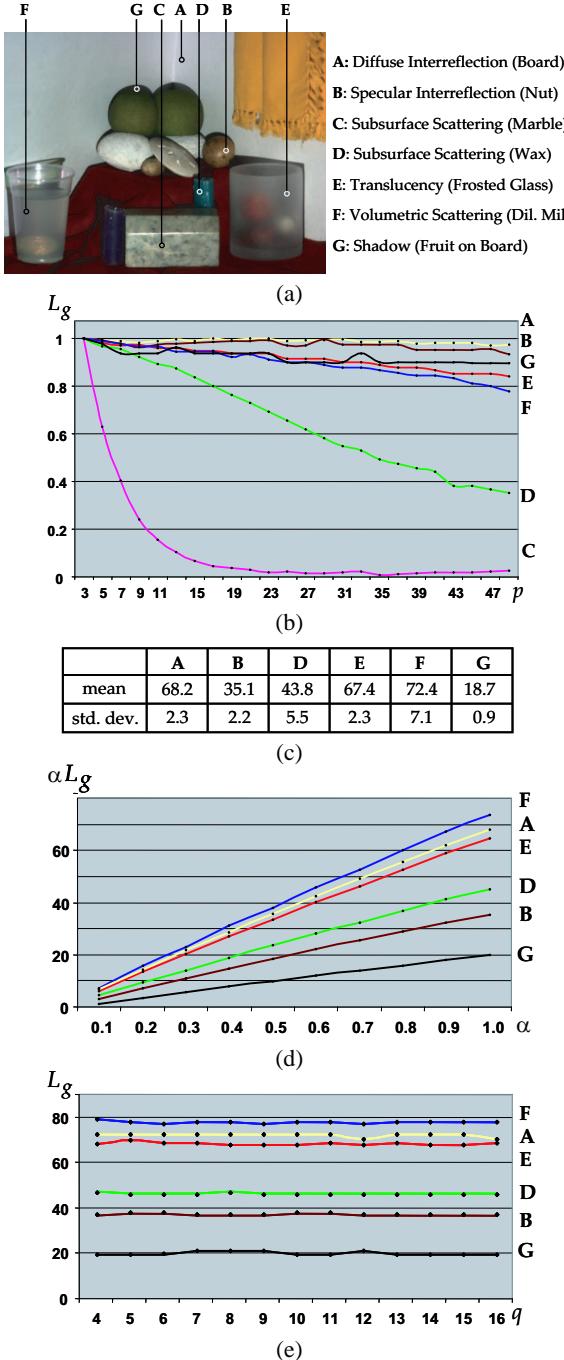


Figure 4: (a) Scene used for the verification experiments. Results are presented for the points A – G, which represent different global illumination phenomena. (b) In this experiment, the scene was fully lit except for square patches of size p (in pixels) centered around the points A – G. The estimated global component L_g is plotted as a function of p . (c) The mean and standard deviation of L_g for the points A – G (excluding C), computed using 100 illumination patterns with 6×6 checkers. In each of the patterns, half the checkers were lit and their locations were selected randomly. (d) In this experiment, the scene was lit with patterns with 6×6 checkers where only a fraction α of the checkers were lit. The measured αL_g varies more or less linearly with α . (e) Here, the scene was lit with a checkerboard pattern with half the checkers lit. L_g is plotted for different checker sizes q (in pixels). In experiments (c)-(e) a 6×6 square patch was kept unlit around each of the points A – G (excluding C).

and their global components were computed for each of the 100 patterns. The mean and standard deviation of the estimated global component are shown in Figure 4(c). We see that the global component estimates are robust to the randomness of the illumination pattern.

Next, we explored the sensitivity of the computed global estimate to the fraction α of lit patches in the illumination pattern. Again, each pattern was made of 6×6 patches, but in this case only a fraction α of the patches were activated and these patches were randomly distributed over the pattern. As before, the 6×6 patch around each of the interest points was kept unlit. The results are shown in Figure 4(d). As expected, the estimated global component αL_g varies linearly with α indicating that even small α values can produce reasonable estimates of the global component.

In the final verification experiment, we used checkerboard illumination patterns of different frequencies. For each pattern, the 6×6 patch around each of the interest points was kept unlit and its global component was measured. Since the size of this unlit region was kept the same for all checkerboard frequencies, the computed global estimates do not include highly local contributions. This is relevant mainly in the case of the candle point (D), which has strong subsurface scattering. In Figure 4(e), the global component is plotted as a function of the checker size q (in projector pixels), which is half the period of the checkerboard. The estimated global components of all points stay more or less constant when q is varied from 4 to 16 pixels. Also, they are very close to the mean global estimates in Figure 4(c), which were obtained using randomized checkerboards.

4 Separation Methods and Novel Images

We now present several methods that use high frequency illumination patterns to perform the separation. In addition, we show how the computed direct and global images can be used to create novel images of the scene. Note that the focus of our work is on the separation and that the novel images we show are only simple examples of what is possible⁷. Moreover, while the novel images we show are physically motivated, they are not claimed to be physically precise.

4.1 Checkerboard Illumination Shifts

As we have seen in Section 3.5, a high frequency checkerboard illumination pattern and its complementary pattern are sufficient to obtain the direct and global components of a scene. Unfortunately, it is difficult to obtain such ideal patterns using an off-the-shelf digital projector. Due to light leakages within the projector optics and custom image processing incorporated by the manufacturer, the lit and unlit checkers have brightness variations within them. Furthermore, due to the limited depth of field of the projector, the checkers can be defocused in some scene regions⁸. To overcome these problems we take a larger number of images than the theory requires. In our experiments, we used a pattern with checkers that are 8×8 pixels in size and shifted the pattern 5 times (by 3 pixels each time) in each of the two dimensions to capture a total of 25 images. The separation steps are illustrated in Figure 5(a), where the images correspond to a small part of the scene in Figure 1. At each pixel, the maximum and minimum measured brightnesses (L_{max}, L_{min}) were used to compute the direct and global estimates (L_d, L_g) using Equation 9.

Figures 6(a)-(e) show separation results for several scenes where each scene has a dominant phenomenon that produces the global illumination. In Figure 6(a) the global image includes the strong diffuse interreflections between the eggs and within the creases of the orange paper. In Figure 6(b) the wooden blocks have a specular component as well and hence the global image has regions of strong

⁷Some of the novel image generations we show here can be done using a single image of a normally illuminated object using recently developed algorithms for specularity removal (see [Mallick et al. 2006]).

⁸In all our experiments, the camera lens was used with an F-number greater than 8. Hence, defocus effects due to the camera were negligible.

color bleeding between the blocks. In Figure 6(c), the appearance of the peppers is dominated by subsurface scattering, as seen from the global image. The direct component mainly includes specular reflections, except in the case of the green stalks that are rather more diffuse reflectors. Similar effects are seen in Figure 6(d), where the colors of the cheeses and the green and purple grapes are mostly due to global illumination by subsurface scattering. The direct component for both sets of grapes are similar and include both specular and diffuse reflections from the surface.

Figure 6(e) shows results for a kitchen sink with strong volumetric scattering effects. The sink includes some objects and is filled with very dilute milk which acts like a participating medium. Note that the direct component appears like an image of the scene in a clear medium (air), while the global component includes the scattering of light by the dilute milk as well as the secondary illumination of the objects by the dilute milk. Comparing the direct and global images, we see that the secondary illumination by the milk is significantly stronger than the direct illumination by the source.

Figure 6(f) shows examples of how novel images of a scene can be computed from the direct and global images. In the case of the wooden blocks, the novel image is just a differently weighted sum of the two component images – the global component is given thrice the weight of the direct component. Although such an image appears unrealistic and is impossible from a physical standpoint, it is interesting as it emphasizes the optical interactions between objects in the scene. In the novel image of the peppers, the peppers have been given different colors by changing their hues in the global image and recombining with the direct image, which includes the specular highlights that have the color of the source. The same process is used to generate the novel image of the grapes. In comparison with the peppers, the direct component of the grapes includes both specular and diffuse reflections from the skin.

In the case of the kitchen sink, the dark regions of the global image were used to estimate the brightness L_m of the milk. L_m was assumed to be constant over the scene and was removed from the global image to obtain the radiance L_{gm} of the objects due to illumination by the milk. The ratios of brightnesses in the direct image L_d and the milk illumination image L_{gm} were tabulated. Then, the direct images of two other objects (green fruit and yellow pot) were separately captured and their milk illumination images were computed using the tabulated ratios. The L_d , L_m and L_{gm} components of these new objects were then added and the objects were inserted into the scene image. Notice how the inserted objects include not only the effects of scattering by the milk but also secondary illumination by the milk. Figure 8 shows a case where our separation method fails. This scene includes a mirror sphere which violates the high frequency illumination assumption we have made. The global illumination functions for points on the mirror sphere as well as points on the diffuse walls are no longer smooth compared to the illumination frequency due to the specular BRDF of the sphere. As a result, we see checker-like artifacts in the measured direct and global images.

4.2 Source Occluders

Thus far, we have used a projector to generate the high frequency illumination patterns. In the case of a simple uncontrollable source, such as the sun, occluders of various kinds can be used to cast high frequency shadows on the scene. For example, a line occluder, such as the stick in Figure 5(b), can be swept across the scene while it is captured by a video camera⁹. If the occluder is thin, its shadow will occupy a small part of the scene and hence we can assume $\alpha = 1$ in Equation (8)¹⁰. Furthermore, if the scene point lies within the

⁹The scanning of a scene using a line occluder has also been used to recover the geometry of the scene [Bouguet and Perona 1999].

¹⁰Only the direct components of other shadowed scene points will not contribute to the global component of each shadowed scene point.

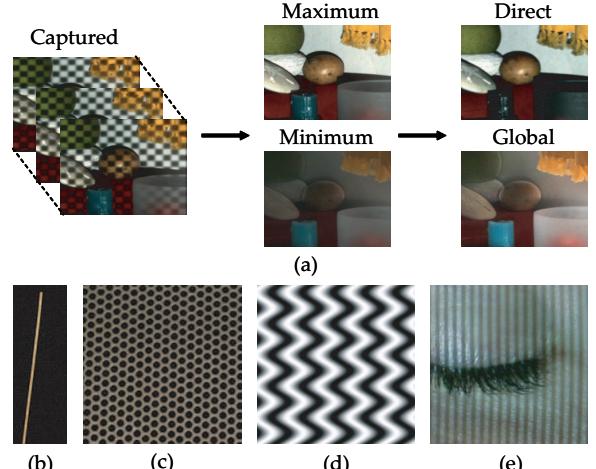


Figure 5: (a) The steps involved in the computation of direct and global images using a set of shifted checkerboard illumination patterns. (b) The line occluder (stick) used to scan scenes lit by a simple source such as the sun. (c) The mesh occluder used to expedite the scanning process. (d) Three shifted versions of this sinusoid-based illumination pattern are sufficient to perform the separation. (e) A magnified part of the face image in Figure 7(d) reveals the stripe pattern used to do the separation from a single image.

umbra of the shadow there will be no direct contribution due to the source and hence $b = 0$ in Equation 8. Let L_{max} and L_{min} be the maximum and minimum brightnesses observed at a scene point in the video captured while sweeping the occluder. Then, $L_{max} = L_d + L_g$, $L_{min} = L_g$ and the two components can be computed. Figure 7(a) shows results obtained for an outdoor scene with red leaves. One of the frames of the captured video is shown on the left (see the shadow cast by the stick). The leaves have a strong global component because they are somewhat translucent. The global component also includes the ambient illumination of the leaves by the sky.

In the case of a line occluder, the captured video must be long enough to ensure that all the scene points have been subjected to the umbra of the shadow. The capture process can be made much more efficient by using a more complex mesh occluder, such as the 2D grid of circular holes shown in Figure 5(c). In this case, only a small circular motion of the occluder is needed to ensure that all the points are captured in and out of shadow. If a fraction β of the grid is occupied by holes, then we have $L_{max} = L_d + \beta L_g$ and $L_{min} = \beta L_g$. The scene in Figure 7(b) includes a mannequin behind a translucent shower curtain with an opaque design on it. The direct image includes specular reflections from the curtain as well as diffuse reflections from the design. The global component includes the mannequin as it is only lit indirectly by diffuse light from the curtain.

The measurement of the direct component has implications for several shape from brightness methods used in computer vision. As shown in [Nayar et al. 1991], interreflections can cause shape from brightness methods to produce incorrect results. In the case of concave Lambertian surfaces, the shape computed by photometric stereo can be refined by using an iterative algorithm. If the direct component can be measured for each of the light sources used to perform photometric stereo, one does not need to rely on such an algorithm. Figure 7(c) shows photometric stereo results for a diffuse painted bowl, obtained using three sources and a line occluder to do the separation. The original image and the corresponding direct and global images for one of the sources are shown on the left. In the middle are shown the albedo images computed using the original images (uncorrected) and the direct images. On the right are the depth profiles computed using the original images and the direct images. As expected, the original images produce an incorrect and shallow shape

[Nayar et al. 1991]. The direct images result in a shape that is very close to the ground truth shape, which was manually measured.

4.3 Other High Frequency Patterns

Although we have discussed two-valued illumination patterns (or shadows) thus far, our separation method is applicable to other high frequency illuminations as well. For instance, it can be easily incorporated into standard structured-light range finders that use coded illumination patterns. In the case of binary coded patterns, some of the patterns will have high frequency stripes. The corresponding images can be used to estimate the direct and global components.

In the case of a projector, any positive illumination function can be generated. A convenient class of functions is based on the sinusoidal function. By using a high frequency pattern that varies over space and/or time as a sinusoid at each projector pixel, the separation can be done using just three patterns. In the first pattern, the brightnesses of all the projector pixels are randomly generated using a uniform distribution between 0 and 1 so that the scene is lit with half the power of the projector to produce a global component of $L_g/2$ at each scene point. Let the brightness of a given projector pixel be $a = 0.5 + 0.5 \sin \phi$, where $0 \leq \phi \leq 2\pi$. Two more illumination patterns are generated by changing the phases of the sinusoids of all the pixels by, say, $2\pi/3$ and $4\pi/3$. Then, the camera brightnesses corresponding to the scene point that is directly lit by the above projector pixel can be written as $L_1 = L_d(0.5 + 0.5 \sin \phi) + L_g/2$, $L_2 = L_d(0.5 + 0.5 \sin(\phi + 2\pi/3)) + L_g/2$ and $L_3 = L_d(0.5 + 0.5 \sin(\phi + 4\pi/3)) + L_g/2$. From these equations, L_d , L_g and ϕ can be found. Note that ϕ gives us the correspondence between camera and projector pixels. Hence, the 3D structure of scene can be computed as well [Wust and Capson 1991].

In the above example, we used a random first pattern and varied its brightness over time as a sinusoid. Alternatively, a pattern can be generated that is sinusoidal with a high frequency in one direction and the phase of the sinusoid can be varied with a high frequency in the other direction. An example of such a function is $\sin(x + \sin y)$, which is shown in Figure 5(d). The phase variation along the y dimension is only used to ensure that the illumination has high frequencies along both spatial dimensions. If three images of the scene are captured using this pattern and two shifted versions of it, where the shifts are in the x dimension and are known, we get three equations as in the previous case and L_d , L_g and ϕ can be found.

4.4 Separation using a Single Image

Thus far, we have presented methods that can produce separation images at the full resolution of the captured images. The direct and global images can be computed at a lower resolution using a single captured image. Consider a scene illuminated by a high frequency binary pattern. We filter each color channel of the captured image to find local peaks and valleys. This is done by assigning a pixel a maximum or minimum label if its brightness is the maximum or minimum within an $n \times m$ window around it. The brightnesses at these peaks and valleys are interpolated to obtain full resolution L_{max} and L_{min} images. Let us assume that the separation results are to be computed at $1/k$ of the resolution of the captured image. Then, we compute L_{max} and L_{min} images at this lower resolution by simply averaging their values within $k \times k$ blocks in the high resolution images. Once this is done, L_d and L_g are computed using Equation 8.

An example of separation using a single image is shown in Figure 7(d). Here, the Sanyo projector used before (with 1024×768 pixels) was used to project a pattern with bright and dark stripes of 2-pixel width on a face. A Canon EOS 20D camera with 3503×2336 pixels was used to capture the image. A magnified region of the captured image is show in Figure 5(e). The separation was done as described above using $n = 11$, $m = 1$ and $k = 4$. We see that the global image captures the color of the face which is primarily due to subsurface scattering, while the direct image has an almost metallic appearance as it mainly includes the surface reflections due to oils and lipids.

Figure 7(e) shows how the oiliness of the face can be modified using the direct and global images. For this, we tone-mapped the direct image to accentuate the effects of the oils and lipids. Then, this image was scaled differently and recombined with the global image to obtain the four novel images. The color of a face has several contributing factors [Tsumura et al. 2003]. Even so, reasonable control over skin tone can be achieved by changing the hue of the global image and adding back the direct image, as shown in the Figure 7(f).

5 Discussion

We have developed efficient methods for separating the direct and global components of a scene lit by a single light source. Our separation approach is applicable to complex scenes with a wide variety of global illumination effects. To our knowledge, this is the first time the direct and global images of arbitrarily complex real-world scenes have been measured. These images reveal a variety of non-intuitive effects and provide new insights into the optical interactions between objects in a scene. In addition, we have shown simple examples of how the separation results can be used to create novel images of a scene. In future work, we hope to show how direct and global images can be used to perform more sophisticated image manipulations.

We are currently exploring ways to incorporate our separation technique into digital cameras by modifying the camera flash to serve as a high frequency source. For our approach to be a practical camera feature, it must be able to handle dynamic scenes. This requires us to do the separation with a minimal number of images, ideally a single image. To this end, our current approach to single-image separation needs to be improved so that the computed component images are of a resolution that is close to the native resolution of the camera. If this can be achieved, a user would be able to edit the appearances of objects in the scene using the measured direct and global components.

Our separation technique only produces the total global component at each scene point and not the photometric coupling between all pairs of scene points. We are looking into ways of using high frequency illuminations to estimate the complete transport matrix associated with a scene. This would lead to deeper insights into the photometric properties of real scenes and perhaps even more powerful methods for creating novel images. We have also assumed that the direct component arises from a single dominant source and that all other sources contribute to the global component. It would be useful to extend our results to the case of multiple sources without having to activate the sources in a sequential fashion.

An interesting direction for future investigation is the interplay between the resolutions of the camera, the source and the scene variations. Like BRDFs and BSSRDFs, our direct and global images are dependent on the scale of illumination and observation. It would be interesting to study how the direct and global images vary with scale and whether the results obtained at one scale can be used at another.

A The Minimum Illumination Frequency

For any realistic scene, it is difficult to derive a closed-form expression for the minimum frequency of the illumination needed to perform the separation. This is because the terms $A[i, j]$ and $L_d[i, j]$ in Equation (3) are complicated functions of surface BRDF and geometry. However, some insights can be gained by viewing these terms as continuous functions and analyzing their sampling by the illumination in frequency domain. Without loss of generality, let us assume the scene is 1D. Let x and y be the continuous versions (defined on the scene surface) of the discrete parameters i and j , respectively. Since we are considering a single surface point x , we can drop this parameter. Then, from Equation (3), we have:

$$L_{gd} = \int A(y)L_d(y)dy. \quad (10)$$

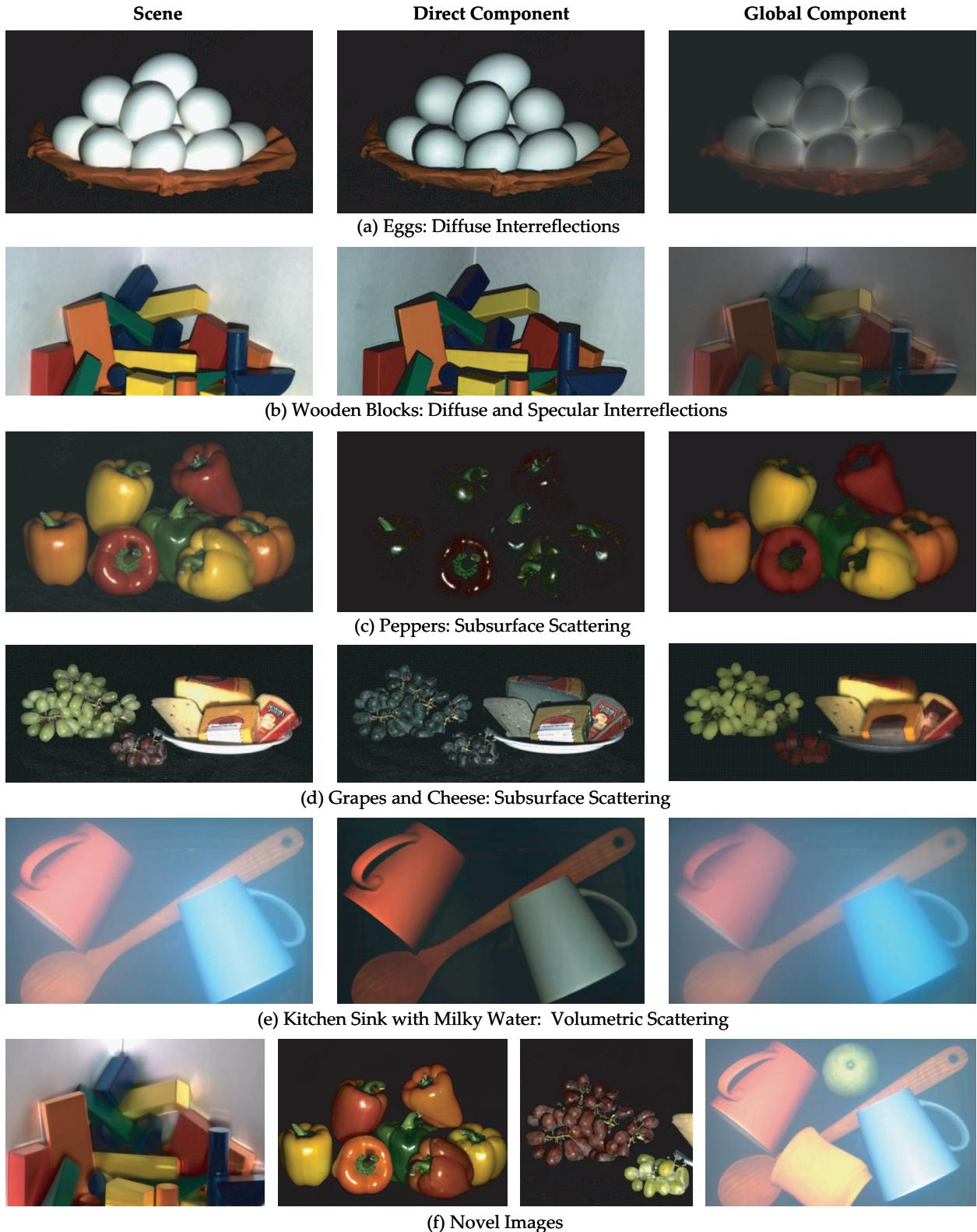


Figure 6: Separation results for different scenes, each with a dominant physical phenomenon that produces the global component. (a) A scene with eggs on a paper napkin that includes diffuse interreflections. (b) A scene with wooden blocks in a corner that includes specular interreflections. (c) A scene with peppers that produce subsurface scattering. (d) A scene with cheeses and grapes that also produce subsurface scattering. (e) A kitchen sink with objects immersed in dilute milk which produces volumetric scattering effects. (f) Novel images computed from the separation results.

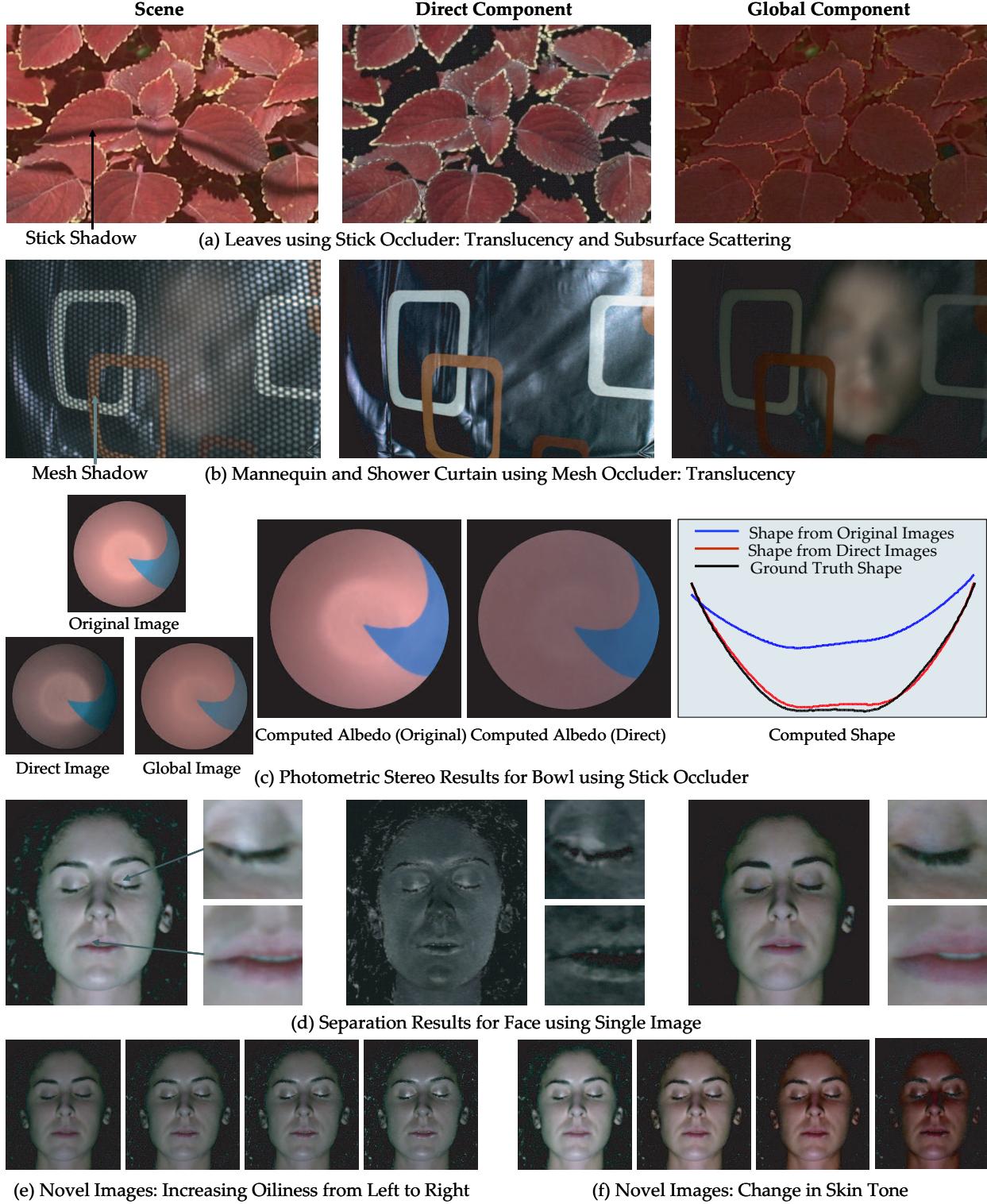


Figure 7: (a) Separation results for leaves that are translucent and produce subsurface scattering effects. The left image is from a video captured while the scene was scanned with a stick occluder under sunlight. (b) Separation results for a mannequin behind a translucent shower curtain. The left image is from a video captured while the scene was scanned using a mesh occluder while lit by a halogen lamp. (c) Photometric stereo results for a colored bowl computed from regular (original) images and from direct images obtained using a stick occluder. (d) The separation of a face into its global and direct components using a single image captured with a high frequency stripe illumination pattern. The direct and global components are used to generate novel images of the face that have (e) different oiliness and (f) different skin tone.

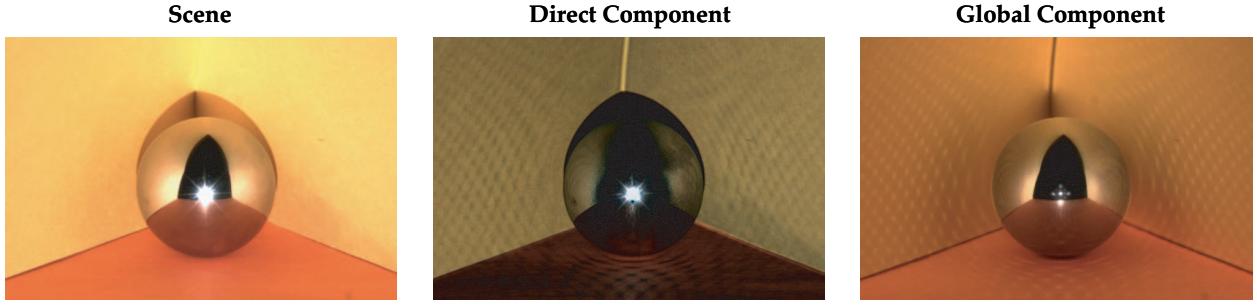


Figure 8: Separation results using checkerboard illumination shifts for a scene that includes a mirror sphere and three diffuse walls around it. The separation method fails in this case as the scene violates our assumption that the global function at each point is smooth compared to the illumination frequency. As a result, we see checker-like artifacts in the measured direct and global images.

Let $A(y)$ and $L_d(y)$ have maximum frequencies of ω_A and ω_L , respectively. Since the product $A(y)L_d(y)$ in spatial domain corresponds to a convolution in frequency domain, it will have a maximum frequency of $\omega_A + \omega_L$. If our goal were to completely reconstruct the function $A(y)L_d(y)$, we would need to sample it with a minimum (Nyquist) frequency of $2(\omega_A + \omega_L)$. However, we are interested in L_{gd} , which is an integral over $A(y)L_d(y)$ and hence equals its zero-frequency (DC) component. To ensure this DC component is not aliased, the signal must be sampled with at least half the Nyquist frequency. Therefore, we need to sample $A(y)$ and $L_d(y)$ with a minimum illumination frequency of $(\omega_A + \omega_L)$ to obtain an accurate estimate of L_{gd} . Global illumination guarantees that the term L_{gg} in Equation (3) will be smoother than L_{gd} . Therefore, the above illumination frequency will also be adequate to obtain an accurate estimate of L_{gg} .

Acknowledgments

This research was supported in parts by a grant from the Office of Naval Research (No. N000140510188) and by Mitsubishi Electric Research Laboratory, Cambridge, MA. The authors thank Vijay Anand for his help with the coding of the separation method.

References

- ARIKAN, O., FORSYTH, D., AND O'BRIEN, J. 2005. Fast and detailed approximate global illumination by irradiance decomposition. *ACM Trans. on Graph.* 24, 3, 1108–1114.
- BOUGUET, J., AND PERONA, P. 1999. 3d photography using shadows in dual-space geometry. *IJCV* 35, 2 (November), 129–149.
- CHANDRASEKHAR, S. 1950. *Radiative Transfer*. Clarendon Press, Oxford, UK. reprinted by Dover Publications, 1960.
- CHUANG, Y., ZONGKER, D., HINDORFF, J., CURLESS, B., SALESIN, D., AND SZELISKI, R. 2000. Environment matting extensions: Towards higher accuracy and real-time capture. In *Proc. of SIGGRAPH*, ACM Press, 121–130.
- DEBEVEC, P., HAWKINS, T., C., T., DUIKER, H., SAROKIN, W., AND SAGAR, M. 2000. Acquiring the reflectance field of a human face. In *Proc. of SIGGRAPH*, ACM Press, 145–156.
- FORSYTH, D., AND ZISSEMAN, A. 1991. Reflections on shading. *IEEE Trans. on PAMI* 13, 7, 671–679.
- GORTLER, S., GRZESZCZUK, R., SZELISKI, R., AND COHEN, M. 1996. The lumigraph. In *Proc. of SIGGRAPH*, ACM Press, 43–54.
- IMMEL, D., COHEN, M., AND GREENBERG, D. 1986. A radiosity method for non-diffuse environments. In *Proc. of SIGGRAPH*, ACM Press, 133–142.
- KAJIYA, J. T. 1986. The Rendering Equation. In *Proc. of SIGGRAPH*, ACM Press, New York, NY, USA, 143–150.
- KOENDERINK, J., AND VAN DOORN, A. 1983. Geometrical modes as a general method to treat diffuse interreflections in radiometry. *JOSA* 73, 6 (June), 843–850.
- LEVOY, M., AND HANRAHAN, P. 1996. Light field rendering. In *Proc. of SIGGRAPH*, ACM Press, 31–42.
- LIN, Z., WONG, T., AND SHUM, H. 2002. Relighting with the reflected irradiance field: Representation, sampling and reconstruction. *IJCV* 49, 2-3 (September), 229–246.
- MALLICK, S. P., ZICKLER, T., BELHUMEUR, P. N., AND KRIEGMAN, D. J. 2006. Specularity removal in images and videos: A PDE approach. In *Proc. of ECCV*.
- NAYAR, S., IKEUCHI, K., AND KANADE, T. 1991. Shape from interreflections. *IJCV* 6, 3, 173–195.
- PEERS, P., AND DUTRÉ, P. 2003. Wavelet environment matting. In *Eurographics Symposium on Rendering*, ACM Press, 157–166.
- SEITZ, S., MATSUSHITA, Y., AND KUTULAKOS, K. 2005. A theory of inverse light transport. In *Proc. of ICCV*, II: 1440–1447.
- SEN, P., CHEN, B., GARG, G., MARSCHNER, S., HOROWITZ, M., LEVOY, M., AND LENSCHE, H. 2005. Dual photography. *ACM Trans. on Graph.* 24, 3, 745–755.
- SHIM, H., AND CHEN, T. 2005. A statistical framework for image-based relighting. In *Proc. of ICASSP*.
- SLOAN, P., KAFTZ, J., AND SNYDER, J. 2002. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. In *Proc. of SIGGRAPH*, ACM Press, 527–536.
- TSUMURA, N., OJIMA, N., SATO, K., SHIRAISHI, M., SHIMIZU, H., NABESHIMA, H., AKAZAKI, S., HORI, K., AND MIYAKE, Y. 2003. Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin. In *Proc. of SIGGRAPH*, ACM Press, 770–779.
- WOODHAM, R. 1980. Photometric Method for Determining Surface Orientation from Multiple Images. *Optical Engineering* 19, 1 (January), 139–144.
- WUST, C., AND CAPSON, D. W. 1991. Surface profile measurement using color fringe projection. *Machine Vision and Applications* 4, 193–203.
- ZHU, J., AND YANG, Y. 2004. Frequency-based environment matting. In *Pacific Conf. on Comp. Graph. and Apps.*, 402–410.
- ZONGKER, D., WERNER, D., CURLESS, B., AND SALESIN, D. 1999. Environment matting and compositing. In *Proc. of SIGGRAPH*, ACM Press, 205–214.