Scene Recovery and Rendering Techniques Under Global Light Transport

Mohit Gupta

CMU-RI-TR-10-37

Submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy in Robotics

The Robotics Institute
School of Computer Science
Carnegie Mellon University

Pittsburgh, Pennsylvania 15213

October 2010

Thesis Committee:
Srinivasa G. Narasimhan, Chair
Takeo Kanade
Martial Hebert

Shree K. Nayar, Columbia University

© 2010 Mohit Gupta.

Abstract

Light interacts with the world around us in complex ways. These interactions can broadly be classified as **direct illumination** – when a scene point is illuminated directly by the light source, or **indirect illumination** – when a scene point receives light that is reflected, refracted or scattered off other scene elements. Several computer vision techniques make the unrealistic assumption that scenes receive only direct illumination. In many real-world scenarios, such as indoors, underground caves, underwater, foggy conditions and for objects made of translucent materials like human tissue, fruits and flowers, the amount of indirect illumination is significant, often more than the direct illumination. In these scenarios, vision techniques that do not account for the indirect illumination result in strong and systematic errors in the recovered scene properties.

The above stated assumption is made because computational models for indirect illumination (also called global illumination or global light transport) are complex, even for relatively simple scenes. The goal of this thesis is to build simple, tractable models of global light transport, which can be used for a variety of scene recovery and rendering applications. This thesis has three contributions. **First**, recovering scene geometry and appearance despite the presence of global light transport. We show that two different classes of shape recovery techniques - structured light triangulation and shape from projector defocus - can be made robust to the effects of global light transport. We demonstrate our approaches on scenes with complex shapes and optically challenging materials. We then investigate the problem of recovering scene appearance in the presence of common poor visibility scenarios, such as murky water, bad weather, dust and smoke. Computer vision systems deployed in such conditions suffer due to scattering and attenuation of light. We show that by controlling the incident illumination, loss of image contrast due to scattering can be significantly reduced. Our framework can be used for improving visibility in a variety of outdoor applications, such as designing headlights for vehicles, both terrestrial and underwater.

Global light transport is *not always noise*. In numerous scenarios, measuring global light transport can actually provide useful information about the scene. The **second** contribution is to recover material and scene properties by measuring global light transport. We present a simple device and technique for robustly measuring the volumetric scattering properties of a broad class of participating media. We have constructed a data-set of the scattering properties, which can be immediately used by the computer graphics community to render realistic images. Next, we model the effects of defocused illumination on the process of measuring global light transport in general scenes. Modeling the effects of defocus is important because projectors, having limited depth-of-field, are increasingly being used as programmable illumination in vision applications. With our techniques, we can sepa-

rate the direct and global components of light transport for scenes whose depth-ranges are significantly greater than the depth of field of projectors (< 0.3m).

The **third** contribution of this thesis is fast rendering of dynamic and non-homogenous volumetric media, such as fog, smoke, and dust. Rendering such media requires simulating the fluid properties (density and velocity fields) and rendering volumetric scattering effects. Unfortunately, fluid simulation and volumetric rendering have always been treated as two disparate problems in computer graphics, making it hard to leverage the advances made in both fields together. In particular, reduced space methods have been developed separately for both fields, which exploit the observation that the associated fields (density, velocity and intensity) can be faithfully represented with a relatively small number of parameters. We develop a unified reduced space framework for both fluid simulation and volumetric rendering. Since both fluid simulation and volumetric rendering are done in a reduced space, our technique achieves computational speed-ups of one to three orders of magnitude over traditional spatial domain methods. We demonstrate complex visual effects resulting from volumetric scattering in dynamic and non-homogenous media, including fluid simulation effects such as particles inserted in turbulent wind-fields.

Contents

Chapter 1

Introduction

There are two kinds of light - the glow that illuminates, and the glare that obscures. $\sim$ *James Thurber*

Light interacts with the world around us in complex ways, resulting in a rich variety of visual sensations: the glow around street-lights on a foggy night, the soft appearance of translucent objects like skin, flowers and marble, different colors of liquids like wine, beer and milk and the splendor of underwater effects (Figure 1.1). These interactions can broadly be classified as **direct illumination** – when a scene point is illuminated directly by the light source, or **indirect illumination** – when a scene point receives light that is reflected, refracted or scattered off other scene elements. It is hard to imagine the world without the effects of indirect illumination, also called global illumination or global light transport. Without these intra-scene optical interactions, it would become hard for us to perceive the three-dimensional shapes of the indoor spaces that we are in. Human skin, flowers and fruits would lose their beautiful, soft appearance, instead looking like plastic. Red wine would look like red paint and the glow around the moon would disappear. Global illumination is such an integral part of our daily visual experience that the human visual system has learnt to account for global illumination effects when perceiving the color and shape of surfaces [8, 34], much like it is known to discount direct illumination from light sources [76, 5, 70].

Computer graphics researchers have strived to simulate global illumination to recreate the visual world around us. The very first images were rendered assuming only direct illumination from the light sources. But now, complex global light transport effects, such as inter-reflections, sub-surface scattering and volumetric scattering can be faithfully rendered, resulting in almost photo-realistic images. Unfortunately, the same can not be said for computer vision. Several computer vision techniques make the unrealistic assumption

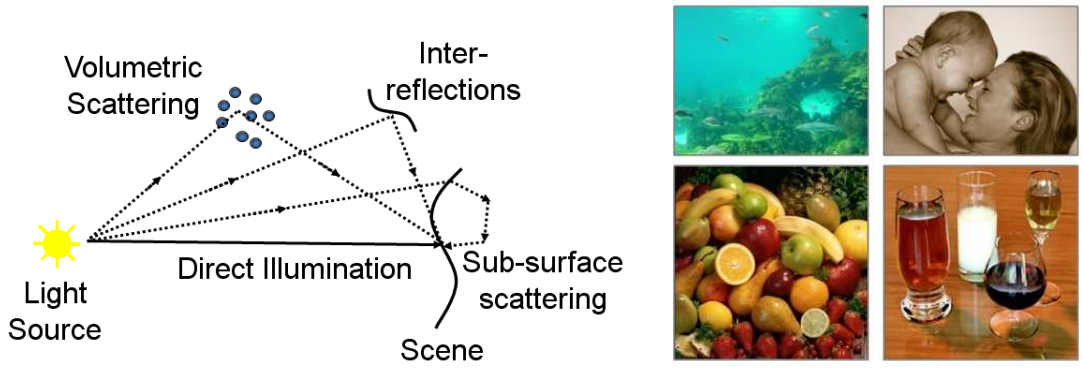


Figure 1.1: Global Illumination. (a) Light interacts in complex ways with the world around us. These complex interactions, in the form of inter-reflections, sub-surface scattering and volumetric scattering are collectively called global illumination. (b) Given that it is almost impossible to find a real world scene without global illumination, it is important to make vision algorithms robust to the presence of global illumination.

that scenes receive only direct illumination. In many real-world scenarios, such as indoors, underground caves, underwater, foggy conditions and for objects made of translucent materials like human tissue, fruits and flowers, the amount of indirect illumination is significant, often more than the direct illumination. In these scenarios, vision techniques that do not account for the indirect illumination result in strong and systematic errors in the recovered scene properties.

This assumption has been necessary because scene recovery using active vision involves modeling and *inverting* light transport models, which can become intractable as the complexity of models increases. The goal of this thesis is to derive simple models of global light transport for a variety of scene recovery and rendering applications. We show that by *actively controlling the illumination*, we can recover scene properties (geometry, appearance and material properties) despite, and in some scenarios, using global illumination. This has been made possible by the observation that for computer vision techniques, often we need to model only the aggregate effects of global light transport. Although modeling individual light rays might be complex, but much simpler models can be made for aggregate effects [87, 43]. In addition, the recent advent of projectors has provided us with an unprecedented flexibility in controlling illumination. With projectors, light can be modulated along multiple dimensions: space, time, wavelength, focus and polarization. Projectors can thus act as powerful probes for understanding light transport and scene recovery.

Broadly, this thesis has three main thrusts:

Recovering scene geometry and appearance in the presence of global light transport: Virtually all active scene recovery techniques, for example shape and appearance from intensity [97, 65, 129, 53, 134, 51], shape from structured light triangulation [102, 56, 12], and shape from projector defocus [131] do not account for global light transport. Given the ubiquity of global light transport, ignoring it can result in strong, systematic biases in the recovered scene properties [38, 86]. In general, separating the effects of global illumination from a shape/appearance cue requires explicit modeling and inverting of global illumination [86]. Given the complexity of light transport though, this can be nearly impossible.

We have built simple aggregate models of global light transport and used them to derive two classes of shape-recovery techniques that are invariant to global light transport. First, we consider the depth-cue of projector defocus [131] (Chapter 2). We have studied the inter-relationship between defocused illumination and global light transport and showed that both these effects manifest themselves as similar low frequency artifacts during image formation. This observation brings the two seemingly different physical phenomena on the same platform - now, we can simply use signal processing tools to analyze and separate both effects. This enables accurate depth recovery in the presence of global light transport, without explicitly modeling or measuring it. The second class of shape recovery techniques that we consider is structured light triangulation, one of the oldest and most widely used 3D shape measurement technique. As with most active vision techniques, it does not model the effects of global illumination, thus resulting in potentially significant errors [16, 15, 130]. We formally analyze the errors caused due to different modes of global light transport. Then, we design structured light illumination patterns that modulate global illumination and prevent the errors at capture time itself (Chapter 5). We have demonstrated 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 deep concavities with strong inter-reflections.

Computer vision systems deployed in the presence of volumetric scattering media such as murky water, bad weather, dust and smoke face a different set of challenges. Images captured in these conditions show severe contrast degradation, blurring and loss of light due to attenuation, making it hard to perform meaningful scene analysis. The ability to improve image contrast of passive methods such as digital post-processing is limited when the quality of acquired images is extremely poor to begin with. In such scenarios, it is thus critical to act *before* the images are captured. We have investigated the problem of how to illuminate the scene for minimizing light attenuation and the loss of contrast due to backscatter, while maximizing the signal (Chapter 5). Our framework can be used for

improving visibility in a variety of outdoor applications, such as designing headlights for vehicles (terrestrial and underwater). We are extending our analysis to multi-camera vision systems deployed in scattering media, for the purpose of shape recovery.

Measuring light transport in scenes and materials: Global light transport is not always *noise*. In numerous scenarios, sensing and measuring global illumination can actually provide useful information about the scene, especially material properties. For instance, measuring the amount of scattering in volumetric media such as milk, beverages and ocean water can provide their optical properties, which in turn can be used to render realistic images of these materials. As before, this is hard because of the complexity of light transport. The key observation that we made is that light transport in scattering media can be simplified by simply diluting the media with water. This is similar in spirit to measuring the reflectance properties of surfaces by measuring intensities of single flat patches. In chapter 4, we present a simple device and technique for robustly estimating the optical properties of a broad class of participating media using active illumination. We have constructed a database of the scattering properties of variety of media, which can be immediately used by the computer graphics community to render realistic images of materials.

In general, measuring the global component separately from the direct component can provide useful insights into various scene properties, which their sum can not. For example, sub-surface scattering is a measure of material translucency. Inter-reflections can be used to infer the geometric lay-out of scenes. It was only recently that Nayar et al presented an efficient way to perform the direct-global separation [87]. This technique assumes 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. This is another important assumption made by most active vision techniques. It has become particularly pertinent as projectors, having limited depth-of-field, are increasingly being used as programmable illumination in vision applications. We have derived simple models for the effects of defocused illumination on the direct-global separation process. With our techniques, we can separate the direct and global components for scenes whose depth-ranges are greater than the depth of field of projectors (< 0.3m).

A future goal is to decompose global light transport further into it's constituent modes: inter-reflections, sub-surface scattering and volumetric scattering. We are particularly interested in building machine vision systems for measuring sub-surface scattering as it is the physical process behind the perception of translucency. Several objects of interest to humans, such as food, plants and skin are translucent. Such a system would have wide ranging applications from food inspection, medicine and surgery to cosmetics industry. We believe that a combination of active illumination and building simple models of translucency would

go a long way towards achieving this goal.

Rendering of scenes under volumetric scattering: Brute-force rendering of volumetric media, although capable of achieving nearly photo-realistic images, can be prohibitively slow (taking CPU-days or even weeks). On the other end of the spectrum are analytic solutions, which although fast, are possible only for homogenous and static media. Most of the interesting real world phenomenon such as smoke, dust and fog are non-homogenous and evolving with time. Despite this complexity, most of these physical processes can be represented with a significantly fewer variables (reduced space) than that required by a finite element methods based computer simulation. This observation has led to the development of a variety of *reduced space methods* in computer graphics. Examples include reduced space methods for fluid simulation [121] and pre-computed radiance transfer for rendering [112]. These methods achieve significant computational speed-ups over spatial-domain finite element based simulations.

Despite significant advances in both the fields of fluid simulation and rendering, these two continue to be addressed separately. In nature, both these processes interact with each other to provide us with a variety of rich visual experiences. This makes a case for both of them to be studied together to better leverage the advances in both fields. We have proposed a unified framework for both fluid simulation and rendering in an analytic reduced space (Chapter 6). We believe that this is an important first step towards bridging the gap between model reduction for fluid simulation and pre-computed radiance transfer for rendering. Since both fluid simulation and rendering are done in a reduced space, our technique achieves computational speed-ups of one to three orders of magnitude over traditional spatial domain methods. We demonstrate complex visual effects resulting from volumetric scattering in dynamic and non-homogenous media, including fluid simulation effects such as particles inserted in turbulent wind-fields.

It is worth noting a few important **difference between volumetric scattering and other modes of light transport**. In volumetric scattering, there is a clear boundary between the scene and the medium. For other modes of light transport, there is no such dichotomy. For all other modes of light transport, there is usually a single significant scattering event along a sensor element's line of sight. On the other hand, for volumetric scattering, there can be multiple significant scattering events along a line of sight. Thus, a significant portion of the irradiance received at the sensor is due to light rays that never reach a scene point (backscatter).

In view of these differences, this thesis is divided into two parts. The first part deals with inter-reflections, sub-surface scattering and illumination defocus. The first chapter builds a

combined theory of defocused illumination, inter-reflections and sub-surface scattering. All these effects are expressed as blur kernels on the incident illumination and an invariant is derived to separate the device dependent effect (defocus) from the global light transport effects (inter-reflections and sub-surface scattering). This invariant is used to recover scene depths using the depth cue of projector defocus despite the presence of global illumination effects. Second chapter presents a structured light based 3D scanning system which is robust to the presence of global illumination effects.

The second part deals with scene recovery and rendering techniques under volumetric scattering. In the third chapter, we present a simple device and technique for robustly estimating the optical properties of a broad class of volumetric media using active illumination. In the fourth chapter, we design active vision systems which can see clearer and farther in various poor visibility scenarios, such as smoke, fog and underwater. In the fifth chapter, we present a fast rendering technique for dynamic and non-homogenous volumetric media. Finally, in the sixth chapter, we discuss future research directions.

Part I

Inter-reflections, sub-surface scattering and defocused illumination

Chapter 2

A Combined Theory of Defocused Illumination and Global Light Transport

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

Office of the future [1] Direct-Global Separation [87] High Speed Vision [82]

Figure 2.1: Projector are being used as programmable illumination for a variety of applications in Computer Graphics, vision, virtual reality and human-computer interfaces. In most of these applications, projectors interact not just with flat screens, but with complex real world scenes around us. Since projectors are limited depth of field devices, this creates an interesting interplay between defocused illumination and global light transport.

The goal of this work [43] is to study the inter-relationship between defocused illumination and global light transport. This relationship is important as projectors are increasingly being used as programmable illumination in a variety of applications in computer graphics, vision, virtual reality and human-computer interfaces (Figure 2.1). In virtually all these scenarios, projectors interact not just with flat screens but with complex, real world objects

around us. Since projectors are limited depth of field devices, this creates an interesting interplay between defocused illumination and global light transport.

It may seem that these two 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, which brings these effects on a common ground, is that both of them manifest as low pass filters on the incident illumination. This enables using signal processing tools to analyze both these effects, without having to explicitly model either of them. Building on 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 transport and significant depth variations. Especially from a computer vision point of view, it is desirable to separate them as both of them encode different information about the scene.

We show applications in two scenarios. **First**, accurate depth recovery in the presence of global light transport. Under global illumination, most active techniques such as photometric stereo [129], shape from shading [53], structured light scanning, shape from projector defocus [131] produce erroneous results. We show that global light transport can be separated from the depth cue of illumination defocus [131] without explicitly modeling or measuring light transport. This significantly reduces the errors caused due to global light transport.

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 ($< 0.3m$) 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*.

2.1 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 [119, 81, 85]. 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 [114] and using it to recover scene depths [88, 127]. 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.

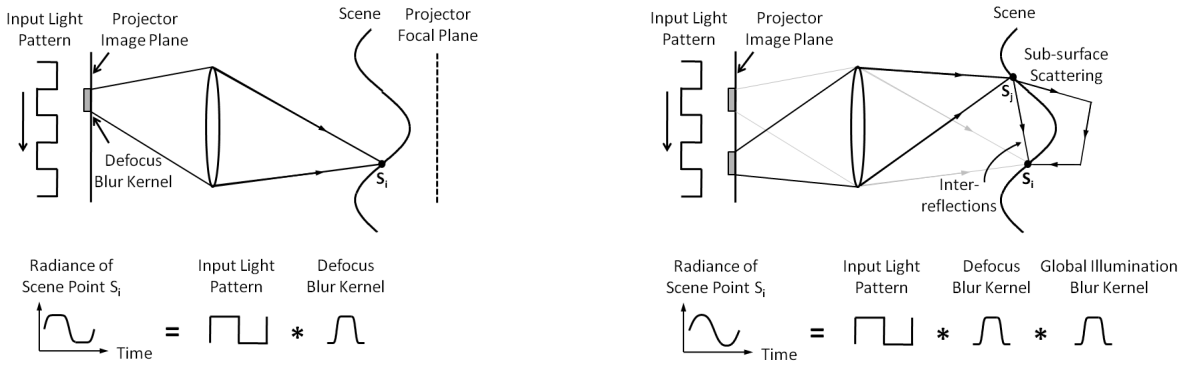
Shape recovery under global light transport: Most existing shape-from-intensity techniques [129, 53, 131] account for only the direct component of light transport. One possibility is to remove the global component a priori using the approach of Nayar et al [87]. However, this approach requires the projector's illumination to be focused on the entire 3D scene, making it unamenable for depth recovery using projector defocus analysis. Nayar et al [86] 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 [109, 30] can be used to remove inter-reflection from images [107]. 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 [38]. Researchers have used polarization [15], modulation with a high-frequency illumination pattern [16] and fluorescence [54] 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) [88, 46] and depth from camera defocus (DFD) [127] techniques can compute complete depth maps[1], but they rely on scene texture for accurate scene recovery. We use a co-located camera-projector setup for data acquisition, as shown in Figure 2.3 (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 [7, 39]. Confocal imaging techniques recover partially transparent volumes by focusing the illumination and sensor simultaneously on slices of the volume [28, 72]. Morris et al [79] and Kutulakos et al [69] 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.

[1] Although DFD and DFF also suffer from occlusion, the effects are not as severe due to a much smaller base-line [105].

2.2 Defocused illumination and global light transport as low pass filters



(a) Direct Illumination

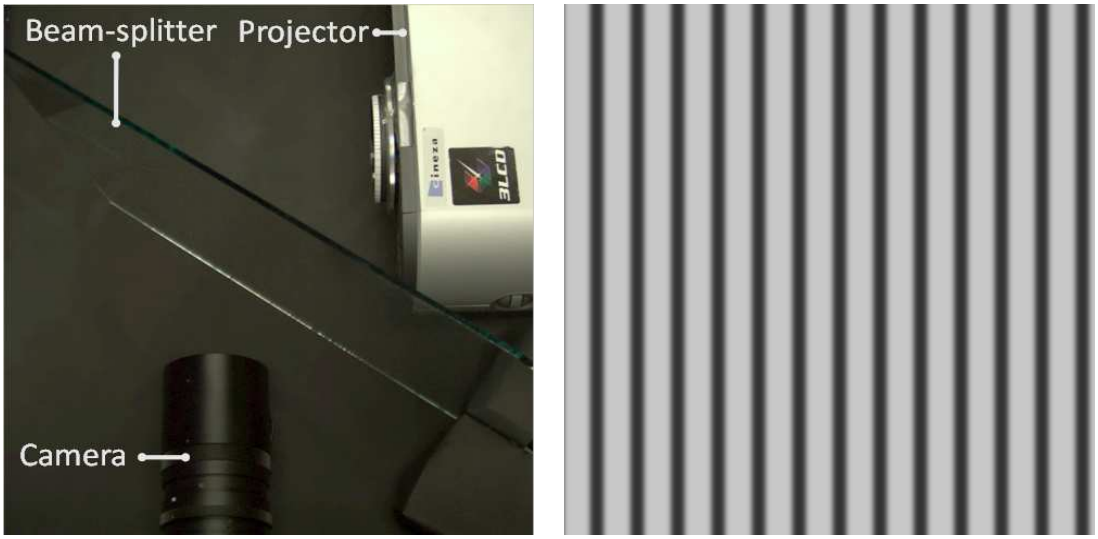
(b) Direct+Indirect Illumination

Figure 2.2: Defocused and Global Illumination as low-pass filters. (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 focus setting. This enables depth recovery even in the presence of global light transport.

Seemingly, defocused illumination and global illumination are very different physical phenomena. Defocus is a result of the optics of the illumination (or imaging) system, and encodes scene depths. On the other hand, global illumination is an intrinsic property of the scene, and encodes material properties and scene geometry. We start with the observation that, interestingly, both defocused and global illumination manifest themselves as **low pass filters** on the incident illumination. An out-of-focus projector projects a blurred image. On the other hand, volumetric scattering around street lights in the presence of fog results in a glow, which again is blurring of the illumination.

In particular, consider a scene being illuminated by a projector with a periodic high frequency pattern. An example pattern is shown in Figure 2.3(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 Figure 2.2(b)).

Direct Illumination: Consider the illustration in Figure 2.2 (a). The direct component of the radiance $e_i^d(t, f)$ at the scene point S_i is the convolution of the illumination pattern,



(a) (b)

Figure 2.3: 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.

$p_i(t)$, and the defocus blur kernel $b_i(t, f)$ at S_i [2] [3]:

$$e_i^d(t, f) \quad = \quad \alpha_i \; p_i(t) * b_i(t, f) \, . \tag{2.1}$$

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 Figure 2.2(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:

$$e_i^g(t, f) \quad = \quad \sum_{S_j \in Scene, j \neq i} m_{ij} \, p_j(t) * b_j(t, f) \, . \tag{2.2}$$

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

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

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

$$e_i(t, f) \;=\; e_i^d(t, f) + e_i^g(t, f)\,. \tag{2.3}$$

We compactly write the expression for radiance at scene point S_i using Eqs. 2.1, 2.2 and 2.3:

$$e_i(t, f) \;=\; \sum_{S_j \in Scene} m_{ij}\, p_j(t) * b_j(t, f)\,. \tag{2.4}$$

We have implicitly included the α_i term with m_{ii}. Taking the Fourier transform of Eq. 2.4:

$$E_i(w, f) \;=\; P(w) \sum_{S_j} m_{ij}\, exp(-I\, w\, \phi_j)\, B_j(w, f)\,, \tag{2.5}$$

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

$$E_i(w, f) \;=\; P(w)\;\; B_i(w, f)\;\; G_i(w, f)\,, \tag{2.6}$$

$$G_i(w, f) \;=\; \sum_{S_j} m_{ij}\, exp(-I\, w\, \phi_j)\, \frac{B_j(w, f)}{B_i(w, f)}\,. \tag{2.7}$$

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 2.6 states that the observed radiance profile at S_i is the convolution of the input pattern with the defocus blur kernel and the global illumination blur kernel. By expressing both these phenomena as blur kernels, now we don't have to explicitly model either of these. This enables us to borrow tools from signal-processing literature to analyze both these effects. Note that the above analysis and the techniques presented in this section do not make any assumption on the particular form of the blur kernels.

We compute $E_i(w, f)$ by taking the Discrete Fourier Transform of the observed radiance profile. We use the third coefficient of the DFT ($w = 3$) as a measure of the amount of blur, as we empirically found it to be the most informative coefficient. 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.

2.3 Relationship between global light transport 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. For a variety of applications, especially those in computer vision, it is desirable to separate the effects of defocus and global illumination. Indeed, both the blurs encode different information about the scene, which can be recovered by separating them. How can the two kernels be separated? Fortunately, we know that the defocus kernel $B(f)$ can be modulated by changing the projector focus settings. In this section, we study the dependence of the global illumination blur kernel on the projector focus setting.

2.3.1 Invariance of global illumination blur to projector defocus

Consider the expression for the global illumination blur at a scene point S_i, as given in Equation 2.7. In the summation, the contribution from points S_j which are distant from S_i is low because the form factors m_{ij} fall rapidly with distance. Additionally, in a local neighborhood, the form factors remain nearly constant. On the other hand, the phase terms vary between -1 and 1, thus canceling out the contributions from a local neighborhood. For points at relatively small distances from S_i, although the form factor can be large, the defocus blur kernel is approximately the same as that of S_i. Consequently, the ratio $\frac{B_j(w,f)}{B_i(w,f)}$ remains nearly constant as the focus setting is changed. Thus, the global illumination blur kernel $G_i(w, f)$ is *nearly invariant* to the projector focus setting.

In the following, we provide empirical validation for the above observation using scenes exhibiting strong inter reflections and sub-surface scattering. In appendix 2.7 at the end of this section, we provide validation using simulations for different distributions of scene points.

2.3.2 Empirical validation of the invariance

For the purpose of validation, we measure $G(f)$ for a wide range of projector focus settings f. For a scene point S_i, we can compute $G_i(f)$ up to a constant scale factor by identifying

another scene point S_j which does not receive any global illumination, and has the same depth as S_i. Using Eq. 2.6 and noting that $B_i(f) = B_j(f)$:

$$\frac{G_i(f)}{\alpha_j} = \frac{E_i(f)}{E_j(f)} . \tag{2.8}$$

Experimental Setup: We use a co-located camera-projector system as shown in Figure 2.3 (a). Our system consists of a Sony Cineza 3-LCD video projector and a Lumenera Lu165C 12-bit camera. The projector focus setting is changed by rotating the focus ring manually. Markings were made on the focus ring to be able to replicate the focus settings. We use the pattern shown in Figure 2.3 (b) to illuminate the scene.[4] This pattern has a period of 24 pixels in the horizontal direction [131]. For each focus setting, we acquire 24 images as the pattern is translated horizontally, one pixel at a time. The total number of images acquired is $24 \times F$, where F is the number of focus settings used. The acquisition time is approximately 1 minute per focus setting.

Validation Results: We design experiments to establish the invariant for both sub-surface scattering and inter-reflections. **For inter-reflections**, we construct a V-groove using two diffuse planes, as shown in Figure 2.4(a). Figures 2.4(b-d) show sample input images for three out of six focus settings. We compute $E(f)$ 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 (Figure 2.4(a)). In this case, the scene point A does not receive any global illumination. Figures 2.4(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 2.4(k-l) show discrete-time Fourier transform $E(\omega, f)$ of (i) and (j). Figure 2.4(m) shows the plot of $E(3, f)$ for points A (no global illumination) and B (with global illumination). The global illumination blur $G(f)$ is computed by taking the point-wise ratio of the two curves, according to Eq. 2.8. Figure2.4(n) shows the plot of scaled $G(3, f)$ at point B. Figures 2.4(k-l) shows 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 Figure 2.4(n), the total variation in $G(f)$ is less than 5% over the entire range of focal plane positions (0.3m-2.5m). It can also be observed in Figures 2.4(k-l) that the global illumination kernels remain nearly constant, while the defocus kernels show significant variation. This validates our claim that the global illumination blur resulting from inter-reflections is insensitive to the projector focus setting. **Second**, we observe that the plots for $E(f)$, with and

[4]Note that any periodic pattern with broad-band frequency response can be used.

without global illumination, achieve maxima at the same focal plane position, as shown in Figure 2.4(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 Figure 2.5 (a). We choose a point B on the exposed part which receives global illumination in the form of sub-surface scattering. Point A, on the same vertical column and lying on the diffuse paper, is at the same depth as B but receives no global illumination. We plot $E(f)$ for A and B in Figure 2.5(i). As before, $G(f)$ 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 global illumination blur kernel remains nearly constant as the projector focus setting is changed significantly.

2.4 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. We follow the frequency domain approach of Zhang et al [131] and derive two depth estimation algorithms. The first algorithm requires a sweep of the focal plane across the scene, acquiring images at multiple focus settings. This is similar in spirit to shape-from-camera-focus techniques which require a sweep of the camera focal plane. Recall that the total blur 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 that we present requires only two focus settings and is similar in spirit to shape-from-camera-defocus methods. Note that our analysis is done independently at each pixel. Hence, we do not impose any smoothness constraints, or require presence/absence of scene texture.

2.4.1 Depth from multiple projector focal planes

In this algorithm, the DFT coefficients $E(f)$ are computed for multiple (≥ 3) focal plane positions f spanning the depth-range of the scene. Since the global illumination blur $G(f)$ is invariant to f, the plot of $E(f)$ against f reflects the behavior of the defocus blur $B(f)$. In other words, $E(f)$ and $B(f)$ attain a maxima at the same focal plane location f, when the corresponding scene point is the best in focus. It follows that scene points at the same depth but receiving different amounts of global illumination share the same maxima location. Two examples are shown in Figures 2.4(m) and 2.5(i). This suggests the maxima location $\overline{f}_i$ as a global-illumination invariant depth measure:

$$\overline{f}_i = \arg\max_f \; E_i(f) \tag{2.9}$$

The resolution of the above depth measure is limited by the number of focal settings used. The resolution can be improved by interpolating the focus measures $E_i(f)$ between the discrete focal plane settings [88]. As a one time calibration step, we compute a one-to-one mapping between scene depths and $\overline{f}_i$ using a planar, diffuse reflective board, whose depths are known a priori (see Figure 2.6 (a)). This mapping, along with the estimates of $\overline{f}$, is used to compute the actual depths for a given scene. This algorithm can be considered a dual to the shape-from-camera-focus technique, where depths are computed by sweeping the camera focal plane across the scene.

2.4.2 Depth from two projector focal planes

In this algorithm, we compute depths as a function of a defocus measure defined using only two focal positions f_1 and f_2. Since $G_i(f)$ is invariant to f, $G_i(f_1) = G_i(f_2)$. Using Eq. 2.6, we define the following ratio measure which is invariant to global illumination:

$$\Omega_i \;=\; \frac{E_i(f_2)}{E_i(f_1)} \;=\; \frac{B_i(f_2)}{B_i(f_1)}\,. \tag{2.10}$$

We compute a mapping (monotonic) between scene depths and Ω_i using a planar calibration board, as shown in Figure 2.6 (b). This mapping, along with the estimates of Ω is used to compute the actual depths for a given scene.

Results Figure 2.7 shows results of our techniques for the V-groove and the candle scenes. The single focal plane algorithm [131] over-estimates the defocus blur due to inter-reflections and sub-surface scattering resulting in incorrect depth estimates near the concavity and for the exposed parts of the candle. Our depth from two planes (Section 2.4.2) and multiple planes (Section 2.4.1) algorithms reconstruct both the shapes accurately. Theoretically, 3 focal planes are sufficient for the multiple focal planes algorithm. For robustness, we used 6 to 8 focal plane positions. Since we compute depths independently at every pixel, fine details such as the wick of the candle are reconstructed as well. The ground truth depths in Figure 2.7 were acquired using a calibration plane with pre-measured depths. We demonstrate our algorithms on scenes with complex shapes and material properties, and significant global illumination, as explained below.

Candle and marble scene (Figure 2.8): 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 [131] 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 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. [5]

Synthetic materials scene (Figure 2.9): 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 (Figure 2.12), 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.

Real and fake materials scene (Figure 2.10): This scene consists of real and fake flowers, real and fake fruits and milk with different fat content. Objects present in this scene exhibit 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. Similar effect can be noticed in the Candles and soaps scene (Figure 2.11), 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 lesser.

Failure case: If a scene point receives strong global illumination contributions from other scene points at significantly different depths, the global illumination blur is not invariant to the projector focus setting. This can happen in a scene with strong, long range inter-reflections, for example due to mirrors. In this case, our techniques will fail to completely account for the errors due to global illumination. This is illustrated in Figure 2.13. The pot receives strong, specular inter-reflections from a distant mirror (not visible in the image). 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. The projector and

[5] 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.

camera were not co-located in this experiment. Depth computation was not performed in the shadow regions.

2.5 Direct-Global Separation under Defocus

The direct component can be considered as the *purest measurement* of a scene's material properties. On the other hand, the global component encodes the optical interactions between different scene objects. Separating the direct component from the global component provides useful insights into various scene properties, which their sum can not.

The algorithm proposed in [87] separates the direct and global components of light transport with a single 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 derive measures of global illumination that are invariant to projector defocus. 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 [87], the max-image, $e^+(f)$, computed by taking pixel-wise maximum, receives *approximately half* the global component. In the presence of defocus blur, the illumination pattern gets blurred. However, since the period of the pattern remains the same, this approximation still holds. Thus, using Eqs. 2.1 and 2.3, we write the expression for $e^+(f)$ in the presence of defocus:

$$e_i^+(f) = \beta_i^+(f)\, e_i^d + 0.5\, e_i^g\,, \tag{2.11}$$

$$\beta_i^+(f) = \max_t \{\, p_i(t) * b_i(t,f)\}\,. \tag{2.12}$$

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, $e^-(f)$:

$$e_i^-(f) = \beta_i^-(f)\, e_i^d + 0.5\, e_i^g\,, \text{ where} \tag{2.13}$$

$$\beta_i^-(f) = \min_t \{\, p_i(t) * b_i(t,f)\}\,. \tag{2.14}$$

These equations are generalizations of the separation equations given in [87], 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 $\beta_i^-(f) = 0$.

2.5.1 Separation using multiple focal planes

In this section, we present a separation technique using multiple focal planes. We use a checker-board illumination pattern as in [87]. Input images (about 25) [6] are acquired at different focus settings. Figure 2.15 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. Using a Gaussian interpolation scheme similar to previous section, we compute $\overline{e}_i^+$ and $\overline{e}_i^-$, the extrema values of $e_i^+(f)$ and $e_i^-(f)$ respectively. An example plot for a point on the candle is shown in Figure 2.14 (a). Note that the curve for $e_i^+(f)$ attains a maximum, while the curve for $e_i^-(f)$ attains a minimum. The computed images $\overline{e}_i^+$ and $\overline{e}_i^-$ are the max and min image respectively *as if* the scene is in perfect focus. Thus, we can write the separation equations as:

$$\overline{e}_i^+ = e_i^d + 0.5\, e_i^g \tag{2.15}$$

$$\overline{e}_i^- = 0.5\, e_i^g \tag{2.16}$$

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

2.5.2 Separation using one plane and a depth map

Here, we present an algorithm to compute separation in the presence of defocus blur using a single focal plane and a depth map of the scene computed using the techniques presented in Section 2.4. For a scene point S_i, the direct and the global component are given using Eqs. 2.11 and 2.13:

$$e_i^d = \frac{e_i^+(f) - e_i^-(f)}{\beta_i^+(f) - \beta_i^-(f)}, \tag{2.17}$$

$$e_i^g = e_i - e_i^d, \tag{2.18}$$

where e_i is the observed intensity when the scene is fully lit. The denominator in Eq. 2.17

[6] Theoretically, only 2 images are required. For robustness, we acquire multiple images

encodes the effects of defocus blur, and needs to be eliminated in order to recover the direct and global components. To this end, we build a mapping between $\left(\beta_i^+(f) - \beta_i^-(f)\right)$ and scene depths using a flat diffuse inclined plane with known depths and no global illumination, as shown in Figure 2.14 (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:

$$\beta_r^+(f) - \beta_r^-(f) \quad = \quad \frac{e_r^+(f) - e_r^-(f)}{e_r}\,, \tag{2.19}$$

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 Eq. 2.17 with Eq. 2.19, to recover the direct and global components.

Experiments and results for direct-global separation: For direct-global separation, we use the same setup as for depth estimation. We illuminate the scene with a 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 2.16 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 [87] 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 (0.3m - 2m), significantly more than the depth of field of the projector, as shown in Figures 2.17. The technique in [87] 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 2.18 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 ($\sim$ 30 cms.). For more results and comparison,

see the project web-page [2].

2.6 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.

Defocused illumination from area light sources: We have discussed defocused illumination in the context of projectors. However, illumination defocus is a more general effect which can be observed whenever an occluder is placed in front of an area light source. If the occluder is a long, thin mask, for example a fence, the shadow would be sharp near the obstacle and becomes more and more blurred as one moves away from the obstacle, as illustrated in Figure 2.19 (a). This effect is similar to defocus observed with projectors.

Formally, the shape of the occluder (mask) acts as the projector-pattern. For example, the fence shown in Figure 2.19 (a) simulates vertical projector stripes. The mask also acts as the focal plane of the setup - scene points close to the mask receive sharp shadow and points further away receive defocused shadow. The area light source acts as the aperture of the projector - larger the source, more the defocus. We can change the *focus setting* by either moving the mask or by changing the size of the area source. With this setup, we can use sun as the light source and port all our techniques to outdoor settings, where the dynamic range of projectors may not be sufficient. Sun has previously been used for outdoor scene recovery [67, 10, 87]. However, illumination defocus effects have not been considered, thus requiring the occluder to be very close to the scene.

Limitations: 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 Figure 2.13.

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.

2.7 Appendix: Validation of invariance 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 Equation 2.7. To account for intensity fall-off, occlusions and multiple bounces, we assume that the transfer coefficients m_{ij} between two points S_i and S_j to be inversely proportional to D_{ij}^2, the square-distance between them. Thus:

$$m_{ij} \propto \frac{1}{D_{ij}^2} \tag{2.20}$$

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

$$m_{ij} \propto \frac{1}{D_{ij}^2} \, exp(-D_{ij}) \tag{2.21}$$

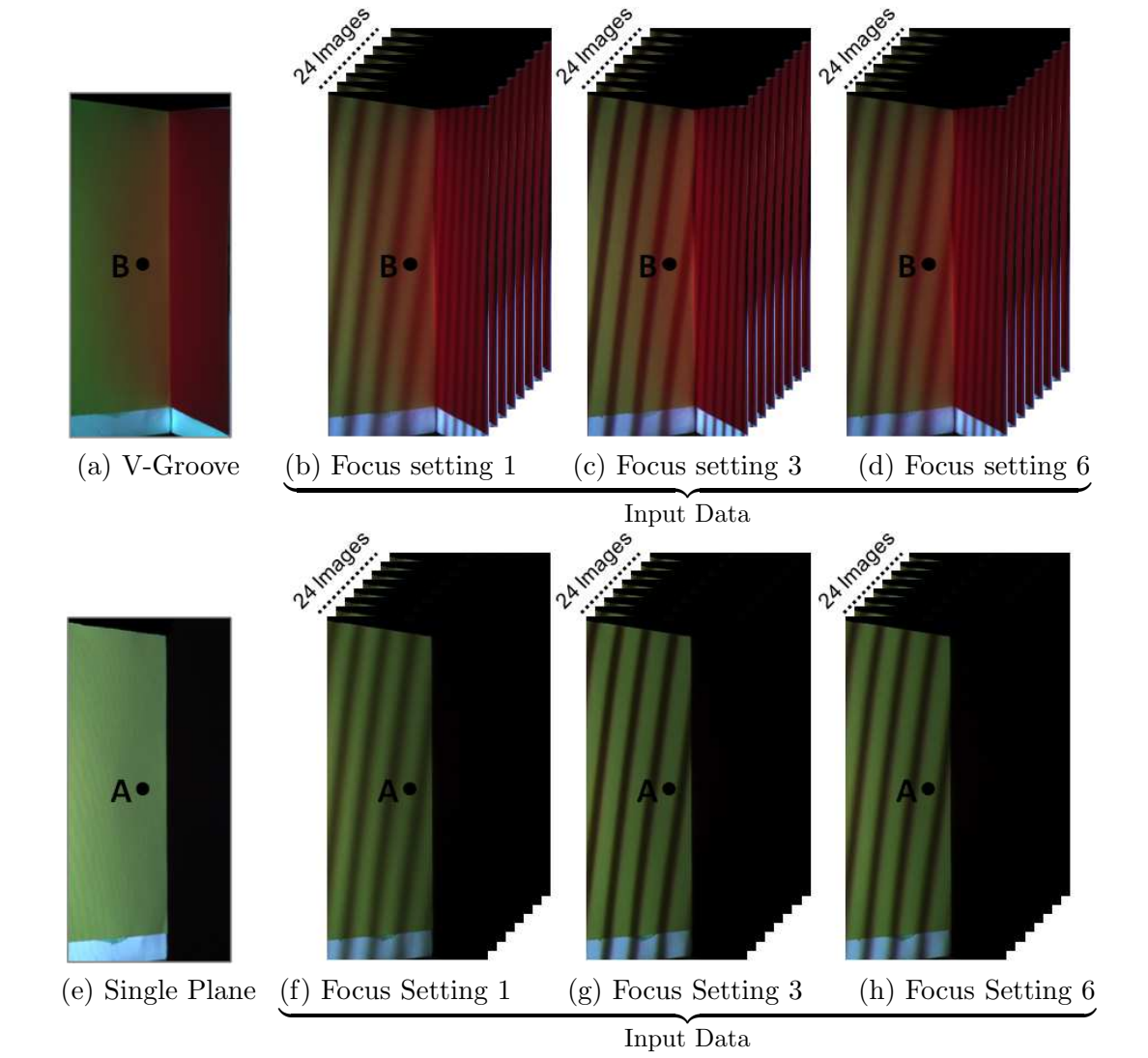
For diffusion, we use the following expression for the form factor:

$$m_{ij} \propto \frac{1}{D_{ij}} \, exp(-D_{ij}) \tag{2.22}$$

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 2.20 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 2.21 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.



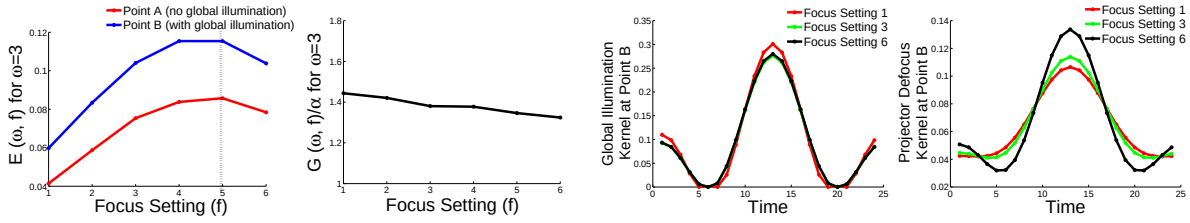


Figure 2.4: Experiments to show the invariance of global illumination blur (due to inter-reflections) to the projector focus setting. (a) A V-groove constructed by placing two planes. (b-d) Sample input images for three out of six focus settings. (e) A single plane after removing the right, red colored plane from the V-Groove. (f-h) Sample input images for three out of six focus settings. (i) Plot of $E(3, f)$ for Points A (no global illumination) and B (with global illumination). (j) Plot of scaled $G(3, f)$ at point B. The relative variation in $G(3, f)$ is less than 5%. (k) Global illumination blur kernels computed at different focus settings. (l) Defocus blur kernels at different focus settings. The global illumination kernels remain nearly constant, while the defocus kernels show significant variation.

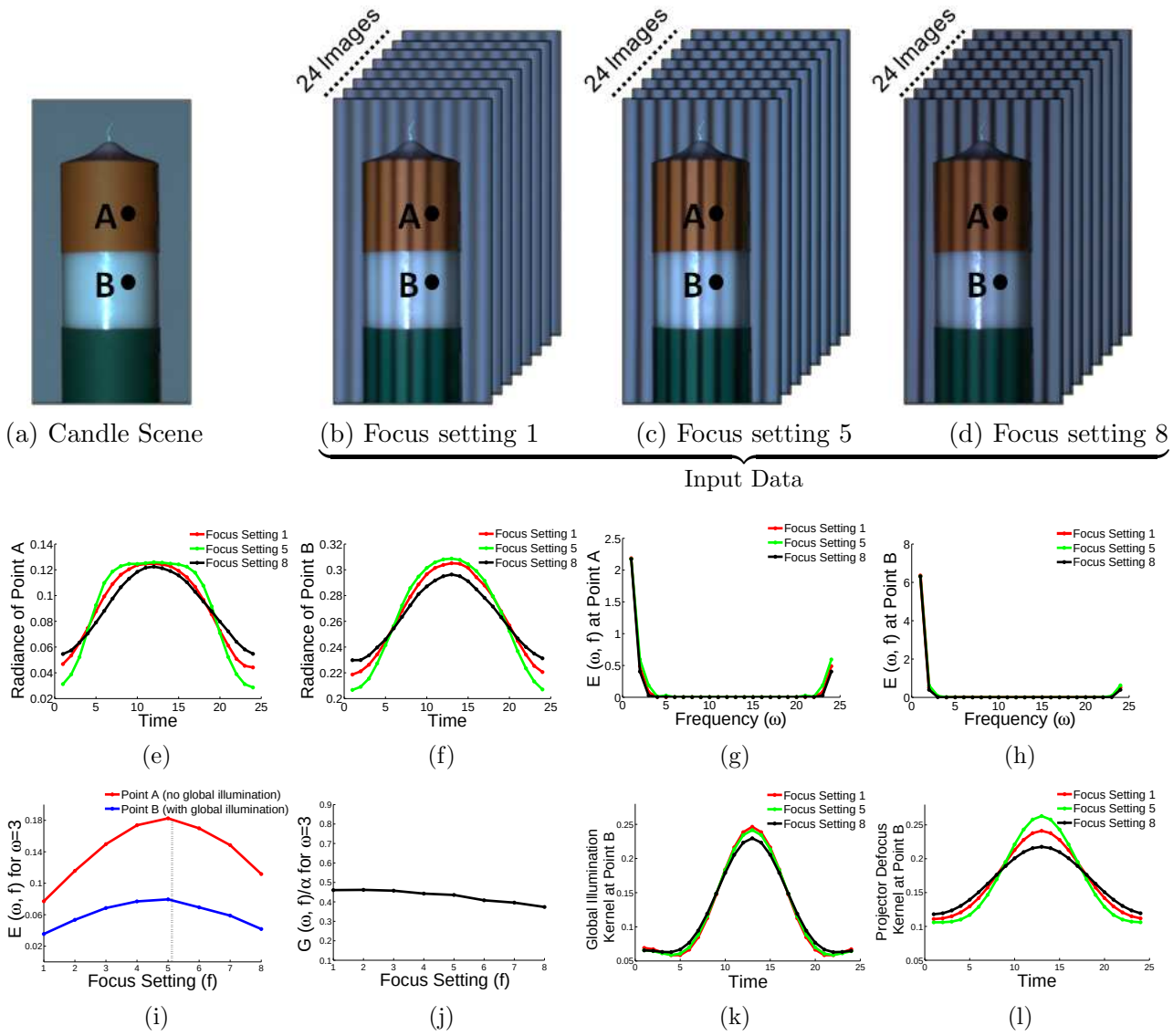


Figure 2.5: Experiments to show the invariance of global illumination blur (due to sub-surface scattering) to the projector focus setting. (a) Candle scene. (b-d) Sample input images for three out of eight focus settings. (e-f) Temporal intensity profiles at Point A and B respectively. (g-h) Discrete-time 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 $G(3, f)$ at point B. This is computed by taking the ratio of the two curves in (d) as in Eq. 2.8. The relative variation in $G(3, f)$ is less than 7% across the range of projector settings. (k) Global illumination blur kernels computed at different focus settings. (l) Defocus blur kernels at different focus settings. The global illumination kernels remain nearly constant, while the defocus kernels show significant variation.

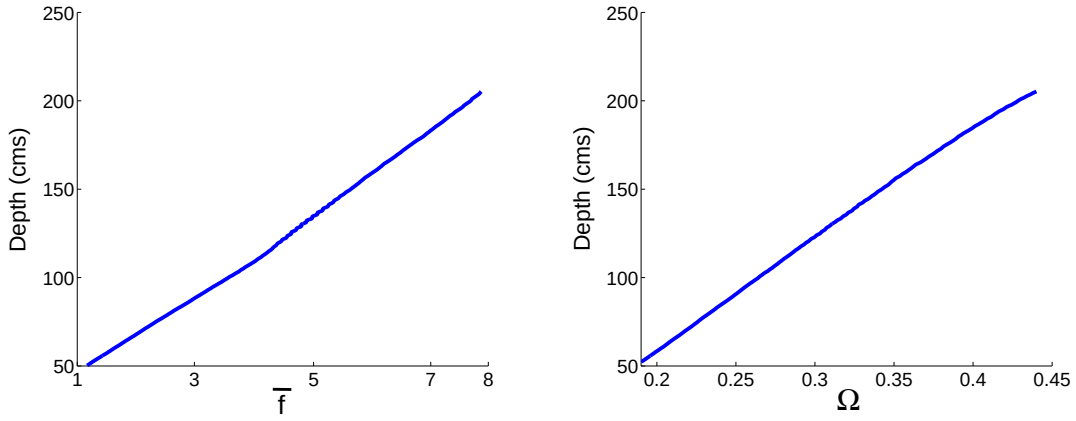


Figure 2.6: Mappings between (a) scene depths and the focus measure $\overline{f}$, (b) scene depths and the defocus measure Ω.

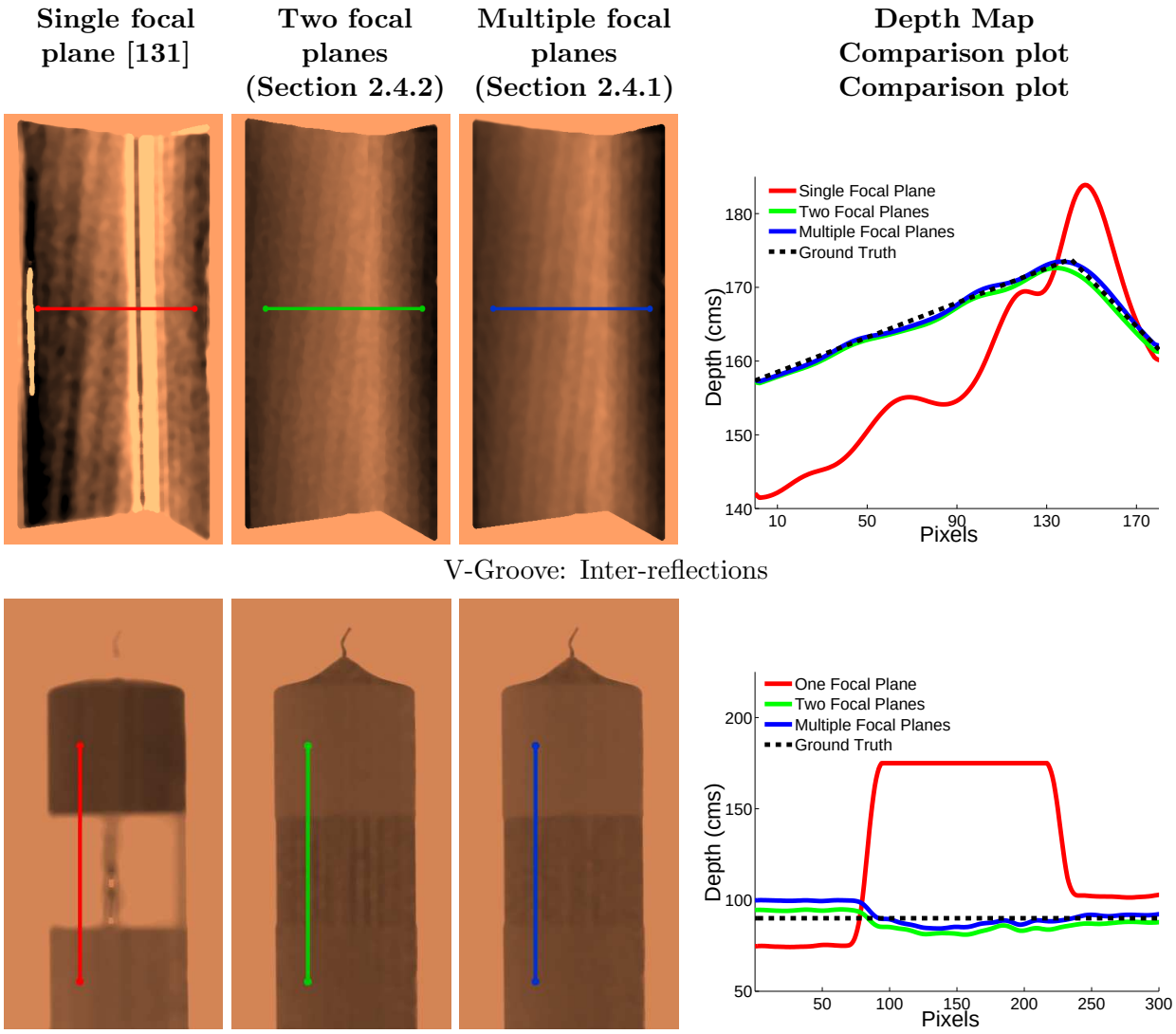
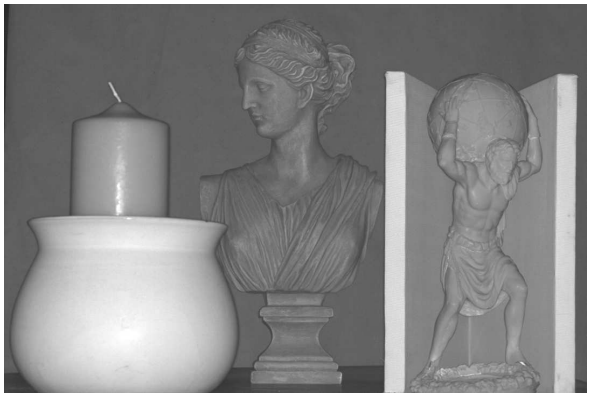


Figure 2.7: **Comparison of the three depth recovery techniques for the V-groove (Fig. 2.4) and the candle (Fig. 2.5) scenes.** The single focal plane algorithm [131] over-estimates the defocus blur in the presence of inter-reflections and sub-surface scattering. This results in incorrect depth estimates near the concavity of the V-groove and for the exposed parts of the candle. On the other hand, the relative RMS error for our algorithms is less than 1% for the V-groove and less than 5% for the candle. The ground truth depths were acquired using a calibration inclined plane with pre-measured depths.

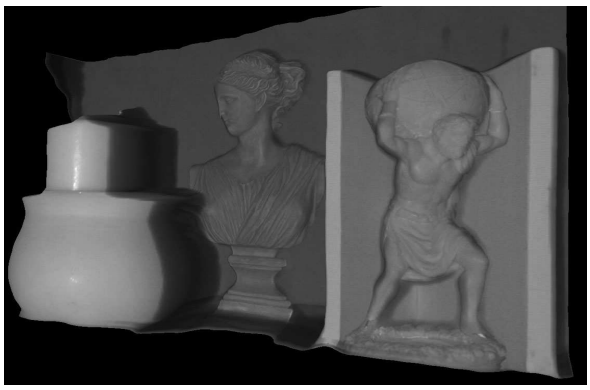


(a) 'Candle and Marble' Scene

(b) Depth using single focal plane [131]

(c) Depth using two focal planes (Sec. 2.4.2)

(d) Depth using multiple focal planes (Sec. 2.4.1)



(e,f) Renderings from alternative view-point using depth map from (d)

Figure 2.8: **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 [131]. 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 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).

(a) 'Synthetic materials' scene

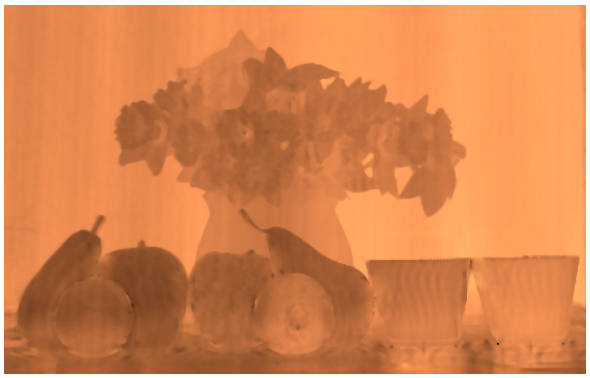
(b) Depth using single focal plane [131]

(c) Depth using two focal planes (Sec. 2.4.2)

(d) Depth using multiple focal planes (Sec. 2.4.1)

Figure 2.9: **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 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. (c,d) Depth maps computed using our techniques.

(a) 'Real and fake materials' scene



(b) Depth using single focal plane [131]

(c) Depth using two focal planes (Sec. 2.4.2)

(d) Depth using multiple focal planes (Sec. 2.4.1)

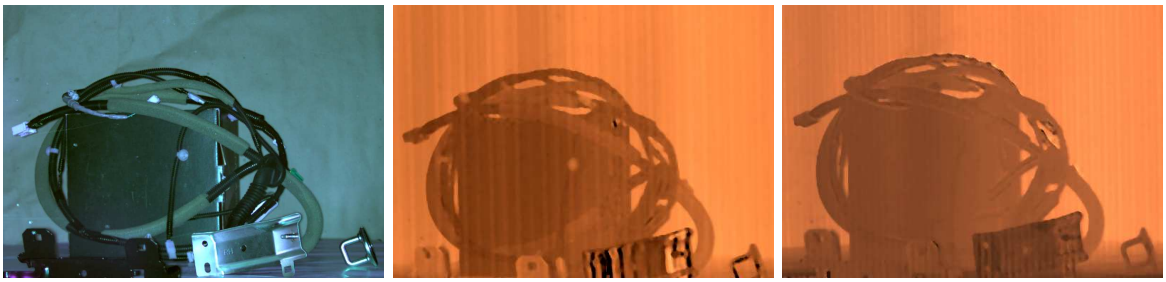
Figure 2.10: **Depth computation for the 'Real and fake materials' scene.** (a) This scene consists of real and fake flowers, real and fake fruits and milk with different fat content. Materials present in this scene exhibit varying degrees of translucency. (b) The single focal plane algorithm does not account for different material properties, thus computing incorrect depths. For instance, in the correct depth map, the two milk glasses should have the same depths. (c,d) Depth maps computed using our techniques.

(a) 'Candles and soaps' scene

(b) Depth using single focal plane [131]

(c) Depth using multiple focal planes (Sec. 2.4.1)

Figure 2.11: **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 algorithm 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.



(a) 'Industrial parts' scene (b) Depth using single focal plane [131] (c) Depth using multiple focal planes (Sec. 2.4.1)

Figure 2.12: **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.

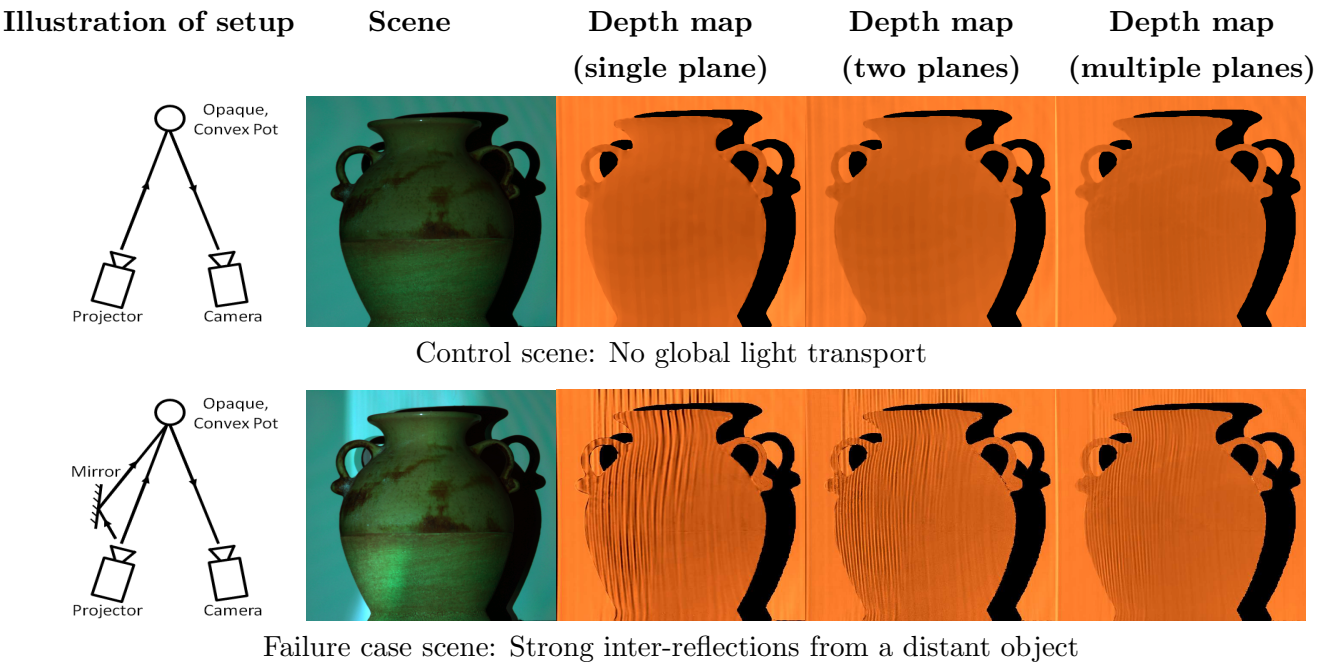
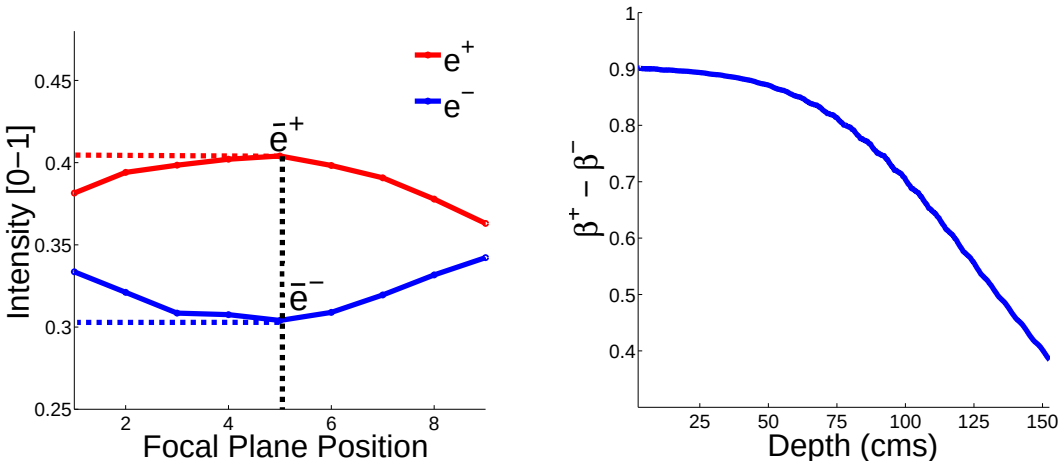
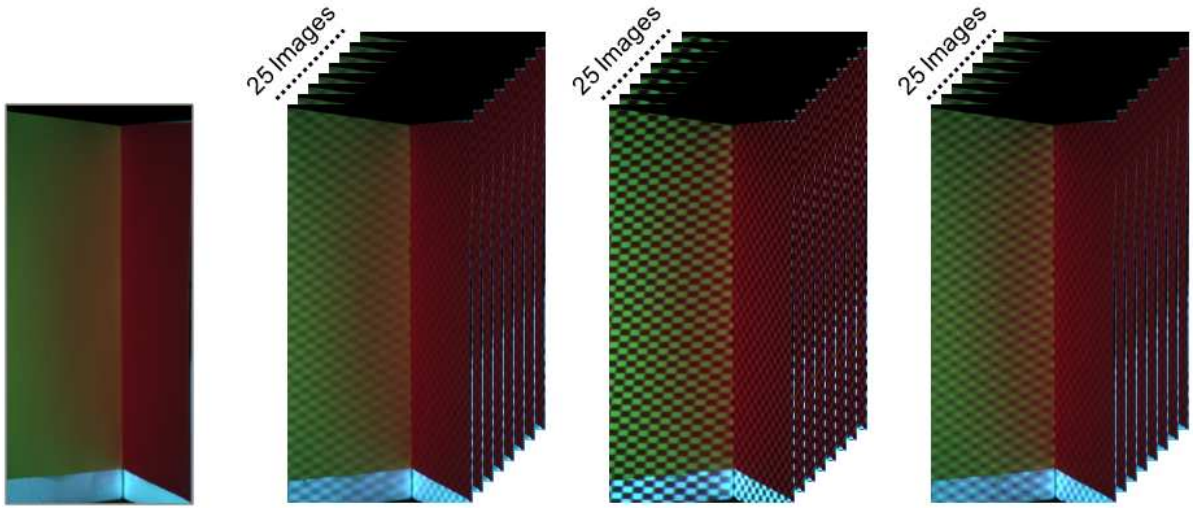


Figure 2.13: **A failure case:** Top row: Control scene. In the absence of global light transport, all three techniques compute accurate depth maps. Bottom row: Failure case scene. The pot receives strong, specular inter-reflections from a distant mirror (not visible in the image). Depth map computed 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.



(a) (b)

Figure 2.14: (a) Separation using multiple focal planes. We compute the extrema values of $e_i^+(f)$ and $e_i^-(f)$ and use them for separation in Eqs. 2.15 and 2.16. (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 Eqs. 2.17 and 2.18.



(a) V-groove (b) Focus setting 1 (c) Focus setting 3 (d) Focus setting 6

Input Data

Figure 2.15: **Input images for direct-global separation.** The multiple focal planes algorithm (Sec. 2.5.1) 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 (Sec. 2.5.2) requires images captured only at a single focal plane, and a depth map of the scene.

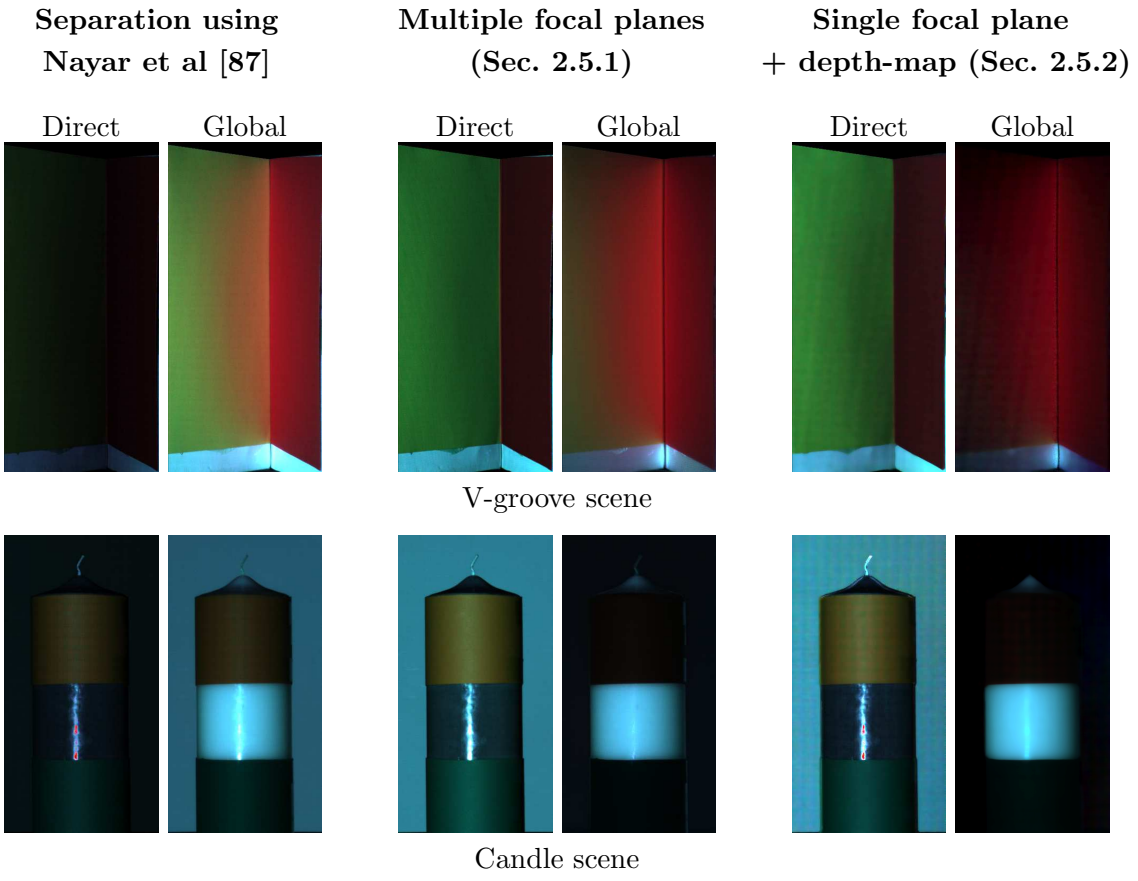
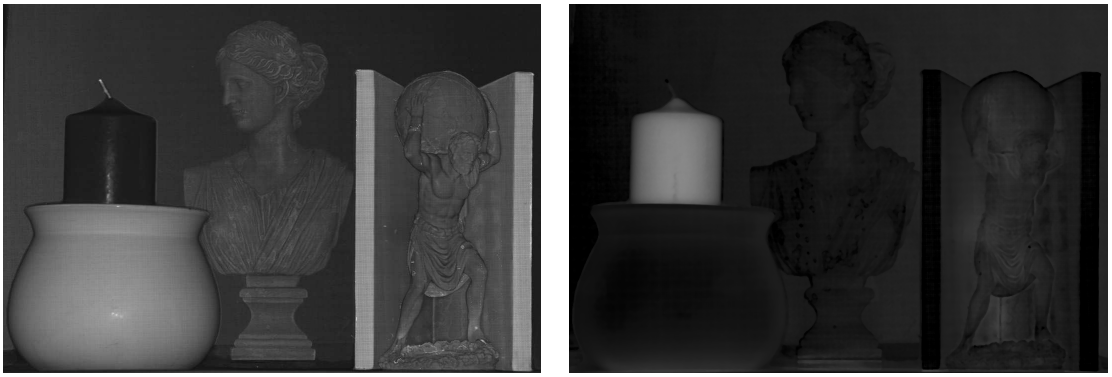


Figure 2.16: **Comparison of the three direct-global separation techniques.** The technique in [87] does not account for illumination defocus and incorrectly estimates the direct and global components. The direct component is underestimated and the global component is over-estimated on the planes of 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.

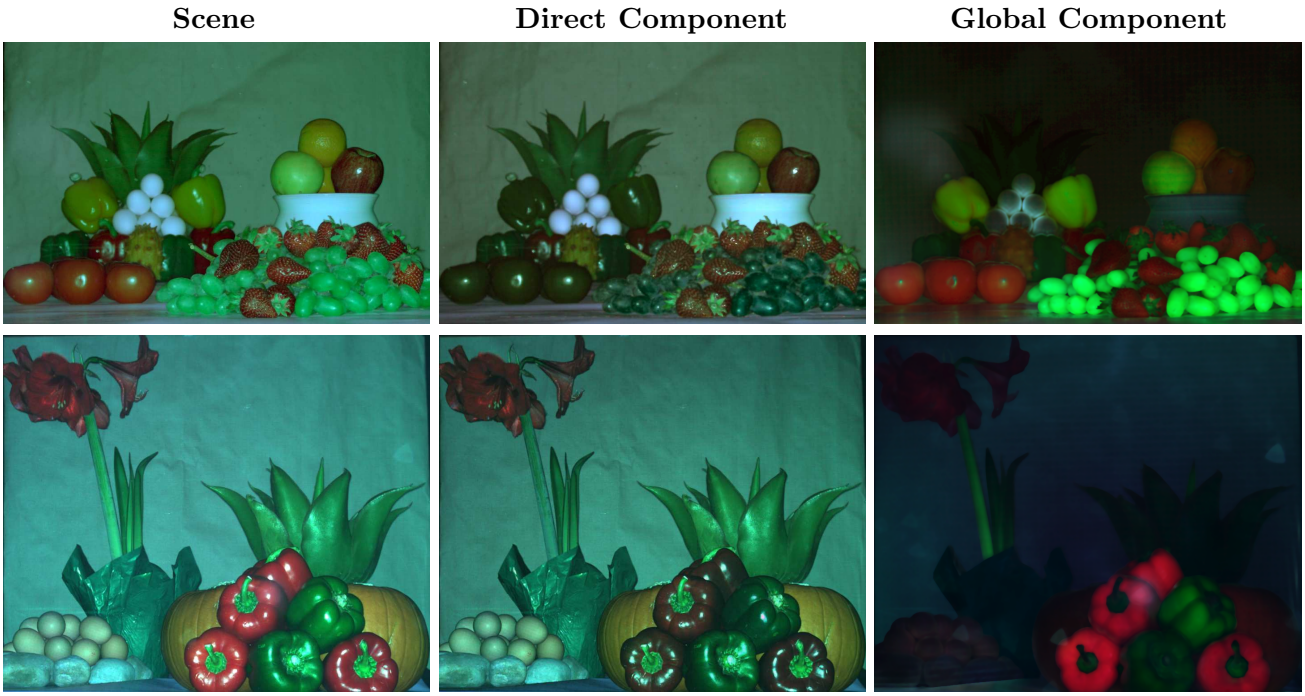
(a) Direct (top) and Global (bottom) components computed at different focal plane positions using [87] From left to right: projector focal plane in front of the scene, in the middle of the scene and behind the scene.

(b) Direct (left) and Global (right) components computed using multiple focal planes method (Sec. 2.5.1)



(c) Direct (left) and Global (right) components computed using single focal plane + depth map method (Sec. 2.5.2)

Figure 2.17: **Comparison of direct-global separation techniques for the marbles and candle scene.** The depth of the scene is approximately 150 cms, larger than the depth of field of projectors ($\sim$ 30 cms.). Thus, illuminating this scene with a projector results in defocused illumination. The technique in [87] incorrectly produces different direct-global separation for different focus settings. On the other hand, our techniques account for defocus while computing the direct-global separation. Notice the large global component on the candle due to sub-surface scattering and inside the V-groove due to inter-reflections.

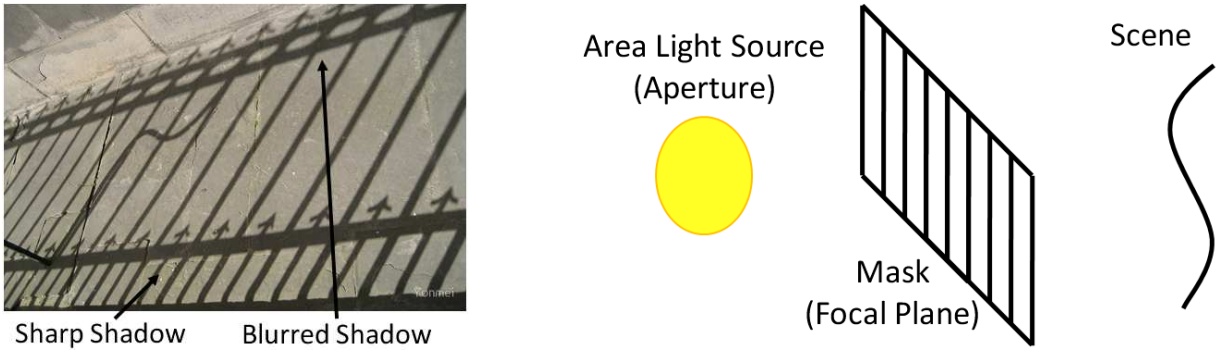


Two scenes consisting of organic materials. Notice the inter-reflections between the eggs.

Scene consisting of translucent candles and soaps.

Scene consisting of synthetic materials.

Figure 2.18: **Direct-Global separation for different scenes computed using the multiple focal planes algorithm (Sec. 2.5.1).** The depth range of all the scenes is larger than 150 cms., more than the depth of field of projectors ($\sim$ 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 [2] for more results.



(a) (b)

Figure 2.19: Simulating illumination defocus using an area light source. (a) Placing an obstacle in front of an area light source (e.g., sun) creates a shadow. The shadow is sharp near the obstacle and becomes more and more blurred as one moves away from the obstacle. (b) Using this observation, we can simulate illumination defocus using an area light source and a mask placed in front of it. The mask acts as the focal plane of the setup and the area light source acts as the aperture. We can change the focus settings by either moving the mask or by changing the size of the area source. With this setup, we can use sun as the light source and port all our techniques to outdoor settings. Image in (a) courtesy *http://images-1.redbubble.net/img/art/*.

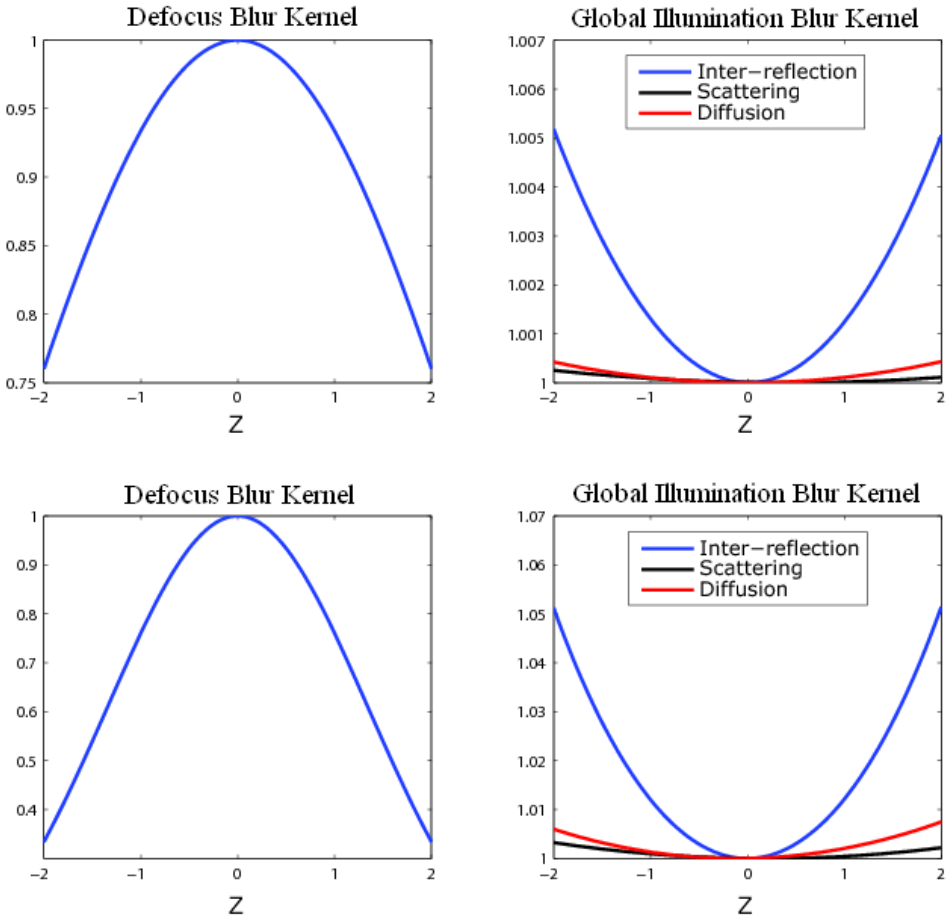


Figure 2.20: Simulation result for symmetric case. The first row is for the first DFT coefficient ($w = \pi/12$ in our setup). The second row is for the second DFT coefficient ($w = \pi/6$). For each row, the first column is the defocus kernel at that frequency, the second column is average global illumination blur over 100 simulations. Blue curve is for inter-reflection, black for sub-surface scattering and red for diffusion.

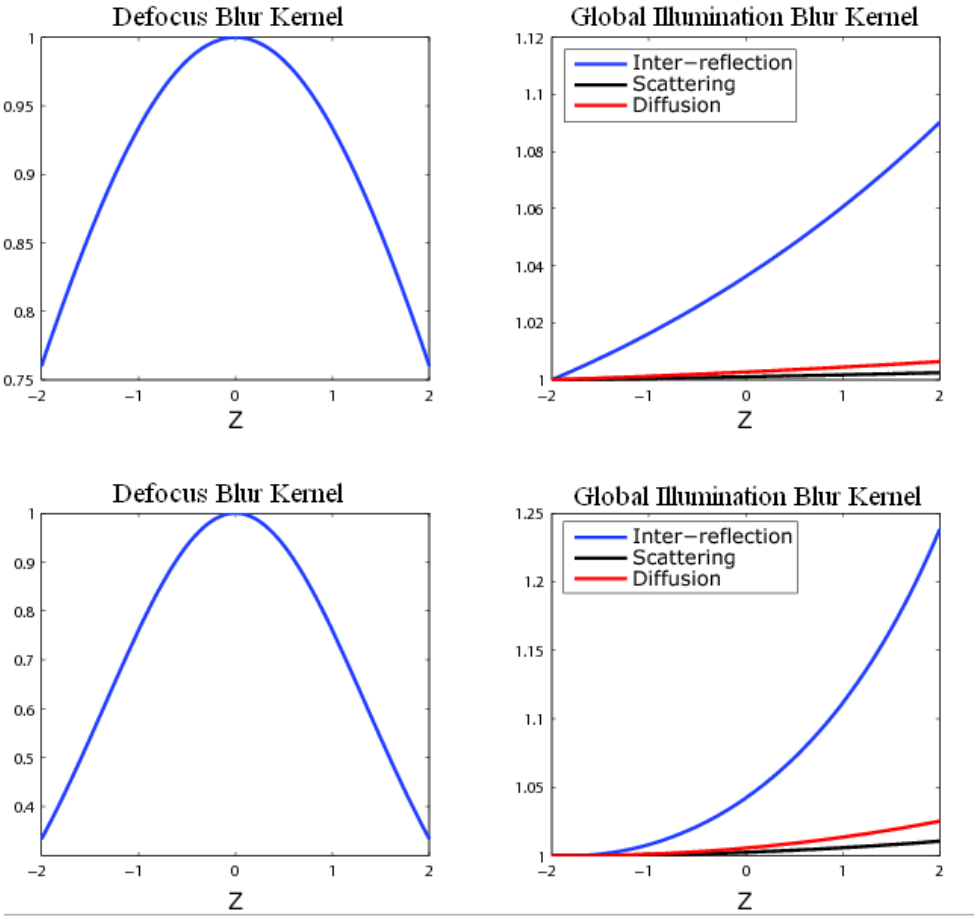


Figure 2.21: Asymmetric case. See the caption of Figure 2.20 for labels. The variation of geometry blur is higher than symmetric case, yet it still remains nearly constant over a large range of f.

Chapter 3

Structured Light 3D Scanning Under Global Light Transport

Structured light triangulation has become the method of choice for shape measurement in several applications including industrial automation, graphics, human-computer interaction and surgery. Since the early work in the field about 40 years ago [128, 78, 95], research has been driven by two factors: reducing the acquisition time and increasing the depth resolution. Significant progress has been made on both fronts (see the survey by Salvi et al [102]) as demonstrated by systems which can recover shapes at close to 1000 Hz. [133] and at a depth resolution better than 30 microns [31].

Despite these advances, most structured light techniques make an important assumption: scene points receive illumination only directly from the light source. For many real world scenarios, this is not true. Imagine a robot trying to navigate an underground cave or an indoor scenario, a surgical instrument inside human body, a robotic arm sorting a heap of metallic machine parts, or a movie director wanting to image the face of an actor. In all these settings, scene points receive illumination indirectly in the form of inter-reflections, sub-surface or volumetric scattering. Such global or indirect illumination[1] effects often dominate the direct illumination and strongly depend on the shape and material properties of the scene. Not accounting for these effects results in large errors in the recovered shape (see Figure 3.1d-e). Moreover, because of the systematic nature of these errors, it is hard to correct them in post-processing.

Our goal is to build an end-to-end system for structured light scanning under a broad range of global illumination effects. First, we formally analyze the errors caused due to different global illumination effects. We show that the types and magnitude of errors depend

[1]Global illumination should not be confused with the oft-used "ambient illumination" in structured light settings. Ambient illumination is subtracted by capturing image with the structured light source turned off.

on the region of influence of global illumination at any scene point. For instance, some scene points may receive global illumination only from a local neighborhood (sub-surface scattering). We call these short-range effects. Some points may receive global illumination from a larger region (inter-reflections or diffusion). We call these long range effects.

The key idea is to design patterns that modulate global illumination and prevent the errors at capture time itself. Short and long range effects place contrasting demands on the patterns. Whereas low spatial frequency patterns are best suited for short range effects, long range effects require the patterns to have high-frequencies. Since most currently used patterns (e.g., binary and sinusoidal codes) contain a combination of both low and high spatial frequencies, they are ill-equipped to prevent errors. We show that such patterns can be converted to those with only high frequencies by applying simple logical operations, making them resilient to long range effects. Similarly, we use tools from combinatorial mathematics to design patterns consisting solely of frequencies that are low enough to make them resilient to short range effects.

But how do we handle scenes that exhibit more than one type of global illumination effect (such as the one in Figure 3.1a)? To answer this, we observe that it is highly unlikely for two different patterns to produce the same erroneous decoding. This observation allows us to project a small ensemble of patterns and use a simple voting scheme to compute the correct decoding at every pixel, without any prior knowledge about the types of effects in the scene (Figure 3.1f). For very challenging scenes, we present an error detection scheme based on a simple consistency check over the results of the individual codes in the ensemble. Finally, we present an error correction scheme by collecting a few additional images. We demonstrate accurate reconstructions on scenes with complex geometry and material properties, such as shiny brushed metal, translucent wax and marble and thick plastic diffusers (like shower curtains).

Our techniques do not require explicit separation of the direct and global components of scene radiance and hence work even in scenarios where the separation fails (e.g., strong inter-reflections among metallic objects) or where the direct component is too low and noisy (e.g., translucent objects or in the presence of defocus). Our techniques consistently outperform many traditional coding schemes and techniques which require explicit separation of the global component, such as modulated phase-shifting [16]. Our methods are simple to implement and can be readily incorporated into existing systems without significant overhead in terms of acquisition time or hardware.

3.1 Related Work

In this section, we summarize the works that address the problem of shape recovery under global illumination. The seminal work of Nayar et al [86] presented an iterative approach for reconstructing shape of Lambertian objects in the presence of inter-reflections. In the previous chapter (ref. [43]), we presented methods for recovering depths using projector defocus [131] under global illumination effects. Chandraker et al [13] use inter-reflections to resolve the bas-relief ambiguity inherent in shape-from-shading techniques. Holroyd et al [52] proposed an active multi-view stereo technique where high-frequency illumination is used as scene texture that is invariant to global illumination. Park et al [93] move the camera or the scene to mitigate the errors due to global illumination in a structured light setup. Hermans et al [50] use a moving projector in a variant of structured light triangulation. The depth measure used in this technique (frequency of the intensity profile at each pixel) is invariant to global light transport effects. In contrast, our focus is on designing structured light systems while avoiding the overhead due to moving components.

Recently, it was shown that the direct and global components of scene radiance could be efficiently separated [87] using high-frequency illumination patterns. This has led to several attempts to perform structured light scanning under global illumination [15, 16]. All these techniques rely on *subtracting or reducing* the global component and apply conventional approaches on the residual direct component. While these approaches have shown promise, there are three issues that prevent them from being applicable broadly: (a) the direct component estimation may fail due to strong inter-reflections (as with shiny metallic parts), (b) the residual direct component may be too low and noisy (as with translucent surfaces, milk and murky water), and (c) they require significantly higher number of images than traditional approaches, or rely on weak cues like polarization. In contrast, we explicitly design ensembles of illumination patterns that are resilient to a broader range of global illumination effects, using significantly less number of images.

3.2 Errors due to Global Illumination

The type and magnitude of errors due to global illumination depends on the spatial frequencies of the patterns and the global illumination effect. As shown in Figures 3.2 and 3.3, long range effects and short range effects result in incorrect decoding of low and high spatial frequency patterns, respectively. In this section, we formally analyze these errors. For ease of exposition, we have focused on binary patterns. The analysis and techniques are easily extended to N-ary codes.

Binary patterns are decoded by binarizing the captured images into projector-illuminated

vs. non-illuminated pixels. A robust way to do this is to capture two images L and $\overline{L}$, under the pattern P and the inverse pattern $\overline{P}$, respectively [2]. For a scene point S^i, its irradiances L^i and $\overline{L^i}$ are compared. If, $L^i > \overline{L^i}$, then the point is classified as directly lit. A fundamental assumption for correct binarization is that each scene point receives irradiance from only a single illumination element (light stripe or a projector pixel). However, due to global illumination effects and projector defocus, a scene point can receive irradiance from multiple projector pixels, resulting in incorrect binarization.

In the following, we derive the condition for correct binarization in the presence of global illumination and defocus. Suppose S^i is directly lit under a pattern P. The irradiances L^i and $\overline{L^i}$ are given as:

$$L^i = L^i_d + \beta L^i_g \ , \tag{3.1}$$

$$\overline{L^i} = (1-\beta) L^i_g \ , \tag{3.2}$$

where L^i_d and L^i_g are the direct and global components of the irradiance at S^i when the scene is fully lit. β is the fraction of the global component under the pattern P.

In the presence of defocus (projector or camera), the projected patterns and the captured image is blurred. Similarly, projectors with imperfect optics also result in blurring of the projected patterns [3]. Using Gray codes, which have a Hamming distance of one between consecutive code-words gives resistance for up to 1 pixel blur. If the blur is larger, it influences the highest frequency patterns, often completely blurring them out. Unlike subsurface scattering, defocus modulates the direct component as well, as shown in [43]. The image intensities are given as:

$$L^i = \alpha L^i_d + \beta L^i_g \ , \tag{3.3}$$

$$\overline{L^i} = (1-\alpha) L^i_d + (1-\beta) L^i_g \ . \tag{3.4}$$

The fractions (α and $1-\alpha$) depend on the projected pattern and the amount of defocus. In the absence of defocus, $\alpha = 1$. For correct binarization, it is required that $L_i > \overline{L_i}$, i.e.

$$\alpha L^i_d + \beta L^i_g > (1-\alpha) L^i_d + (1-\beta) L^i_g \tag{3.5}$$

[2] The inverse image $\overline{L}$ can be generated by subtracting the image L from the image of the fully lit scene.

[3] For example, pico-projectors are increasingly getting popular for structured light applications in industrial assembly lines. We have noticed that due to imperfect optics, they can not faithfully resolve a striped pattern of 2-pixel width.

This condition is satisfied in the absence of global illumination ($L_g^i = 0$) and defocus ($\alpha = 1$). In the following, we analyze the errors in the binarization process due to various global illumination effects and defocus, leading to systematic errors[4].

Long range effects (diffuse and specular inter-reflections): Consider the scenario when S^i receives a major fraction of the global component when it is not directly lit ($\beta \approx 0$), and the global component is larger than the direct component ($L_d^i < L_g^i$) as well. Substituting in the binarization condition (Eqn. 3.5), we get $L^i < \overline{L^i}$, which results in a binarization error. Such a situation can commonly arise due to long-range inter-reflections, when scenes are illuminated with low-frequency patterns. For example, consider the v-groove concavity as shown in Figure 3.2. Under a low frequency pattern, several scene points in the concavity are brighter when they are not directly lit, resulting in a binarization error. Since the low frequency patterns correspond to the higher-order bits, this results in a large error in the recovered shape.

Short-range effects (sub-surface scattering and defocus): Short range effects result in low-pass filtering of the incident illumination. In the context of structured light, these effects can severely blur the high-frequency patterns, making it hard to correctly binarize them. This can be explained in terms of the binarization condition in Eqn 3.5. For high frequency patterns, $\beta \approx 0.5$ [87]. If the difference in the direct terms $|\alpha\, L_d^i - (1-\alpha)\, L_d^i|$ is small, either because the direct component is low due to sub-surface scattering ($L_d^i \approx 0$) or because of severe defocus ($\alpha \approx 0.5$), the pattern can not be robustly binarized due to low signal-to-noise-ratio (SNR). An example is shown in Figure 3.3. For conventional Gray codes, this results in a loss of depth resolution, as illustrated in Figure 3.4.

In summary, in the presence of long-range effects, low-frequency patterns are susceptible to incorrect binarization. On the other hand, for short-range effects, high-frequency patterns are susceptible to coding errors.

3.3 Patterns for Error Prevention

Errors due to global illumination are systematic, scene-dependent errors that are hard to eliminate in post-processing. In this section, we design patterns that modulate global illumination and prevent errors from happening at capture time itself. In the presence of only long range effects and no short-range effects, high-frequency binary patterns (with equal off and on pixels) are decoded correctly because $\beta \approx 0.5$ [87], as shown in Figures 3.2(f-i). On the other hand, in the presence of short-range effects, most of the global illumination comes from a local neighborhood. Thus, for low frequency patterns, when a scene point is

[4]Errors for the particular case of laser range scanning of translucent materials are analyzed in [38]. Errors due to sensor noise and spatial mis-alignment of projector-camera pixels were analyzed in [123].

directly illuminated, most of its local neighborhood is directly illuminated as well. Hence, $\alpha \geq 0.5$ and $\beta \geq 0.5$. Thus, if we use low frequency patterns for short-range effects, the global component actually helps in correct decoding even when the direct component is low (Figure 3.3).

Because of the contrasting requirements on spatial frequencies, it is clear that we need different codes for different effects. For long range effects, we want patterns with only high frequencies (low maximum stripe-widths). For short-range effects, we want patterns with only low frequencies (high minimum stripe-widths). But most currently used patterns contain a combination of both low and high spatial frequencies. How do we design patterns with only low or only high frequencies? We show that by performing simple logical operations, it is possible to design codes with only high frequency patterns. For short range effects, we draw on tools from the combinatorial maths literature to design patterns with large minimum stripe-widths.

3.3.1 Logical coding-decoding for long range effects

We introduce the concept of logical coding and decoding to design patterns with only high frequencies. An example of logical coding-decoding is given in Figure 3.2. The important observation is that for structured light decoding, *the direct component is just an intermediate representation*, with the eventual goal being the correct binarization of the captured image. Thus, we can bypass explicitly computing the direct component. Instead, we can model the binarization process as a scene-dependent function from the set of binary projected patterns ($\mathbb{P}$) to the set of binary classifications of the captured image ($\mathbb{B}$):

$$f : \mathbb{P} \Rightarrow \mathbb{B} . \tag{3.6}$$

For a given pattern $P \in \mathbb{P}$, this function returns a binarization of the captured image if the scene is illuminated by P. Under inter-reflections, this function can be computed robustly for high-frequency patterns but not for low-frequency patterns. For a low frequency pattern P_{lf}, we would like to decompose it into two high-frequency patterns P^1_{hf} and P^2_{hf} using a pixel-wise binary operator $\odot$ such that:

$$f(P_{lf}) = f\left(P^1_{hf} \odot P^2_{hf}\right) = f\left(P^1_{hf}\right) \odot f\left(P^2_{hf}\right) \tag{3.7}$$

If we find such a decomposition, we can robustly compute the binarizations $f\left(P^1_{hf}\right)$ and $f\left(P^2_{hf}\right)$ under the two high frequency patterns, and compose these to achieve the correct

binarization $f(P_{lf})$ under the low frequency pattern. Two questions remain: (a) What binary operator can be used? (b) How can we decompose a low frequency pattern into two high frequency patterns? For the binary operator, we choose the logical XOR ($\oplus$) because it has the following property:

$$P^2_{hf} \oplus P^1_{hf} = P_{lf} \quad \Rightarrow \quad P^2_{hf} = P_{lf} \oplus P^1_{hf} \tag{3.8}$$

This choice of operator provides a simple means to decompose P_{lf}. We first choose a high-frequency pattern P^1_{hf}. The second pattern P^2_{hf} is then computed by simply taking the pixel-wise logical XOR of P_{lf} and P^1_{hf}. We call the first high frequency pattern the *base* pattern. Instead of projecting the original low frequency patterns, we project the base pattern P^1_{hf} and the second high-frequency patterns P^2_{hf}. For example, if we use the last Gray code pattern (stripe width of 2) as the base pattern, all the projected patterns have a maximum width of 2. We call these the *XOR-02* codes (Figure 3.5). In contrast, the original Gray codes have a maximum stripe-width of 512. Note that there is no overhead introduced; the number of projected patterns remains the same as the conventional codes. Similarly, if we use the second-last pattern as the base-plane, we get the *XOR-04* codes (Figure 3.5).

3.3.2 Maximizing the minimum stripe-widths for short-range effects

Short-range effects can severely blur the high-frequency base plane of the logical XOR codes. The resulting binarization error will propagate to all the decoded patterns. Thus, for short-range effects, we need to design codes with large minimum stripe-width (min-SW). It is not feasible to find such codes with a brute-force search as these codes are extremely rare[5].

Fortunately, this problem has been well studied in combinatorial mathematics. There are constructions available to generate codes with large min-SW. For instance, the 10-bit Gray code with the maximum known min-SW (8) is given by Goddyn et al [37] (Figure 3.5). In comparison, conventional Gray codes have a min-SW of 2. Kim et al [64] used a variant of Gray codes with large min-SW called the antipodal Gray codes to mitigate errors due to defocus. For conventional Gray codes, although short-range effects might result in incorrect binarization of the lower-order bits, the higher-order bits are decoded correctly. Thus, these codes can be used in the presence of short-range effects as well.

[5]On the contrary, it is easy to generate codes with small maximum stripe-width (9), as compared to 512 for the conventional Gray codes, by performing a brute-force search

3.3.3 Ensemble of codes for general scenes

Global illumination in most real world scenes is not limited to either short or long range effects. Codes optimized for long-range effects would make errors in the presence of short-range effects and vice versa. In general, it is not straight-forward to identify which code to use without knowing the dominant error-inducing mode of light transport, which in turn would require a priori knowledge about scene. We show that by projecting a small ensemble of codes optimized for different effects, we can minimize the errors automatically without a priori knowledge about the scene.

We project four codes optimized for different effects: two for long-range (the *XOR-04* codes and the *XOR-02* codes), and two for short-range (the Gray codes with maximum min-SW and the conventional Gray codes). The key idea is that errors made by different codes are nearly random. We can find the correct depth value by comparing the depth values computed using the individual codes. If any two agree within a small threshold, that value is returned. If only the two Gray codes agree, we return the value computed by the Gray code with maximum min-SW, because they result in a better depth resolution. Since the chosen set of codes are optimized for different effects, they cover a large class of optically challenging scenes.

Results: Figure 3.1 shows a scene consisting of a bowl on a marble slab. For our ensemble codes, we project a total of 41 patterns - 10 patterns for each of the 4 codes and 1 all-white pattern. For phase-shifting, we project 18 patterns (3 frequencies, 6 shifts for each frequency). For modulated phase-shifting [16], we project 162 patterns (9 modulated patterns for each phase-shifting pattern). Interestingly, by analyzing the errors made by the individual codes, we can get qualitative information about light-transport. Scene points where only the logical codes agree (marked in light-blue) indicate strong inter-reflections. On the other hand, scene points where only the two Gray codes agree (green) correspond to translucent materials (sub-surface scattering). Scene points where all the codes agree (maroon) do not receive much global illumination.

The scenes in Figure 3.6 has inter-reflections inside the fruit basket and strong sub-surface scattering on the fruits. Modulated phase-shifting performs poorly on translucent materials, whereas conventional Gray codes and phase-shifting produce errors in the presence of inter-reflections. Reconstruction produced using our ensemble of codes has significantly reduced errors. Similarly, figures 3.1 and 3.8 have global illumination in the form of inter-reflections, sub-surface scattering and diffusion. As before, conventional Gray codes and phase-shifting produce errors in the presence of long-range effects (inter-reflections and diffusion) whereas modulated phase-shifting results in errors due to short-range sub-surface scattering. Our ensemble of codes minimize the errors while keeping the number of acquired

images low.

Figure 3.9 shows shape measurement of two translucent materials. Because of significant sub-surface scattering, modulated phase-shifting produces poor reconstructions. In Figures 3.10 and 3.11, the goal is to reconstruct the lamp and the shower curtain respectively. The correct shape of the curtain is planar, without any ripples. Light diffuses through the curtain and is reflected from the background, creating long-range interactions. Similarly, light diffuses through the lamp, is reflected from the interior and comes back out, creating long-range optical interactions. Conventional Gray codes and phase-shifting result in large errors. In the case of the shower-curtain, with only the logical codes optimized for long-range interactions, a nearly error-free reconstruction is achieved, with the number of images the same as the conventional codes.

Our techniques consistently outperform many existing schemes (Gray codes, phase-shifting, and modulated phase-shifting [16]). For some moderately difficult scenes, modulated phase-shifting performs as well as our methods, but requires an order of magnitude more images (162 vs. 11). In these cases, it is sufficient to use only one of our patterns, instead of the full ensemble.

3.4 Error detection and correction

The patterns presented in the previous section can successfully prevent a large fraction of errors. For highly challenging scenes, however, some errors might still be made. For building a reliable shape measurement system, it is critical to detect and correct these residual errors. In this section, we present strategies for error detection and correction.

Error detection: Traditionally, error detection and correction strategies from communication theory have been adopted in the context of structured light. An example is the Hamming error correcting codes used by Minou et al [78]. These techniques treat structured light coding-decoding as a signal transmission problem. Although good for handling random sensor/illumination noise, these codes can not handle the systematic errors made due to global illumination.

We show that the consistency check proposed in the previous section, in addition to preventing errors, can also be used for detecting errors. For a pixel, if none of the four codes agree, we mark it as an *error* pixel. This is illustrated in figure 3.12. Error pixels are marked with red color. It is possible that one of the four values might be the correct value. However, as there is an error correction stage, we take a conservative approach and classify such pixels as error pixels. Since no extra patterns need to be projected, the error detection stage does not place any overhead in terms of acquisition time. Park et

al [93, 92] use similar consistency checks across range scans acquired from different view points. By registering different scans and comparing the values from different scans, they remove spurious measurements due to specular inter-reflections. In contrast, our technique does not require moving the acquisition setup or the object.

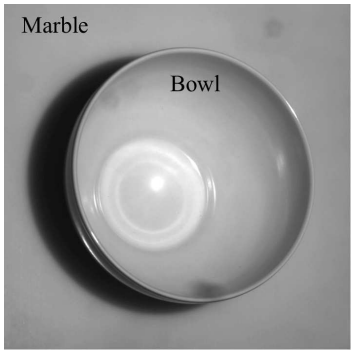
Error correction: To correct the errors, we iteratively collect additional images while illuminating only the scene points corresponding to the error pixels. This technique, based on the work of Xu et al [130], progressively reduces the amount of global illumination, resulting in reduction of the error pixels. The concave lamp made of shiny brushed metal given in figure 3.12 is an extremely challenging object because of high-frequency inter-reflections. Conventional Gray codes can not reconstruct a large portion of the object. Separation based modulated phase-shifting [16] can not account for the high-frequency inter-reflections, resulting in large errors. Our ensemble of codes, while reducing the errors, can not reconstruct the object completely. By acquiring images in 2 extra iterations, we achieve a nearly perfect reconstruction. In this case, we projected only the logical codes in subsequent iterations, thus requiring 81 images in total.

It is important to note that for this error correction strategy to be effective, our error prevention and detection stages are critical. Since our techniques correctly reconstruct a large fraction of the scene in the first iteration itself, we require only a small number of extra iterations (typically 1-2) even for challenging scenes. In comparison, Xu et al's approach requires a large number of iterations (10-20) and images (500-800), even for moderately difficult scenes. This is because they use conventional Gray codes, which do not prevent errors in the first place. Secondly, their error detection technique, based on direct-global separation, is highly conservative. For translucent materials, their error detection mechanism always results in error, and the technique does not converge.

3.5 Limitations

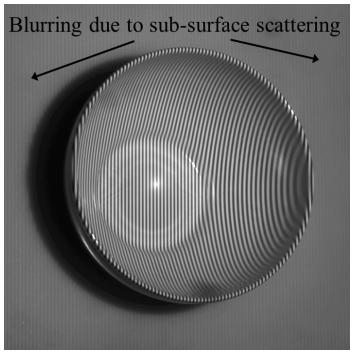
Our techniques can not handle objects with perfectly mirror inter-reflections. The frequency of these optical interactions is higher than what our codes can handle. Additionally, our methods assumes a single dominant mode of light transport for every scene point. If a scene point receives both large short-range and long-range effects, for example, inside of a translucent concave bowl, all our individual codes will result in errors. Since neither code produced the correct result, our combination technique and further error correction steps will not be able to retrieve the correct result. Currently, we have not considered the effects of volumetric scattering. Volumetric scattering is challenging as it results in both short-range and long-range interactions. Narasimhan et al [85] presented a technique for reconstructing the shape of under-water scenes. Their scanning technique is slow and assumes a single

scattering model. A future direction would be to design fast systems for reconstructing shape under a general model of volumetric scattering.

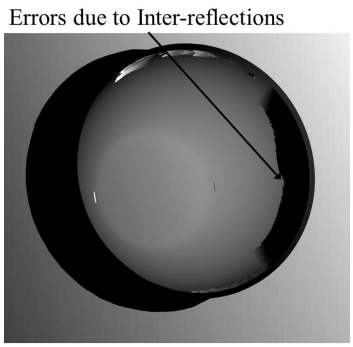


(a) Bowl on a marble slab

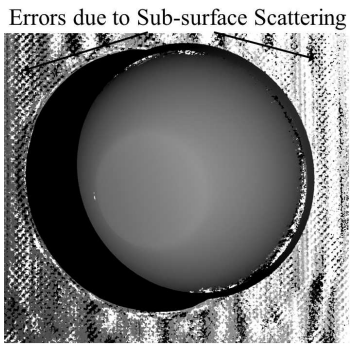
(b) Low-frequency illumination



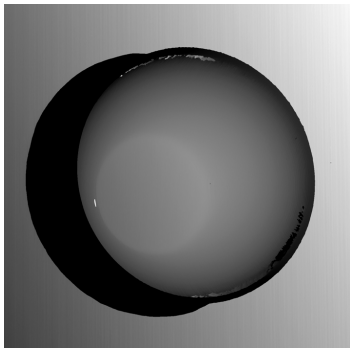
(c) High-frequency illumination



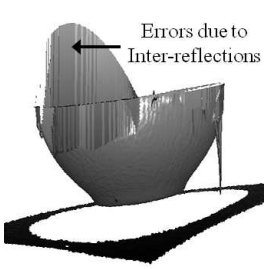
(d) Depth using conventional Gray codes (11 images)

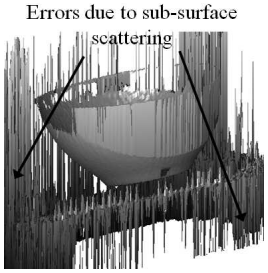


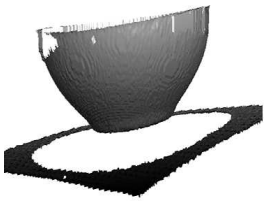
(e) Depth map using modulated phase shifting [16] (162 images)



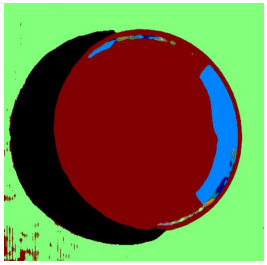
(f) Depth map using our ensemble codes (41 images)







(g-i) 3D visualizations for (d-f) respectively



(j) Qualitative light-transport map

Figure 3.1: **Measuring shape for the 'bowl on marble-slab' scene.** This scene is challenging because of strong inter-reflections inside the concave bowl and sub-surface scattering on the translucent marble slab. Images captured under low frequency patterns (b) and high-frequency patterns (c) are binarized incorrectly due to long range and short-range effects respectively. (d-f) Shape reconstructions. Parentheses contain the number of input images. (d) Conventional Gray codes result in incorrect depths due to inter-reflections. (e) Modulated phase-shifting results in errors on the marble-slab because of low direct component. (f) Our technique uses an ensemble of codes optimized for individual light transport effects, and results in the best shape reconstruction. (j) By analyzing the errors made by the individual codes, we can infer qualitative information about light-transport. Points marked in green correspond to translucent materials. Points marked in light-blue receive heavy inter-reflections. Maroon points do not receive much global illumination.

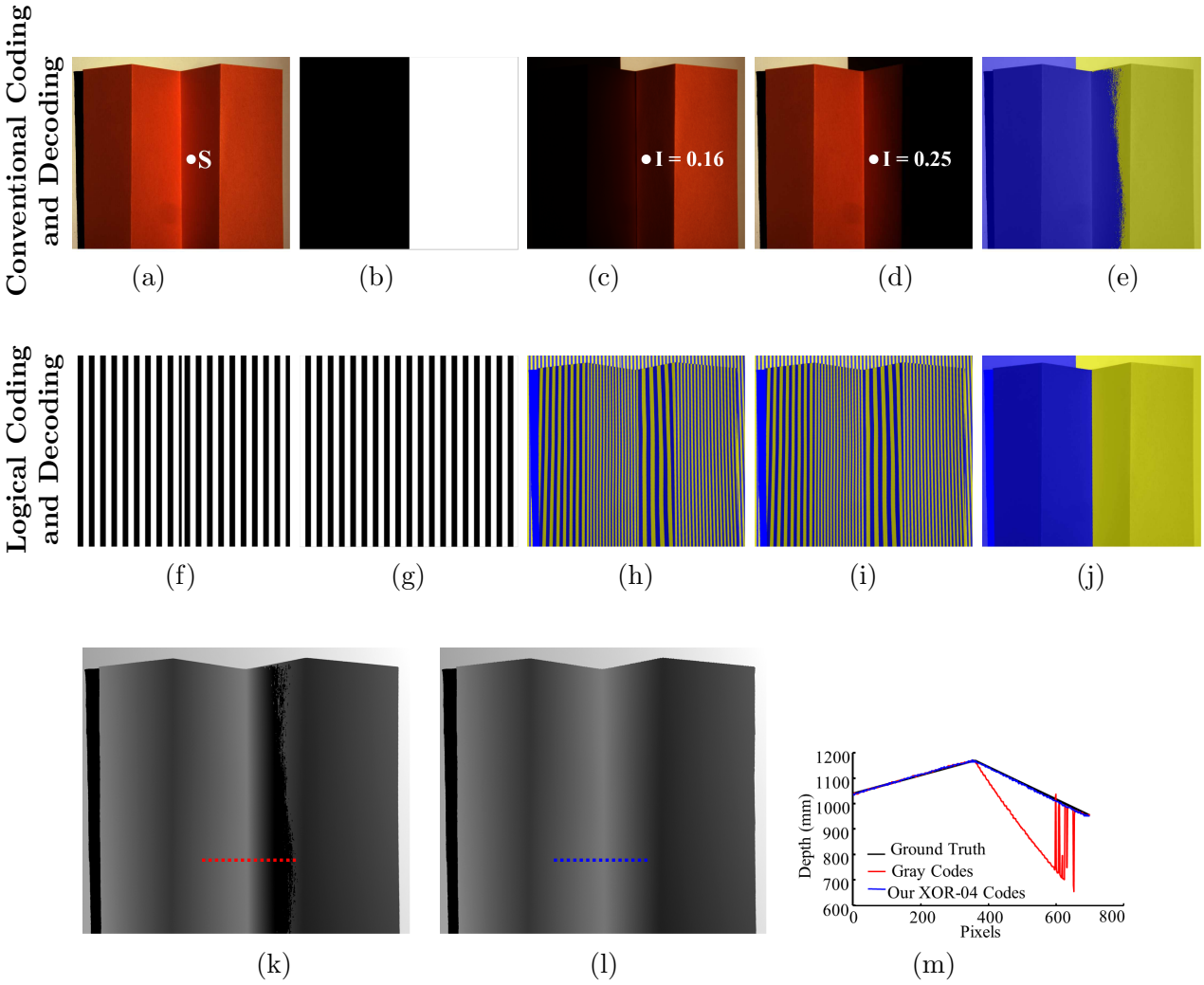
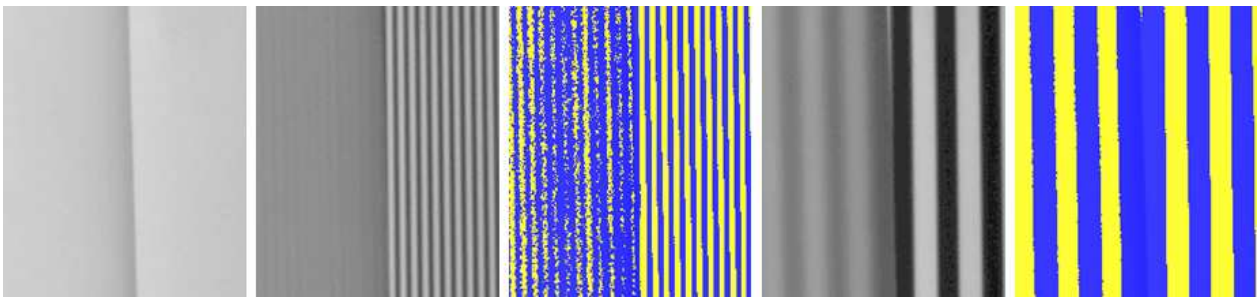
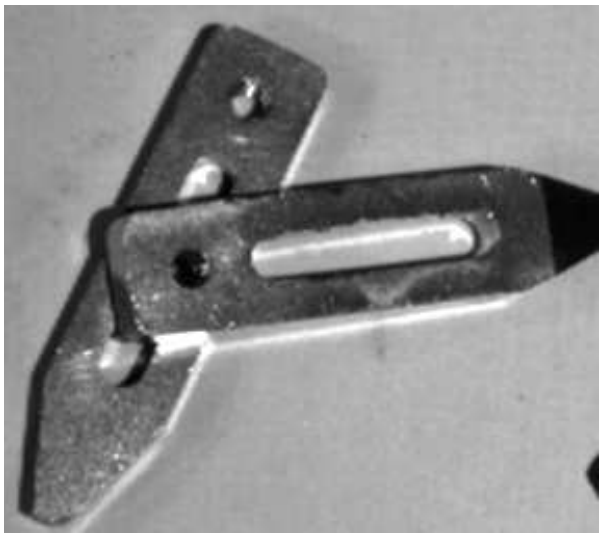


Figure 3.2: **Errors due to inter-reflections:** Top row: Conventional coding and decoding. (a) A concave V-groove. The center edge is concave. (b) Low frequency pattern. (c-d) Images captured with pattern (b) and its inverse respectively. Point S is directly illuminated in (c). However, because of inter-reflections, its intensity is higher in (d), resulting in a decoding error. (e) Decoded bit plane. Points decoded as one (directly illuminated) and zero (not illuminated) are marked in yellow and blue respectively. In the correct decoding, only the points to the right of the concave edge should be one, and the rest zero. (k) Depth map computed with the conventional codes. Because of incorrect binarization of the low frequency patterns (higher-order bits), depth map has large errors. **Second row: Logical coding and decoding (Section 3.3.1)**. (f-g) Pattern in (b) is decomposed into two high-frequency patterns. (h-i) Binarization of images captured with (f-g) respectively. (j) Binary decoding under (b) computed by taking pixel-wise XOR of (h) and (i). (l) Depth map computed using logical coding-decoding. The errors have been completely removed. (m) Comparison along the dotted lines in (k-l).

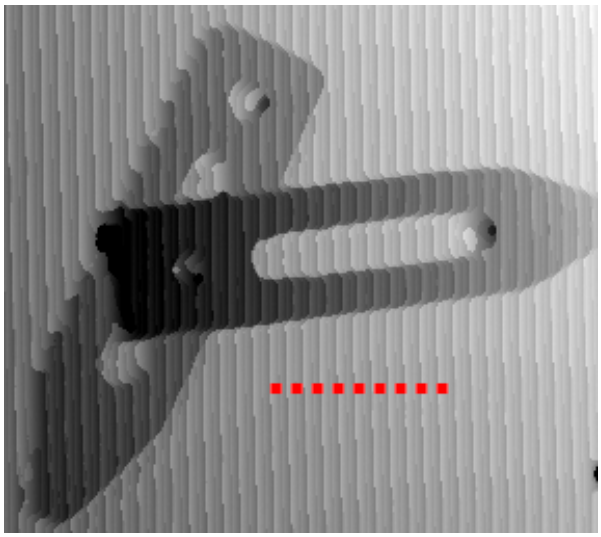


(a) (b) (c) (d) (e)

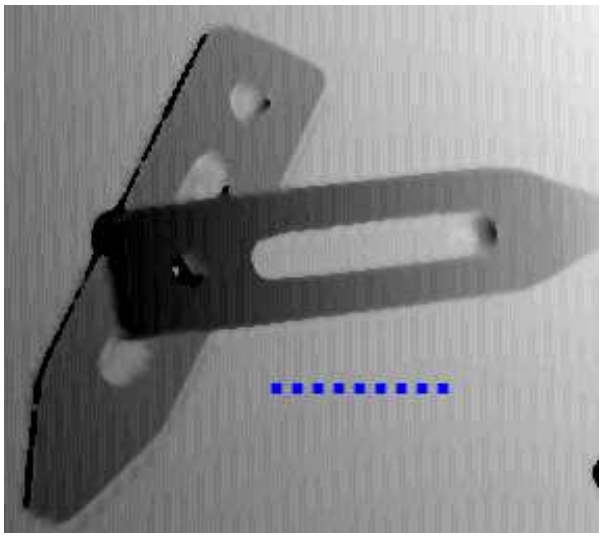
Figure 3.3: **Errors due to sub-surface scattering:** (a) This scene consists of a translucent slab of marble on the left and an opaque plane on the right. (b) A high frequency pattern is severely blurred on the marble, and can not be binarized correctly (c). Image captured (d) under a low-frequency pattern can be binarized correctly (e).



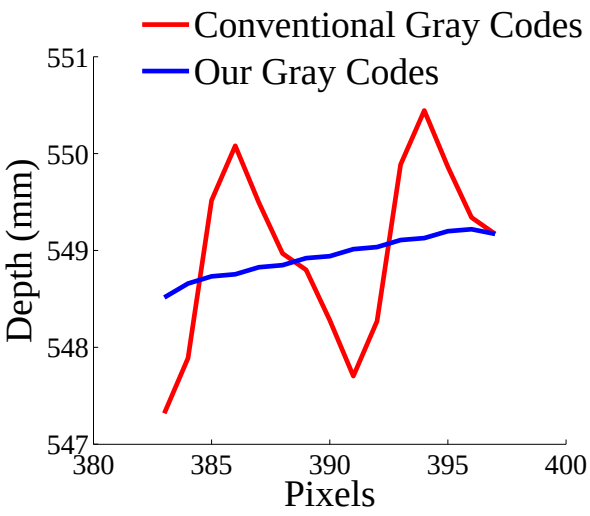
(a) Scene



(b) Conventional Gray codes (11 images)

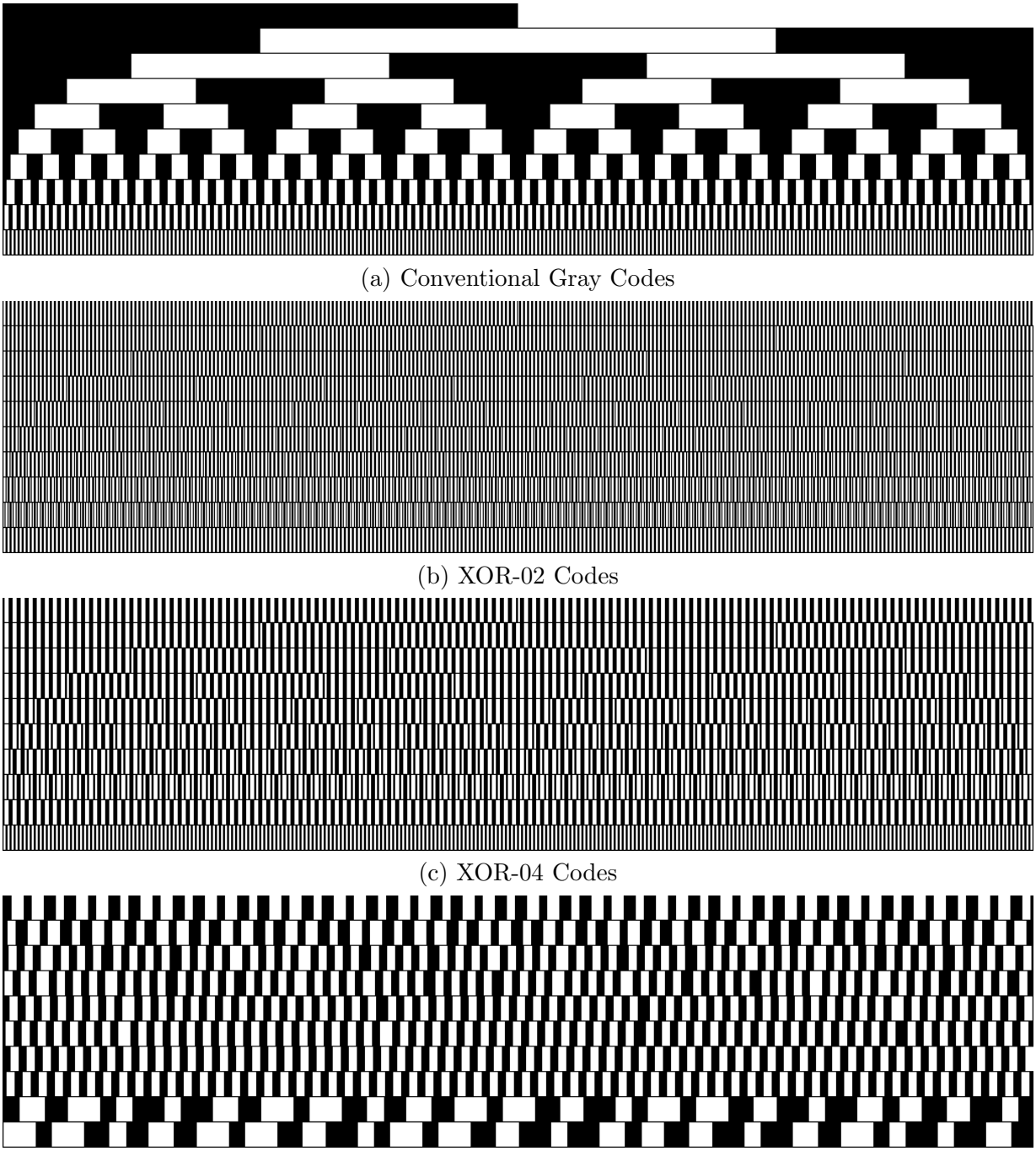


(c) Large min-SW Gray codes (11 images)



(d) Comparison

Figure 3.4: **Depth computation under defocus:** (a) A scene consisting of industrial parts. (b) Due to defocus, the high frequency patterns in the conventional Gray codes can not be decoded, resulting in a loss of depth resolution. Notice the quantization artifacts. Parentheses contain the number of input images. (c) Depth map computed using Gray codes with large minimum stripe-width (min-SW) does not suffer from loss of depth resolution. The number of input images is the same as conventional Gray codes.



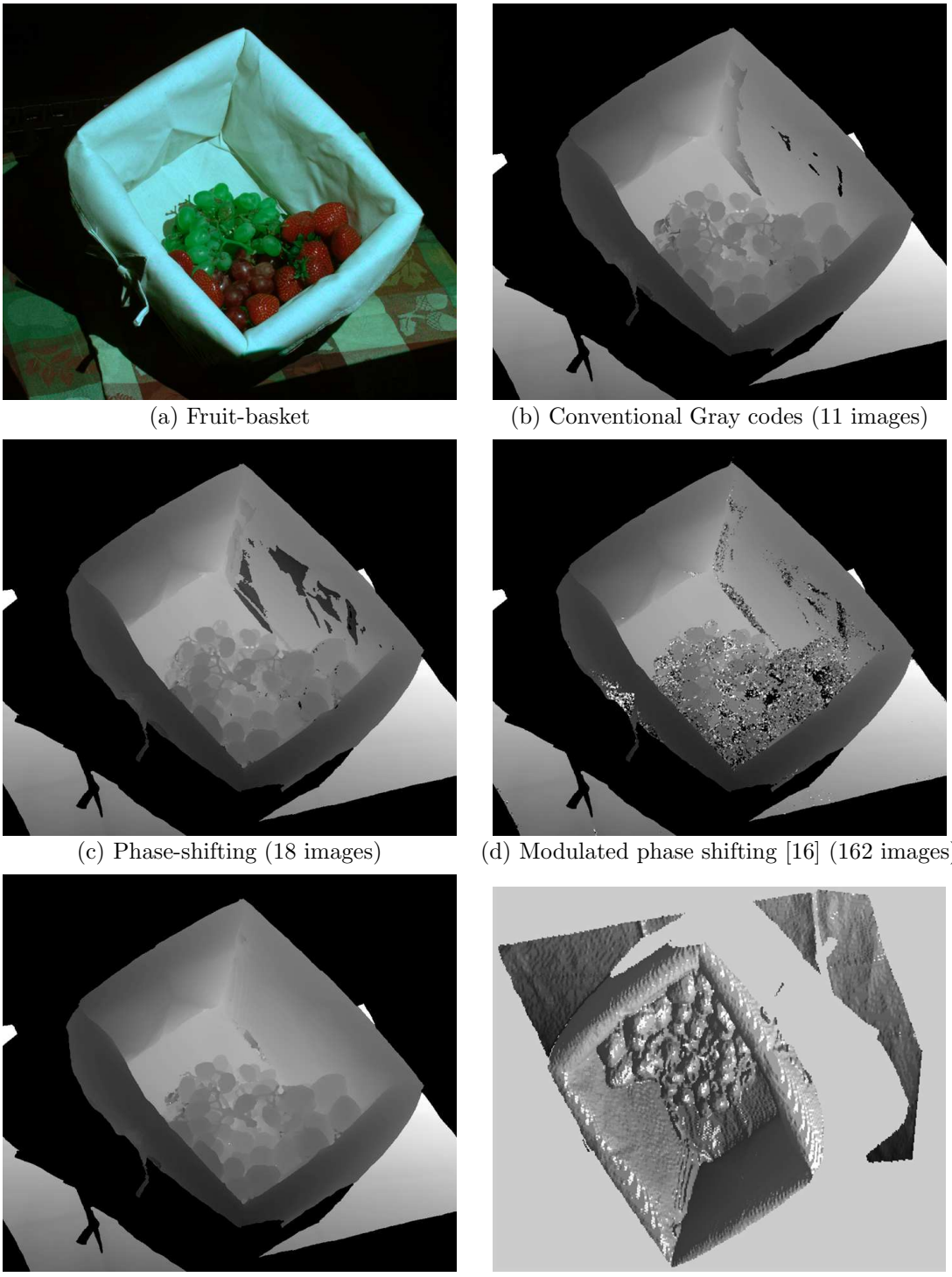
(a) Conventional Gray Codes

(b) XOR-02 Codes

(c) XOR-04 Codes

(d) Gray codes with maximum min-SW

Figure 3.5: **Binary codes for structured light:** The range of stripe-widths for conventional Gray codes is [2, 512]. For XOR-02 and XOR-04 codes (optimized for long range effects), the ranges are [1, 2] and [2, 4] respectively. On the other hand, for Gray codes with maximized mins-SW (optimized for short-range effects), the range is [8, 32].

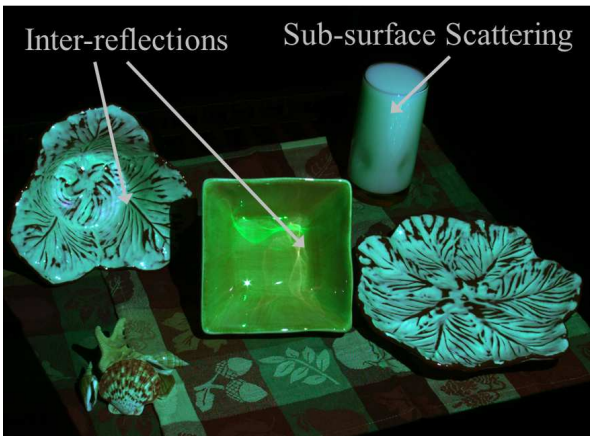


(a) Fruit-basket (b) Conventional Gray codes (11 images)

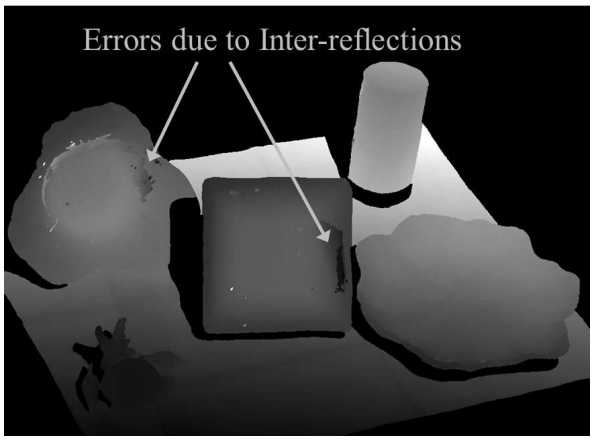
(c) Phase-shifting (18 images) (d) Modulated phase shifting [16] (162 images)

(e) Our ensemble codes (41 images) (f) Visualization for (e)

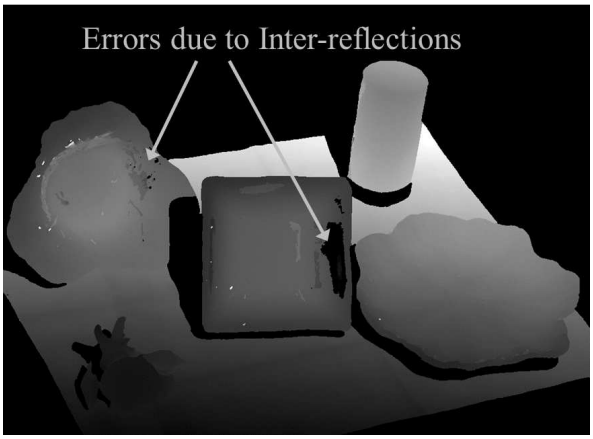
Figure 3.6: **Depth map computation for the fruit-basket scene.** Parentheses contain the number of input images. Conventional Gray codes (b) and phase-shifting (c) result in errors at points receiving inter-reflections. Modulated phase-shifting (d) produces errors on translucent fruits, due to low direct component. (e) Our ensemble of codes significantly reduces the errors.



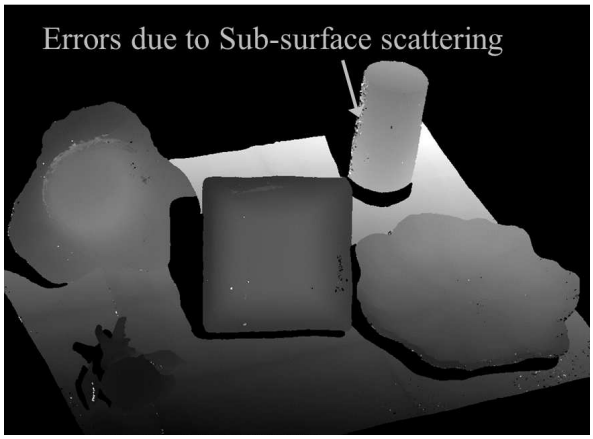
(a) Bowls-and-milk scene



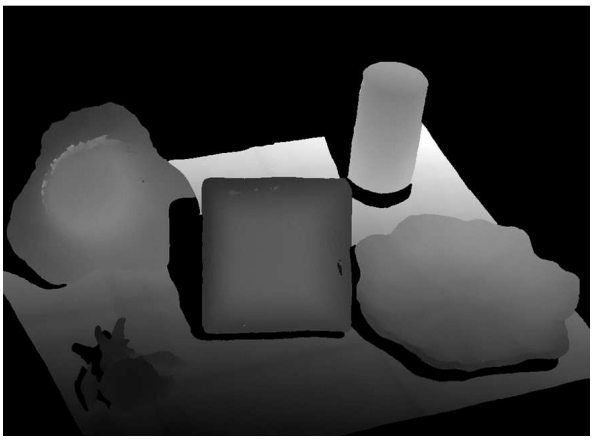
(b) Conventional Gray codes (11 images)



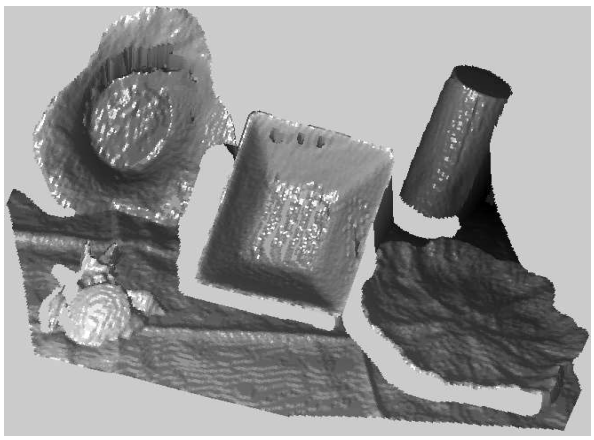
(c) Phase-shifting (18 images)



(d) Modulated phase-shifting [16] (162 images)

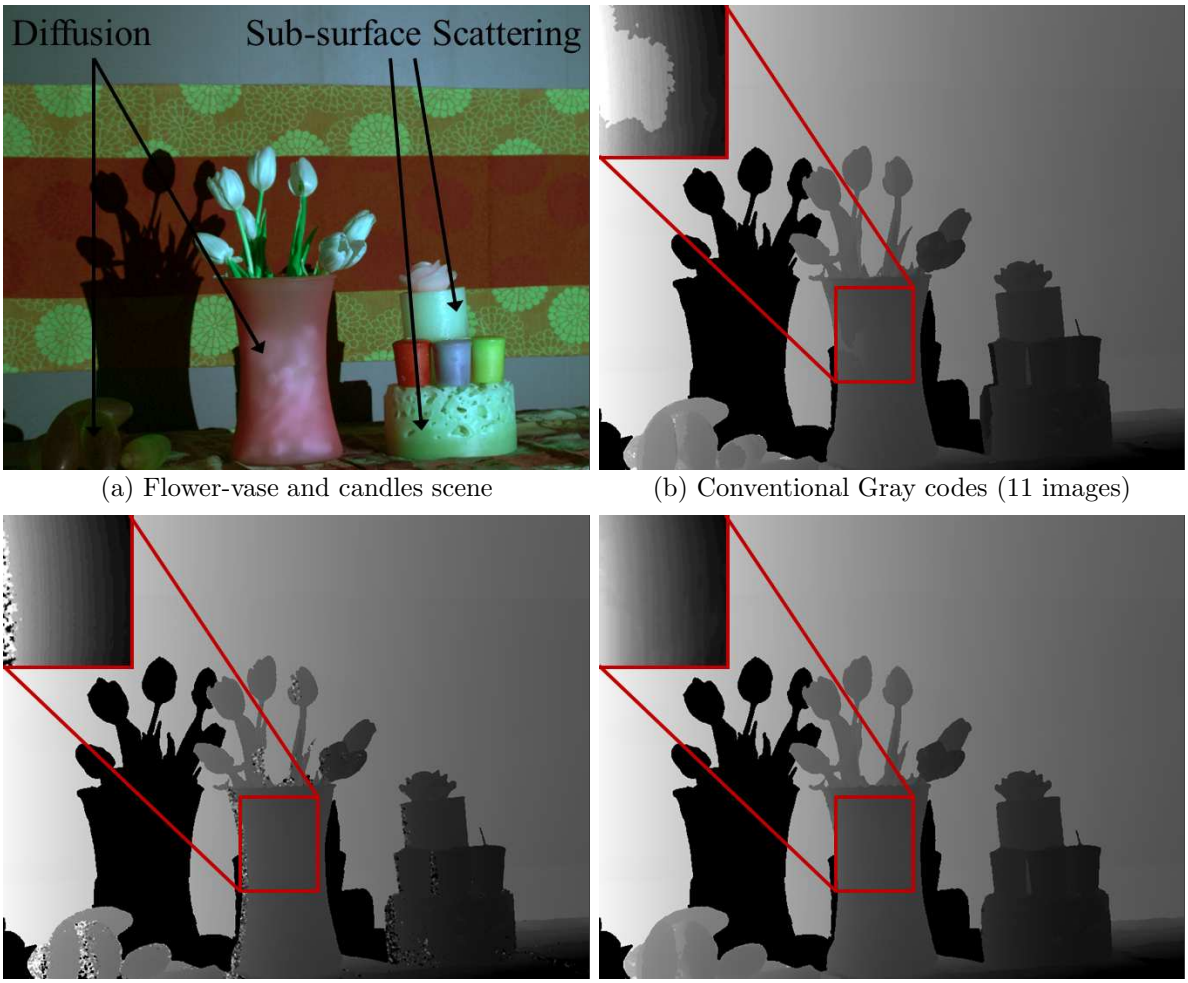


(e) Our ensemble codes (41 images)



(f) 3D visualization of (e)

Figure 3.7: **Depth map computation for the bowls-and-milk scene.** Parentheses contain the number of input images. Conventional Gray codes (b) and phase-shifting (c) result in errors at points receiving inter-reflections. Modulated phase-shifting (d) produces errors on the boundary of the translucent milk glass, due to low direct component. (e) Our ensemble of codes significantly reduces the errors.



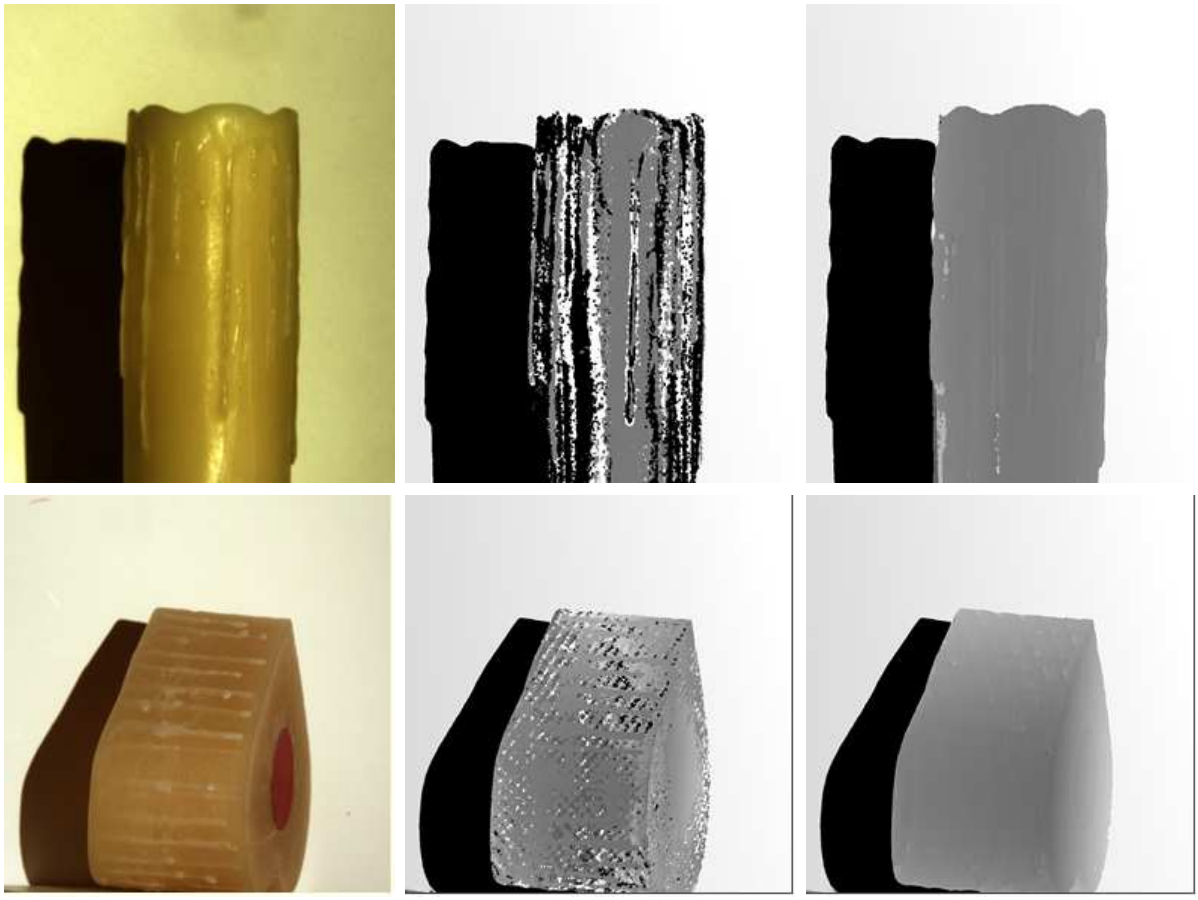
(a) Flower-vase and candles scene

(b) Conventional Gray codes (11 images)

(c) Modulated phase-shifting [16] (162 images)

(d) Our ensemble codes

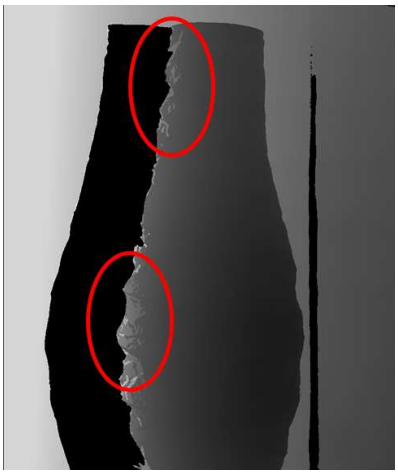
Figure 3.8: **Depth map computation for the flower-vase and candles scene.** Light diffuses inside the flower-vase, gets reflected and seeps out, creating long-range interactions. Consequently, phase-shifting (b) result in errors at points on the vase. Modulated phase-shifting (c) produces errors on translucent candles and stones, due to low direct component. (e) Our ensemble of codes significantly reduces the errors.



(a) Translucent wax objects (b) Modulated phase-shifting [16] (162 images) (c) Our ensemble codes (41 images)

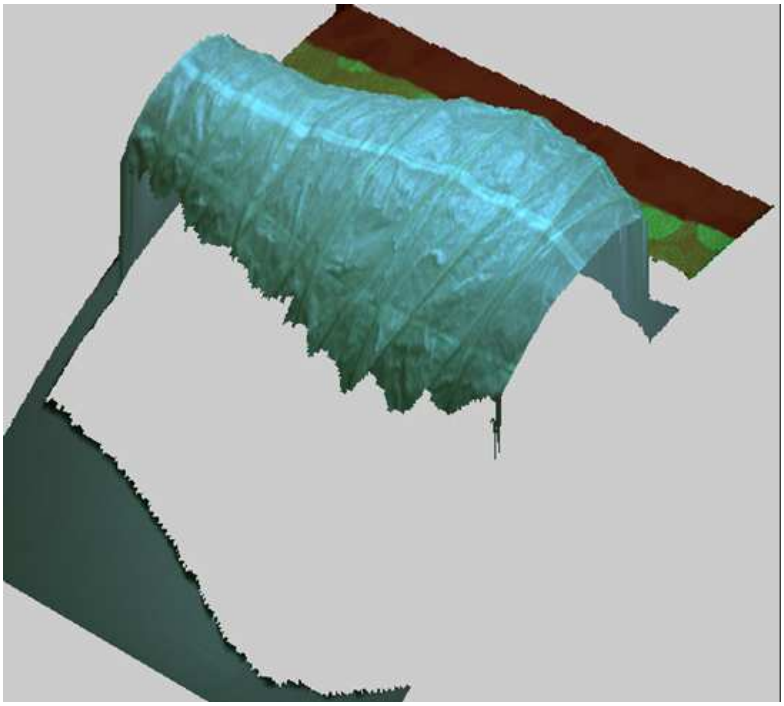
Figure 3.9: **Depth map computation for translucent objects.** Parentheses contain the number of input images. Modulated phase-shifting (b) produces errors due to low direct component. (c) Results with our ensemble of codes.

(a) Ikea lamp



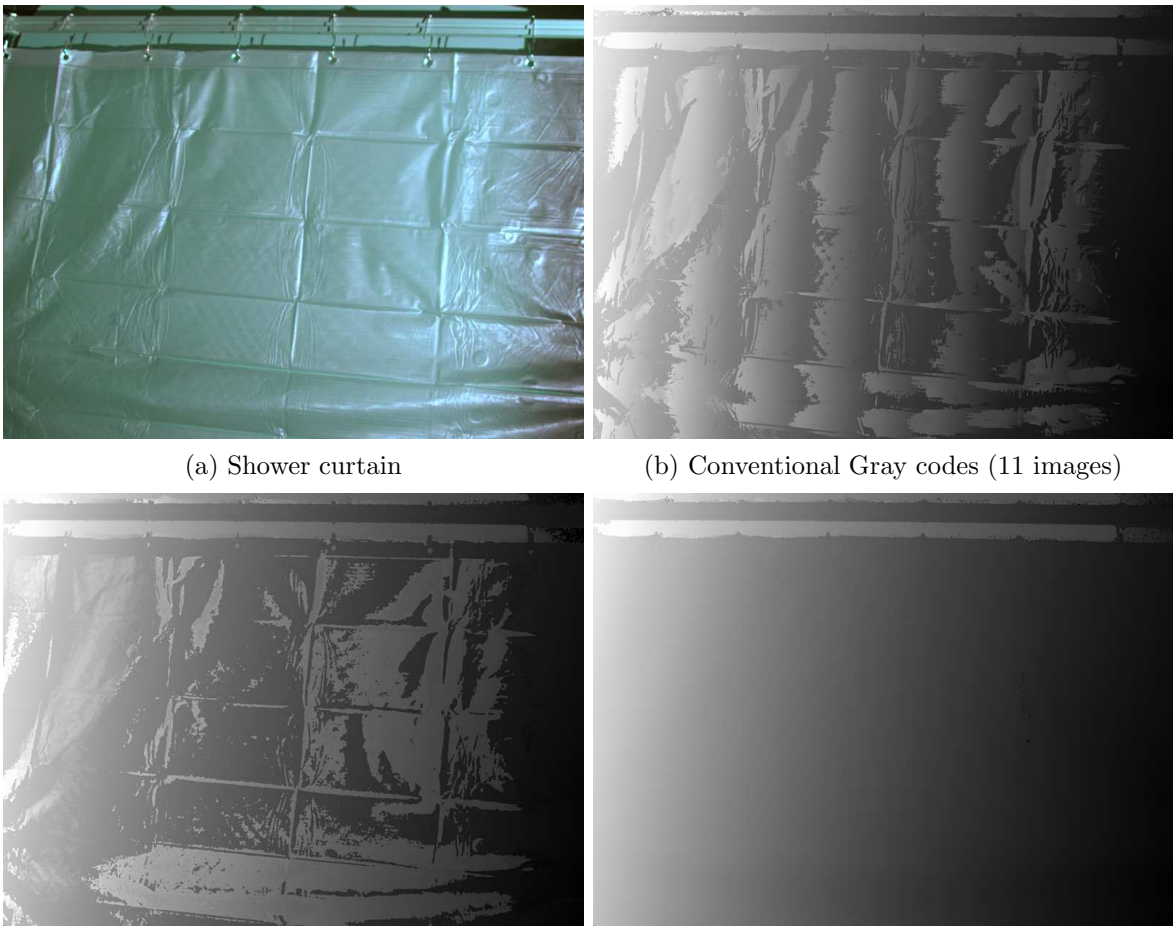
(b) Conventional Gray codes (11 images)

(c) Our ensemble codes (41 images)



(d) 3D visualization of (c)

Figure 3.10: **Depth map computation for an Ikea lamp.** Parentheses contain the number of input images. Light diffuses through the thin outer wall of lamp, gets reflected inside and comes back out, creating long range interactions. Consequently, conventional Gray codes (b) produce erroneous depths. (c) Errors are significantly mitigated in the depth map computed using our ensemble of codes.



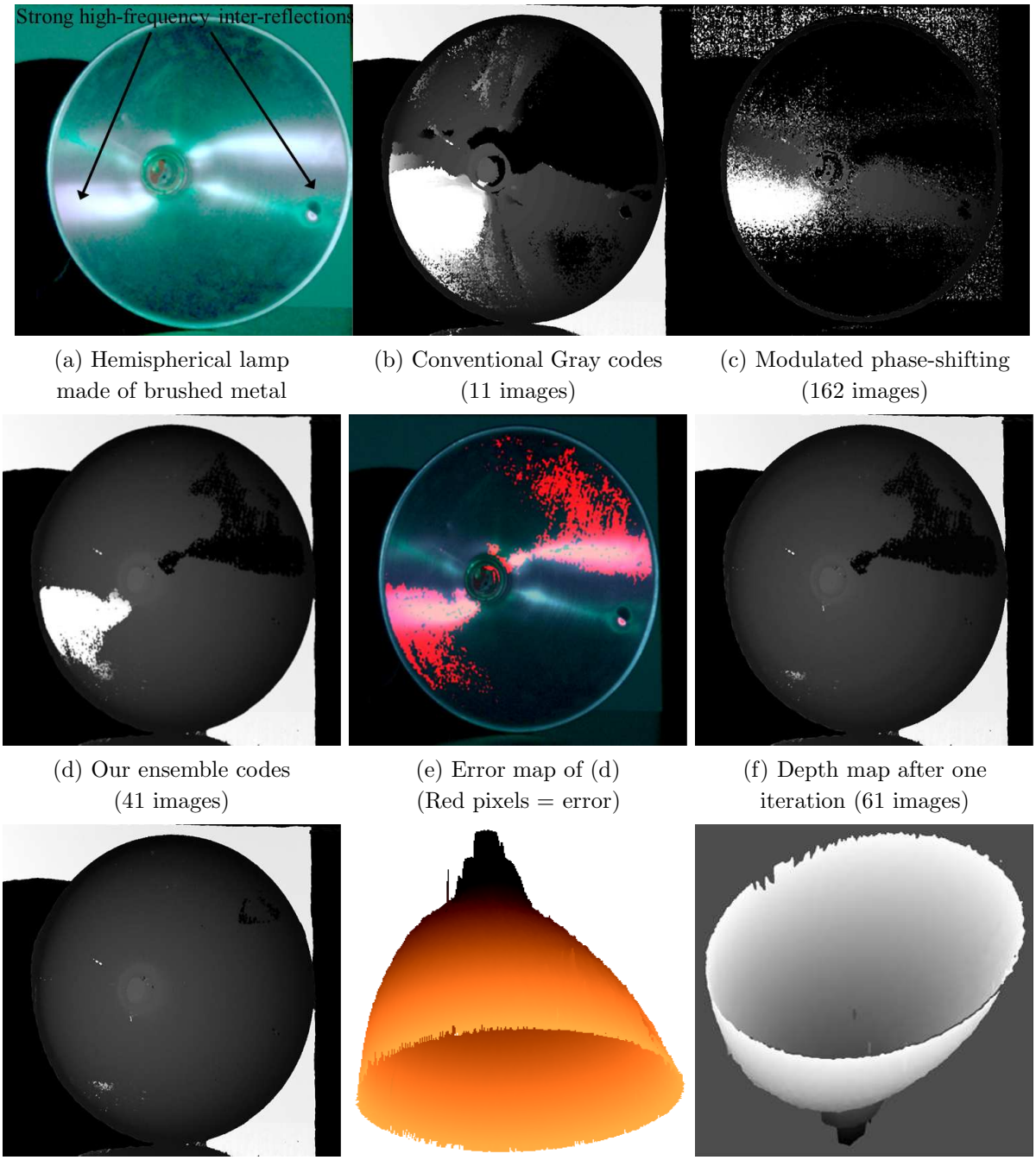
(a) Shower curtain

(b) Conventional Gray codes (11 images)

(c) Phase-shifting (18 images)

(d) XOR-04 (11 images)

Figure 3.11: **Reconstructing a shower-curtain:** The correct shape of the curtain is planar, without ripples. Light diffuses through the curtain and is reflected from the background. (b-c) Conventional Gray codes and phase-shifting result in large errors due to long range optical interactions. (d) Reconstruction using logical codes is nearly error-free, with same number of input images as conventional codes.



(a) Hemispherical lamp made of brushed metal

(b) Conventional Gray codes (11 images)

(c) Modulated phase-shifting (162 images)

(d) Our ensemble codes (41 images)

(e) Error map of (d) (Red pixels = error)

(f) Depth map after one iteration (61 images)

(g) Depth map after two iterations (81 images)

3D visualizations of (g)

Figure 3.12: **Depth computation for the shiny metal lamp.** (a) A hemispherical lamp made of brushed shiny metal. (b-c) Both conventional Gray codes and modulated phase-shifting perform poorly due to strong and high-frequency inter-reflections. (d) Our ensemble codes reduce the errors significantly. (e) Detection of error pixels. Our error detection does not require capturing additional images. (f-g) Using a small number of additional iterations and captured images, we progressively reduce the errors and get a nearly perfect reconstruction. Numbers in parentheses indicate the cumulative total number of acquired images.

Part II

Volumetric Scattering

Chapter 4

Measuring scattering properties of volumetric media

http://www.cs.cmu.edu/~ILIM/projects/LT/dilution/

Very often in our daily lives, we see participating media such as fluids (juices, beverages, milks) and underwater impurities (natural ocean, river and lake waters). The propagation of light through these media results in a broad range of effects, including softer appearance of milk, coloring of wines and juices, the transformation of appearances when liquids are mixed (coffee with milk, and cocktails), the brilliant caustics from glasses containing these liquids, and low visibility in underwater situations. These effects inherently depend on several physical properties of the media such as scattering nature, sizes, shapes, and densities of particles [55, 14]. Rendering these effects accurately is critical to achieving photo-realism in computer graphics.

In the past few years, there has been a considerable effort towards developing efficient and accurate rendering algorithms for participating media, based on Monte Carlo simulation and analytic approximations. All these algorithms and models contain parameters (scattering coefficient, absorption coefficient, phase function) that directly or indirectly represent the physical properties of the medium. In order to faithfully render the effects of any participating medium, the right parameters must be input. Given the progress in developing rendering algorithms, the quality of images is now often limited by the quality of these input parameters. Since there has so far been relatively little work in measuring or estimating scattering properties of media relevant to computer graphics, the parameters are currently often set in an ad-hoc manner.

This situation is similar in some ways to that of standard surface rendering. In that case, global illumination algorithms have progressed to the point of creating almost photo-realistic

images, leaving the realism limited by the quality of the reflectance models, and leading to much recent effort on measuring BRDFs. [73, 19, 74]. However, existing methods for directly measuring physical properties for media usually require very expensive equipment, such as the particle sizing apparatus used in colloidal chemistry [24, 58], resulting in little usable data for graphics.

(a) Acquired photographs

(b) Rendering at low concentrations

(c) Rendering at natural concentrations

Figure 4.1: (a) Photographs of our simple setup consisting of a glass tank and a bulb, filled with diluted participating media (from top, Merlot, Chardonnay, Yuengling beer and milk). The colors of the bulb and the glow around it illustrate the scattering and absorption properties in these media. At low concentrations, single scattering of light is dominant while multiple scattering of light is negligible. From a single HDR photograph, we robustly estimate all the scattering properties of the medium. Once these properties are estimated, a standard volumetric Monte Carlo technique can be used to create renderings at any concentration and with multiple scattering, as shown in (b) and (c). While the colors are only slightly visible in the diluted setting in (b), notice the bright colors of the liquids - deep red and golden-yellow wines, soft white milk, and orange-red beer - in their natural concentrations. Notice, also the differences in the caustics and the strong interreflections of milk onto other liquids.

Earlier efforts to estimate scattering properties from images of media have often yielded ill-conditioned and non-unique results, because of the difficulties of solving the inverse light transport problem. The reasoning for the ill-conditioning of the inverse problem is mainly due to multiple scattering, which blurs the incident light field and results in significant loss of information [77, 6]. This is analogous to the ill-conditioning of BRDF estimation under complex illumination [98]. In this work [81], we take a completely different approach. The key idea is to estimate properties of media by acquiring the data in a state where multiple scattering effects are negligible. Instead, the data is acquired when single scattering (which does not degrade the incident light significantly) is the dominant effect. This is achieved

by diluting the material to low concentrations.

We present a simple and inexpensive experimental setup, along with a robust and accurate technique for measuring the scattering properties of a broad class of participating media that can be either (a) diluted in water such as juices, beverages, paints and cleaning supplies, or (b) suspended in natural waters such as impurities and organisms, or even (c) dissolved in water such as powders and sugar or salt crystals. These media collectively have a wide range of scattering and absorption properties. We first derive a simple image formation model for single scattering of light in our setup. Through extensive simulations of both our model and ground truth (with multiple scattering), we then determine the space of concentrations and scattering properties of media for which single scattering is dominant. Within this regime of valid concentrations, we conduct simulations to demonstrate that our estimation technique uniquely solves the inverse single scattering light transport problem. Finally, we present a simple experimental procedure to determine the best concentration (dilution) for any material despite no prior knowledge of its scattering properties.

We have used our approach to create a dataset of scattering parameters for **forty commonly found materials**, which can be directly used for computer graphics rendering. Once the scattering parameters have been estimated, they can be used to render realistic images of arbitrary concentrations of the material with multiple scattering, using a standard physically based volumetric rendering algorithm. Figure 4.1 shows two renderings of a scene with four liquids in their natural high density states and their diluted states. The scattering parameters of each material were computed using a single HDR photograph of our setup. Notice the bright saturated colors obtained despite the murky appearance of the diluted states. We can also create realistic images of mixtures of the original measured materials, thus giving the user a wide flexibility in creating realistic images of participating media.

4.1 Related Work

Figure 4.2 shows the most common properties of participating media including the scattering and absorption coefficients, and the phase function (angular scattering distribution represented by the Henyey-Greenstein (H-G) model [49]). The scattering and absorption coefficients are proportional to the concentration or volume fraction of the particulate medium. We will briefly review some of the representative works on the direct measurement and indirect estimation of these parameters.

Estimation based on analytic approximations to light transport. Surprisingly, little work has been done in computer graphics on the measurement of scattering properties

Medium Property	Notation
Concentration or Volume Fraction	C
Scattering Coefficient (mm^{-1})	β
Absorption Coefficient (mm^{-1})	κ
Extinction Coefficient (mm^{-1})	$\sigma = \beta + \kappa$
Single Scattering Albedo	$\omega = \beta/\sigma$
Scattering Angle	θ
Henyey-Greenstein (H-G) Parameter	g
H-G Phase Function	$P(g,\theta) = \frac{1}{4\pi} \frac{1-g^2}{(1+g^2-2g\cos\theta)^{3/2}}$

Figure 4.2: The different scattering properties of a participating medium and their notations used in this section. Light transport equations are usually written in terms of three parameters σ, β and g. We estimate these parameters for participating media based on single scattering.

of media. A diffusion model for sub-surface scattering was proposed in [61]. They present a measurement of a number of translucent materials. However, the diffusion approximation assumes multiple scattering for optically dense media, so that only a limited amount of information on the scattering parameters can be estimated. For instance, this approximation is independent of the phase function of the medium, and therefore this important property cannot be estimated. Furthermore, the diffusion is a poor approximation when scattering is comparable to absorption [96]. The analytic multiple scattering model presented in [84] has also been used to estimate properties of only purely scattering media (visibility and type of weather such as fog and mist). Our focus is somewhat different in considering fluids like juices or beverages, instead of subsurface scattering in translucent solids like marble and skin, or weather conditions such as fog. Nevertheless, our approach is valid for media with the entire range of absorbing and scattering properties, significantly extending the class of measurable media for graphics.

[47] measure the extinction coefficient of optically thin smoke from the exponential attenuation of a laser beam in a tank. They also use a separate mirror setup to directly measure the phase function (see below). In contrast, our setup uses divergent beams from a simple bulb to include more light in the volume (than a single laser beam) for robust measurements, and requires only a single photograph to measure all scattering properties shown in Figure 4.2.

Numerical solution to inverse light transport: In cases where there are no analytic solutions to light transport, several works have taken a numerical approach to estimate scattering properties [6]. However, it is widely known, that inverse problems in radiative transfer that take into account multiple scattering are ill-conditioned and require regularizing assumptions to obtain reliable estimates. See the reports and critiques in [77]. Furthermore, the computational complexity of such inverse estimation techniques make it

hard for measuring large sets of media for computer graphics or vision applications. Our focus here is on estimating scattering properties of media that can be measured in a state where multiple scattering is negligible.

The observation that single scattering is dominant for optically thin media has been made by [47, 116]. We exploit this observation and apply the single scattering model for the first time to a large class of materials which exhibit significant multiple scattering in their natural states of existence. We also determine the exact range of optical thicknesses for which single scattering is dominant for media with arbitrary scattering properties, and propose an experimental procedure to ensure the dominance of single scattering in real data.

Goniophotometry is often used to directly measure the phase function. Here, several detectors measure radiance in different directions after being scattered by a very small volume of the medium. [29] use thin laser light sheet microscopy for detecting and localizing microorganisms in ocean waters. [9, 91] investigate the relationship of light scattering at a single angle and the extinction coefficient using specialized receivers and transmitters. However, all these techniques assume that there is no attenuation of light through the sample and require expensive devices with precise alignment of detectors and transmitters. In contrast, our setup is extremely simple (consisting of a glass tank and an off the shelf bulb), and our technique robustly estimates all properties from only a single photograph, thus making it inexpensive and easy to measure a large number of participating media.

4.2 Single Scattering in Dilute Media

Our approach is to measure media in a state where single scattering is dominant and multiple scattering is negligible. This is achieved by diluting the otherwise optically thick media, such as fluids, in water. The process of dilution does not usually corrupt the inherent scattering properties of media[1] since the scattering and absorption of pure water itself is negligible for very small distances (less than 50 cm) [115]. We begin by presenting our acquisition setup and an image formation model for single scattered light transport within the measurement volume. We will then present extensive simulations of this model and compare with traditional Monte-Carlo approaches that include multiple scattering, to derive a valid space of scattering parameters over which single scattering is dominant. Based on this simulation, we design a simple experimental procedure to choose the best concentration for any particular medium. Later, we will describe our algorithm to estimate the scattering parameters using our image formation model.

[1]When crystals are dissolved in water, they may exhibit different scattering properties due to ionization.

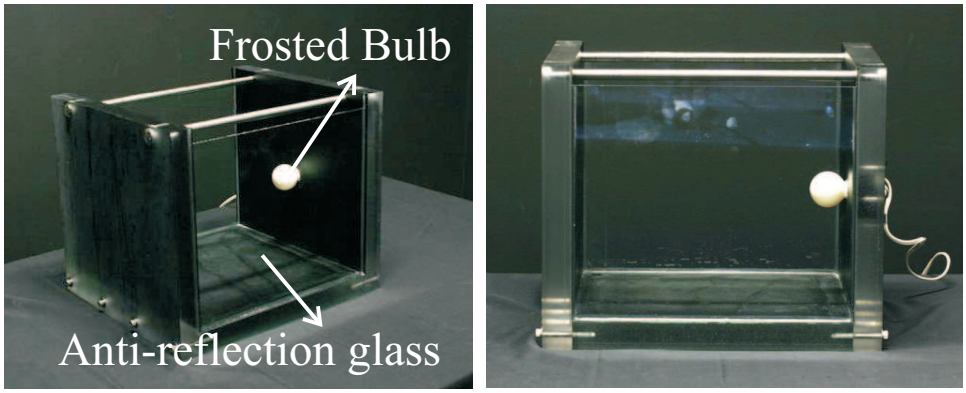


Figure 4.3: Two views of the apparatus used to measure scattering properties of water-soluble media. A glass tank with rectangular cross-section is fitted with a small light bulb. The glass is anti-reflection coated. Different volumes of participating media are diluted with water in the tank, to simulate different concentrations. A camera views the front face of the tank at normal incidence to avoid refractions at the medium-glass-air boundaries.

4.2.1 Acquisition Setup

The measurement apparatus, shown in Figure 4.3, consists of a $25 \times 30 \times 30$ cm^3 tank that is filled with the diluted scattering medium. The depth of the tank is large enough to ensure the scattering angles are adequately covered (0 to 175 degrees). The volume of the tank is designed to be large enough to dilute concentrated media such as milk. Two sides of the tank are constructed using anti-reflection glass and the other sides using diffuse black coated acrylic. A small frosted (diffuse) glass bulb fixed to a side of the tank illuminates the medium. A Canon EOS-20D 12-bit 3504x2336 pixel digital camera with a zoom lens is placed five meters away from the tank and observes a face of the tank at normal incidence. The field of view occupied by the tank in the image is three degrees and is therefore approximately orthographic. Orthographic projection avoids the need for modeling refractions of light rays at the medium-glass-air interfaces. In all our experiments, about 25 different exposures (1/500s to 10s) were used to acquire HDR images.

4.2.2 Image Formation Model

Although the basic principles of single scattering are well known, the exact nature of the image formation model depends on the geometry of the volume and the locations of the source and the camera. Figure 4.4 illustrates the illumination and measurement geometry based on our acquisition setup. For simplicity, we will assume that the medium is illuminated by an isotropic point light source (later we extend the analysis to area sources) of intensity I_0 that is located at the coordinates $(0, B, H)$.

Consider the path of one single-scattered light ray (thick ray in Figure 4.4) in the medium

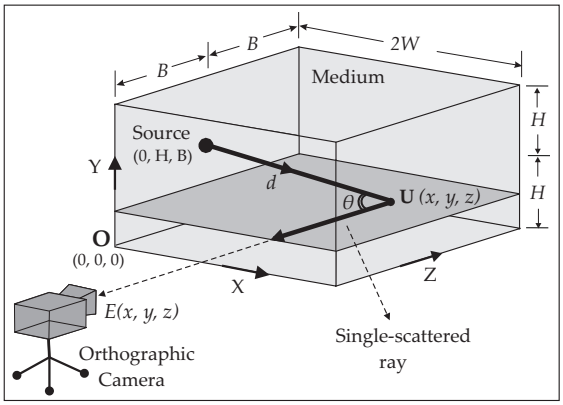


Figure 4.4: A volume filled with a homogeneous participating medium and illuminated by an isotropic point light source. A camera views the front face of the volume at normal incidence. The path of one single-scattered ray as it travels from the source to the camera is shown. This ray is first attenuated in intensity over a distance d, is then scattered at an angle $\pi - \theta$, and finally, is attenuated again over a distance z, before reaching the camera. The irradiances due to all the rays that scatter into a viewing direction must be integrated to obtain the final camera irradiance.

as it travels from the source to the camera. This ray is first exponentially attenuated in intensity for a distance d. At location $\mathbf{U}$ (x, y, z), depending on the phase function P, a fraction of the light intensity is scattered at an angle $\pi - \theta$. Finally, the ray is attenuated again for a distance z, before it reaches the camera. Mathematically, the irradiance at the camera produced by this ray is written as [116],

$$\begin{aligned} E(x,y,z) &= \frac{I_0}{d^2} \;.\; e^{-\sigma d} \;.\; \beta\, P(g, \pi-\theta) \;.\; e^{-\sigma z}\,. \\ d &= \sqrt{x^2 + (y-H)^2 + (z-B)^2} \quad , \; \cos\theta \;=\; (z-B)/d\,. \end{aligned} \tag{4.1}$$

Here, $P(g, \pi - \theta)$ is the Henyey-Greenstein (H-G) phase function, and β and σ are the scattering and extinction coefficients (Figure 4.2). Then, the total irradiance E at pixel (x, y) in the camera is obtained by integrating intensities due to all rays that are scattered at various angles along the pixel's line of sight (Z-direction),

$$\begin{aligned} E(x,y) &= \int_0^{2B} E(x,y,z)\, dz \\ &= \beta \int_0^{2B} \frac{I_0\, e^{-\sigma(z+\sqrt{x^2+(y-H)^2+(z-B)^2})}}{x^2+(y-H)^2+(z-B)^2}\; P(g,\pi-\theta)\; dz\,. \end{aligned} \tag{4.2}$$

The above equation relates the camera irradiances as a function of the three medium parameters, σ, β and g. Although obtaining an analytic (closed-form) solution to the above

integral is hard [116], it is straightforward to evaluate it numerically.

4.2.3 Space of valid medium parameters

Different materials have their own natural densities and scattering properties, which are all unknown before experimentation. So, how do we know if single scattering is dominant at a particular concentration for a given material? Note that the scattering β, absorption κ and extinction σ, coefficients are proportional to the concentration (fraction of volume diluted in water) of the medium. Thus, we performed exhaustive simulations to derive the complete space of parameters for which the above image formation model is accurate[2]. For ground truth, we simulated the irradiances obtained using multiple scattering for the same set of parameter values, using a standard volumetric Monte Carlo technique. Figure 4.5 shows a plot of the differences between energies captured by the single scattering and multiple scattering simulations for a set of parameter values. From the RMS errors in the plot, we can define the upper bounds on the parameters κ and $\sigma = \beta + \kappa$ as those for which the energy differences between our model and the ground truth are less than five percent. For example, the valid domain where single scattering is dominant, is approximately $\sigma < 0.04$ for $\kappa < 0.004$.

4.2.4 How to choose the best concentration?

Based on the simulations, we present an experimental method to determine the best concentration for our measurements. Figure 4.6 shows images acquired of different concentrations of milk and MERLOT. Which among these images should we use to measure the scattering properties? Several heuristics may be used to decide on a particular concentration. For instance, the extent of blurring of the light source provides us a good clue to determine whether multiple scattering is significant (rightmost image in Figure 4.6). A better heuristic is to compute an approximation to the extinction coefficient σ from the attenuated brightness of the light source. Under single scattering, the radiance in the direction of the source (distance d) can be approximated using exponential attenuation as:

$$E(0) \approx \left(\frac{I_0}{d^2}\right) \; e^{-\hat{\sigma}\, d}\,, \tag{4.3}$$

where $\hat{\sigma}$ is an estimate of the extinction coefficient σ. In the absence of multiple scattering, this estimate is closer to the true value of σ (and varies linearly with concentration), whereas, in the presence of multiple scattering, this estimate is called *diffuse or reduced attenuation coefficient* [57] and is usually much lesser than σ. Thus, we can determine

[2]This extends the simulations in [116], where a small part of the possible parameter space (pure isotropic scattering) was considered.

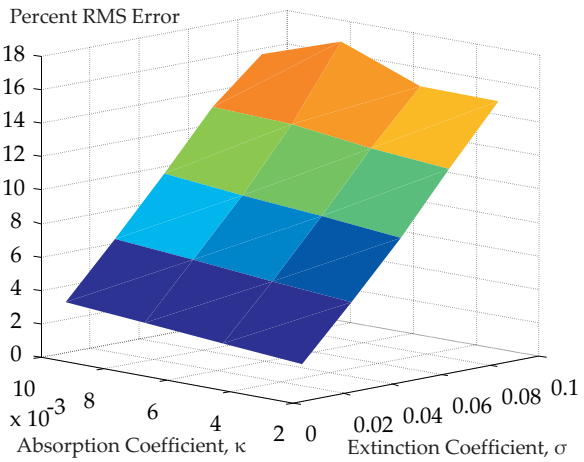


Figure 4.5: Plot showing the differences between irradiances obtained by simulating single scattering and multiple scattering (ground truth) models, for a large space of parameter values σ and $\kappa = \sigma - \beta$. An upper bound on the differences of, say, 5%, can be used to define the range of parameters for which single scattering is a valid approximation. From the plot, the valid range is approximately $\sigma < 0.04$ for $\kappa < 0.004$.

whether the concentration can be used for measurement by observing the plot (Figure 4.7 of $\hat{\sigma}$ versus the volume fraction of the medium diluted with water). Figure 4.7 shows that after a certain amount of milk is added to water, the $\hat{\sigma}$ no longer remains linear with concentration (dashed line), and must not be used for measurements. For a purely absorbing liquid like wine (MERLOT), the plot is completely linear and any image that has the best signal-to-noise ratio may be used. Similarly, the plot shows that coke scatters, albeit weakly, and ESPRESSO coffee scatters light strongly. We use this simple procedure to try several concentrations and observe where the linearity in the plot fails to determine the best concentration. As a further test, we check if the estimated parameters from this concentration lie within the valid space of parameters simulated above.

4.3 Estimating Medium Properties based on Single Scattering

In this section, we present a non-linear minimization algorithm to estimate the properties of the medium (σ, β and g), from the measured image irradiances $E(x, y)$ (see Equation (4.2)). We then demonstrate the accuracy of the algorithm through extensive simulations.

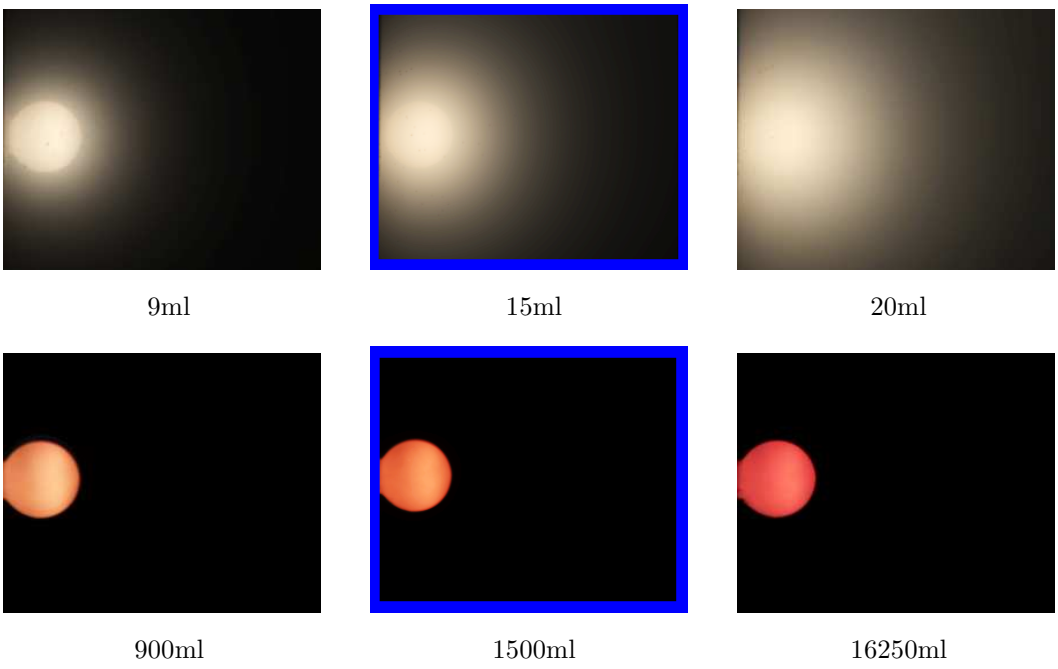


Figure 4.6: Images illustrating different degrees of scattering and absorption. [Top row] Images of milk at various concentrations. Since milk is a highly scattering liquid, we observe an increase in blurring with increasing concentration. [Bottom Row] Images of red wine at various concentrations. Red wine is a highly absorbing liquid, showing only a saturation of the bulb color with increasing concentration, and no blurring. The highlighted images are chosen for estimating the parameters.

4.3.1 Formulating the Error Function

The error at each pixel is written as the difference between the measured irradiance $E(x, y)$ and the irradiance predicted by the model in equation 4.2,

$$\mathcal{F}(x, y) = E(x, y) - RHS(x, y)\,. \tag{4.4}$$

Here $RHS(x, y)$ is the numerically evaluated right hand side integral in the model of equation 4.2.

Then, the parameters σ, β and g can be estimated by computing the global minimum of the sum of squares of the errors of all the pixels, as,

$$\min_{\beta,\sigma,g} \sum_{y} \sum_{x} \mathcal{F}^2(x, y)\ . \tag{4.5}$$

The above function essentially requires a 3-parameter search. However, note that the parameter β is a global scale factor. Thus, we can eliminate β by defining a normalized error

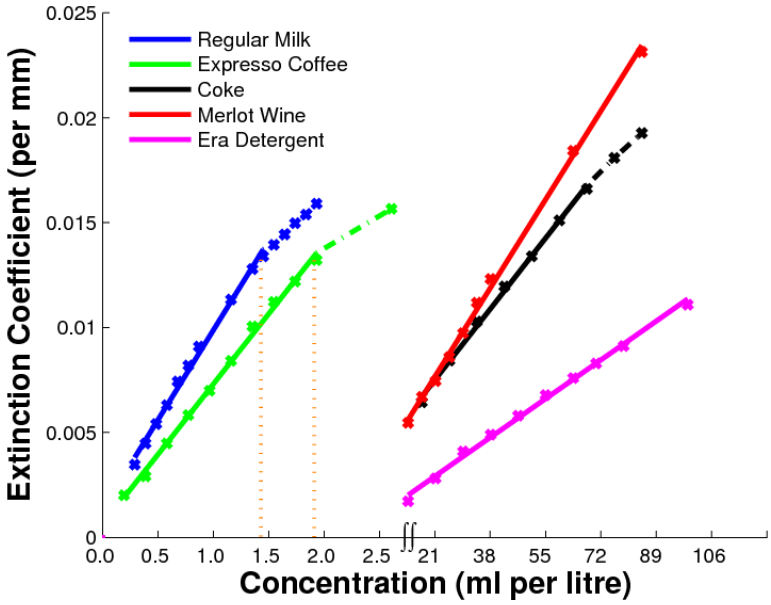


Figure 4.7: Plot of extinction coefficient estimate $\hat{\sigma}$ as a function of the volume of the media diluted in water in the measurement apparatus. The plots are linear when multiple scattering is negligible and single scattering is dominant. As the concentrations of media (and hence multiple scattering) increase, the estimated $\hat{\sigma}$ is less than the true extinction coefficient σ. For a highly scattering medium such as milk, the linearity fails at very low concentrations, while for an absorbing medium such as MERLOT, the linearity is always preserved.

function as,

$$\mathcal{F}_{norm}(x,y) = \frac{E(x,y)}{\max\limits_{x,y} E(x,y)} - \frac{RHS(x,y)}{\max\limits_{x,y} RHS(x,y)} \,. \tag{4.6}$$

Now, instead of requiring a 3-parameter search, the above problem can be reduced to a 2-parameter search that minimizes the normalized objective function to estimate σ and g:

$$\min_{\sigma,g} \sum_{y} \sum_{x} {\mathcal{F}^2}_{norm}(x,y) \,. \tag{4.7}$$

Then, the scale factor β can be recovered using the original function $\mathcal{F}$. To compute the global minimum, we use Nelder-Meade search implemented by the Matlab^{TM} function "fminsearch".

4.3.2 Estimation Accuracy using Simulations

Fortunately, since the space of the possible parameters is small (see Section 4.2.3), exhaustive simulation of the above algorithm is possible. We only show the correctness of the

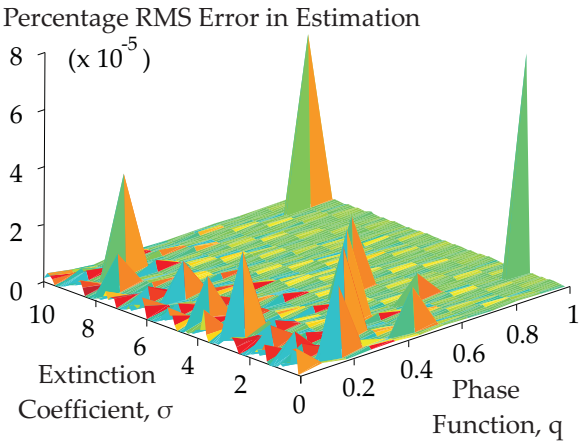


Figure 4.8: Plot showing the errors in reconstruction of the single scattering parameters σ and $q = |g|$, where $-1 < g < 1$, compared to ground truth values. The low errors indicate the accuracy of our estimation technique. The maximum of the errors for positive or negative g is shown.

estimated parameters σ and g, using Equation (4.7). The estimation of the scale factor β then follows trivially. Gaussian noise of unit standard deviation was added in all our simulations. The non-linear search was initialized randomly for both the parameters σ and g. The plot in Figure 4.8 shows the error in the estimated parameters as compared to ground truth values. In all the cases, the estimation errors were less than 0.0001%, and the number of iterations required for convergence was less than 100. Since the numerical evaluation of the integral is very fast, the time for convergence is usually of the order of a few minutes. This demonstrates that the inverse estimation is fast and results in unique and correct parameters.

4.3.3 Implementation Issues

We present two issues that need careful implementation for our algorithm to be successful on real images.

Calibrating the area source: Our method does not rely on isotropic point sources but requires only a calibrated divergent source to take advantage of the different phase angles measured in the same view and hence, any off-the-shelf bulb suffices. For our real setup, we have implemented a spherical diffuse area source. To compute the irradiance at any point P within the tank, we sample (using roughly 10x10 samples) the hemisphere of the bulb that is visible to that point P. The non-uniform directional intensities and intensity fall-off were calibrated carefully by using a light meter at discrete 3D locations within the tank. The camera also measures a pure water image (without any scattering or absorption) to give the image irradiance of each source element (sample). This irradiance along with the fall-off value and the pixel solid angle is used to determine the intensity without scattering.

Instabilities in the H-G phase function for highly absorbing media: The H-G phase function was designed for scattering media and is not defined for purely absorbing media. However, for highly absorbing media, the scattering coefficient β is very low and the average cosine $g \approx 1$ since rays only pass straight through, much like highly forward scattering media. Even though this was not a problem in simulations, the instability for $g > 0.95$ can be high in real experiments. For this special case, we simply use a truncated legendre polynomial expansion of the H-G phase function as $P(g,\theta) = \sum_i (2i+1)\, g^i\, L_i(\theta)$, and truncate to less than 100 terms. As an undesirable byproduct the fits may show some "ringing" at the tail of the phase function. However, this truncated function still fits higher brightness well and thus does not affect appearance strongly. Despite this instability, the H-G phase function is flexible enough to model the scattering behavior of all our materials.

4.4 Actual Measurements and Validation

Using our approach, we have measured the scattering properties of a broad class of **forty** commonly found participating media that can be either (a) diluted in water such as juices (for example, apple, strawberry, orange), beverages (for example, coffee, soft drinks, milks, wines, beers), cleaning supplies (detergents), or (b) suspended in natural waters such as impurities and organisms, or even (c) dissolved in water such as powders and sugar, salt crystals. In addition to liquids available at the usual supermarkets, we have also collected four samples from different locations and depths in the Pacific ocean. We then present detailed validation by showing that our parameters extrapolate correctly to higher concentrations as well, where multiple scattering is prominent.

A subset of nine photographs of the diluted set of liquids contained in the glass tank is shown in Figure 4.9, similar to the four in Figure 4.1. Together, these include representative types of media such as highly scattering, highly absorbing and moderate levels of absorption and scattering. The images show a high dynamic range of brightness and are enhanced to show the scattering effects. The set of scattering parameters for all the media is shown in Table 4.1. The extinction (σ) and scattering (β) coefficients are given for each of the three color channels, red, green and blue. The phase function parameter g is also shown for the three color channels. Note that all the extinction and scattering coefficients are less than 0.04 in accordance with our simulations in Section 4.2.3. Also, as expected, in all cases, the scattering coefficient does not increase with wavelength.

The second column in the table lists the volumes V of the materials dissolved in $23 - V$ litres of water to achieve the desired levels of dilution where single scattering is dominant. These parameters can be proportionately scaled to any other volume V_n, using a scale factor

of V_n/V. The percentage RMS errors (obtained over all pixels) quantify the accuracy of fits achieved with the estimated parameters to the measured intensity profiles. Errors for all the highly scattering media are less than 3%. For low-scattering materials, the total intensity of profiles is relatively low, thus making the estimation more sensitive to noise. Even for such low-scattering media, the errors are less than $5-6\%$. The last four rows are the parameters for various ocean water samples at their original concentrations. The **time elapsed** between the collection of samples and the image acquisition is listed in the parentheses. Since the suspended particles in ocean water settle down with time, we observe a small decrease in scattering coefficients in the sample for which 8 hours had been elapsed as compared to the one which was imaged just 30 minutes after collection. Note that all the extinction and scattering coefficients are less than 0.04 in accordance with our simulations in Section 4.2.3. As expected, the scattering coefficients do not decrease with wavelength. The scattering albedos (ratio of scattering coefficients to the extinction coefficients) is much higher for the scattering media (milk, coffee, orange powder) as compared to the absorbing ones (coke, wine). For materials that have $\beta = 0$, the phase function parameter g is undefined. As seen from the values of g which are closer to 1, several media are predominantly forward scattering. The parameters for the milks match those in [61] up to a scale factor (due to the different fat contents in the milks used), providing further support for our estimation.

4.4.1 Fits to Measured Brightness Profiles

We demonstrate the accuracy of our technique by reconstructing the photographs using the estimated parameters. Although we considered the brightness at all pixels in the captured photographs, for illustration purposes we show only the profile of intensity values in the direction that is radially outward from the source. Figure 4.10 shows the good fits obtained using the estimated parameters compared against the measured profiles for a subset of six materials of varying degrees of scattering and absorption properties. When there is no scattering (pure absorption), fitting a scattering model can induce some "ringing" effect in the dark tail end of the profile. We can detect this special case and use the attenuation model to compute the absorption coefficient ($\kappa = \sigma$).

4.4.2 Extrapolation to higher concentrations

The extinction and scattering coefficients are proportional to the concentration of the medium. Thus, if β_1 and σ_1 are estimated at concentration c_1, then the coefficients β_2 and σ_2 at another concentration c_2 can be extrapolated using:

$$\beta_2 = \beta_1 \left(\frac{c_2}{c_1}\right) \ , \ \sigma_2 = \sigma_1 \left(\frac{c_2}{c_1}\right) . \tag{4.8}$$

Note, however, that g is independent of the medium concentration. While we estimate the parameters from lower concentrations, it is important to ensure that the parameters can be scaled to any concentration (where multiple scattering cannot be ignored) to produce accurate scattering effects. We show an example validation using fits obtained in comparison to the measured brightness profiles of chocolate milk at various concentrations. Figure 4.11 shows the fits in this validation experiment. First, we estimate the parameters from the photograph of only 8ml of chocolate milk diluted in water, where single scattering is dominant. In (a), we show the fit obtained compared against the measured intensity profile. However, for higher concentrations of 50ml, 100ml and 150ml, multiple scattering cannot be ignored. For these cases, we scaled the coefficients (σ and β) by factors of $\{50/8, 100/8, 150/8\}$ (see Equation 4.8) and use them in a standard volumetric Monte Carlo renderer that includes multiple scattering. The plots in (b) - (d) demonstrate the strong fits obtained. This demonstrates that our parameters are robust enough to be extrapolated to higher concentrations. In fact, we will show renderings of most of the liquids at their natural concentrations (Section 4.5) despite measuring the parameters at significantly dilute states.

4.5 Example Volumetric Renderings

The scattering properties estimated in this work can be input to any volumetric rendering algorithm to create visual effects of participating media. Here, we chose brute-force volumetric Monte-Carlo path tracing since it can be used to render arbitrary materials[3]. We use photon mapping for rendering caustics. For display purposes, we have applied a tone-mapping operator [126] to the renderings. Indices of refraction (IOR) of these media are also important for rendering. In initial experiments, we found the IOR to be between 1.33 (water) and 1.42 (milk) and varying linearly with concentrations, by using location of total internal reflection from the top of the water surface in the tank. In current renderings, we have simply used an IOR proportionate to the medium concentrations between 1.33 and 1.42, since this does not alter the visual appearance of the liquid drastically. We wish to perform thorough experiments in the future.

Figure 4.12 shows a mosaic of images of liquids rendered in their natural concentrations, partially filled in a cognac glass and illuminated by the "Kitchen Environment Map" [20]. These include two different types of wine (deep red MERLOT and golden-yellow CHARDONNAY), dark brown coffee ESPRESSO, and the golden-orange YUENGLING beer. Notice the color differences between MERLOT (no scattering) and ESPRESSO (mod-

[3]Under-sampling of path-traces can cause speckle noise seen in the renderings, and is not an artifact of our estimation.

erate scattering) even though both of them are dark liquids. Observe that while beer and CHARDONNAY are very clear liquids, coffee is noticeably more opaque. Similarly, Figure 4.13 shows a mosaic of predominantly bright colored liquids such as the deep blue ERA detergent, the reddish strawberry shampoo, and powders dissolved in water such as the "pinkish" strawberry lemonade and orange powders. These images are illuminated only by a strong directional source to illustrate the bright caustics whose colorings are primarily due to absorption. We also present different types of novel visual effects obtained by changing or blending the parameters of different media to create realistic images of dilutions and mixtures of the original measured materials.

Effect of changing concentrations: Figure 4.14 illustrates the effects of changing concentrations of media in water. The top row shows a transition from pure water to MERLOT, obtained by scaling parameters of wine as in Equation 4.8. Notice the changes in caustics and the gradual deepening of the red color of the liquid. Note that as the transition occurs, the liquid remains clear even though the color changes; this is due to the pure absorbing nature of wine, as depicted by our parameters. The bottom row shows the effect of changing milk concentration in water. Since milk is a highly scattering medium, as expected, the appearance quickly changes from murky whitish water to soft and thick white milk. This is because the scattering albedo β/σ is high and the phase function parameter g is such that a significant amount of light diffuses into different directions.

Blending parameters for mixtures of media: For example, what are the properties of a mixture of ESPRESSO and milk, or otherwise known as *light coffee*? Consider a medium containing a mixture of two types of media, A and B. The properties of the individual media are denoted with the subscripts A and B. The scattering coefficient of the mixture is obtained by a weighted average,

$$\beta_{mix} = \frac{V_A\beta_A + V_B\beta_B}{V_A + V_B}\,. \tag{4.9}$$

The absorption and extinction coefficients are similarly defined. Unlike above where we just changed the scattering and absorption coefficients, here a new phase function parameter must be defined for the mixture as the weighted average [63],

$$g_{mix} = \frac{g_A\beta_A + g_B\beta_B}{\beta_{mix}}\,. \tag{4.10}$$

These equations can be used to render mixtures of participating media or morph from one medium into another. Figure 4.15 shows mixing of different proportions of milk and wine. The second example shows a more common mixing of milk and coffee. Such mixing between materials, for the first time, gives a user the flexibility to create novel renderings

of participating media.

4.6 Discussion

Rendering the rich visual effects of participating media, like fluids or underwater impurities, requires precise measurements of their scattering properties. We have developed a simple device and method for accurately estimating the scattering properties of a variety of media that can be diluted in water. Our approach only requires a single high dynamic range photograph. By diluting the medium, we work in the single scattering regime, where the inverse light transport problem is well conditioned—however, we can later render at arbitrary concentrations and even mix materials. We have presented a database of scattering parameters for 40 commonly found materials. This database is the first of its kind, and enables computer graphics practitioners to accurately render a wide variety of participating media, rather than having to set parameters in an ad-hoc fashion. In the future, we would like to improve this work by investigating different phase functions and measuring indices of refraction more accurately.

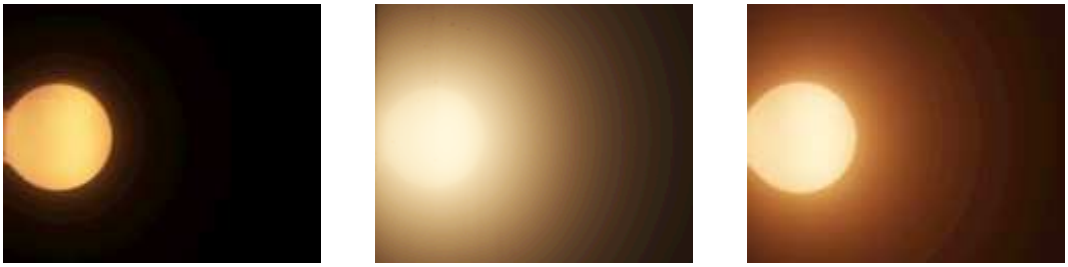
This work is a step towards data-driven measurements and models of scattering, just as there has been much recent work on data-driven reflectance and texture. In the future, we expect more work in this direction—for example, we currently mostly use a parametric phase function, but in future, we can develop an arbitrary non-parametric expansion in Legendre polynomials. We also wish to measure scattering properties of inhomogeneous media, just as there is interest in spatially-varying reflectance. In general, we believe that the accurate simulation of participating media is critical for photorealistic renderings, and accurate scattering properties of real materials are crucial to achieving this goal.

During the course of the entire experimentation, we have learnt a few important lessons which we believe will allow us to obtain even better datasets in the future. Most importantly, we wish to use a multi-spectral camera that measures radiance at finer resolution wavelength bands. Off-the-shelf cameras measure the three wide-band color channels, R, G, and B, while scattering properties can vary strongly within each wavelength band. This results in a somewhat lower color saturation for thick highly scattering liquids. This problem is in some ways analogous to the saturated colors of surface interreflections. Finally, we wish to explore more sophisticated phase functions, borrowing from literature in colloidal chemistry.

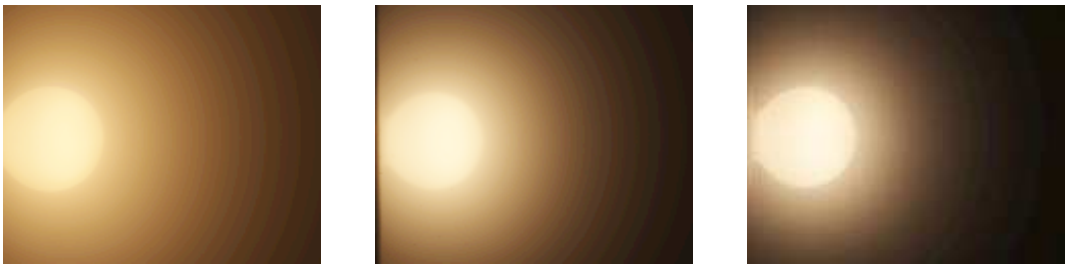
Material Name	Volume	Extinction Coefficient (σ) ($\times 10^{-2}$ mm^{-1})			Scattering Coefficient (β) ($\times 10^{-2}$ mm^{-1})			Average Cosine (g)		
		R	G	B	R	G	B	R	G	B
Milk (lowfat)	16ml	0.9126	1.0748	1.2500	0.9124	1.0744	1.2492	0.932	0.902	0.859
Milk (reduced)	18ml	1.0750	1.2213	1.3941	1.0748	1.2209	1.3931	0.819	0.797	0.746
Milk (regular)	15ml	1.1874	1.3296	1.4602	1.1873	1.3293	1.4589	0.750	0.714	0.681
Coffee (espresso)	8ml	0.4376	0.5115	0.6048	0.2707	0.2828	0.2970	0.907	0.896	0.880
Coffee (mint mocha)	6ml	0.1900	0.2600	0.3500	0.0916	0.1081	0.1460	0.910	0.907	0.914
Soy Milk (lowfat)	16ml	0.1419	0.1625	0.2740	0.1418	0.1620	0.2715	0.850	0.853	0.842
Soymilk (regular)	12ml	0.2434	0.2719	0.4597	0.2433	0.2714	0.4563	0.873	0.858	0.832
Chocolate Milk (lowfat)	10ml	0.4282	0.5014	0.5791	0.4277	0.4998	0.5723	0.934	0.927	0.916
Chocolate Milk (regular)	16ml	0.7359	0.9172	1.0688	0.7352	0.9142	1.0588	0.862	0.838	0.806
Soda (coke)	1600ml	0.7143	1.1688	1.7169	0.0177	0.0208	0.0000	0.965	0.972	–
Soda (pepsi)	1600ml	0.6433	0.9990	1.4420	0.0058	0.0141	0.0000	0.926	0.979	–
Soda (sprite)	15000ml	0.1299	0.1283	0.1395	0.0069	0.0089	0.0089	0.943	0.953	0.952
Sports Gatorade	1500ml	0.4009	0.4185	0.4324	0.2392	0.2927	0.3745	0.933	0.933	0.935
Wine (chardonnay)	3300ml	0.1577	0.1748	0.3512	0.0030	0.0047	0.0069	0.914	0.958	0.975
Wine (white zinfandel)	3300ml	0.1763	0.2370	0.2913	0.0031	0.0048	0.0066	0.919	0.943	0.972
Wine (merlot)	1500ml	0.7639	1.6429	1.9196	0.0053	0.0000	0.0000	0.974	–	–
Beer (budweiser)	2900ml	0.1486	0.3210	0.7360	0.0037	0.0069	0.0074	0.917	0.956	0.982
Beer (coorslight)	1000ml	0.0295	0.0663	0.1521	0.0027	0.0055	0.0000	0.918	0.966	–
Beer (yuengling)	2900ml	0.1535	0.3322	0.7452	0.0495	0.0521	0.0597	0.969	0.969	0.975
Detergent (Clorox)	1200ml	0.1600	0.2500	0.3300	0.1425	0.1723	0.1928	0.912	0.905	0.892
Detergent (Era)	2300ml	0.7987	0.5746	0.2849	0.0553	0.0586	0.0906	0.949	0.950	0.971
Apple Juice	1800ml	0.1215	0.2101	0.4407	0.0201	0.0243	0.0323	0.947	0.949	0.945
Cranberry Juice	1500ml	0.2700	0.6300	0.8300	0.0128	0.0155	0.0196	0.947	0.951	0.974
Grape Juice	1200ml	0.5500	1.2500	1.5300	0.0072	0.0000	0.0000	0.961	–	–
Ruby Grapefruit Juice	240ml	0.2513	0.3517	0.4305	0.1617	0.1606	0.1669	0.929	0.929	0.931
White Grapefruit Juice	160ml	0.3609	0.3800	0.5632	0.3513	0.3669	0.5237	0.548	0.545	0.565
Shampoo (balancing)	300ml	0.0288	0.0710	0.0952	0.0104	0.0114	0.0147	0.910	0.905	0.920
Shampoo (strawberry)	300ml	0.0217	0.0788	0.1022	0.0028	0.0032	0.0033	0.927	0.935	0.994
Head & Shoulders	240ml	0.3674	0.4527	0.5211	0.2791	0.2890	0.3086	0.911	0.896	0.884
Lemon Tea Powder	5tsp	0.3400	0.5800	0.8800	0.0798	0.0898	0.1073	0.946	0.946	0.949
Orange Powder	4tbsp	0.3377	0.5573	1.0122	0.1928	0.2132	0.2259	0.919	0.918	0.922
Pink Lemonade Powder	5tbsp	0.2400	0.3700	0.4500	0.1235	0.1334	0.1305	0.902	0.902	0.904
Cappuccino Powder	0.25tsp	0.2574	0.3536	0.4840	0.0654	0.0882	0.1568	0.849	0.843	0.926
Salt Powder	1.75cup	0.7600	0.8685	0.9363	0.2485	0.2822	0.3216	0.802	0.793	0.821
Sugar Powder	5cup	0.0795	0.1759	0.2780	0.0145	0.0162	0.0202	0.921	0.919	0.931
Suisse Mocha Powder	0.5tsp	0.5098	0.6476	0.7944	0.3223	0.3583	0.4148	0.907	0.894	0.888
Mission Bay Surface Water (1-2 hours)		3.3623	3.2929	3.2193	0.2415	0.2762	0.3256	0.842	0.865	0.912
Pacific Ocean Surface Water (1 hour)		3.3645	3.3158	3.2428	0.1800	0.1834	0.2281	0.902	0.825	0.914
Mission Bay 10ft deep Water (30 min)		3.4063	3.3410	3.2810	0.0990	0.1274	0.1875	0.726	0.820	0.921
Mission Bay 10ft deep Water (8 hours)		3.3997	3.3457	3.2928	0.1018	0.1033	0.1611	0.929	0.910	0.945

Table 4.1: Scattering properties for 40 different water-soluble materials estimated using our technique. The second column in the table lists the volumes V of the materials dissolved in $23 - V$ litres of water to achieve the desired levels of dilution where single scattering is dominant. These parameters can be proportionately scaled to any other volume V_n, using a scale factor of V_n/V. The last four rows are the parameters for various ocean water samples at their original concentrations. The **time elapsed** between the collection of samples and the image acquisition is listed in the parentheses.

Grape Juice | ERA Detergent | Strawberry Shampoo



Lemon Tea Powder | Chocolate milk (regular) | Pink Lemonade Powder



Cappuccino Powder | Coffee Espresso | Low Fat Milk

Figure 4.9: Captured photographs of a variety of water-soluble media illustrating different degrees of scattering and absorption. For highly scattering media such as milk, chocolate milk and espresso, we observe a significant blur around the bulb. For highly absorbing media such as grape juice, there is very little scattering. All the images have wide dynamic range of intensities and hence, we have tone-mapped them for illustration. **Please see the project web-page [3] for more images.**

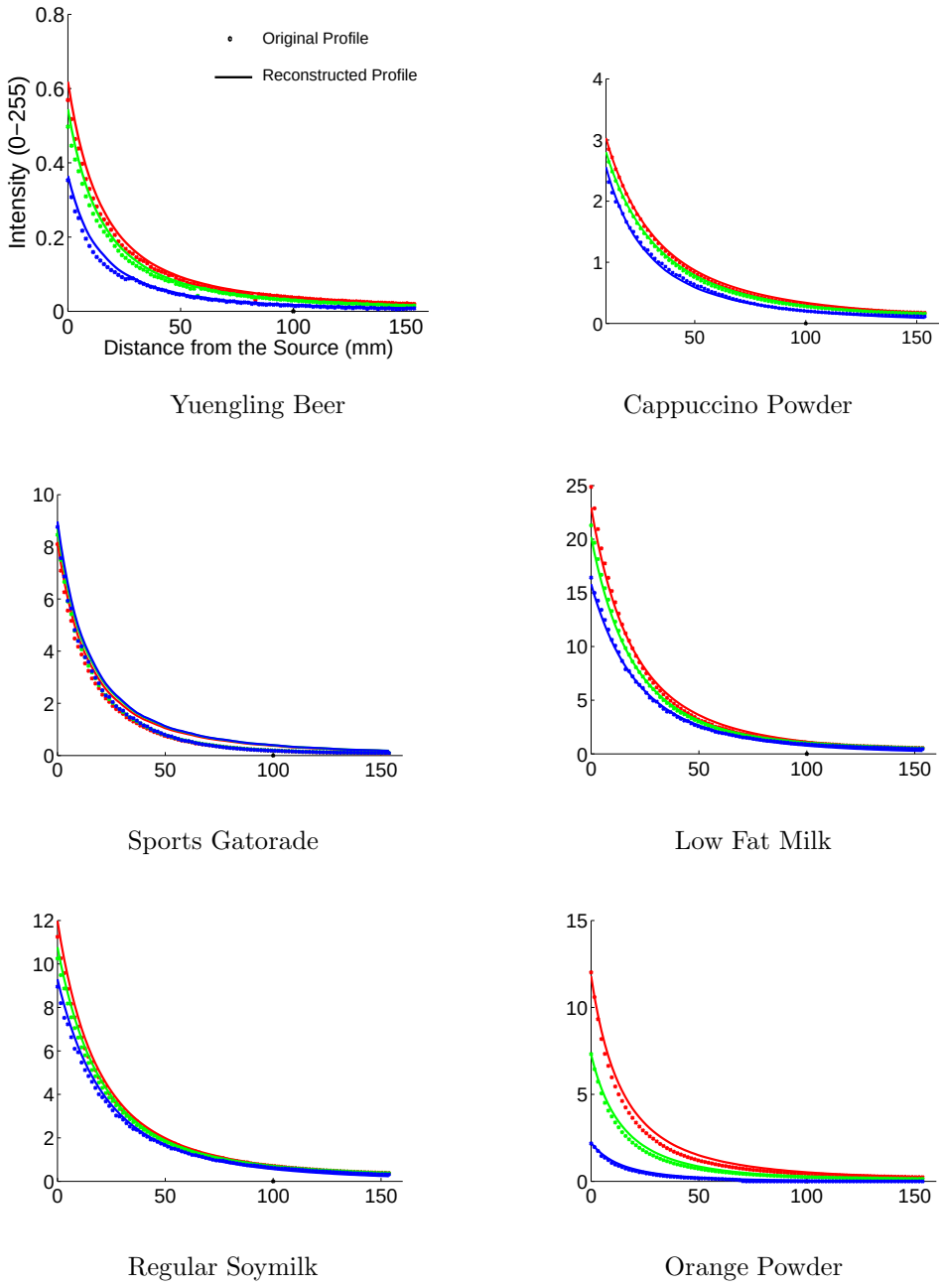


Figure 4.10: Fits obtained using the estimated parameters as compared against the corresponding measured brightness profiles in the captured photographs. The brightness profile is measured radially outward from the source in the image. The red, green and blue plots correspond to the three color channels of the camera. The match between the estimated and measured data demonstrates the accuracy of the estimation technique. The fits for six (out of 40) representative materials with varying degrees of absorption and scattering are shown. **Please see the project web-page [3] for more plots.**

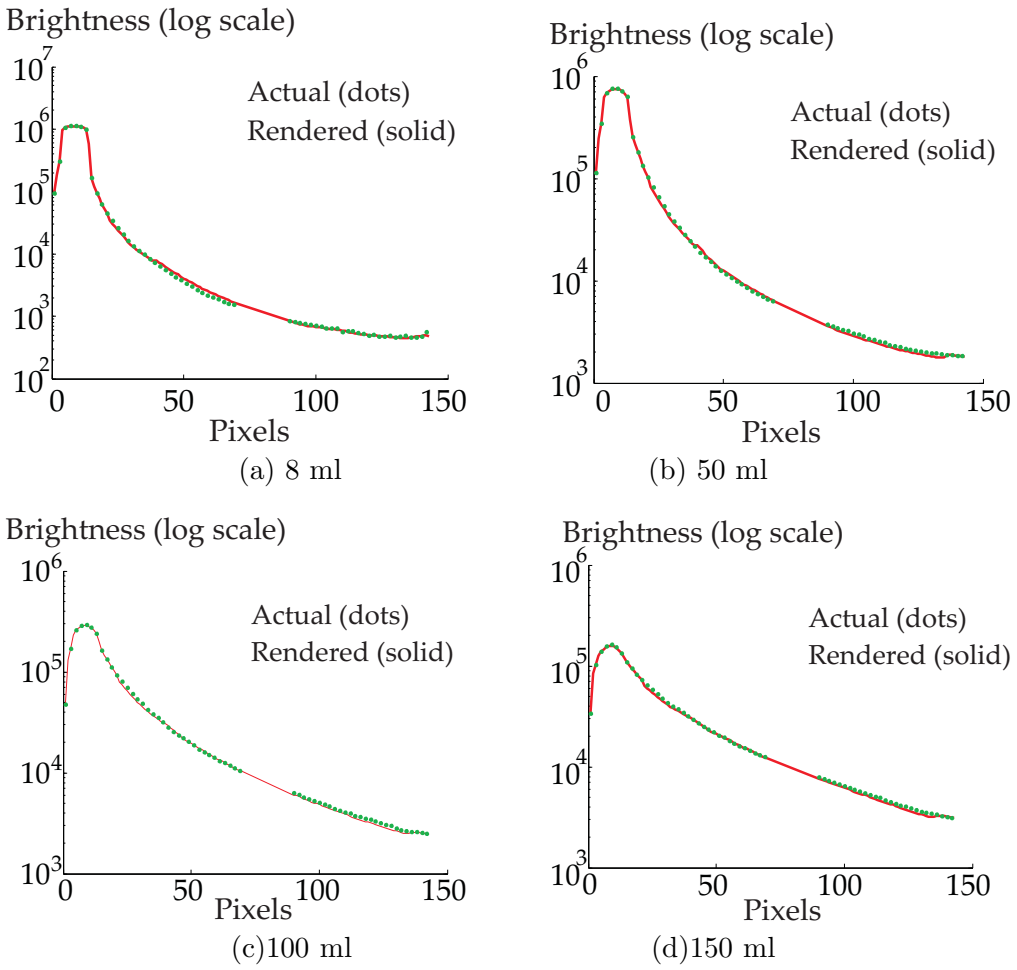


Figure 4.11: Extrapolation of parameters to higher concentrations with multiple scattering. (a) 8 ml of chocolate milk is diluted in water and the parameters are estimated using the measured brightness profile. (b) - (d) The parameters estimated in (a) are scaled to higher concentrations (50ml, 100ml and 150ml) where multiple scattering cannot be ignored. Plots show a good fit between the brightness profile obtained by extrapolating our estimated parameters with a Monte Carlo renderer, and the ground truth measurements. The fits are shown in logarithmic scale.

MERLOT Wine

CHARDONNAY Wine

ESPRESSO Coffee

YUENGLING Beer

Figure 4.12: Rendered scenes of liquids in a cognac glass under complex lighting. The KITCHEN environment map [20] was used for the lighting. The natural colors, shading and caustics indicate the high accuracy of our parameters.

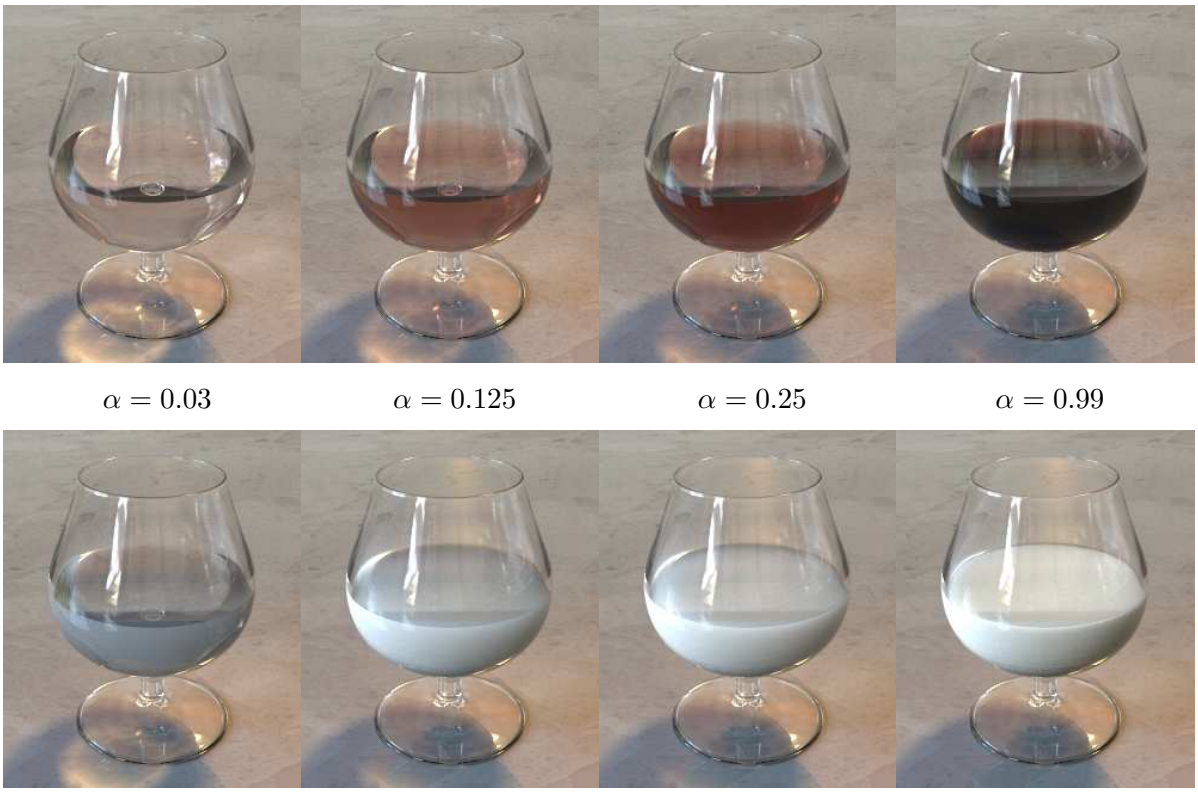
Pink Lemonade Powder

ERA Detergent

Strawberry Shampoo

Orange Powder

Figure 4.13: Rendered scenes of liquids and powders in a cognac glass illuminated with a single directional white light source. The bright caustics show the colors transmitted through the media.



$\alpha = 0.03$ $\alpha = 0.125$ $\alpha = 0.25$ $\alpha = 0.99$

$\alpha = 0.03$ $\alpha = 0.125$ $\alpha = 0.25$ $\alpha = 0.99$

Figure 4.14: Effect of changing concentrations of a highly absorbing (MERLOT) and a highly scattering (milk) liquid. In the case of wine, notice that while the color gradually becomes deep red, the liquid remains clear, due to the lack of scattering. In the case of milk, however, we see a quick transition from a murky appearance to a soft white appearance, due to the high scattering albedo of milk.

50% Milk + 50% Coffee

75% Milk + 25% Coffee

50% Wine + 50% Milk

75% Wine + 25% Milk

Figure 4.15: Mixing two liquids - milk and coffee (top) and milk and wine (bottom), in different proportions. The wine-milk combination produces a soft pink appearance while the ESPRESSO-milk combination produces soft but brown appearance. (Minor noise due to Monte-Carlo under-sampling.)

Chapter 5

Illuminating the scene in poor visibility environments

http://graphics.cs.cmu.edu/projects/LightTransport/

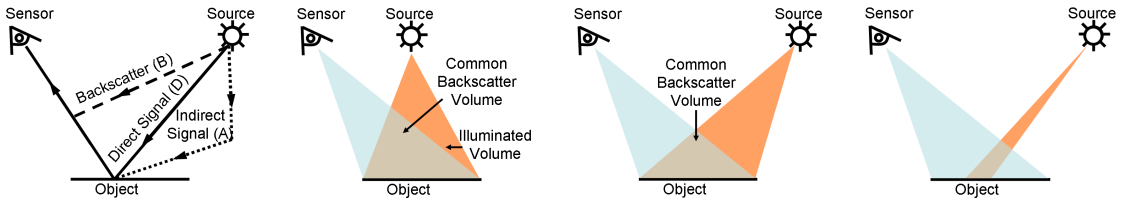


Figure 5.1: Light transport in scattering media for different source and sensor configurations. (a) Illustration of the three light transport components. (b) The backscatter B reduces the image contrast. The amount of backscatter increases with the common backscatter volume. (c) By changing the relative placement of the sensor and source, we can modulate the light transport components for increasing the image contrast. (d) The common backscatter volume can be reduced by using light stripe scanning as well.

Computer vision systems are increasingly being deployed in domains such as surveillance and transportation (terrestrial, underwater or aerial). To be successful, these systems must perform satisfactorily in common poor visibility conditions including murky water, bad weather, dust and smoke. Unfortunately, images captured in these conditions show severe contrast degradation and blurring, making it hard to perform meaningful scene analysis.

Passive methods for restoring scene contrast [83, 104, 118] and estimating 3D scene structure [18, 80, 132] rely on post-processing based on the models of light transport in natural lighting. Such methods do not require special equipment and are effective under

moderate visibility [80], but are of limited use in poor visibility environments. Very often, there is simply not enough useful scene information in images. For example, in an 8-bit camera, the intensity due to dense fog might take up 7 bits, leaving only 1 bit for scene radiance. Active systems, on the other hand, give us flexibility in lighting and/or camera design, allowing us to control the light transport in the environment for better image quality.

While propagating within a medium such as murky water or fog, light gets absorbed and scattered. Broadly speaking, light transport [14] can be classified based on three specific pathways: (a) from the light source to the object, (b) from the object to the sensor and (c) from the light source to the sensor without reaching the object (see Figure 5.1). Of these, the third pathway causes loss of contrast and effective dynamic range (for example, the backscatter of car headlights in fog), and is thus undesirable.

The goal of this research [42] is to design computer vision systems which can see clearer and farther in poor visibility conditions. One way to achieve this is to actively control illumination and choose sensing modalities that maximize light transport along the first two pathways while simultaneously minimizing transport along the third. To this end, we exploit some real world observations. For example, while driving in foggy conditions, flood-lighting the road ahead with a *high-beam* may reduce visibility due to backscatter. Underwater divers realize that maintaining a good separation between the source and the camera reduces backscatter, and improves visibility [111, 59]. On the other hand, polarization filters have also been used to reduce contrast loss due to haze and murky water [110, 106, 33]. Based on these observations, we attempt to address two key questions:

(1) First, which illumination and sensing modality allows us to modulate the three light transport pathways most effectively? We present an active imaging technique called *polarized light striping* and show that it performs better than previous techniques such as flood-lighting, unpolarized light striping [66, 85, 59], and high frequency illumination based separation of light transport components [87].

(2) Second, what is the "optimal" placement of the source and the sensor? We derive a numerical approach for computing the *optimal relative sensor-source position* in poor visibility conditions. We consider a variety of illumination and sensing techniques, while accounting for the limits imposed by sensor noise. Our model can be used for improving visibility in different outdoor applications. It is useful for tasks such as designing headlights for vehicles (terrestrial and underwater). We validate our approach in real experiments.

Finally, we extend the analysis to multi-camera vision systems for the purpose of recovering shape of underwater scenes. We propose a technique for recovering both scene depths and normals simultaneously. It is based on the principle of Helmholtz reciprocity ([48], p. 231). Our techniques can be applied in a variety of outdoor applications, such as **designing headlights** for vehicles (terrestrial and underwater).

(a) (b)

Figure 5.2: Visibility degradation in scattering media. (a) An underwater scene shot in the Mediterranean Sea. The glow at the top and bottom of the image corresponding to backscatter from the two light sources has reduced image contrast. The image is taken from Treibitz et al [119]. (c) Backscatter due to car headlights reduces visibility in foggy conditions.

5.1 Light Transport in scattering media

While propagating through a scattering medium, light gets absorbed and scattered (Figure 5.2 (a)). The image irradiance at a pixel is given as a sum of the three components, the direct signal (D), the indirect signal (A) and the backscatter (B):

$$E(x,y) \quad = \quad \underbrace{D(x,y) + A(x,y)}_{\text{Signal}} + \quad \underbrace{B(x,y)}_{\text{Backscatter}} \quad . \tag{5.1}$$

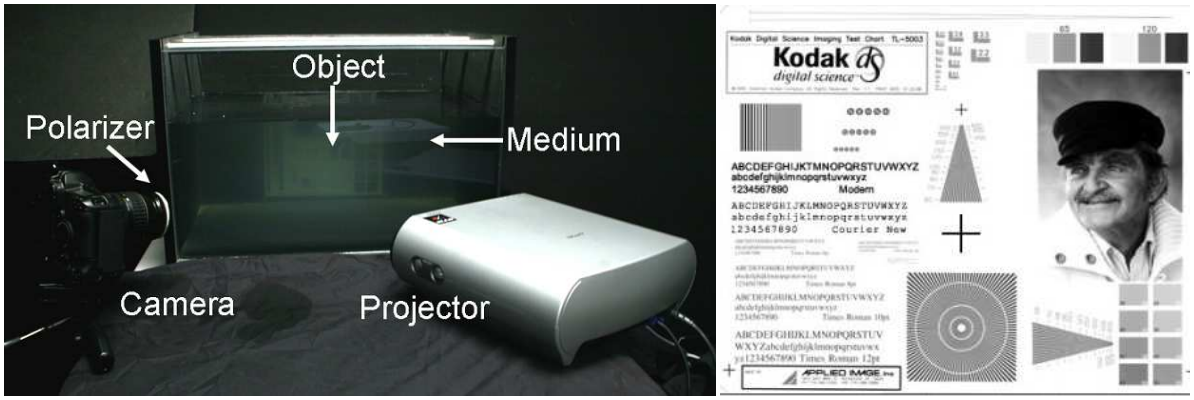
The signal component S is

$$S(x,y) \quad = \quad D(x,y) + A(x,y)\,. \tag{5.2}$$

The backscatter component B reduces image contrast thus degrading visibility. For example, Figure 5.2 (a) illustrates an underwater scene shot in the Mediterranean Sea. The glow at the top and bottom of the image corresponding to backscatter from the two light sources has reduced image contrast. Similarly, everybody who has driven in foggy conditions has observed that flood-lighting the road ahead with a high-beam can be counter-productive due to backscatter.

5.2 How to Illuminate and Capture the Scene?

The goal of an active vision system deployed in poor visibility conditions should be to modulates the components of light transport effectively. Specifically, we want to maximize the signal S, while minimizing the backscatter B. We present an active imaging technique,



Experimental Setup Kodak Contrast Chart

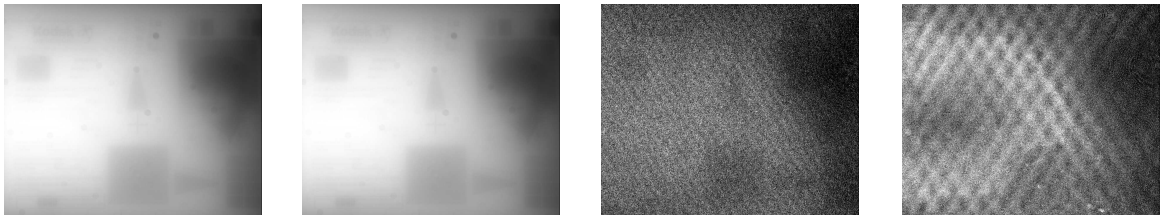
Figure 5.3: Our experimental setup consisting of a glass tank, filled with moderate to high concentrations of milk (four times as those in [85]). An LCD projector illuminates the medium with polarized light. The camera (with a polarizer attached) observes a contrast chart through the medium.

polarized light striping, which combines the advantages of different imaging and illumination modalities. We also analyze the relative merits of different existing techniques, and show that polarized light striping outperforms them.

We begin by analyzing a representative set of existing techniques in laboratory experiments. Our experimental setup consists of a $60 \times 60 \times 38\,cm^3$ glass tank filled with dilute milk (see Figure 5.3). The glass facades are anti-reflection coated to avoid stray reflections.[1] The scene consists of objects immersed in murky water or placed behind the glass tank. A projector illuminates the scene and a camera fitted with a polarizer observes the scene. We use a Sony VPL-HS51A, Cineza 3-LCD video projector. The red and the green light emitted from the projector are inherently polarized channels. If we want to illuminate the scene with blue light, we place a polarizer in front of the projector. We use a 12-bit Canon EOS1D Mark-II camera, and a Kodak contrast chart as the object of interest to demonstrate the contrast loss or enhancement for different techniques.

High-frequency illumination: Ref. [87] presented a technique to separate direct and global components of light transport using high frequency illumination, with good separation results for inter-reflections and sub-surface scattering. What happens in the case of light transport in volumetric media? Separation results in the presence of moderate volumetric scattering are illustrated in Figure 5.4. The direct component is the direct signal (D), whereas the global component is the sum of indirect signal (A) and the backscatter (B), as shown in Figure 5.1. Thus, this method seeks the following separation:

[1]Imaging into a medium through a flat interface creates a non-single viewpoint system. The associated distortions are analyzed in [120].



(a) Maximum image (b) Global component (c) Direct component (d) Direct component (low freq illumination)

Figure 5.4: Limitations of the high frequency illumination based method. A shifting checkerboard illumination pattern was used with the checker size of 10×10 pixels. (a) Maximum image (b) Minimum image (global component) (c) Direct component (d) Direct component obtained using lower frequency illumination (checker size of 20×20 pixels). The direct component images have low SNR in the presence of moderate to heavy volumetric scattering. The global image is approximately the same as a flood-lit image, and hence, suffers from low contrast. This experiment was conducted in moderate scattering conditions, same as the second row of Figure 5.6.

$$E(x,y) \quad = \quad \underbrace{D(x,y)}_{\text{Direct}} + \underbrace{A(x,y) + B(x,y)}_{\text{Global}} . \tag{5.3}$$

However, to achieve the best contrast, we wish to separate the signal $D + A$ from the backscatter B. As the medium becomes more strongly scattering, the ratio $\frac{D}{S}$ falls rapidly due to heavy attenuation and scattering, as illustrated in Figure 5.5. This plot was estimated using numerical simulations using the single scattering model of light transport.[2] Consequently, for moderate to high densities of the medium, the direct image suffers from low signal-to-noise-ratio (SNR), as shown in Figure 5.4. Further, the indirect signal (A) remains unseparated from the backscatter B, in the global component. Thus, the global image is similar to a flood-lit image, and suffers from low contrast.

Polarized flood-lighting: Polarization imaging has been used to improve image contrast [106, 119, 33] in poor visibility environments. It is based on the principle that the backscatter component is partially polarized, whereas the scene radiance is assumed to be unpolarized. Using a sensor mounted with a polarizer, two images can be taken with two orthogonal orientations of the polarizer:

$$E_{\mathrm{b}} = \frac{D + A}{2} + \frac{B(1-p)}{2} \tag{5.4}$$

[2]With multiple scattering, the ratio falls even more sharply.

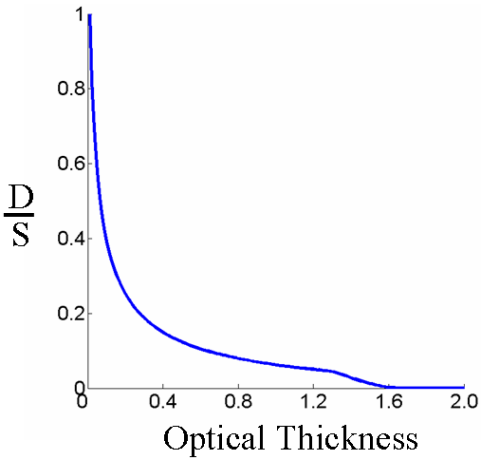


Figure 5.5: The relative direct component of the signal reduces with increasing optical thickness of the medium. This plot was calculated using simulations, with a two-term Henyey-Greenstein scattering phase function [44] for a parameter value of 0.8.

$$E_{\mathrm{w}} = \frac{D + A}{2} + \frac{B(1+p)}{2}\,, \tag{5.5}$$

where p is the degree of polarization (DOP) of the backscatter. Here, E_{b} and E_{w} are the 'best-polarized image' and the 'worst-polarized image', respectively. Thus, using optical filtering alone, backscatter can be removed partially, depending on the value of p. Further, it is possible to recover an estimate of the signal $\hat{S}$ in a post-processing step [106]:

$$\hat{S} \;\; = \;\; E_{\mathrm{b}} \left(1 + \frac{1}{p}\right) + E_{\mathrm{w}} \left(1 - \frac{1}{p}\right)\,. \tag{5.6}$$

However, in optically dense media, heavy backscatter due to flood-lighting can dominate the signal, making it impossible for the signal to be recovered. This is illustrated in Figure 5.6, where in the case of flood-lighting under heavy scattering, polarization imaging does not improve visibility.

Light stripe scanning: Here, a thin sheet of light is scanned across the scene. In comparison to the above approaches, the common backscatter volume is considerably reduced (see Figure 5.1d). The sheet of light intersects the object to create a stripe that is detected using a gradient operator.[3] All stripes are then mosaiced to create a composite image CI [66, 85, 59]. Alternatively, the composite image can be obtained by simply selecting the maximum value at each pixel over all the individual light stripe images SI_k:

[3]In our particular implementation, the projector illuminates a single plane and has low power. We compensate for this by increasing the exposure time of the camera.

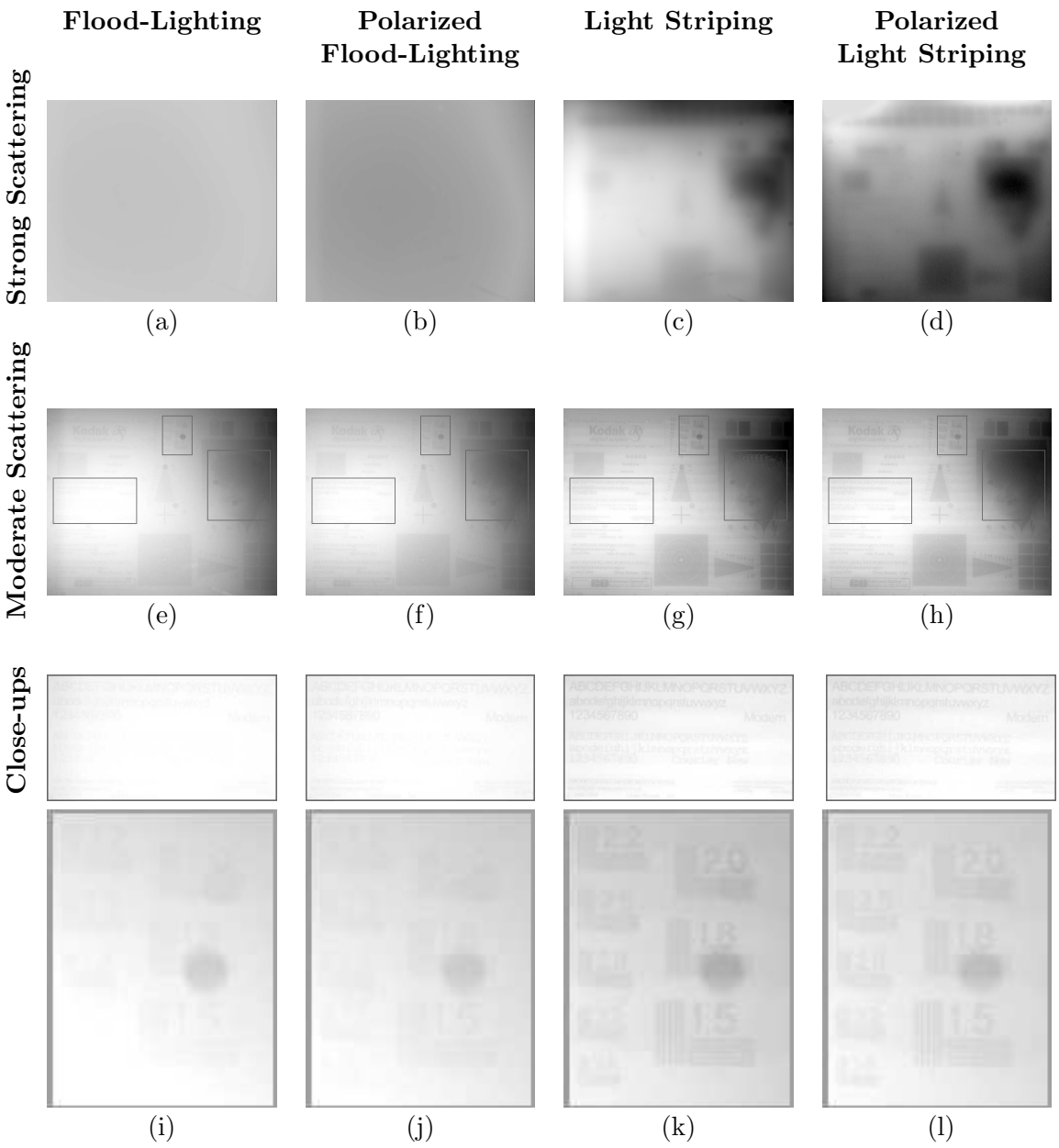


Figure 5.6: Comparison of various illumination and sensing techniques (zoom into the marked areas to better assess the image quality). Flood-lit images suffer from a severe loss of contrast, specially in the presence of heavy scattering (a,b). *Polarized light striping* achieves a significant increase in image contrast, even in the presence of heavy scattering (a-d). In moderate scattering, fine details (text) are recovered more reliably in (g) and (h), as compared to (e). See (i), (j), (k) and (l) for close-ups of the marked areas in (e), (f), (g) and (h) respectively. The moderate scattering experiment was conducted under the same conditions as the experiment in Figure 5.4.

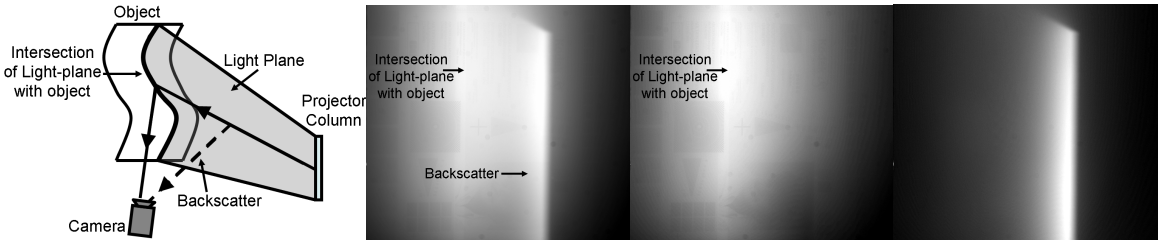
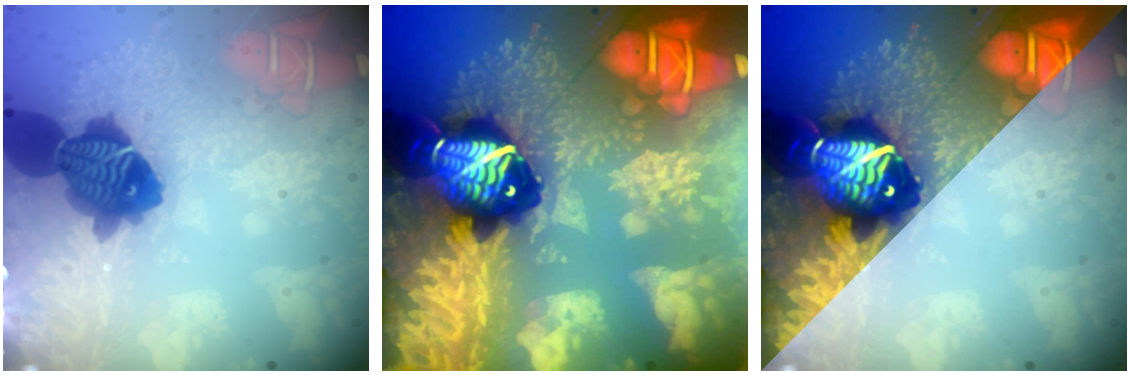


Figure 5.7: Unpolarized versus polarized light stripe scanning. (a) Ray diagram illustrating light stripe scanning, adapted from [85]. (b) The camera observes a light stripe (1 out of 30) without a polarizer. The visible light plane is the backscatter and impedes reliable detection of the object stripe. (c) Through a polarizer, there is a considerable reduction in backscatter. The light plane-object intersection becomes more distinct, thus enabling its reliable delineation. (d) The removed backscatter (difference of (b) and (c)). Video of a complete scan can be downloaded from the project web-page [2].



Flood Lighting | Polarized Light Striping | Comparison

Figure 5.8: Comparison between the performance of flood-lighting and light-stripe scanning.

$$\mathrm{CI}(x,y) \quad = \quad \max_k \{\mathrm{SI}_k(x,y)\}\,. \tag{5.7}$$

5.2.1 Improving image contrast using polarization + light stripe scanning

We propose polarized light striping as a technique that combines the advantages of polarization imaging and light striping, and thus, is applicable for an extended range of medium densities. Earlier, we demonstrated that light striping reduces the amount of backscatter. However, reliable localization of the object stripes (by using gradient operator or by selecting the maximum pixel value, as in Eq. 5.7) is severely impeded due to strong backscatter. This is illustrated in Figure 5.7.

To enable reliable detection of the object stripes even in the presence of strong scattering, we use polarization imaging in conjunction with light striping. A high DOP of the backscatter is essential for removing the backscatter using polarization filtering (Eq. 5.4), or to recover a reliable estimate of the signal using post-processing (Eq. 5.6). In our experiments, the camera observes the scene through a polarization filter and the light sheets irradiating the scene are polarized. Since the incident illumination is completely polarized, the DOP of the backscatter is high (see appendix). This results in a significant reduction in the amount of backscatter, and thus, enables reliable detection of the stripes.[4] This is shown in Figure 5.7. We compare the results of polarized light striping versus previous illumination and sensing techniques in Figure 5.6. Notice especially the differences in the contrast under strong scattering. Another result is shown in Figure 5.8.

5.3 Optimal Camera-Source Placement

Conventional wisdom from the underwater imaging literature suggests maximizing the sensor-source separation to reduce the backscatter, and hence, increase the image contrast [59, 111] (see Figure 5.1). However, this does not take into account the limitations posed by measurement noise. Placing the source and the sensor far from each other or the scene results in strong attenuation of light, and a low Signal-to-Noise-Ratio (SNR). We study this trade-off and derive a numerical approach for computing the optimal sensor-source placement for many different illumination and sensing modalities.

5.3.1 Image Quality Measures

In order to formalize the notion of "optimal", we define various image quality measures for different imaging and illumination techniques. These quality measures serve as objective functions which can be maximized to find the optimal placement of the source and the camera.

Contrast Quality Measure: A major goal of an imaging system is to maximize the image contrast. Analogous to [11, 32], we define the contrast quality measure, $\mathrm{CQM}(x, y)$ as the ratio of the signal $S(x, y)$ to the total intensity $E(x, y)$:

$$\mathrm{CQM}(x, y, p) \quad = \quad \frac{S}{S + B(1 - p)} \cdot \tag{5.8}$$

This measure takes polarization imaging into account by defining the total intensity as that of the *best polarized image*, as in Eq. (5.4). In the absence of a polarizer, $p = 0$.

[4] Polarization imaging was previously used with phase-shifted structured illumination for improved reconstruction of translucent objects [15].

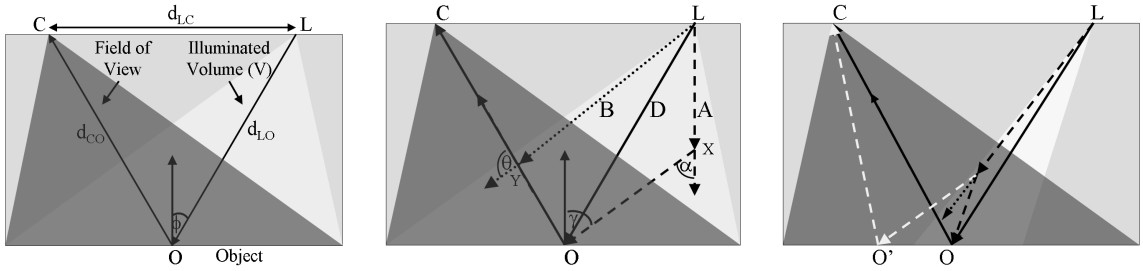


Figure 5.9: Simulating image formation for finding the optimal sensor-source configuration. (a) A schematic view of the volume. We use a point light source (L) and a pinhole camera (C). The object is Lambertian, with reflectance R. (b) We calculate D, A and B according to Eqs. (5.11-5.13). (c) In the case of light striping, the point O' is not getting directly irradiated by the source. Also, the viewing ray from O' does not intersect the common backscatter volume. Thus, the direct component and the backscatter component at O' are null. This results in a brightness gradient across the stripe edge. The strength of the gradient is given by Eq. 5.9.

Delineation of light plane-scene intersection: Success of light striping in scattering media relies on reliable delineation of the object stripe. One scheme is to detect a brightness discontinuity in the intensity profile across the stripe edge. Thus, for a light stripe scanning system, we define a gradient quality measure (GQM) along the edge of the stripe in terms of the strength of gradient across the stripe edge. Consider Figure 5.9c; since the scene point O' does not have the direct component D or the backscatter component $B(1-p)$, the normalized difference in intensity of O and O' is given as:

$$\mathrm{GQM}(x, y, p) \quad = \quad \frac{D + B(1-p)}{D + A + B(1-p)} \; . \tag{5.9}$$

SNR dependent weighting: An image with high contrast but low overall intensity may result in a low SNR, and hence be of limited use. Thus, we define an SNR dependent weight W as a monotonically increasing function of the total signal value S. The quality measures (CQM and GQM) are weighted by W so that signal values in the low SNR range are penalized. For example, W can be a linear function of S. For more flexibilty, we use a sigmoid function of S:

$$\mathrm{W}(x, y) \quad = \quad \frac{1}{1 + e^{-\left(\frac{S-\mu}{z}\right)}} \; , \tag{5.10}$$

where μ is the shift and z is the steepness of the sigmoid. For example, μ can be the dark current offset. Similarly, if the noise is derived from a Gaussian distribution, z can be the standard deviation. In addition, we should account for the effect of post-processing on

image noise [103].

Simulations

Consider an underwater scenario where a remote operated vehicle (ROV) wants to capture images at a given distance. Given an approximate estimate of the object albedo, medium scattering parameters [81] and sensor noise, we can simulate the image formation process. To illustrate the concept, we simulate the image formation process for our experimental setup. The Lambertian object reflectance was assumed to be 0.6. For different source-camera configurations, we compute the appropriate quality measure described above. Then, the optimal configuration is the one that maximizes the quality measure.

Figure 5.9 illustrates the image formation geometry. In our experiments and simulations, the scene and camera remain fixed, while the source is moved to vary the sensor-source separation d_{LC}. Point O on the object is being observed by the camera. Points X and Y are in the medium. The distances $d_{LO}, d_{CO}, d_{LX}, d_{XO}, d_{CO}, d_{LY}$ and d_{YC}, and the angles $\phi, \alpha, \gamma, \theta$ are as illustrated in Figure 5.9. To keep our simulations simple, we assume a single scattering model of light transport and a homogeneous medium. The individual components of light transport are then given by:

$$D = \frac{I_0}{d_{LO}^2} e^{-\sigma(d_{LO}+d_{CO})} R(\phi) \tag{5.11}$$

$$A = \int_V \frac{I_0}{d_{LX}^2} e^{-\sigma(d_{LX}+d_{XO}+d_{CO})} F(\alpha) R(\gamma) dV \tag{5.12}$$

$$B = \int_O^C \frac{I_0}{d_{LY}^2} e^{-\sigma(d_{LY}+d_{YC})} F(\theta) dY \,, \tag{5.13}$$

where I_0 is the source radiance, σ is the extinction coefficient, R is the Lambertian object reflectance, F is the scattering phase function (we use the two-term Henyey-Greenstein function [44]) and V is the illuminated volume.

Polarized images, E_{b} and E_{w} are simulated according to Eqs. (5.4-5.5). This requires knowledge of the DOP of the backscatter p. Using our experimental setup, we estimated p to be approximately 0.8, from the regions of the image without any object. We can also compute p analytically, given the dependence of the DOP of scattered light on the scattering angle, such as given in the Appendix.

Optimal configuration for flood-lighting: Let us find the configuration that is optimal in terms of both image contrast and noise. We plot the product of the CQM and W versus the sensor-source separation d_{LC} (Figure 5.10a). The tradeoff between contrast and SNR results in a local maximum. Notice that polarization improves image quality as compared

to unpolarized imaging. However, since the DOP (and hence, the amount of contrast enhancement) is similar for all sensor-source positions, the location of the peak remains the same. The curve for the ideal case of zero noise increases monotonically. However, for real world scenarios, where measurement noise places limits on the sensor's abilities, our approach can yield an optimal placement. This is illustrated in Figure 5.10 (b-c). The image taken using the optimal separation (40 cms) has high contrast and low noise. On the other hand, notice the significant noise in the image taken using a large separation (60 cms).

Optimizing the light stripe scan: The case of light stripe scanning is more interesting. Instead of illuminating the whole scene at once, we illuminate it using one sheet of light at a time. We want to find the optimal light stripe scan. Should we scan the scene (a) by rotating the source, (b) by translating it, or (c) a combination thereof? To answer this, we plot the product of the GQM and the W for our setup (Figure 5.11). We observe different optimal separations for different (3 out of 30) stripe locations. Figure 5.11 (e) shows the high-contrast image acquired using the results of the simulations. The camera and the projector were placed at a small distance from the facade of the glass tank in real experiments. By carefully choosing the light rays, we can simulate a light source and a sensor placed on the glass facade, as assumed in the simulations. The optimal scan for polarized light striping is the same as unpolarized light striping, but results in better image quality.

5.4 Recovering 3D structure of underwater scenes

We are investigating the problem of how to illuminate the scene for multi-camera vision systems deployed in scattering media. The goal is to recover scene shape - both depths and normals. First, we consider depth recovery using binocular stereo. We discuss the issues which need to be addressed while designing a stereo system under scattering media. Second, we present a technique for recovering both surface depths and normals simultaneously. This technique is based on the principle of Helmholtz reciprocity ([48], p. 231), and uses ideas from the Helmholtz Stereopsis technique proposed by Zickler et al [134].

5.4.1 Binocular Stereopsis in scattering media

In the presence of scattering media, due to attenuation and backscatter, the relative placement of the source with respect to the sensors is an important consideration.Consider the imaging geometry in Figure 5.12. This setup is used to acquire a stereo image pair. The two cameras are places at locations A and B, and the light-source is placed at location L.

For binocular stereo, we assume the scene to be Lambertian. We will relax this assumption later in the section. Let P be a scene point, with its normal given by $\hat{n_P}$.

For low concentrations, we can assume a single scattering model of light transport in participating media. Additionally, we can assume the indirect signal component to be negligible as compared to the direct signal component [42]. Let I_l be the total image intensity for scene point P when the camera is at the left location (A). The total intensity has two components, the direct signal component D_l and the backscatter B_l :

$$I_l \;=\; D_l + B_l \;. \tag{5.14}$$

Let σ be the extinction coefficient of the medium. Let f_P be the albedo at the scene point P. Then, the direct signal component D_l is given as:

$$D_l \;=\; \frac{I_o\, e^{-\sigma d_{LP}}}{d_{LP}^2}\;\cos(\beta)\;\frac{f_P}{\pi}\;e^{-\sigma d_{AP}}\;. \tag{5.15}$$

Let the scattering phase function is given by $\Omega(g,\gamma)$, where g is the phase function parameter, and γ is the angle between the incident and the scattered ray. Then, the backscatter component B_l (single scattering only) is given as:

$$B_l \;=\; \int_0^{d_{AP}} \frac{I_o\, e^{-\sigma a}}{a^2}\;\Omega(g,\gamma)\;e^{-\sigma x}\;dx\;. \tag{5.16}$$

Similarly, let I_r be the total image intensity when the camera is at the right location (B):

$$I_r \;=\; D_r + B_r \;. \tag{5.17}$$

The direct signal and the global components, D_r and B_r are given as:

$$D_r \;=\; \frac{I_o\, e^{-\sigma d_{LP}}}{d_{LP}^2}\;\cos(\beta)\;\frac{f_P}{\pi}\;e^{-\sigma d_{BP}}\;. \tag{5.18}$$

$$B_r \;=\; \int_0^{d_{BP}} \frac{I_o\, e^{-\sigma b}}{b^2}\; \Omega(g,\rho)\; e^{-\sigma y}\; dy\;. \tag{5.19}$$

How to place the source relative to the sensors?

In the presence of scattering media, due to attenuation and backscatter, the relative placement of the source with respect to the sensors is an important consideration. The first row of Figure 5.13 illustrates 4 different sensor-source configurations. The second and the third rows show the corresponding images (left camera and right camera respectively) simulated using the single scattering model. The scene consists of a textured sphere, placed at a distance of $1m$ from the source and the sensor. The sensor-source baseline is $0.5m$. The extinction coefficient $\sigma = 1.0m^{-1}$. Following is a comparison of the four configurations in terms of the contrast and SNR of the acquired images:

- **Configuration 1 - Light in the middle:** In this configuration, the sensor-source separation for both the cameras is low, resulting in high backscatter and low contrast for both the images.
- **Configuration 2 - Light to the side:** The sensor-source separation for the left camera is high, resulting in high contrast but low SNR. Right image has low contrast due to high backscatter.
- **Configuration 3 - Light at the apex:** We can have *optimal* sensor-source separation for both the cameras, resulting in good contrast and SNR for both the images.
- **Configuration 4 - Swap source and sensor:** In this configuration, the left image is taken by placing the source at the right sensor location. The right image is taken by placing the source at the left sensor location. Similar to configuration 3, we can have the *optimal* sensor-source placement for both the images. Note that both the images have good contrast and SNR.

Binocular stereo matching constraint

In clear air, the binocular stereo matching constraint is simply:

$$I_l \;=\; I_r\;, \tag{5.20}$$

i.e., the images of a scene point in the two cameras should have equal intensities. However, in the presence of scattering media, the image intensities of the corresponding scene points in the two images are different due to different amounts of backscatter and attenuation. Thus, the matching constraint has to account for backscatter and attenuation. From Eqs. 5.15 and 5.18:

$$D_l \; e^{\sigma d_{AP}} \;\; = \;\; D_r \; e^{\sigma d_{BP}} \; . \tag{5.21}$$

By substituting $D_l = I_l - B_l$ and $D_r = I_r - B_r$, we derive the binocular stereo matching constraint in the presence of scattering media:

$$(I_l - B_l) \;\; e^{\sigma d_{AP}} - (I_r - B_r) \;\; e^{\sigma d_{BP}} \;\; = \;\; 0 \; . \tag{5.22}$$

In this equation, I_l and I_r are measurements obtained from the camera. For geometrically calibrated cameras, d_{AP} and d_{BP} are functions of the depth d of the scene point. Also, the backscatter terms B_l and B_r are functions of the depth d. Additionally, the backscatter terms depend on the medium scattering parameters: the phase function parameter g, the extinction coefficient σ and the scattering coefficient. The LHS of Eqn. 5.22 can be written as $s(d, \sigma, g)$ to denote the fact that it is a function of the scene depth and medium parameters:

$$s(d, \sigma, g) \;\; = \;\; (I_l - B_l) \;\; e^{\sigma d_{AP}} - (I_r - B_r) \;\; e^{\sigma d_{BP}} \; .$$

Thus, given the knowledge of the medium parameters, scene depths can be computed by minimizing $s(d, \sigma, g)$ as a function of d:

$$d^* \;\; = \;\; \arg\min_d \{ s(d, \sigma, g) \} \; . \tag{5.23}$$

Figure 5.14 illustrates the depth recovery results on simulated images using our binocular stereo matching constraint. The input image pairs were simulated using the single scattering model of Eqs. 5.14- 5.19. Few example input images are shown in Figure 5.13. The three shapes considered are a fronto-parallel plane, an inclined plane and a half-sphere. The ground truth shapes are shown in the top row. Depths recovered using our new matching constraint are shown in the middle row. Depths recovered using the *clear air* matching

constraint (Eq. 5.20) are shown in the bottom row. We can notice that the depths recovered without accounting for attenuation and backscatter are incorrect.

Recovering the medium parameters

We need to recover the medium scattering parameters to estimate depths using the stereo matching constraint, as in the previous sub-section. We can recover the medium parameters from a single image of the medium (without any scene). An example simulated image is shown in Figure 5.15. The light source is to the left of the camera. Both the light source and the camera are pointed in the same direction, perpendicular to the baseline. We added Gaussian noise with a standard deviation of 0.03 times the mean intensity of the image. By fitting the image formation model to the observed image [81], we could recover the medium parameters accurately and robustly despite the presence of noise.

5.4.2 Helmholtz Stereopsis in scattering media

In this section, we present a technique for recovering both the scene depths and normals simultaneously in the presence of scattering media. Our technique is based on the principle of Helmholtz reciprocity. Zickler et al [134] used the reciprocity principle in the context of surface reflection to recover surface shape. Recently, in astro-physics literature, the principle of reciprocity has been extended to volumetric scattering [35, 36]. We extend the Helmholtz Stereopsis technique proposed by Zickler et al to account for volumetric scattering effects. Following are the advantages of Helmholtz stereo over binocular stereo:

- The input images for Helmholtz Stereo are acquired by swapping the source-sensor positions, as shown in the last column of Figure 5.13. With this configuration, we can maintain the *optimal* sensor-source separation for all the input images, resulting in high image quality, both in terms of contrast and SNR.
- Most underwater scenes do not have much scene texture. In the absence of scene texture, binocular stereo can produce erroneous depth-estimates. Helmholtz stereo uses photometric information to recover scene normals in addition to the depths. The normals can then be integrated to compute a more reliable estimate of scene depths.
- With Helmholtz stereo, we can relax the assumption of the scene being Lambertian, and handle a much more general class of BRDFs.

Image Formation Model

Now we describe the image formation model for the Helmholtz stereo image acquisition setup. Consider the imaging geometry in Figure 5.16. This setup is used to acquire a reciprocal image pair. The first image is taken by placing the camera at the left location (A), and the light at the right location (B), as shown in Figure 5.16 (a). The second image is taken by switching the light and the camera positions, as shown in Figure 5.16 (b). **Note the difference with Figure 5.12, where the light source remains stationary for both the views.** Let P be a scene point, with its normal given by $\hat{n_P}$. Let $\hat{d_{AP}}$ and $\hat{d_{BP}}$ be the unit vectors from the scene point P to the locations A and B respectively, with d_{AP} and d_{BP} being the respective vector-lengths.

As before, we assume a single scattering model of light transport and a negligible indirect signal component [42]. Let I_l be the total image intensity for scene point P when the camera is at the left location (A). The total intensity has two components, the direct signal component D_l and the backscatter B_l :

$$I_l \;\; = \;\; D_l + B_l \; . \tag{5.24}$$

Let σ be the extinction coefficient of the medium. Let $f_P(\theta_i, \theta_r)$ be the BRDF at the scene point P, where θ_i and θ_r are the angles that the normal makes with the incident and the reflected ray respectively. Then, the direct signal component D_l is given as:

$$D_l \;\; = \;\; \frac{I_o \, e^{-\sigma d_{BP}}}{d_{BP}^2} \;\; \cos(\beta) \;\; f_P(\beta, \alpha) \;\; e^{-\sigma d_{AP}} \; . \tag{5.25}$$

Let the scattering phase function is given by $\Omega(g, \gamma)$, where g is the phase function parameter, and γ is the angle between the incident and the scattered ray. Then, the backscatter component B_l (single scattering only) is given as:

$$B_l \;\; = \;\; \int_0^{d_{AP}} \frac{I_o \, e^{-\sigma a}}{a^2} \;\; \Omega(g, \gamma) \;\; e^{-\sigma x} \;\; dx \; . \tag{5.26}$$

Similarly, let I_r be the total image intensity when the camera is at the right location (B), and the light is at the left location (A):

$$I_r \;=\; D_r + B_r \;, \tag{5.27}$$

The direct signal and the global components, D_r and B_r are given as before:

$$D_r \;=\; \frac{I_o\, e^{-\sigma d_{AP}}}{d_{AP}^2}\;\cos(\alpha)\;\; f_P(\alpha,\beta)\;\; e^{-\sigma d_{BP}}\;. \tag{5.28}$$

$$B_r \;=\; \int_0^{d_{BP}} \frac{I_o\, e^{-\sigma b}}{b^2}\;\; \Omega(g,\rho)\;\; e^{-\sigma y}\;\; dy\;, \tag{5.29}$$

Helmholtz stereo matching constraint

Now we derive the matching constraint for the input images acquired using the Helmholtz setup. The principle of Helmholtz reciprocity states that the BRDF is symmetric about the incoming and the outgoing directions, i.e., $f_P(\alpha,\beta) = f_P(\beta,\alpha)$. Using this principle, we derive the following relation from Equations 5.25 and 5.28:

$$D_l\;\frac{d_{BP}^2}{\cos(\beta)} \;=\; D_r\;\frac{d_{AP}^2}{\cos(\alpha)}\;. \tag{5.30}$$

Since $D_l = I_l - B_l$ and $D_r = I_r - B_r$:

$$(I_l - B_l)\;\frac{d_{BP}^2}{\cos(\beta)} \;=\; (I_r - B_r)\;\frac{d_{AP}^2}{\cos(\alpha)}\;. \tag{5.31}$$

By substituting $\cos(\alpha) = \hat{d_{AP}} \,.\, \hat{n_P}$ and $\cos(\beta) = \hat{d_{BP}} \,.\, \hat{n_P}$, and re-arranging the terms, we get:

$$\left((I_l - B_l)\;\frac{\hat{d_{AP}}}{d_{AP}^2} \;-\; (I_r - B_r)\;\frac{\hat{d_{BP}}}{d_{BP}^2}\right) \,.\, \hat{n_P} \;=\; 0 \tag{5.32}$$

In this equation, I_l and I_r are measurements obtained from the cameras. Also, for

geometrically calibrated cameras, the values d_{AP}, d_{BP} and the vector $\hat{d_{AP}}$ and $\hat{d_{BP}}$ are functions of the scene depth d. As before, the backscatter terms B_l and B_r are functions of the depth d and the medium scattering parameters: the phase function parameter g, the extinction coefficient σ and the scattering coefficient, which can be estimated, as shown in Section 5.4.1. We can rewrite Eq. 5.32 compactly as:

$$w(d, g, \sigma) \; . \; \hat{n_P} \quad = \quad 0 \; . \tag{5.33}$$

Note that the vector w lies in the epipolar place defined by points P, A and B. The above equation provides a constraint on the pixel intensities of the corresponding image points. For each scene point, we have three unknowns: the surface normal n_P and the depth d. Thus, we need at least 3 such reciprocal image pairs to solve for both the depths and the normals. We acquire such pairs by changing the orientation of the source-sensor arm. An example set of simulated input images is shown in Figure 5.17.

For each camera-light pair, w lies on the intersection of the epipolar plane (E) and the plane perpendicular to the normal. We define $W(d, g, \sigma)$ as the matrix formed by stacking all the w vectors row-wise. Then, it follows that the rank of the matrix should be ≤ 2. We can find the correct depth by minimizing the rank of the $W(d)$ matrix:

$$d^* \quad = \quad \arg\min_{d} \{rank(W(d, \sigma, g))\} \; . \tag{5.34}$$

The normal n_P can then be computed by finding the null-space of the matrix $W(d^*)$. Once we have the normals, the scene depths can also be estimated by integrating the normal field. Figure 5.18 shows the depth recovery results using Helmholtz stereo. We can notice that the depths estimated using normals are smoother and more reliable. The depth estimates using the previous Helmholtz stereo algorithm [134] doesn't account for backscatter, thus producing erroneous depth estimates.

5.5 Discussion

With existing techniques for measurement of medium scattering [81] and polarization properties [125], our simulation-based approach can be used to adapt the illumination and sensing system in-situ. Post-processing approaches are expected to recover the scene when applied to the images acquired using our system.

For depth recovery under scattering media, we have extended binocular and Helmholtz

stereo to account for volumetric scattering. We have analyzed the performance of both algorithms using simulated images. Currently, we are building a rig to capture real, underwater images for testing our techniques.

Appendix: Degree of Polarization of Scattering

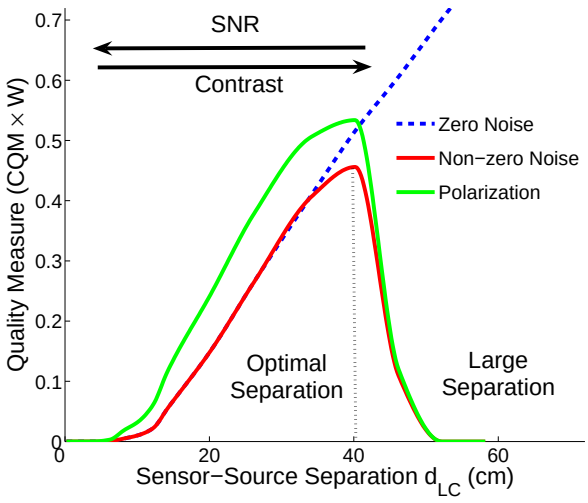
In this appendix, we study the dependence of the DOP of the scattered light, $\mathrm{DOP_B}$, on the scattering angle and the DOP of the incident light, $\mathrm{DOP_L}$. We consider only the vertical and horizontal polarized components of linearly polarized light. Hence, we consider the first 2×2 sub-matrix of the full 4×4 Mueller matrix. Polarization properties of scattered light can be characterized by the Mueller matrix [124]:

$$\begin{bmatrix} I_\mathrm{B} \\ Q_\mathrm{B} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} I_\mathrm{L} \\ Q_\mathrm{L} \end{bmatrix}, \tag{5.35}$$

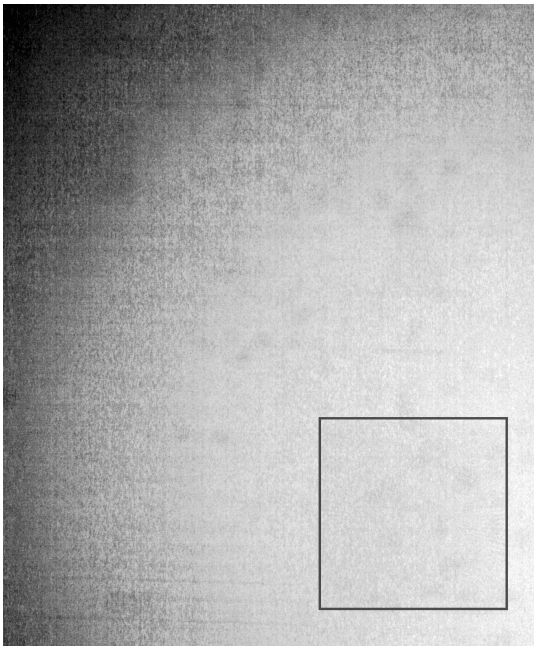
where I_L is the sum, and Q_L is the difference of the horizontal and vertically polarized components of the incident light. Similarly, I_B and Q_B are the sum and difference respectively of the scattered light. Note that $\mathrm{DOP} = \frac{Q}{I}$. Consequently, based on Eq. (5.35):

$$\mathrm{DOP_B} \;=\; \frac{m_{21} + m_{22}\,\mathrm{DOP_L}}{m_{11} + m_{12}\,\mathrm{DOP_L}}\,. \tag{5.36}$$

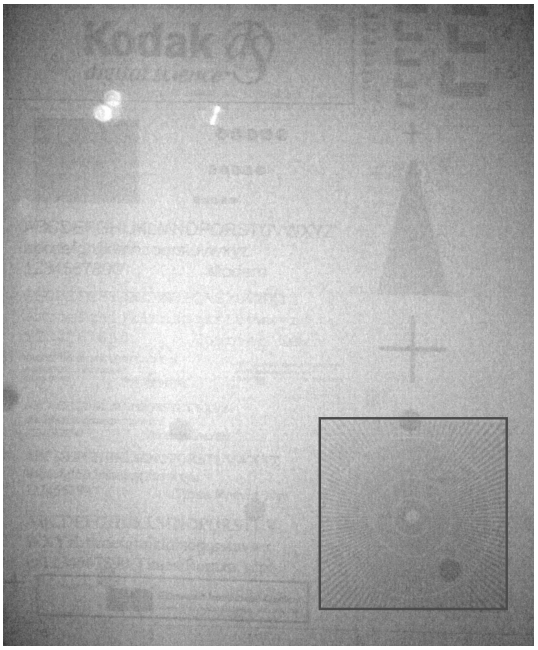
Using the above equation and the measured Mueller matrix data for ocean water [125], we plot $\mathrm{DOP_B}$ versus the scattering angle in Figure 5.19. For comparison, we also plot the behavior for Rayleigh scattering. For low values of $\mathrm{DOP_L}$ (natural light), the curve qualitatively resembles that of Rayleigh scattering. On the other hand, for a completely polarized source (for example, an LCD projector), the curve is flatter, with an average value of 0.8 for backscattering angles. Interestingly, this agrees with the observation made in [119] as well.



(a)



(b) Large Separation



(c) Optimal Separation

Figure 5.10: Optimal sensor-source configuration for flood-lighting. (a) Plot of CQM × W versus d_{LC} for our experimental setup. The tradeoff between contrast and SNR results in a maximum. (b) Large separation (60 cms) results in heavy image noise (c) Optimal separation (40 cms) results in a high contrast, low noise image (zoom into the marked area). Both the frames were captured with the same exposure time.

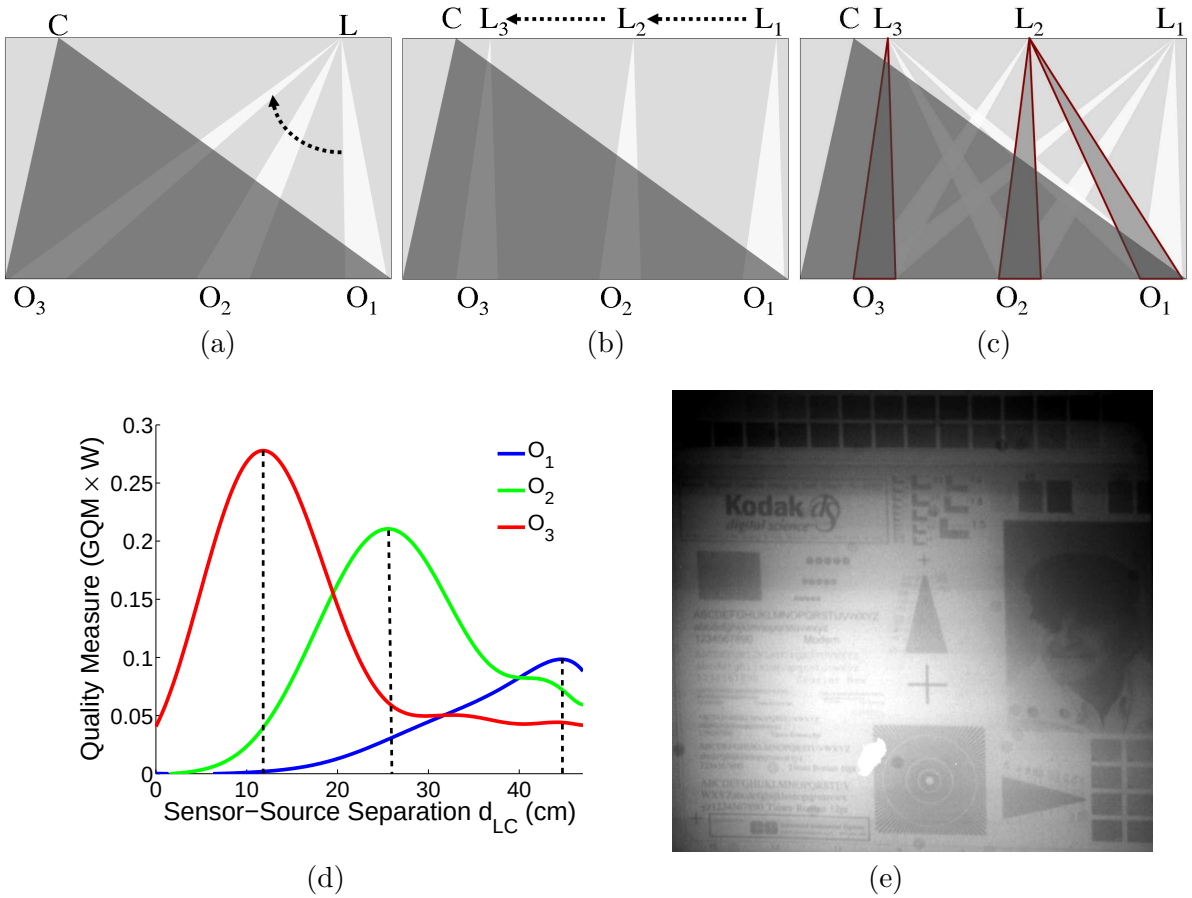


Figure 5.11: We can scan the scene (a) by rotating the source, (b) by translating it, or (c) a combination thereof. (d) Plot of GQM × W versus d_{LC} for different stripe locations O_1, O_2 and O_3, for our setup. We can notice different optimal separations for these stripe locations. (e) A high contrast image resulting from the optimal light stripe scan designed using simulations.

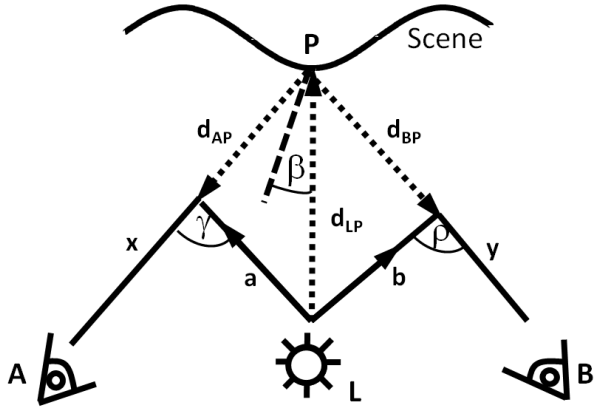


Figure 5.12: Image formation model for Binocular Stereo under scattering media.

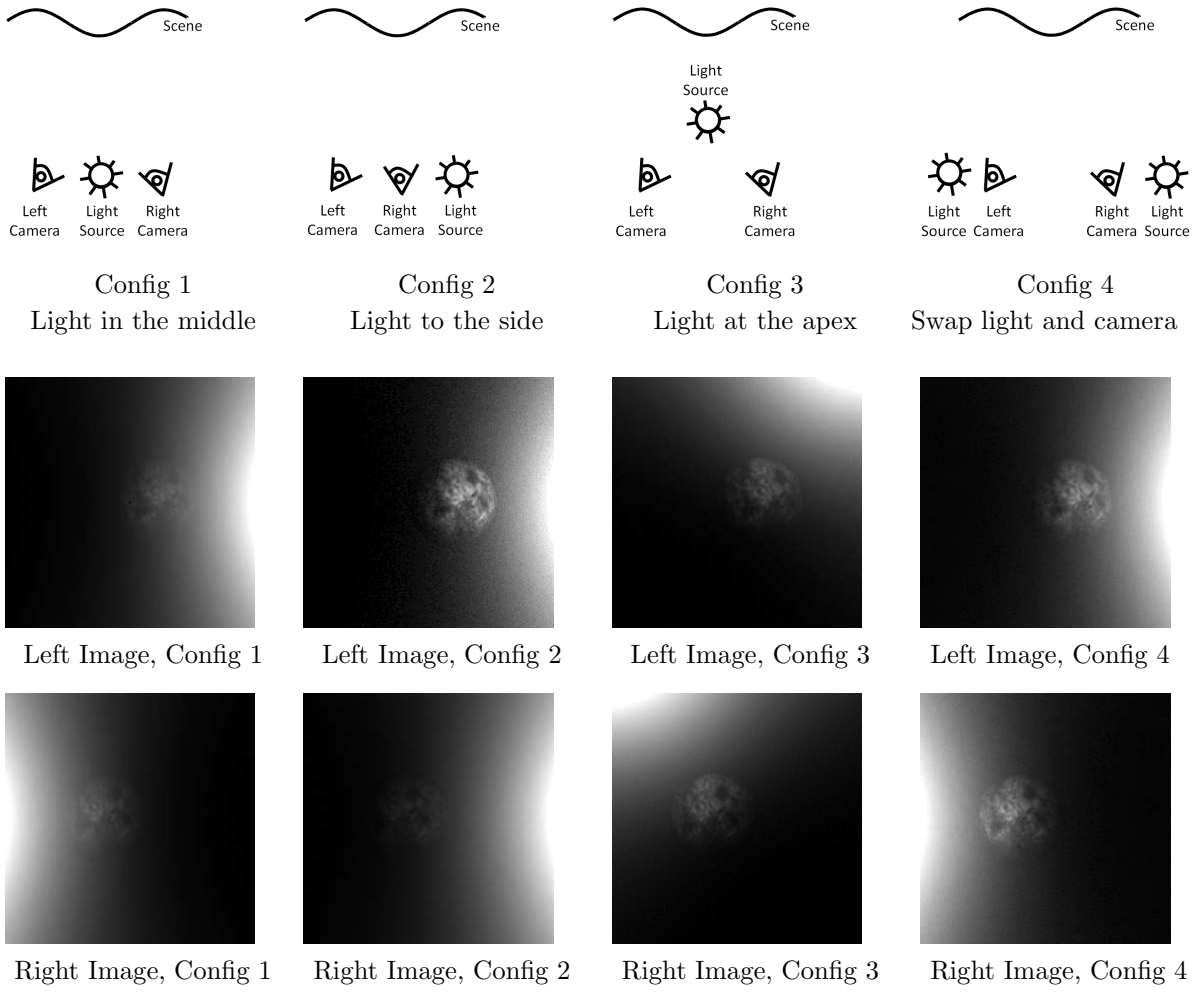
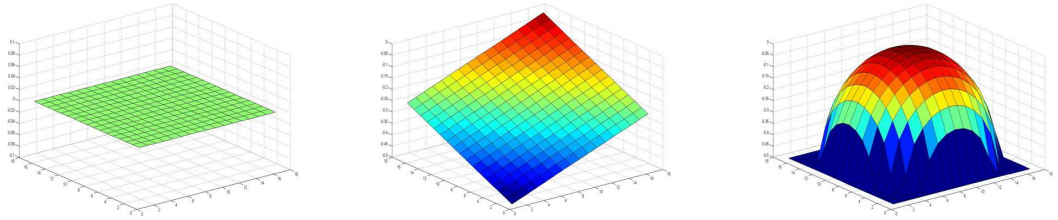
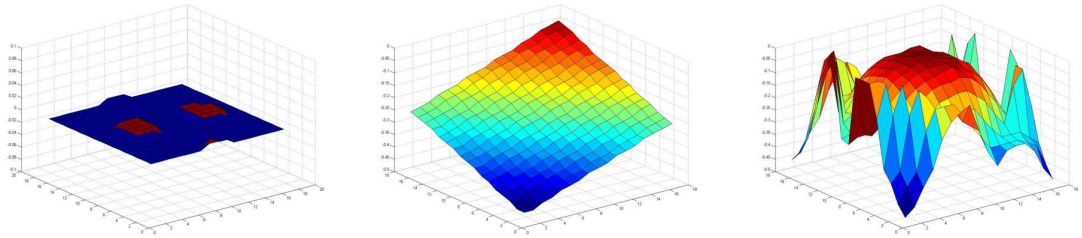


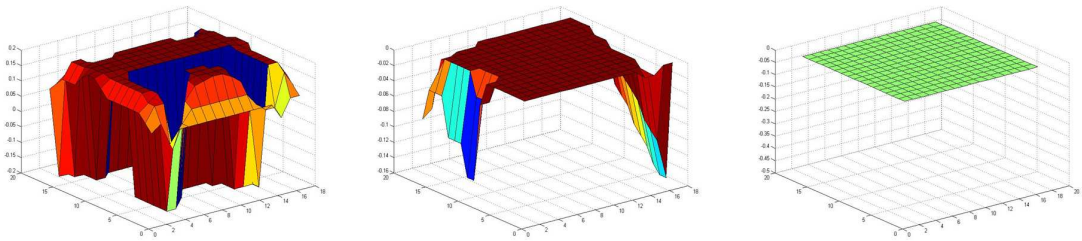
Figure 5.13: Different placements of the source relative to the sensors. The first row of Figure 5.13 illustrates 4 different sensor-source configurations. The second and the third rows show the corresponding images simulated using the single scattering model for the left camera and the right camera respectively.



Flat Plane Ground Truth Inclined Plane Ground Truth Sphere Ground Truth



Shapes recovered using our binocular stereo constraint (Eq. 5.22) - **our method**



Shapes recovered using the original clear air stereo constraint (Eq. 5.20) - **previous method**

Figure 5.14: Depth recovery results using binocular stereo. The ground truth shapes are shown in the top row. Depths recovered using our new matching constraint are shown in the middle row. Depths recovered using the *clear air* matching constraint (Eq. 5.20) are shown in the bottom row. We can notice that the depths recovered without accounting for attenuation and backscatter are incorrect.

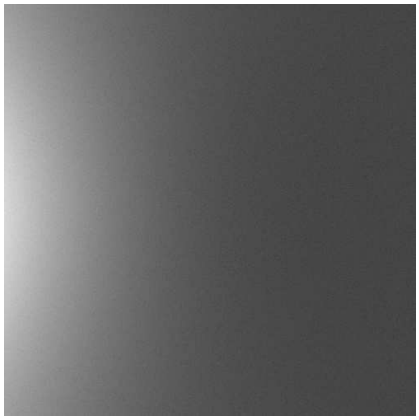
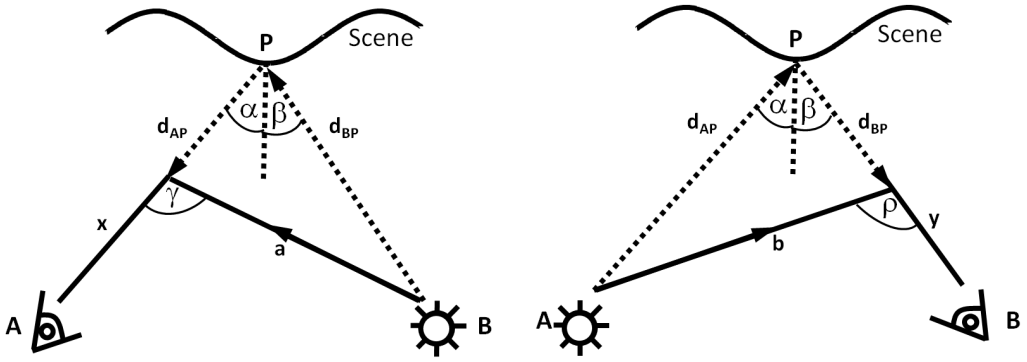
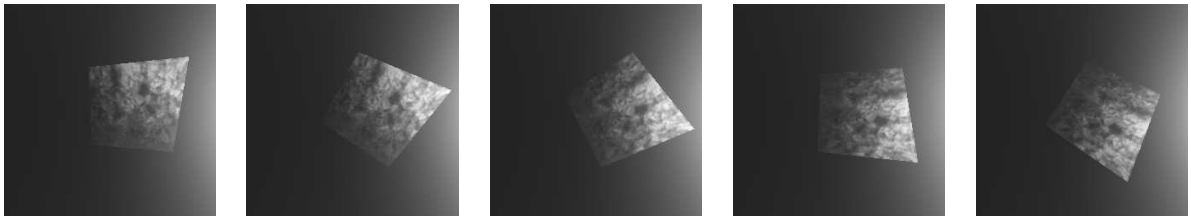


Figure 5.15: A simulated image of the medium (without scene). Such an image can be acquired and used to recover the medium scattering parameters.

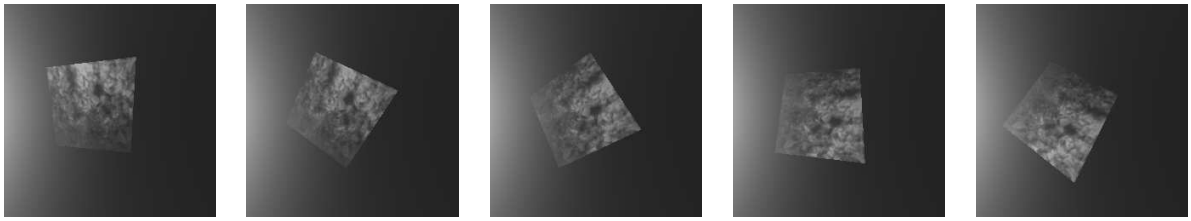


(a) Camera left, Illumination right (b) Camera right, Illumination left

Figure 5.16: Image formation model for Helmholtz stereo setup.

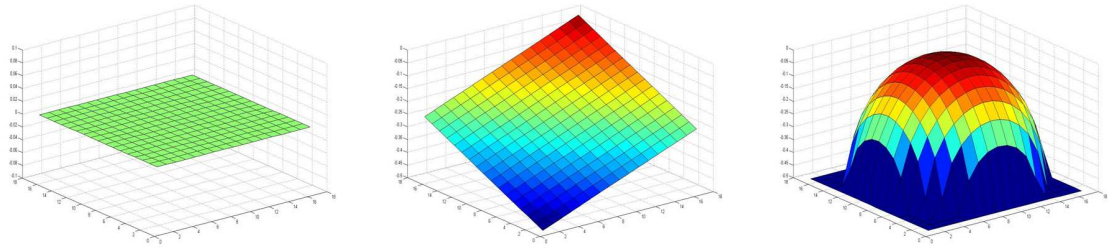


Simulated input images from the left camera

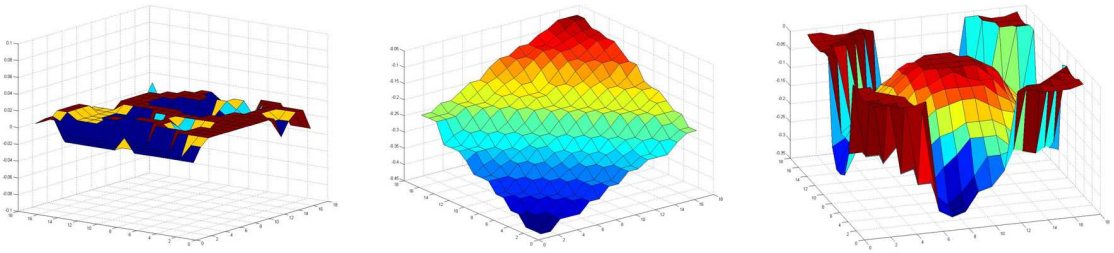


Simulated input images from the right camera

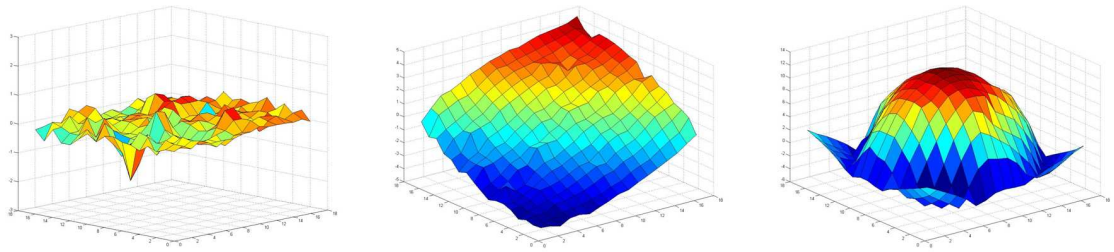
Figure 5.17: Simulated input images for Helmholtz stereo setup. Here we show 5 (out of 18) pairs of simulated images for the inclined plane for different orientations of the source-sensor arm.



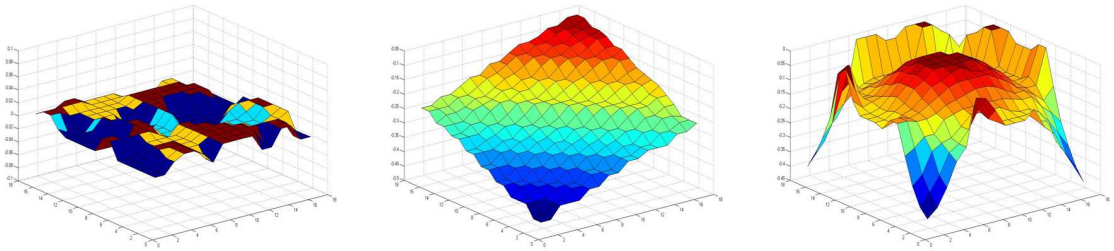
Flat Plane Ground Truth | Inclined Plane Ground Truth | Sphere Ground Truth



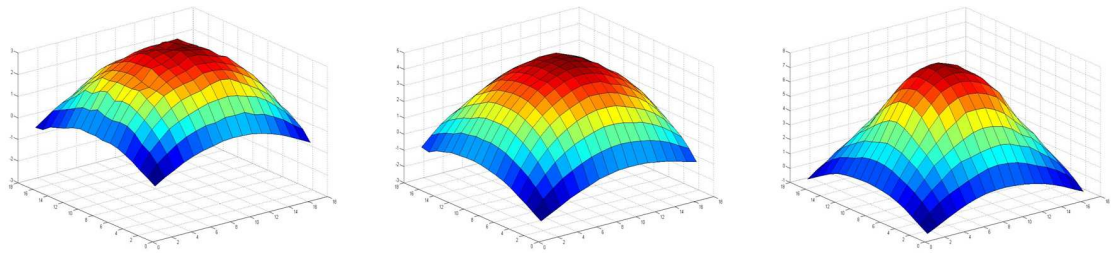
Shapes recovered by minimizing the rank - **our method** (Eq. 5.32)



Shapes recovered by integrating the normals - **our method** (Eq. 5.32)



Shapes recovered by minimizing the rank - **previous method** [134]



Shapes recovered by integrating the normals - **previous method** [134]

Figure 5.18: Depth recovery using Helmholtz stereo. Depths estimated using normals (third row) are smoother and more reliable as compared to the depths estimated by minimizing the ranks (second row). The previous Helmholtz stereo algorithm [134] doesn't account for backscatter, thus producing erroneous depth estimates (fourth and fifth row).

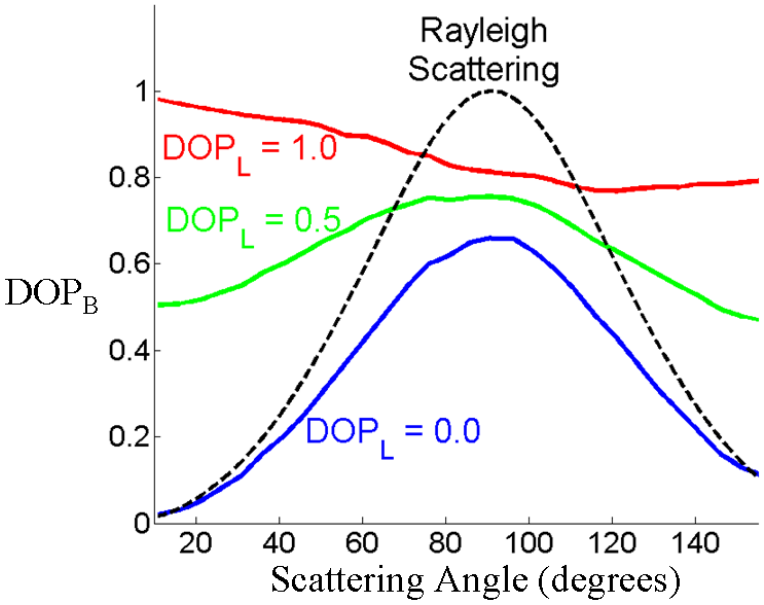


Figure 5.19: Variation of the DOP of the scattered light, $\mathrm{DOP_B}$, on the scattering angle. For low values of $\mathrm{DOP_L}$, the curve qualitatively resembles that of Rayleigh scattering. For a completely polarized source, the curve is flatter, with an average value of ≈ 0.8 for backscattering angles ($> 90°$).

Chapter 6

Real-time rendering of dynamic, non-homogenous volumetric media

http://graphics.cs.cmu.edu/projects/LegendreFluids/

Clear Weather | Snow | Snow and mist

Figure 6.1: **Legendre domain 3D fluid simulation and rendering:** In this example, we have 3000 snow flakes being carried by a wind field (Legendre domain fluid simulation). We add mist to the scene using Legendre domain rendering for participating media. Notice further objects appearing brighter due to the air-light effect, and distant snow-flakes becoming invisible as the mist density is increased. The clear weather Christmas image was downloaded from www.survivinggrady.com/2005_12_01_archive.html

Most of the computer generated imagery today in video games, movies and scientific simulations are of scenes on clear days or nights. However, it is important to simulate the visual effects resulting from global light transport for providing realism in movies and games. Besides digital entertainment, such effects also need to be simulated for training human operators in safety, medical and hazardous situations - pilots landing through fog, soldiers conducting reconnaissance in dusty desert terrain, divers exploring ocean depths, and doctors looking for cancerous tissue. In the absence of such effects, a rendering would

appear unnatural. Also, many applications like games require interactive changes in lighting, view-point and the medium properties. For such applications, it is imperative to achieve these visual effects in real-time.

Brute-force rendering of volumetric media can be prohibitively slow (taking CPU-days or even weeks). On the other end of the spectrum are analytic solutions, which although fast, are possible only for homogenous and static media. However, most of the interesting real world phenomenon such as smoke, dust and fog are non-homogenous and evolving with time. Despite this complexity, most of these physical processes can be represented in a space with significantly lower dimensionality (reduced space) than that required by a finite element methods based computer simulation. This observation has led to the development of a variety of *reduced space methods* in computer graphics. Examples include reduced space methods for fluid simulation [121] and pre-computed radiance transfer for rendering [112]. These methods achieve significant computational speed-ups over spatial-domain finite element based simulations.

Despite significant advances in both the fields of fluid simulation and rendering, these two continue to be addressed separately. In nature, both these processes really interact with each other to provide us with a variety of rich visual experiences. This makes a case for both of them to be studied together to better leverage the advances in both fields. **The goal of this research [41]** is fast rendering of complex visual effects involving dynamic and non-homogenous media, including fluid simulation effects such as particles inserted in turbulent wind-fields. We propose a unified framework for both fluid simulation and rendering in an analytic reduced space. We believe that this is an important first step towards bridging the gap between model reduction for fluid simulation and pre-computed radiance transfer for rendering. Since both fluid simulation and rendering are done in a reduced space, our technique achieves computational speed-ups of one to three orders of magnitude over traditional spatial domain methods. We demonstrate several visual effects resulting from volumetric scattering in time-varying participating media. Figure 6.1 shows frames from a sequence where we add mist and snow-flakes to a clear day scene.

Analytic Space: Most previous reduced space techniques are *data-driven*; a low-dimensional space is constructed from previously observed/simulated data. While achieving considerable speed-ups, these approaches might not generalize to novel instances. We propose a completely analytic reduced space defined by **Legendre polynomials [14] basis**. Such a space has the advantage of generalizing to novel instances while providing computational speed-ups as well. To keep the analysis tractable, we will focus on optically thin media where single scattering is the dominant form of light transport [116, 81]. Under these conditions, the common Legendre polynomial basis for different fields allows us to analytically

solve both the Navier-Stokes and light transport equations in the reduced Legendre space. It turns out that this solution requires us to analyze triple product integrals of Legendre polynomials and their sparsity (see Appendix A at the end of the chapter), similar in spirit to the triple product wavelet integrals for relighting [89].

6.1 Related Work

Fluid simulation: The Navier-Stokes equations for incompressible fluid flow [17] provide a differential model for simulating the density and velocity fields. Explicit analytic solutions to Navier-Stokes equations are hard to obtain and hence, a number of works that employ numerical finite difference methods (FDM) have been proposed [26, 27, 113, 25, 23, 90, 108]. Although simple to implement, such schemes require high spatial resolution to minimize the finite differencing numerical errors, placing serious demands on memory and compromising speed. [122] develop an approach for key-framing of fluid flows that alleviate the discretization errors. However, their approach becomes computationally prohibitive for large grid sizes. A data-driven approach was proposed in [121] to simulate the velocity fields using a reduced dimensional PCA basis. This approach achieves considerable speed-ups and produces impressive results, but at the cost of high memory requirements and lengthy pre-computation. Furthermore, as the authors mention, it is unclear whether the approach generalizes to new fluid flows that are not represented in the pre-computation.

Rendering of participating media: Rendering of participating media requires modeling the intensity fields resulting from volumetric scattering. The intensity field of the participating medium can be rendered by solving the light transport equation [14]. Analogous to fluid modeling, many works that numerically solve the light transport equation based on FDMs have been proposed [62, 60, 94, 22, 101, 71, 100]. As such, many of the issues related to numerical errors must be addressed here as well. While these methods can produce impressive visual effects, they are too slow for interactive applications. Recent hardware-accelerated techniques [21, 99, 45] can significantly decrease the running times of numerical simulations, although they are specialized to particular phenomena.

In addition, the intensity fields depend on the illumination and viewing geometry as well as the scattering properties [81, 47] of the participating medium. Moreover, the lighting, viewpoint and the densities of the medium may change with time. Thus, the pre-computations required are too prohibitive for data driven approaches to be applied to intensity fields. For the special case of homogeneous media, many previous analytic approaches [75, 116, 61, 84] may be used to render the effects of scattering in real-time. However, we the assumption of homogenous medium is too restrictive for several applications. We wish to render non-homogenous and dynamic media, like smoke, dust and

$\mathbf{u}$	velocity field
$u_\diamond$	$\diamond-$component of velocity field
r	density field
$\mathbf{b}$	external force field
$b_\diamond$	$\diamond-$component of force field
E^d	direct transmission intensity field
E^s	scattered intensity field
ν	kinematic viscosity
σ	extinction coefficient
β	scattering coefficient
θ	scattering angle
$\Omega(\theta)$	scattering phase function
$\omega^{\mathbf{d}}$	lighting direction
$\omega^{\mathbf{s}}$	viewing direction

Figure 6.2: Notation used in this section. $\diamond$ stands for either x, y or z.

fog.

6.2 Physical Models for Participating Media

Dynamic and non-homogenous participating media can be characterized by density, velocity and intensity fields, that vary across both space and time. Whereas Navier-Stokes equations for incompressible fluid flow model the evolution of the density and velocity fields over time, the intensity fields are rendered using light transport equations. The time evolution of the velocity field $\mathbf{u}$ is given by [17]:

$$\frac{\partial \mathbf{u}}{\partial t} = -(\mathbf{u}.\nabla)\mathbf{u} - \nu\nabla^2\mathbf{u} + \nabla p + \mathbf{b}, \qquad s.t. \quad \nabla.\mathbf{u} = 0, \tag{6.1}$$

where, ν is the kinematic viscosity, p is the pressure field and $\mathbf{b}$ denotes the external forces (the notation used in this section is given in Figure 6.2). Following [113, 17], Equation 6.1 can be written as:

$$\frac{\partial \mathbf{u}}{\partial t} = \underbrace{P}_{projection}\left(\underbrace{-(\mathbf{u}.\nabla)\mathbf{u}}_{advection} + \underbrace{\nu\nabla^2\mathbf{u}}_{diffusion} + \underbrace{\mathbf{b}}_{forces}\right) \tag{6.2}$$

Here, P is a linear operator which projects a vector field to its divergence free component. Equation 6.2 can be resolved by splitting the right hand side into four sequential steps: (i) advection, (ii) diffusion, (iii) external forces and (iv) projection [113]. Similarly, the time evolution of the density field r is given by:

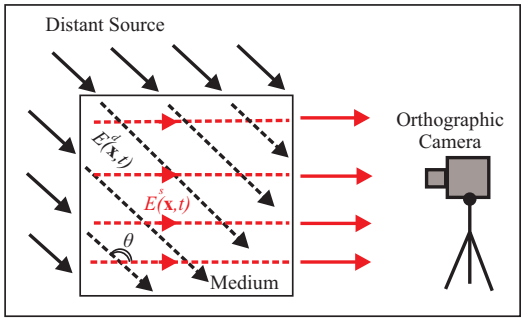


Figure 6.3: The participating medium is illuminated by a distant light source and is viewed by an orthographic camera. Under the single scattering assumption, the intensity field within the medium volume can be split into two sets of light rays: the pre-scattering (direct transmission) intensity field $E^d(x,t)$ and post-scattering intensity field $E^s(x,t)$ (shown using red rays).

$$\frac{\partial r}{\partial t} = \underbrace{-(\mathbf{u}.\nabla)r}_{advection} - \underbrace{\kappa\nabla^2 r}_{diffusion} + \underbrace{-\alpha r}_{dissipation} + \underbrace{S_r}_{source}, \tag{6.3}$$

where, κ is the diffusion constant, α is the dissipation rate and S_r is the source term for density.

Using the density field r, we can render the intensity fields for any configuration of illumination and viewing geometry. In this work, we consider optically thin media where **single scattering** is the dominant form of light transport. Figure 6.3 shows an orthographic camera viewing a participating medium that is illuminated by a distant light source. Then, we can split the intensity fields into two components: the *pre-scattering* (direct transmission) intensity field $E^d(\mathbf{x},t)$, and the *post-scattering* intensity field $E^s(\mathbf{x},t)$. Mathematically, these intensity fields can be written as [14]:

$$\left(\omega^d.\nabla\right) E^d = -\sigma r \cdot E^d \tag{6.4}$$

$$\left(\omega^s.\nabla\right) E^s = -\sigma r \cdot E^s + \beta r \cdot \Omega(\theta) \cdot E^d \tag{6.5}$$

Here, σ and β are the extinction and scattering coefficients respectively and $\Omega(\theta)$ is the phase function. When the camera is outside the medium, the acquired **2D image** of the medium is simply the boundary of the 3D intensity field E^s(Figure 6.3).

6.2.1 Compact Analytic Representation using Legendre Polynomials Basis

The key idea is to represent the 3D spatial variation of the density, velocity and intensity fields using the same analytic basis. We choose to use Legendre polynomials as basis functions. In many situations, natural effects such as mist, outdoor smoke and dust are

Orthogonality	$\int_{-1}^{1} L_i(x)L_j(x)dx = \delta_{ij}$
Derivative	$L_i'(x) = \sum_k c_{ik}L_k(x)$
Integral	$\int L_i(x)dx = \sum_k b_{ik}L_k(x)$

Figure 6.4: Properties of Legendre Polynomials [14].

smooth (low frequency) phenomena, and can be compactly represented by a small number of coefficients. Legendre polynomials are orthogonal, have global support (non-zero over the entire domain), and have analytic derivatives and integrals (Figure 6.4). As a result, they find wide application in mathematical physics literature in conjunction to solving differential equations [14].

A function $f(x)$ can be represented as a linear combination of Legendre polynomials L_k of different orders $f(x) = \sum_k F_k L_k(x)$, where the Legendre domain coefficients $[F_k]$ can be computed analytically as:

$$F_k \;=\; \int_{-1}^{1} f(x)L_k(x)dx\,. \tag{6.6}$$

In 3D, we represent a field $f(x, y, z)$ that is smooth in x-,y- and z-directions as:

$$f(x, y, z) = \sum_{ijk} F_{ijk}L_i(z)L_j(y)L_k(x)\,. \tag{6.7}$$

For notational ease, Equation 6.7 is written as $f(x, y, z) \Leftrightarrow [F_{ijk}]$. The Legendre representations for the various fields are given in Figure 6.5.

6.3 Analytic Operators in Legendre Domain

In this section, we derive the Legendre space formulations for various operators and establish that they are compact, computationally efficient, and completely analytic in nature. For ease of exposition, we illustrate the concepts with $1D$ examples; analysis in $2D$ and $3D$ follows in an exactly similar manner.

Derivative Operator

Observe that spatial derivatives appear both in the Navier-Stokes and the light transport equations (6.2, 6.3, 6.4, 6.5) in the form of gradient and Laplacian operators. Using the property that derivative of a legendre polynomial can be expressed in terms of lower order legendre polynomials (Figure 6.4), we derive the derivative operator in legendre domain,

Field	Spatial $\Leftrightarrow$ Legendre
Density Field	$r \Leftrightarrow [R]$
Velocity Field	$u_\diamondsuit \Leftrightarrow [U_\diamondsuit]$
Divergence free Velocity Field	$\widehat{u}_\diamondsuit \Leftrightarrow [\widehat{U}_\diamondsuit(t)]$
External Force Field	$b_+ \Leftrightarrow [B_+(t)]$
Direct Transmission Intensity Field	$E^d \Leftrightarrow [I^d]$
Scattered Intensity Field	$E^s \Leftrightarrow [I^s]$

Figure 6.5: Legendre representations of various fields, where $\diamondsuit$ stands for x,y or z. In Figures 6.5 and 6.6, sub-scripts and arguments have been dropped for brevity. For example, d and $[D]$ should be read as $d(x, y, z, t)$ and $[D_{ijk}(t)]$ respectively.

Operation	Operand	Result	Complexity
Derivative	$g \Leftrightarrow [G]$	$\frac{\partial}{\partial \diamondsuit} g \Leftrightarrow D_\diamondsuit \cdot [G]$	$O(K^2)$
Integral	$g \Leftrightarrow [G]$	$\int g d\diamondsuit \Leftrightarrow \hat{I}_\diamondsuit \cdot [G]$	$O(K^2)$
Product	$g \Leftrightarrow [G]$ $h \Leftrightarrow [H]$	$g \cdot h \Leftrightarrow M^G \cdot [H]$	$O(K^3)$
Truncation	$[G]$	$[G^T] = T \cdot [G]$	$O(K^2)$
Legendre to Spatial	$[G]$	g	$O(NK)$

Figure 6.6: Legendre Space Operators ($\diamondsuit$ stands for x,y or z). N is the size of the spatial grid. K is the size of legendre coefficient representation.

which is **completely analytic**, and hence, devoid of the numerical errors resulting from the Finite Difference approximation:

$$f(x) = \sum_i F_i L_i(x) \qquad \Rightarrow f'(x) = \sum_i F_i L_i'(x) \tag{6.8}$$

$$\begin{aligned} \Rightarrow f'(x) &= \sum_k \left(\sum_i F_i * c_{ik} \right) L_k(x) \qquad \ldots \qquad (Figure\ 6.4) \\ &= \sum_k F'(k) L_k(x) \end{aligned}$$

where $F'(k) = \sum_i F_i * c_{ik}$. We can write this equation in matrix form, with $[F_k']$ and $[F_i]$ as the coefficient vectors corresponding to the derivative and the original function respectively. The *derivative* operator (x-direction) in Legendre Domain is thus given by the matrix $D_x(i,k) = c_{ik}$:

$$[F_k'] = D_x * [F_i] \tag{6.9}$$

Derivatives in y and z and the integral operator can be defined likewise. Figure 6.6 lists all the legendre space operators that we derive, along with the corresponding time complexity. Given K as the size of legendre space representation $[F_i]$, the matrix-vector multiplication require $O(K^2)$ computations. Building the derivative and integral matrices is a one time operation, and takes $O(K^2)$ time.

Product Operator in Legendre Domain

The advection term in the Navier-Stokes equation (6.2, 6.3) as well as the single scattering equation for rendering (6.4, 6.5) entail multiplication of two fields to compute a third one. This motivates investigating the general problem of multiplying two functions, $h(x) = f(x).g(x)$, where both the functions and the result are represented in the Legendre Basis:

$$f(x) = \sum_j F_j L_j(x) \qquad g(x) = \sum_k G_k L_k(x) \qquad h(x) = \sum_i H_i L_i(x)$$

To compute the i^{th} basis coefficient for the result, we use orthogonality of Legendre Polynomials (see Figure 6.4)

$$
\begin{aligned}
H_i &= \int_{-1}^{1} L_i(x)h(x)dx = \int_{-1}^{1} L_i(x)f(x)g(x)dx \\
&= \int_{-1}^{1} L_i(x)\left(\sum_j F_j L_j(x)\right)\left(\sum_k G_k L_k(x)\right)dx \\
&= \sum_{jk} F_j G_k TI_{ijk}
\end{aligned}
$$

where $TI_{ijk} = \int_{-1}^{1} L_i(x)L_j(x)L_k(x)dx$ is the Legendre Polynomial triple product integral, and can be pre-computed apriori. As with the derivative and integral case, we can write the above equation in matrix form as follows:

$$
[H_i] = M^G * [F_j] = M^F * [G_k] \tag{6.10}
$$

where, $M^G(i,j) = \sum_k G_k TI_{ijk}$ and $M^F(i,k) = \sum_j F_j TI_{ijk}$. Given the size of legendre representations as K, the multiplication matrix has $O(K^2)$ entries. For each entry, $O(K)$ computations are required. Thus, we need $O(K^3)$ computations to build the multiplication matrix and $O(K^2)$ time for the matrix-vector multiplication. Therefore, total time complexity of legendre space multiplication is $O(K^3)$. However, we show that the 3D tensor TI is sparse using the **Legendre Polynomials Triple Product Integrals theorem** (see Appendix A). Using the theorem, we show that approximately $\frac{3}{4}$ of the entries of the TI tensor are exactly zero. We exploit this sparsity to achieve computational speed-ups in the advection and the rendering stages. Indeed, the time required to construct the multiplication matrix can be reduced by a factor of 4 in $1D$ and by $4^3 = 64$ in the $3D$ case.

Lower Order Approximation: Note that multiplying two polynomials of degree K each results in a polynomial of degree $2K$. Therefore, given two functions, each with Legendre representation of size K, the Legendre representation of the product will have size $2K$. For computational savings, it is desirable to keep the size of the Legendre representation constant. To this end, we devise a simple approximation scheme using the **Chebyshev Polynomials** to truncate a given Legendre representation from $2K$ terms to K terms, while keeping the approximation error low under the L_∞ norm (see Appendix B at the end of this chapter). We define the **Truncation Matrix Operator T in legendre space**, such that

$$
\underbrace{[F_i^T]}_{K\times 1} = \underbrace{T}_{K\times 2K} * \underbrace{[F_i]}_{2K\times 1}
$$

where $[F_i]$ is the legendre representation of size $2K$, and $[F_i^T]$ is the corresponding truncated representation of size K. As with derivatives and integrals, truncation requires a matrix multiplication with a time complexity of $O(K^2)$. Building the truncation matrix T is a one time operation requiring $O(K^2)$ operations.

6.4 Fluid Simulation in Legendre Domain

Using the Legendre representations for fields and the operators (derivative, multiplication, truncation), we solve the Navier-Stokes equations (6.2, 6.3) in the Legendre domain. For velocity simulation, we decompose Equation 6.2 into the 4 sequential steps of advection, diffusion, external forces and projection [113]. Now we show how each of these steps can be simulated in the Legendre domain:

Advection

In the spatial domain, the conservation form of the advection equation is given by:

$$\frac{\partial}{\partial t}u_\diamondsuit = -\nabla\cdot(\mathbf{u}u_\diamondsuit) \tag{6.11}$$

$$= -\left(\frac{\partial}{\partial x}u_x u_\diamondsuit + \frac{\partial}{\partial y}u_y u_\diamondsuit + \frac{\partial}{\partial z}u_z u_\diamondsuit\right) \tag{6.12}$$

Subscript $_\diamondsuit$ denotes either x,y or z direction. This form implicitly assumes a divergence free velocity field, i.e. $\nabla\cdot\mathbf{u}=0$.

Legendre space advection equation is then derived by substituting the legendre representations of the fields (Figure 6.5), along with the legendre space derivative and multiplication operators (Figure 6.6) in Equation 6.12:

$$\boxed{\frac{\partial}{\partial t}[U_\diamondsuit(t)] = -A\cdot[U_\diamondsuit(t)]} \tag{6.13}$$

$$\text{where } A = \underbrace{T}_{truncation}\cdot\left(\sum_\diamondsuit \underbrace{D_\diamondsuit}_{derivative}\cdot\underbrace{M^{U_\diamondsuit s}}_{multiplication}\right)$$

$\sum_\diamondsuit(\cdot)_\diamondsuit$ is short-hand for $(\cdot)_x+(\cdot)_y+(\cdot)_z$. For example, $\sum_\diamondsuit D_\diamondsuit\cdot M^{U_\diamondsuit}$ is expanded as $D_x\cdot M^{U_x}+D_y\cdot M^{U_y}+D_z\cdot M^{U_z}$. We update the legendre representations of the velocity field by computing the eigen decomposition of $A = A_E\cdot\bigwedge\cdot A_E^{-1}$ [121]:

$$[U_\diamondsuit(t+\triangle t)] = \left(A_E \cdot e^{\triangle t \cdot \Lambda} \cdot A_E^{-1}\right) \cdot [U_\diamondsuit(t)] \tag{6.14}$$

A similar approach can be used to update the density field as well. Since it uses the multiplication operator, the time complexity of Legendre advection is $O(K^3)$ (Section 6.3), where K is the number of coefficients. In addition to the computational speed-up, using the completely analytic Legendre domain derivative operator reduces the numerical dissipation inherent in the FDM based approximations of the derivative operator (Figure 6.7).

(a) Original Field (b) Legendre Advection (c) Spatial Advection

Figure 6.7: Comparison between Legendre and Spatial domain advection (high intensities signify higher values of the field). Notice, that the field after advection in the spatial domain (c) has lower energy than the field resulting from analytic legendre domain advection (b). Spatial advection results in dissipation of energy due to discretization of the gradient operators. The grid size used for spatial advection was 500^2, while 144 coefficients were used for legendre advection.

Diffusion

For the diffusion step, we solve the implicit form of diffusion equation:

$$\left(I_{N\times N} - \nu \triangle t \, \nabla^2\right) \; u_\diamondsuit\left(\mathbf{x},\; t+\triangle t\right) = u_\diamondsuit\left(\mathbf{x}, t\right) \tag{6.15}$$

where N is the total number of simulation grid voxels. The implicit form of diffusion equation is more stable than the explicit form. However, one drawback of the implicit form is that it requires solving a large system of linear equations. Fortunately, in our case, this issue is addressed by solving the diffusion equation in the reduced legendre space. Once again, we use the legendre representation of the fields and the operators (Figures 6.5 and

6.6) to obtain the **legendre space diffusion equation:**

$$\boxed{\left(I_{K\times K} - \nu\triangle t\ D^2\right)\ [U_\diamondsuit\,(t+\triangle t)] = [U_\diamondsuit\,(t)]} \tag{6.16}$$

where, $D^2 = (D_x)^2 + (D_y)^2 + (D_z)^2$ is the legendre space **Laplacian** operator. Since we solve a $K \times K$ linear system, the time complexity of legendre space diffusion is $O(K^3)$. This is a considerable speed-up over solving the $(N \times N)$ system in spatial domain.

External Forces

External forces are handled by adding their legendre representation (Figure 6.5) to that of the velocity field:

$$\boxed{[U_\diamondsuit(t+\triangle t)] = [U_\diamondsuit(t)] + [B_\diamondsuit]\cdot\triangle t} \tag{6.17}$$

Projection

This step ensures that the velocity field is divergence free, which is required to satisfy mass-conservation. For the projection step, we use the implicit definition of the projection operator P:

$$\nabla^2 q\ =\ \nabla\cdot\mathbf{u} \qquad\qquad \widehat{\mathbf{u}} = P\mathbf{u} = \mathbf{u} - \nabla q \tag{6.18}$$

This step requires solving the following Poisson system of equation for the scalar field q: $\nabla^2 q\ =\ \nabla\cdot\mathbf{u}$. $\widehat{\mathbf{u}}$, the divergence free component of $\mathbf{u}$ ($\nabla\cdot\widehat{\mathbf{u}} = 0$), is then computed by subtracting the gradient of q from $\mathbf{u}$. The Poisson equation can be formulated as a linear system of equations by discretizing the ∇^2 operator in the spatial domain. Analogously, we can define P_L, the projection operator in the legendre space implicitly as follows:

$$\boxed{D^2\cdot[Q] = \sum_{\diamondsuit} D_\diamondsuit\cdot[U_\diamondsuit(t)]} \tag{6.19}$$

$$\boxed{[\widehat{U}_\diamondsuit(t)] = P_L\cdot[U_\diamondsuit(t)] = [U_\diamondsuit(t)] - D_\diamondsuit\cdot[Q]} \tag{6.20}$$

Hence, in legendre space projection step, we need to solve the linear system of equations in the unknown vector $[Q]$ (Equation 6.19), requiring $O(K^3)$ time. As with diffusion, this is a considerable speed-up over solving the $(N \times N)$ linear system in spatial domain. **As an additional advantage**, using the analytic definitions of the derivative operators in all the simulation steps alleviates the numerical errors resulting from spatial finite difference

approximations.

Density Dissipation

For density simulation, Equation 6.3 is solved in the legendre space. The advection, diffusion and source terms are handled in a way similar to velocity simulation. The dissipation term is then solved in the legendre space as follows:

$$\boxed{(1 + \triangle t\alpha) \cdot [R(t + \triangle t)] = [R(t)]} \tag{6.21}$$

Size of the Legendre representation: Figure 6.8 illustrates the time-evolution of 2D density and velocity fields for different sizes of Legendre representations. We start with the same low frequency density and velocity fields and apply the same forces throughout the 3 different Legendre domain simulations. We can observe that more coefficients allow for higher frequencies and vorticities as the density and velocity fields evolve. In Figures 6.9, 6.10 and 6.11, we also provide theoretical and empirical computational complexity of our framework as a function of the size of the Legendre representation (K). A user can use these as a guide for choosing the Legendre representation size that best addresses the demands (speed/ high frequency detail) of a particular application.

6.5 Rendering in Legendre Domain

Rendering requires solving the light transport equations (6.4,6.5) in the Legendre domain using techniques similar to those used for the Navier-Stokes equations.

Direct Transmission intensity field

As earlier, substituting Legendre representations of various fields and Legendre operators (Figures 6.5 and 6.6) into Equation 6.4, we get:

$$\left(\sum_{\diamondsuit} \omega_{\diamondsuit}^{\mathbf{d}} D_{\diamondsuit}\right) \cdot [I^d] \quad = \quad -\sigma T \cdot M^R \cdot [I^d]$$

$$\boxed{\Rightarrow L_{\omega^{\mathbf{d}}} \cdot [I^d(t)] = \mathbf{0}} \tag{6.22}$$

where $L_{\omega^{\mathbf{d}}} = \left(\sum_{\diamondsuit} \omega_{\diamondsuit}^{\mathbf{d}} D_{\diamondsuit} + \sigma T \cdot M^R\right)$.

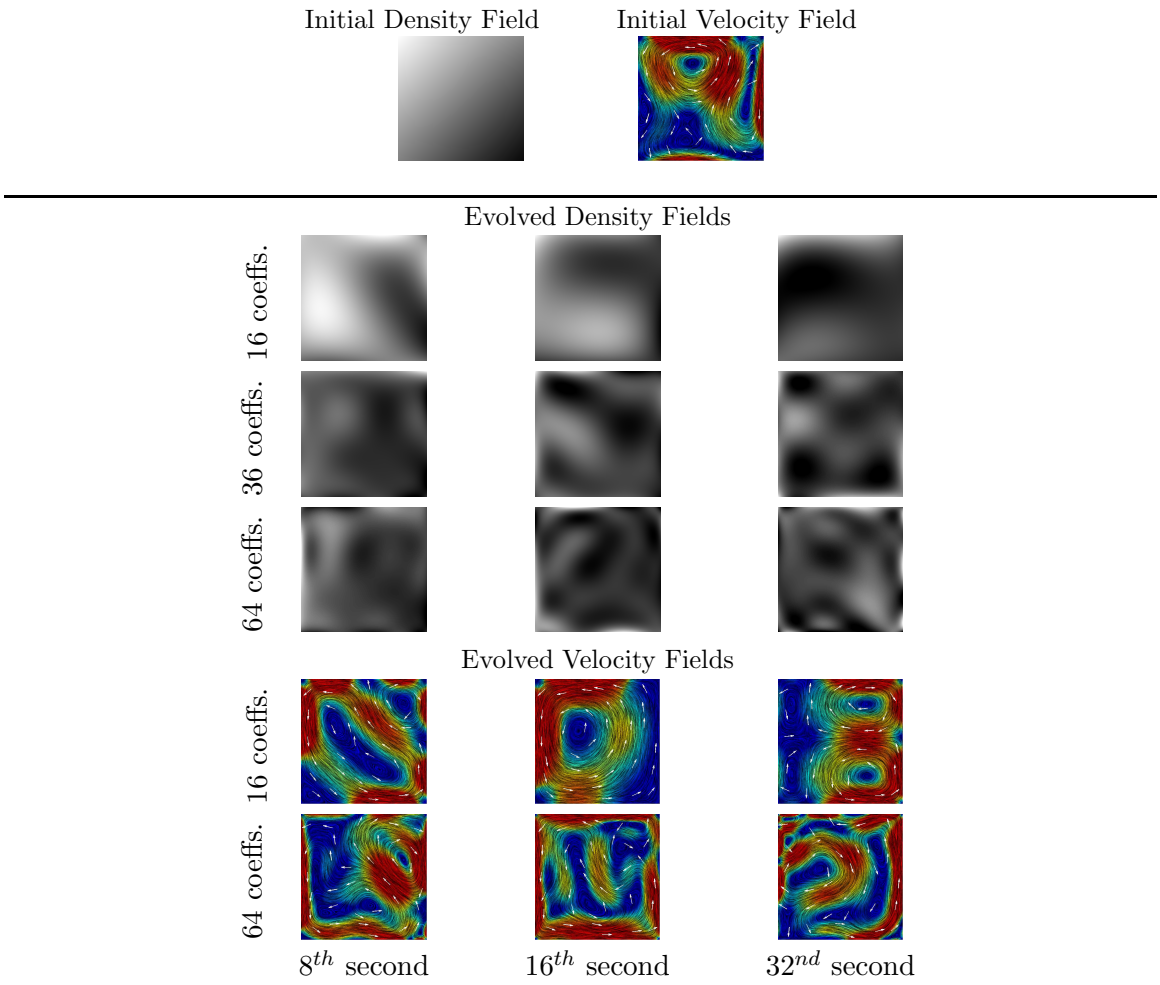


Figure 6.8: **2D Legendre domain Simulation results:** Evolution of density and velocity for different number of Legendre coefficients. More coefficients allow higher frequencies and *vorticities* in the density and velocity fields.

Scattered intensity field

Similarly, we can project Equation 6.5 into the Legendre domain:

$$\left(\sum_{\diamondsuit} \omega^{\mathbf{s}}_{\diamondsuit} D_{\diamondsuit}\right) \cdot [I^s] = -\sigma T \cdot M^R \cdot [I^s] \; + \; \beta\Omega(\theta) \cdot T \cdot M^R \cdot [I^d]$$

$$\boxed{\Rightarrow L_{\omega^{\mathbf{s}}} \cdot [I^s] = \beta\Omega(\theta) \cdot T \cdot M^R \cdot [I^d]} \tag{6.23}$$

For every time step:	
• Update Velocity and Density Fields	
Advection (6.13) and Diffusion (6.16)	$O(K^3)$
Forces/Source (6.17)	$O(K^2)$
Projection(6.19, 6.20)	$O(K^3)$
• Update Intensity Fields	
Direct Transmission(6.22), Scattered (6.23)	$O(K^3)$
• Compute Image (6.24)	$O(SK)$

Figure 6.9: **Legendre domain Rendering algorithm:** K is the size of Legendre space representations and S is the image resolution.

where $L_{\omega^{\mathbf{s}}} = \left(\sum_{\diamondsuit} \omega^{\mathbf{s}}_{\diamondsuit} D_{\diamondsuit} + \sigma T \cdot M^R \right)$.

In the legendre space, both the light scattering equations are thus formulated as linear systems of equations in the unknowns $[I^d]$ (6.22) and $[I^s]$ (6.23). Along with the **boundary conditions**, which can be formulated as additional linear constraints, these systems can be solved in $O(K^3)$ time.

Imagine a camera observing the medium from the outside (Figure 6.3). Then, the **image** recorded is given by the scattering intensity field E_s at the domain boundary:

$$\boxed{E_s(x,y,z,t) = \sum_{ijk} I^s_{ijk} L_i(x) L_j(y) L_k(z)} \tag{6.24}$$

If the image resolution is S, then time-complexity of image computation is $O(SK)$. Note that the image computation step is output-sensitive, and can easily be parallelized. Our Legendre domain modeling and rendering framework is summarized in Figure 6.9.

6.6 Results

Our results show that Legendre polynomials can express a variety of interesting density and force distributions compactly, thereby letting the user manipulate the densities, velocities and forces globally to produce the desired effects.

Particles immersed in dynamic fluid media: Figure 6.12 and Figure 6.1 show simulations of 500 pieces of confetti and 3000 snow-flakes respectively being carried by a wind field simulated using 216 Legendre coefficients each. We can notice vorticities being created in the confetti example due to the turbulent behavior of the wind field. On the other hand, the snow flakes are carried by a more gentle, *breeze-like* wind. Reader is referred to the project web-page [4] for the animation results.

Number of Legendre Coefficients

Spatial Grid Size	16	36	64	144
200^2	500X	250X	75X	10X
300^2	1250X	625X	187X	25X
400^2	2500X	1250X	375X	50X
500^2	5000X	2500X	750X	100X

Figure 6.10: Typical computational speed-ups for 2D simulation and rendering in Legendre domain as compared to the spatial domain.

Number of Legendre Coefficients

Spatial Grid Size	64	125	216
20^3	100X	25X	8X
30^3	1300X	325X	105X
40^3	7500X	1875X	600X

Figure 6.11: Typical computational speed-ups for 3D simulation and rendering in Legendre domain as compared to the spatial domain.

Simulation of smoke and advection of scattering albedos: Figure 6.14 shows a vertically upwards axial impulse applied to a *vase shaped* smoke density field. Since the impulse is applied for a short duration, the density field dissolves towards the end of the simulation. For the first time, we also show advection of the optical properties of the medium (scattering albedos), in addition to the physical properties (densities and velocities), resulting in completely new colors and appearances as the medium evolves under external forces.

Single Scattering based rendering of participating media: We demonstrate the visual effects of both relighting the medium under the single scattering model, and varying the viewpoint and scattering albedos, as the medium evolves under user defined forces. We also show interesting effects of shadowgrams that are cast by the medium on a background plane (Figures 6.13 and 6.14).

3D Visual effects resulting from volumetric scattering in non-homogenous and dynamic participating media: In the examples of Figure 6.15 and Figure 6.1, we add non-homogenous mist to scenes with large depth variation. Notice how distant objects appear brighter due to the airlight [68] effect. Reproducing such effects accurately, particularly for non-homogenous media, is critical for achieving photo-realism while rendering 3D

scenes. Finally, in Figure 6.16, we add non-homogenous and dynamic fog to a clear day fly-through of Swiss Alps.

Computational Speed-ups: Due to the compact representations of fields in the Legendre domain, we can achieve computational speed-ups of one to three orders of magnitude, depending on the number of Legendre coefficients (Figures 6.10 and 6.11). The comparison is made with our implementation of the Stable Fluids [113] algorithm in the spatial domain. However, our technique places a restriction on the size of the simulation time-step; adding higher frequencies will require a progressively smaller time-step owing to stability considerations given by the CFL condition [27]. On the other hand, the Stable Fluids technique can support arbitrarily large time-steps. All our implementation was done in MATLAB on a $3.2GHz$ P-4 PC with 2 GB of RAM.

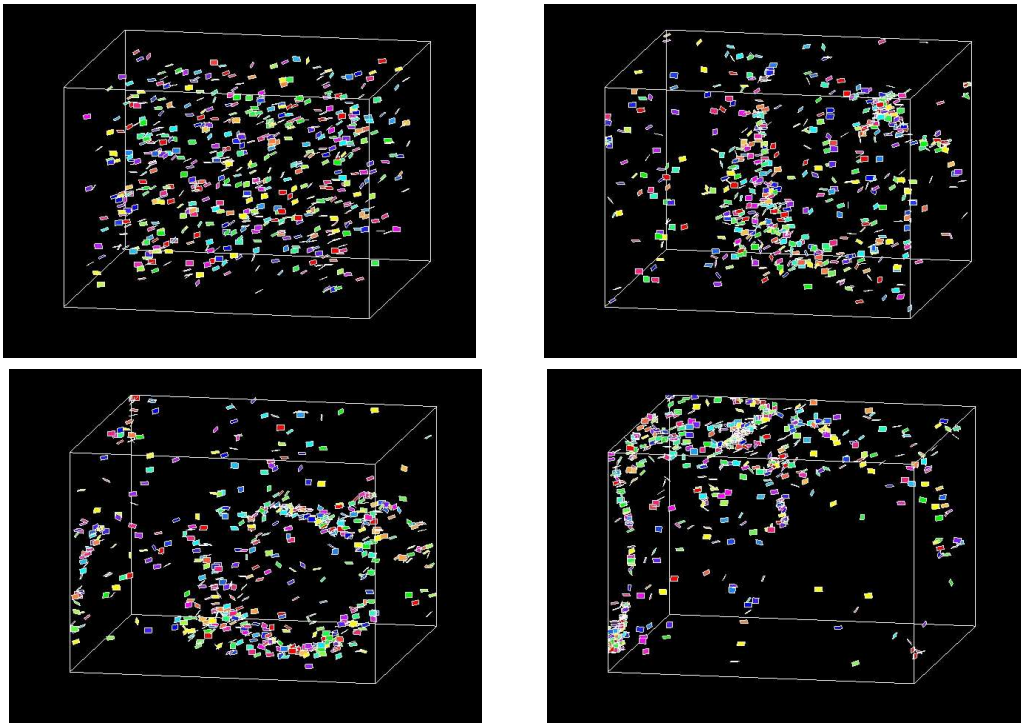
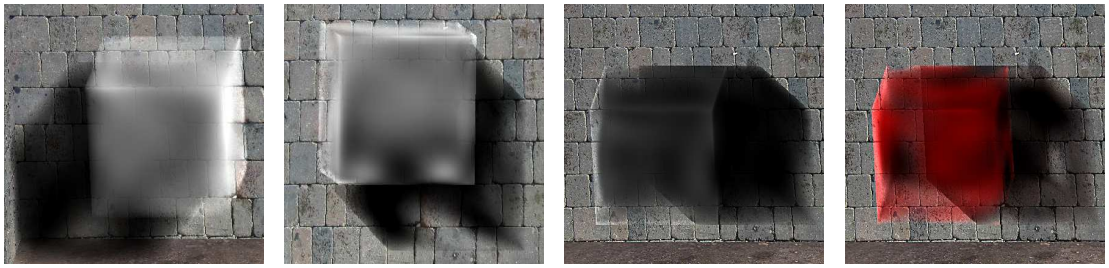


Figure 6.12: **Legendre domain Simulation result:** 500 pieces of confetti being carried by a turbulent wind field simulated using 216 Legendre coefficients.

6.7 Discussion

Our goal is fast rendering of non-homogenous and dynamic participating media. We achieve this by representing the spatio-temporally varying intensity (rendering), as well as density and velocity (simulation) fields in a reduced analytic Legendre space. This results in a sin-



Light Source and View Point Variation Variation in Scattering Properties

Figure 6.13: **3D Legendre domain Rendering:** Here we consider a *smoke-cube* illuminated by distant light source(s). The image is formed at an orthographic viewer observing the scene. Since the whole of our pipe-line is in the reduced Legendre domain, the user can control the view-point, lighting and scattering albedo interactively. Notice the varying shadow-gram patterns on the wall as the smoke evolves. The smoke and the shadow become darker as we decrease the albedo. Colored smoke and shadows can be created by varying the scattering properties differently across the color channels. This example required 64 coefficients for density and velocity, and 216 coefficients for intensity fields.

gle scattering based rendering technique for smooth non-homogenous and dynamic media, a significant improvement over similar techniques which make the severely limiting assumption of homogenous medium densities [116]. We believe this is the first work that provides a unified framework for both modeling and rendering in an analytic reduced space, and hope this can help bridge the gap between model reduction in fluids and pre-computed radiance transfer in rendering.

Limitations: The speed and analytic nature of our technique come at the cost of its limited ability to handle high frequency fluid phenomena. Indeed, using only a global Legendre Polynomials basis offers limited local control and allows only for the **box-boundary conditions**, making it difficult to account for complex effects like local vorticities, turbulence and objects inside the medium. Also, being a **global sub-space method**, it offers low flexibility on the domain boundaries. Also, currently our techniques can not handle objects within the medium, and it assumes single scattering, orthographic viewing and distant lighting.

These limitations can be addressed by augmenting the global Legendre polynomials basis, which capture the majority of the energy of the fluid flow, with a local-support basis such as Haar-Wavelets or spatial voxels, thus accounting for the spatially sparse 'residual energy'. This is similar in spirit to adding local high frequency turbulence, or vorticities [23] to counter the dampening caused by the Stable Fluids semi-Lagrangian technique. Also, high frequency details in a particular dimension can be captured by keeping the full spatial representation and using Legendre expansion in the remainnig directions. Using such *hybrid bases* can provide the desired local control in addition to computational speed-ups,

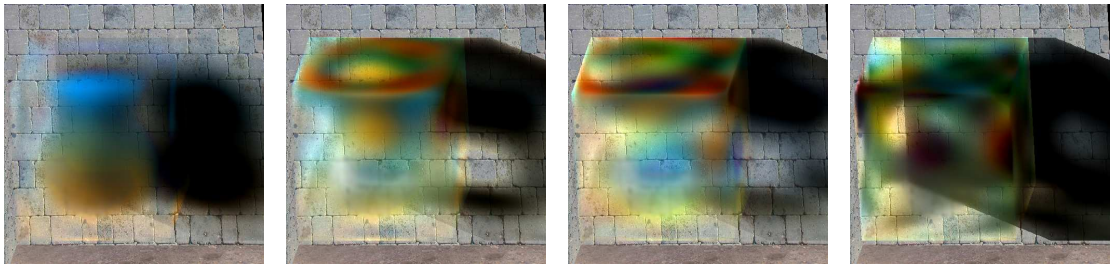


Figure 6.14: **3D Legendre domain simulation and advection of optical properties:** 3D Simulation results for a vertically upwards axial impulse applied to a *vase shaped* smoke density field.. Also, we advect the optical properties of the media (scattering albedos) along with the densities and velocities to create the effect of mixing of different media. This example required 216 Legendre coefficients for density and velocity fields (simulation) and 512 coefficients for intensity fields (rendering).

Clear Weather | Homogenous mist | Non-homogenous mist | Attenuation

Figure 6.15: **Rendering of Non-homogenous participating media:** Our technique can be used to render non-homogenous media as well under the single scattering model efficiently. Here we add mist to a clear weather scene (Images courtesy Google Earth). Non-homogenous density distributions, for example the high mist density over the lake provides for more realism as compared to homogenous mist. Also, notice how distant objects appear brighter due to the air-light effect, whereas distant objects appear darker in the attenuation-only image.

and in our opinion, forms a very promising direction for future research. Since we also make assumptions of single scattering, orthographic viewing and distant lighting, extending our system to perspective viewer and more general, near-field lighting is another research direction worth exploring.

6.8 Appendix A: Sparsity of Legendre polynomials triple product integrals

We have analyzed Legendre polynomials triple product integral. Such integrals arise whenever two functions are multiplied, with both the operands and the result represented in the Legendre polynomial basis. We derive a recurrence relation to calculate these integrals

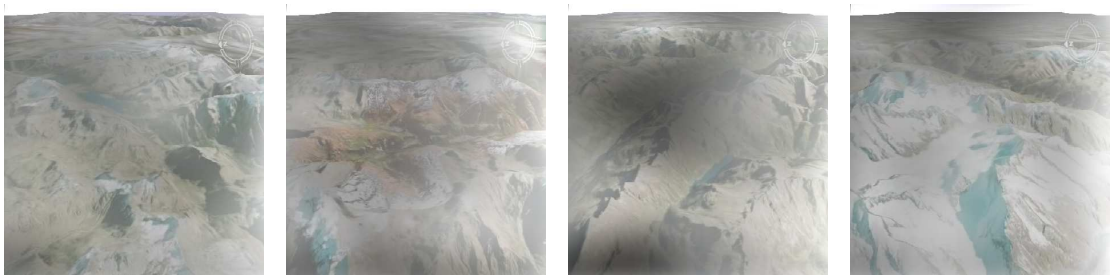


Figure 6.16: Snapshots from a fly-through of Swiss Alps with Non-homogenous and dynamic fog added (Images courtesy Google Earth). Images have been tone-mapped to high-lite the non-homogeneity of the medium. Complete video is included with the supplemental material.

analytically. We also establish the sparsity of the triple product integral tensor, giving the exact closed form expression for the sparsity structure. We show that approximately $\frac{3}{4}$ of the triple-product integral values are exactly zero. We exploit this sparsity to achieve computational speed-ups in addition to that achieved using sparse representation. We believe that this analysis of triple product integrals for Legendre Polynomials can be useful in any scenario where functions are represented in terms of their Legendre polynomials coefficients.

Legendre Polynomial Triple Product Integral Given two functions $f(x)$ and $g(x)$, suppose we want to find the product $h(x) = f(x) * g(x)$, where all the functions are represented in Legendre Polynomials basis:

$$f(x) = \sum_i a_i L_i(x), \quad g(x) = \sum_i b_i L_i(x), \quad h(x) = \sum_i c_i L_i(x),$$

where L_i is the i^{th} Legendre Polynomial.

$$\Rightarrow \sum_k c_k L_k(x) = (\sum_i a_i L_i(x)) * (\sum_j b_j L_j(x))$$

$$\Rightarrow \sum_k c_k L_k(x) = \sum_i \sum_j a_i b_j L_i(x) L_j(x)$$

$c_k's$ are the unknowns here. To simplify, we exploit the orthogonality of Legendre Polynomials. Multiplying by $L_k(x)$ and integrating from -1 to 1:

$$c_k = \tfrac{2k+1}{2} \sum_i \sum_j a_i b_j TI(i,j,k)$$

where $TI(i,j,k) = \int_{-1}^{1} L_i(x) L_j(x) L_k(x) dx$ is defined as the Legendre Polynomial Triple Product Integral.

Given K terms each in the expansion of $f(x)$ and $g(x)$, $h(x)$ will have $2K$ terms. Thus, we have a total of $2K \times K \times K = 2K^3$ computations to calculate all the c_k. **However, looking at the slices** of the TI tensor (3D), one can observe that a large fraction of the

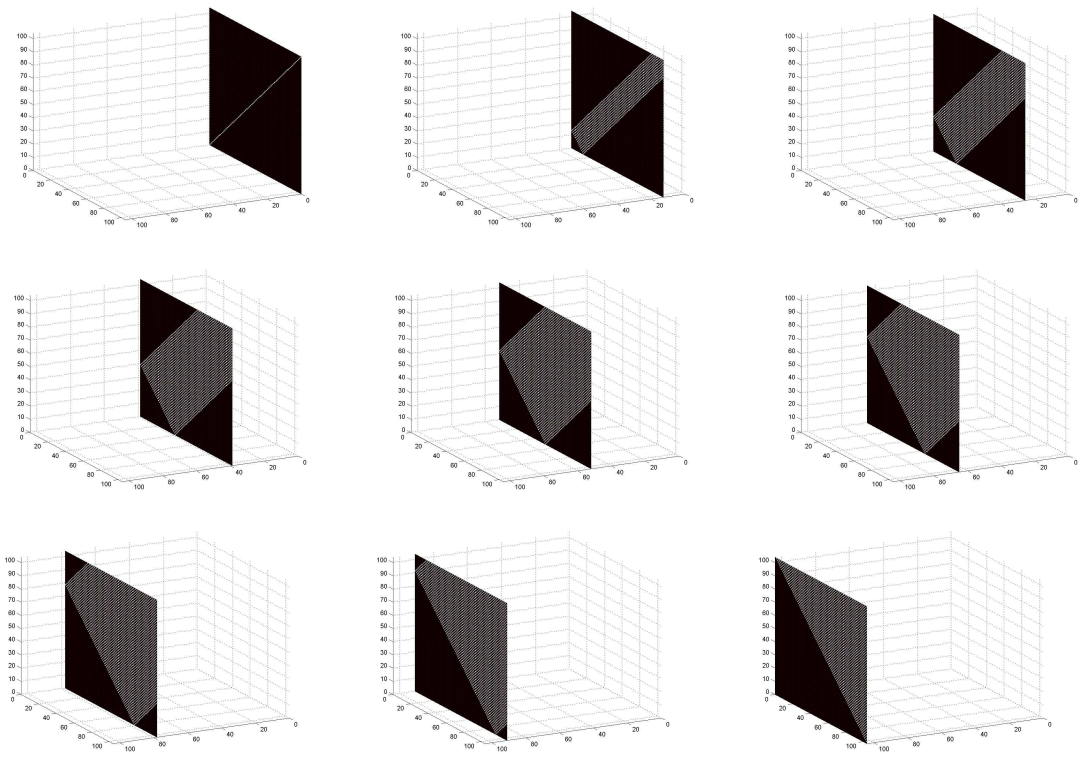


Figure 6.17: Various Slices of the TI tensor. Area in black is zero. We can observe that a significant fraction of the TI tensor is zero. Also, the sparsity of the TI tensor seems to have a particular structure. We establish the sparsity formally, and find a closed form expression for the sparsity structure.

entries are zero (Figure 6.17). In the following, we first establish the **sparsity** of the TI tensor – prove that only a small fraction of elements are non-zero. Next, we also find out the exact distribution of non-zero elements, i.e. find the closed analytical form of the function $N(i, j, k)$, which returns 1 if the element is non-zero, and 0 otherwise.

Previous Work Ng et al [17] provide an analysis of triple product integrals for basis like Haar Wavelets, Spherical Harmonics and Fourier Series. However, to the best of our knowledge, there is no such previous work on Legendre Polynomials. Legendre Polynomials form a system of basis polynomials with a **wide support**, as compared to Haar Wavelets which provide only **compact support**. Thus, in practise, a function can be represented using a relatively small number of Legendre Polynomials. We believe that a thorough analysis of triple product integrals for Legendre Polynomials would be useful, given the generality of the problem (multiplication of two functions in Legendre Coefficients domain).

Recurrence relation for the triple integral

As earlier, $TI(i,j,k) = \int_{-1}^{1} L_i(z)L_j(z)L_k(z)dz$. We will now try to calculate the triple integrals by formulating a recurrence relation for the same:

$$
\begin{aligned}
TI(i,j,k) &= \int_{-1}^{1} \underbrace{L_i(z)L_j(z)}_{Term1}\underbrace{L_k(z)}_{Term2}dz \qquad \ldots\ldots \qquad \text{(Integration By Parts)} \\
&= \left[L_iL_j\int L_k\right]_{-1}^{1} - \int_{-1}^{1}(L_iL_j)'\int L_k \\
&= \frac{1}{2k+1}\left[L_iL_j(L_{k+1}-L_{k-1})\right]_{-1}^{1} - \frac{1}{2k+1}\int_{-1}^{1}(L_iL_j)'(L_{k+1}-L_{k-1}) \\
&= 0 - \frac{1}{2k+1}\int_{-1}^{1}(L_i'L_j + L_iL_j')(L_{k+1}-L_{k-1}) \\
&= -\frac{1}{2k+1}\int_{-1}^{1}\left(L_j\sum_{r=1}^{\lceil\frac{i}{2}\rceil}(2i+3-4r)*L_{i+1-2r} + L_i\sum_{r=1}^{\lceil\frac{j}{2}\rceil}(2j+3-4r)*L_{j+1-2r}\right)*(L_{k+1}-L_{k-1}) \\
&= -\frac{1}{2k+1}\sum_{r=1}^{\lceil\frac{i}{2}\rceil}(2i+3-4r)*[TI(i+1-2r,j,k+1) - TI(i+1-2r,j,k-1)] \\
&\quad -\frac{1}{2k+1}\sum_{r=1}^{\lceil\frac{j}{2}\rceil}(2j+3-4r)*[TI(i,j+1-2r,k+1) - TI(i,j+1-2r,k-1)]
\end{aligned}
$$

Thus, we have expressed TI's in terms of a summation of TI's of lower order. We can calculate the $3D$ tensor of TI's using Dynamic Programming in time $O(K^4)$. We can also exploit the symmetry $TI(i,j,k) = TI(i,k,j) = TI(j,k,i) = \ldots$ to make the computations fast and numerically stable.

Sparsity of the triple integral

We use the recurrence relation derived above to prove the sparsity of the TI tensor, as given by the following **Legendre Polynomials Triple Product Integral** theorem:

Theorem 1. $TI(i,j,k) = 0$ if either of the following two conditions hold:

- The triplet (i,j,k) doesn't satisfy **triangle inequality**, i.e. either $i+j < k$, or $i+k < j$ or $j+k < i$
- $(i+j+k)\ mod\ 2 \neq 0$.

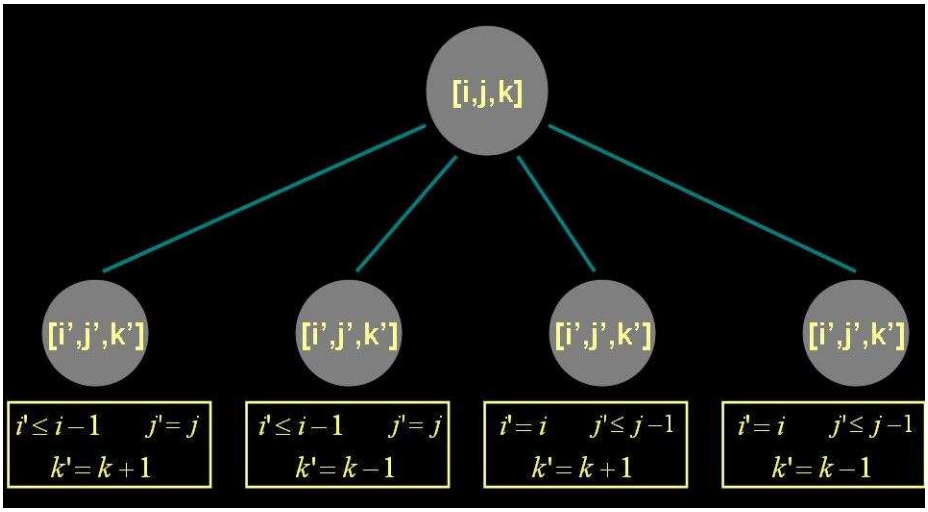


Figure 6.18: First Level of the recurrence tree

Proof. We will prove the two cases separately:

- **The triplet (i,j,k) doesn't satisfy triangle inequality:** Let the triplet (i, j, k) doesn't satisfy triangle inequality. Without loss of generality, lets assume that $i + j < k$. Consider the recurrence tree for $TI(i, j, k)$. The first level is shown in Figure 6.18. We will first prove that if the root triplet (i, j, k) doesn't satisfy triangle inequality, then neither of the triplets in the corresponding sub-tree does.

  Let a child of $TI(i, j, k)$ be $TI(i', j', k')$. We consider two cases:

  - $k' = k + 1$: Since either $i' < i, j' = j$ or $i' = i, j' < j$, we have $i' + j' < k'$
  - $k' = k - 1$: Since either $i' <= i - 1, j' = j$ or $i' = i, j' <= j - 1$, we still have $i' + j' < k$

  Hence, given a root triplet (i, j, k) not satisfying triangle inequality, neither of the triplets in the corresponding sub-tree does.

  Now we look at the leaves of such a sub-tree, or the **base cases**. We stop recursing when one of the indices (i,j or k) becomes 0. This is when we have reached a leaf of the tree. Let the triplet at a leaf be i_b, j_b, k_b. Without loss of generality, let $i_b = 0$. Now, $i_b + j_b < k_b$ (by above argument). Since $i_b = 0$, we have $j_b \neq k_b$. Thus, the value at the leaf is given by:

$$TI(i_b, j_b, k_b) = \int_{-1}^{1} L_0(x) L_{j_b}(x) L_{k_b}(x) dx$$
$$\Rightarrow TI(i_b, j_b, k_b) = \int_{-1}^{1} L_{j_b}(x) L_{k_b}(x) dx = 0$$

So, if $i + j < k$ for the root of the recurrence tree, $TI = 0$ for all the leaves in the sub-tree. Since the value at the root is a linear combination of all the leaves, the value at the root is zero too.

- $(i + j + k)\ mod\ 2 \neq 0$**:** Consider the recurrence tree for $TI(i, j, k)$ (Figure 6.18). The sum of indices of all the children of $TI(i, j, k)$ is $(i+j+k)+(2-2r)$ or $(i+j+k)-2r$, for a given positive integer r. The same will hold for all the recursion sub-trees as well. This implies that like the *Triangle Inequality-ness*, the *even-ness* of the sum of indices $(i + j + k)$ is preserved into the sub-tree, starting from the root. The sum will be even for all the nodes in the tree if it is even for the root and will be odd if it is odd for the root.

  Now we again look at the leaves of the tree, or the **base cases**. We stop recursing when one of the indices (i,j or k) becomes 0. If the sum of indices of the root of the tree is odd, sum of indices for all the leaves will be odd too (from above). Since one of the indices is zero, the other two won't be the same, and hence, as above, values at all the leaves in the tree would be zero. Thus, the value at the root, $TI(i, j, k) = 0$ if $(i + j + k)\ mod\ 2 \neq 0$.

□

Using the above theorem, the following result can be derived:

Result 1. Let $1 \leq i, j, k \leq K$. The total number of TI(i,j,k) is K^3. The number of **non-zero** TI's $\leq \lceil \frac{1}{4}k^3 + \frac{3}{8}k^2 + \frac{1}{4}k \rceil$.
It is easy to derive the above number by using Theorem 1 and simple counting. **This result implies that only about $\frac{1}{4}$ of the TI entries are non-zero.**

6.9 Appendix B: Polynomial truncation scheme using Chebyshev polynomials

In many settings, we are interested in approximating a polynomial by a lower degree polynomial, while keeping the approximation error low. For example, if we use K terms for the Legendre expansion of two polynomials $f(x)$ and $g(x)$, then the product polynomial will have $2K$ terms. For a variety of computational considerations (memory, speed), it would be desirable to keep the degree of the product polynomial the same as the operands, i.e. K.

We derive a truncation scheme to approximate a polynomial with a lower degree polynomial, while keeping the approximation error low under the L_∞ norm. We use the Chebyshev

$T_0(x)$	1
$T_1(x)$	x
$T_2(x)$	$2x^2 - 1$
$T_3(x)$	$4x^3 - 3x$
$T_4(x)$	$8x^4 - 8x^2 + 1$
$T_5(x)$	$16x^5 - 20x^3 + 5x$
$T_6(x)$	$32x^6 - 48x^4 + 18x^2 - 1$

Figure 6.19: First few Chebyshev Polynomials

polynomials to derive our truncation scheme. Chebyshev Polynomials are a set of orthogonal polynomials, and are denoted by $T_n(x)$. First few Chebyshev polynomials are given in Figure 6.19.They are defined using the following recurrence:

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \qquad \text{with } T_0(x) = 1,\ T_1(x) = x$$

We present empirical results which suggest that the approximation error is quite low, even for fairly low degree approximations. For computing a product of two polynomials in the Legendre domain, this scheme can help in keeping the degree of the product the same as that of the operands.

Truncation Scheme

Chebyshev polynomials find wide use in approximating polynomials with a lower degree one. In particular, the polynomial

$$p_{n-1}(x) = x^n - 2^{1-n}T_n(x)$$

is the best $n-1$ degree approximation for $f(x) = x^n$ on the interval $[-1, 1]$ under the L_∞ norm, with the maximum deviation being 2^{1-n}. We call $p_{n-1}(x)$ as the *degree* $n-1$ *mini-max polynomial*, as it minimizes the maximum deviation.

We extend the result by coming up with an approximation for the function $f(x) = x^{n+1}$ defined on the interval $[-1, 1]$ with a polynomial of degree k, for any k, $0 \leq k \leq n$. It is a useful result since we can approximate x^{n+1} with a polynomial of any lower degree.

Suppose we want a degree k approximation for x^n with $0 \leq k \leq n-1$. We start with the expression for $p_n(x)$, the degree $n-1$ approximation. Since, it is a degree $n-1$ polynomial, it can be written as:

$$p_n(x) = c_{n-1}x^{n-1} + c_{n-2}x^{n-2} + \ldots + c_m x^k + \ldots + c_0$$

Now, we replace x^{n-1} with degree $n-2$ mini-max polynomial in the above expression, which in turns gives us a degree $n-2$ approximation for x^n. Similarly, we cascade down, each time replacing the leading power of x in successive approximations with the one-lower degree mini-max polynomial. In each step, we reduce the degree by one; we repeat until we have a degree k approximation.

Note that although we don't have any theoretical error bounds, we present empirical results which suggest that the approximation error is quite low (Figure 6.20) , even for fairly low degree approximations.

Truncation as a matrix operation

In this section, we formulate the degree truncation operation as a matrix multiplication operation. We represent a polynomial $p(x)$ by a vector P of its coefficients. Let the degree k polynomial approximation of $f(x) = x^n$, according to our scheme, is given by $t_k^n(x)$, and the corresponding coefficient vector by T_k^n (size $k+1$).

Then, given a degree n polynomial $p_n(x)$, and its coefficient vector P_n, we can calculate $p_k(x)$ (P_k), its degree k approximation ($k < n$) as a matrix multiplication:

$$\underbrace{P_k}_{k+1\times 1} = \underbrace{M_{nk}}_{k+1\times n+1} * \underbrace{P_n}_{n+1\times 1}$$

where,

$$M_{nk} = \begin{bmatrix} T_k^n & T_k^{n-1} & \ldots & T_k^{k+1} & I_{k\times k} \end{bmatrix} \quad , \quad I_{k\times k} \text{ is a } k \times k \text{ identity matrix.}$$

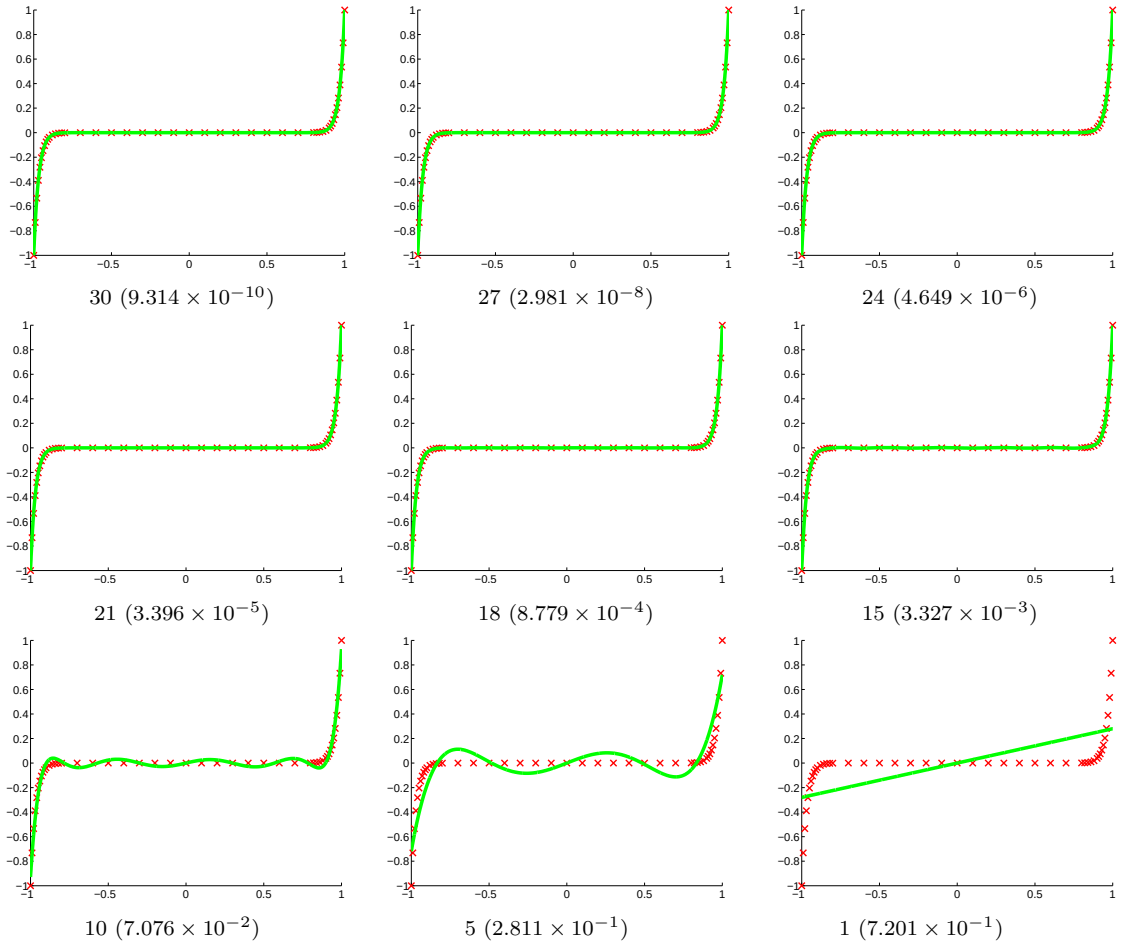


Figure 6.20: Lower Degree Polynomial Approximations of x^{31} with our algorithm in the interval $[-1, 1]$. Green plot is the approximation, and red crosses indicate the original function. Sub-Captions of plots are the degrees of approximation m. The numbers in parentheses are the maximum approximation error (L_∞ error norm) over the interval $[-1, 1]$. We can observe that for $m > 15$, the two plots are almost indistinguishable, which is remarkable, since we are approximating x^{31}. This establishes the accuracy of the algorithm in an empirical sense. For $m < 15$, we start noticing some error.

Chapter 7

Future Directions

In this thesis, we have shown that it is possible to build simple computational models for global light transport. By using these models, combined with active control of illumination, we have made a variety of computer vision techniques robust to the effects of global illumination. On the other hand, again, by simplifying the light transport models and using active illumination, we have used global light transport for recovering scene/material information as well. We believe that this is an important step towards increasing the scope of computer vision systems to challenging real world settings.

The study of global light transport effects in computer vision is still a relatively nascent field. We believe that there are numerous avenues for future research. A few of them are listed below.

Separating global light transport further into inter-reflections and scattering: The technique given in [87] efficiently separates the direct and the global components of light transport. A future goal is to decompose global light transport further into it's constituent modes of inter-reflections, sub-surface scattering and volumetric scattering. Both inter-reflections and sub-surface scattering result in blurring of the incident illumination. However, we have observed that the blurs associated with the two processes are different. The blur at each scene point could be easily measured by illuminating the scene with only a few illumination patterns. The measured blurs could be used to classify the dominant mode of light transport at each scene point further into sub-surface scattering vs. inter-reflections. Such a classification provides more information than the direct-global separation. For instance, it could be used to infer a qualitative geometric model (from inter-reflections) and a translucent material map (from sub-surface scattering) of the scene.

Time varying global light transport: So far, we have considered global light transport as a static phenomenon. However, real world scenes evolve with time. Most of the objects of

interest to humans, such as fruits, plants, skin, biological tissues change material properties. Many physical processes such as weathering change the geometric lay-out of scenes as well, either at a macro or micro structure scale. The variation in appearance of such scenes can be captured by modeling and acquiring time varying global light transport. This can be thought of as generalizing time varying models for surface appearance [40, 117]. Such models will be useful in computer graphics for rendering time varying appearance of a wide range of scenes.

Photometric and Helmholtz stereo under global light transport: Active shape recovery techniques can be classified three ways into triangulation based techniques using structured light, shape from illumination defocus and photometric methods such as photometric and Helmholtz stereo [134]. Each class of techniques has a different set of merits and demerits, and is best suited in different scenarios. In this thesis, we have made the first two resistant to errors due to global light transport. As with the first two techniques, we plan on building simple aggregate models of global illumination which could easily be separated from the shape cues used by photometric and Helmholtz stereo techniques.

Shape recovery for transparent objects: Light transport through transparent objects follows the laws of refraction. Since most traditional shape-recovery techniques are designed for objects that scatter light, they can not be used to reconstruct the shape of transparent objects. Recently, Kutulakos et al [69] presented a technique for reconstructing transparent objects by reconstructing a number of individual 3D light paths, that intersect at a common vertex lying on the object. The number of camera views required grows with the complexity of the object. For instance, reconstructing an object with 2 refractions requires at least 3 camera views if the refractive index is known, and at least 4 views if the refractive index is unknown. In practice, however, many more views (~ 7) are required for robustness, since each point on the surface is reconstructed individually.

We observe that a large collection of light rays can potentially intersect on the object, forming a light network. Instead of reconstructing each light path individually, a large network of 3D rays can be reconstructed together. This offers two advantages. First, it decouples the number of required camera views from the number of light paths needed. For the example mentioned above, an object with 2 refractions can be reconstructed using only 2 camera views, even in the case of unknown refractive index. Second, a light network can potentially pass through a large cluster of points on the surface of the object. Each path in the network constrains the degree of freedom of the network. Interestingly though, the total number of degrees of freedom of the network remains the same irrespective of the number of light paths. By reconstructing the complete network, all the light paths (therefore all the points) can be reconstructed together, thus significantly improving the quality of

reconstruction. This is akin to solving for a small number of unknowns with a large number of equations. Since the robustness is coming from considering the light network, we don't need to take more camera views than the minimum required theoretically.

Machine vision algorithms for material recognition and computational models for sub-surface scattering: While there is significant research effort in computer vision for object recognition and scene understanding, the problem of material recognition has received little attention. We believe that recognizing materials in images is an important stepping stone towards the goals of object recognition and scene understanding. Several objects of interest to humans, such as food, plants and skin are translucent. Perception of translucency results from the physical process of sub-surface scattering. Understanding sub-surface scattering can provide useful cues for recognizing translucent materials in images.

We wish to build machine vision systems for measuring sub-surface scattering in objects. Unlike surface reflectance, which depends only on the local shape, sub-surface scattering in an object is a complex function of both local and global shape, along with material properties, illumination and imaging conditions. If these factors could be untangled, sub-surface scattering could prove to be a rich source of information about the scene. By building factorizable computational models, each of these factors could be recovered by controlling the others. For instance, we plan to investigate the possibility of recovering shape and material properties from measuring sub-surface-scattering using controlled illumination. Such systems would have wide ranging applications in food inspection, digitizing cultural heritage, medicine, surgery and cosmetics industry.

Bibliography

[1] Office of the future web-page. *http://www.cs.unc.edu/Research/stc/index.html*.

[2] Webpage for Chapter 2. *http://graphics.cs.cmu.edu/projects/DefocusGlobal/*.

[3] Webpage for Chapter 4. *http://www.cs.cmu.edu/∼ILIM/projects/LT/dilution/*.

[4] Webpage for Chapter 6. *http://graphics.cs.cmu.edu/projects/LegendreFluids/*.

[5] E. H. Adelson. On seeing stuff: The perception of materials by humans and machines. *Proc. SPIE*, 4299:1–12, 2001.

[6] S. Antyufeev. *Monte Carlo Method for Solving Inverse Problems of Radiative Transfer*. Inverse and Ill-Posed Problems Series, VSP Publishers, 2000.

[7] B. Atcheson, I. Ihrke, W. Heidrich, A. Tevs, D. Bradley, M. Magnor, and H. Seidel. Time-resolved 3D capture of nonstationary gas flows. *ACM Trans. Graph.*, 27(3), 2008.

[8] M. G. Bloj, D. Kersten, and A. Hurlbert. Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature*, 402:877–879, 1999.

[9] E. Boss and W. S. Pegau. Relationship of light scattering at an angle in the backward direction to the backscattering coefficient. *Applied Optics*, 40 (30):5503–5507, 2001.

[10] J.-Y. Bouguet and P. Perona. 3d Photography on Your Desk. In *In Proc. IEEE ICCV*, 1998.

[11] F. M. Caimi, F. R. Dalgleish, T. E. Giddings, J. J. Shirron, C. Mazel, and K. Chiang. Pulse versus CW laser line scan imaging detection methods: Simulation results. *In Proc. IEEE OCEANS*, pages 1–4, 2007.

[12] D. Caspi, N. Kiryati, and J. Shamir. Range imaging with adaptive color structured light. *IEEE PAMI*, 20(5):470–480, 1998.

[13] M. K. Chandraker, F. Kahl, and D. J. Kriegman. Reflections on the generalized bas-relief ambiguity. In *Proc. CVPR*, 2005.

[14] S. Chandrasekhar. *Radiative Transfer*. Oxford Univ. Press, 1960.

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

[16] T. Chen, H.-P. Seidel, and H. P. A. Lensch. Modulated phase-shifting for 3D scanning. In *Proc. CVPR*, 2008.

[17] A. Chorin and J. Marsden. *A Mathematical Introduction to Fluid Mechanics*. Springer-Verlag. Texts in Applied Mathematics 4. Second Edition., New York, 1990.

[18] F. Cozman and E. Krotkov. Depth from scattering. *In Proc. IEEE CVPR*, pages 801–806, 1997.

[19] K. Dana, S. Nayar, B. van Ginneken, and J. Koenderink. Reflectance and texture of real-world surfaces. In *Proc CVPR*, pages 151–157, 1997.

[20] P. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. *Proc. SIGGRAPH 98*, pages 189–198, 1998.

[21] Y. Dobashi, T. Yamamoto, and T. Nishita. Interactive rendering of atmospheric scattering effects using graphics hardware. In *Graphics Hardware Workshop*, pages 99–109, 2002.

[22] D. S. Ebert and R. E. Parent. Rendering and animation of gaseous phenomena by combining fast volume and scanline a-buffer techniques. In *Proceedings of SIGGRAPH*, 1990.

[23] R. Fedkiw, J. Stam, and H. W. Jensen. Visual simulation of smoke. In *Proceedings of SIGGRAPH*, pages 15–22, 2001.

[24] E. G. Finsy and J. Joosten. Maximum entropy inversion of static light scattering data for the particle size distribution by number and volume. In *Advances in measurements and control of colloidal processes. Butterworth-Heineman, Ch. 30*, 1991.

[25] N. Foster and R. Fedkiw. Practical animation of liquids. In *Proceedings of SIGGRAPH*, pages 23–30, 2001.

[26] N. Foster and D. Metaxas. Realistic animation of liquids. *Graph. Models Image Process.*, 58(5):471–483, 1996.

[27] N. Foster and D. Metaxas. Modeling the motion of a hot, turbulent gas. In *Proceedings of SIGGRAPH*, 1997.

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

[29] E. Fuchs and J. S. Jaffe. Thin laser light sheet microscope for microbial oceanography. *OPTICS EXPRESS*, 10 (2):145–154, 2002.

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

[31] J. Ghring. Dense 3-D surface acquisition by structured light using off-the-shelf components. *Videometrics and Optical Methods for 3D Shape Measurement*, 4309, 2001.

[32] T. E. Giddings, J. J. Shirron, and A. Tirat-Gefen. EODES-3: An electro-optic imaging and performance prediction model. *In Proc. IEEE OCEANS*, 2:1380–1387, 2005.

[33] G. D. Gilbert and J. C. Pernicka. Improvement of underwater visibility by reduction of backscatter with a circular polarization technique. *Applied Optics*, 6(4):741–746, 1967.

[34] A. L. Gilchrist. Perceived lightness depends on perceived spatial arrangement. *Science*, 195:185, 1977.

[35] L. D. Girolamo. Reciprocity principle applicable to reflected radiance measurements and the searchlight problem. *Applied Optics*, 38, 1999.

[36] L. D. Girolamo, T. Varnai, and R. Davies. Apparent breakdown of reciprocity in reflected solar radiances. *Journal of Geophysical Research*, 103(D8), 1998.

[37] L. Goddyn and P. Gvozdjak. Binary gray codes with long bit runs. *The electronic journal of combinatorics*, 10, 2003.

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

[39] J. Gu, S. K. Nayar, E. Grinspun, P. N. Belhumeur, and R. Ramamoorthi. Compressive structured light for recovering inhomogeneous participating media. In *Proc. ECCV*, 2008.

[40] J. Gu, C. Tu, R. Ramamoorthi, P. Belhumeur, W. Matusik, and S. K. Nayar. Time-varying Surface Appearance: Acquisition, Modeling, and Rendering. *ACM Trans. on Graphics (also Proc. of ACM SIGGRAPH)*, Jul 2006.

[41] M. Gupta and S. G. Narasimhan. Legendre fluids: a unified framework for analytic reduced space modeling and rendering of participating media. In *Proceedings of the 2007 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 17–25.

[42] M. Gupta, S. G. Narasimhan, and Y. Y. Schechner. On controlling light transport in poor visibility environments. In *IEEE CVPR '08*.

[43] M. Gupta, Y. Tian, S. G. Narasimhan, and L. Zhang. (De) focusing on global light transport for active scene recovery. In *Proc. IEEE CVPR*, 2009.

[44] V. I. Haltrin. One-parameter two-term henyey-greenstein phase function for light scattering in seawater. *Applied Optics*, 41(6):1022–1028, 2002.

[45] M. Harris and A. Lastra. Real-time cloud rendering. In *Eurographics*, pages 76–84, 2001.

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

[47] T. Hawkins, P. Einarsson, and P. Debevec. Acquisition of time-varying participating media. *ACM Trans. Graph.*, 24(3):812–815, 2005.

[48] H. V. Helmholtz. Treatise on physiological optics. 1, 1925.

[49] L. Henyey and J. Greenstein. Diffuse radiation in the galaxy. volume 93, pages 70–83, 1941.

[50] C. Hermans, Y. Francken, T. Cuypers, and P.Bekaert. Depth from sliding projections. In *Proc. CVPR*, 2009.

[51] A. Hertzmann and S. M. Seitz. Shape and materials by example: A photometric stereo approach. In *Proc. IEEE CVPR*, 2003.

[52] M. Holroyd, J. Lawrence, and T. Zickler. A coaxial optical scanner for synchronous acquisition of 3d geometry and surface reflectance. *ACM Trans. Graph.*, 29(3), 2010.

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

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

[55] V. D. Hulst. *Light Scattering by small Particles*. John Wiley and Sons, 1957.

[56] I. Ishii, K. Yamamoto, K. Doi, and T. Tsuji. High-speed 3d image acquisition using coded structured light projection. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 925–930.

[57] A. Ishimaru. *Wave Propagation and Scattering in Random Media. Volume 1: Single Scattering and Transport Theory*. Academic Press, 1978.

[58] D. Jaeger, H. Demeyere, R. Finsy, R. SNEYERS, J. Vanderdeelen, P. Van-Der-Meeren, and M. Van-Laethem. Particle sizing by photon correlation spectroscopy. part i: Monodisperse latices: influence of scattering angle and concentration of dispersed material. In *Part. Syst. Charact. 8, 179*, 1991.

[59] J. Jaffe. Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, 15(2):101–111, 1990.

[60] H. W. Jensen. *Realistic image synthesis using photon mapping*. A. K. Peters, Ltd., Natick, MA, USA, 2001.

[61] H. W. Jensen, S. R. Marschner, M. Levoy, and P. Hanrahan. A practical model for subsurface light transport. In *Proceedings of SIGGRAPH*, pages 511–518, 2001.

[62] J. T. Kajiya and B. P. V. Herzen. Ray tracing volume densities. *SIGGRAPH Comput. Graph.*, 18(3):165–174, 1984.

[63] J. R. Key. Streamer: User's guide. *Tech Report, NOAA/NESDIS, Madison, Wisconsin*, 2005.

[64] D. Kim, M. Ryu, and S. Lee. Antipodal gray codes for structured light. In *Proc. ICRA*, 2008.

[65] D. A. Kleffner and V. S. Ramachandran. On the perception of shape from shading. *Perception and Psychophysics*, 52:18–36, 1992.

[66] D. M. Kocak and F. M. Caimi. The current art of underwater imaging with a glimpse of the past. *MTS Journal*, 39:5–26, 2005.

[67] S. J. Koppal and S. G. Narasimhan. Shadow cameras: Reciprocal views from illumination masks. In *ICCV*, 2009.

[68] H. Koschmieder. Theorie der horizontalen sichtweite. beitr. In *Phys. Freien Atm.*, pages 171–181, 1924.

[69] K. N. Kutulakos and E. Steger. A theory of refractive and specular 3d shape by light-path triangulation. *IJCV*, 76(1), 2008.

[70] E. H. Land and J. J. McCann. Lightness and retinex theory. *J. Opt. Soc. Am.*, 61:1–11, 1971.

[71] E. Languenou, K. Bouatouch, and M. Chelle. Global illumination in presence of participation media with general properties. In *Fifth Eurographics Workshop on Rendering*, 1994.

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

[73] S. Marschner. Inverse rendering for computer graphics. *PhD Thesis, Cornell University*, 1998.

[74] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Trans. on Graphics (SIGGRAPH)*, 22(3):759–769, 2003.

[75] N. L. Max. Atmospheric illumination and shadows. In *Proceedings of SIGGRAPH*, pages 117–124, 1986.

[76] J. J. McCann. Do humans discount the illuminant? *Proc. SPIE*, 5666:9–16, 2005.

[77] N. J. McCormick. A critique of inverse solutions to slab geometry transport problems. *Prog. Nucl. Energy*, 8, 1981.

[78] M. Minou, T. Kanade, and T. Sakai. A method of time-coded parallel planes of light for depth measurement. *Transactions of IECE Japan*, 64(8), 1981.

[79] N. J. W. Morris and K. N. Kutulakos. Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. In *ICCV*, 2007.

[80] S. G. Narasimhan. Models and algorithms for vision through the atmosphere. In *Columbia Univ. Dissertation*, 2004.

[81] S. G. Narasimhan, M. Gupta, C. Donner, R. Ramamoorthi, S. K. Nayar, and H. W. Jensen. Acquiring scattering properties of participating media by dilution. *ACM Trans. Graph.*, 25(3):1003–1012, 2006.

[82] S. G. Narasimhan, S. J. Koppal, and S. Yamazaki. Temporal dithering of illumination for fast active vision. In *ECCV*, 2008.

[83] S. G. Narasimhan and S. K. Nayar. Contrast restoration of weather degraded images. 25(6):713–724, 2003.

[84] S. G. Narasimhan and S. K. Nayar. Shedding light on the weather. In *Proc. CVPR*, volume 1, pages 665 – 672, 2003.

[85] S. G. Narasimhan, S. K. Nayar, B. Sun, and S. J. Koppal. Structured light in scattering media. In *Proc. ICCV*, 2005.

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

[87] S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Trans. Graph.*, 25(3):935–944, 2006.

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

[89] R. Ng, R. Ramamoorthi, and P. Hanrahan. Triple product wavelet integrals for all-frequency relighting. *ACM Trans. Graph.*, 23(3):477–487, 2004.

[90] D. Q. Nguyen, R. Fedkiw, and H. W. Jensen. Physically based modeling and animation of fire. In *Proceedings of SIGGRAPH*, pages 721–728, 2002.

[91] T. Oishi. Significant relationship between the backward scattering coefficient of sea water and the scatterance at 120 degrees. *Applied Optics*, 29 (31):4658–4665, 1990.

[92] J. Park and A. C. Kak. Multi-peak range imaging for accurate 3D reconstruction of specular objects. In *Proc. ACCV*, 2004.

[93] J. Park and A. C. Kak. 3D modeling of optically challenging objects. *IEEE Transactions on Visualization and Computer Graphics*, 14(2), 2008.

[94] S. Pattanaik and S. Mudur. Computation of global illumination in a participating medium by monte carlo simulation. *Journal of Vis. and Computer Animation*, 4(3):133–152, 1993.

[95] J. Posdamer and M. Altschuler. Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*, 18(1), 1982.

[96] S. A. Prahl. Light transport in tissue. *PhD Thesis, University of Texas at Austin*, 1988.

[97] V. S. Ramachandran. Perception of shape from shading. *letters to nature*, 331:163–166, 1988.

[98] R. Ramamoorthi and P. Hanrahan. A signal processing framework for inverse rendering. *Proc. SIGGRAPH 01*, pages 117–128, 2001.

[99] K. Riley, D. Ebert, M. Kraus, J. Tessendorf, and C. Hansen. Efficient rendering of atmospheric phenomena. In *EuroGraphics Symposium on Rendering*, 2004.

[100] H. E. Rushmeier and K. E. Torrance. The zonal method for calculating light intensities in the presence of a participating medium. In *Proceedings of SIGGRAPH*, pages 293–302, 1987.

[101] G. Sakas. Fast rendering of arbitrary distributed volume densities. In *Eurographics*, pages 519–530, 1990.

[102] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, 43(8):2666–2680, 2010.

[103] Y. Y. Schechner and Y. Averbuch. Regularized image recovery in scattering media. *IEEE Trans. PAMI*, 29(9):1655–1660, 2007.

[104] Y. Y. Schechner and N. Karpel. Recovery of underwater visibility and structure by polarization analysis. *IEEE Journal of Oceanic Engineering*, 30(3):570–587, 2005.

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

[106] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar. Polarization-based vision through haze. *Applied Optics*, 42(3):511–525, 2003.

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

[108] A. Selle, N. Rasmussen, and R. Fedkiw. A vortex particle method for smoke, water and explosions. *ACM Trans. Graph.*, 24(3):910–914, 2005.

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

[110] W. A. Shurcliff and S. S. Ballard. *Polarized Light*, pages 98–103. Van Nostrand, Princeton, N.J., 1964.

[111] B. Skerry and H. Hall. *Successful Underwater Photography.* New York: Amphoto books, 2002.

[112] P.-P. Sloan, J. Kautz, and J. Snyder. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *SIGGRAPH*, 2002.

[113] J. Stam. Stable fluids. In *Proceedings of SIGGRAPH*, pages 121–128, 1999.

[114] M. Subbarao and M.-C. Lu. Computer modeling and simulation of camera defocus. In *Machine Vision and Applications*, pages 277–289, 1992.

[115] S. A. Sullivan. Experimental study of the absorption in distilled water, artificial sea water, and heavy water in the visible region of the spectrum. *JOSA*, 53, 1963.

[116] B. Sun, R. Ramamoorthi, S. G. Narasimhan, and S. K. Nayar. A practical analytic single scattering model for real time rendering. *ACM Trans. Graph.*, 24(3), 2005.

[117] B. Sun, K. Sunkavalli, R. Ramamoorthi, P. Belhumeur, and S. Nayar. Time-Varying BRDFs. *IEEE Transactions on Visualization and Computer Graphics*, Mar 2007.

[118] K. Tan and J. P. Oakley. Physics-based approach to color image enhancement in poor visibility conditions. *JOSA A*, 18(10):2460–2467, 2001.

[119] T. Treibitz and Y. Y. Schechner. Instant 3Descatter. In *Proc. IEEE CVPR*, volume 2, pages 1861–1868, 2006.

[120] T. Treibitz, Y. Y. Schechner, and H. Singh. Flat refractive geometry. In *Proc. IEEE CVPR*, 2008.

[121] A. Treuille, A. Lewis, and Z. Popovic. Model reduction for real-time fluids. *ACM Trans. Graph.*, 25(3), 2006.

[122] A. Treuille, A. McNamara, Z. Popovic, and J. Stam. Keyframe control of smoke simulations. *ACM Trans. Graph.*, 22(3):716–723, 2003.

[123] M. Trobina. Error model of a coded-light range sensor. *Technical Report*, 1995.

[124] H. van de Hulst. *Light Scattering by Small Particles.* Chapter 5. Wiley, New York, 1957.

[125] K. J. Voss and E. S. Fry. Measurement of the mueller matrix for ocean water. *Applied Optics*, 23:4427–4439, 1984.

[126] Ward-Larson, H. Rushmeier, and Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Trans. on Visualization and Computer Graphics*, 3(4):291–306, 1997.

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

[128] P. M. Will and K. S. Pennington. Grid coding: A preprocessing technique for robot and machine vision. *Artificial Intelligence*, 2(3-4), 1971.

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

[130] Y. Xu and D. Aliaga. An adaptive correspondence algorithm for modeling scenes with strong interreflections. *IEEE TVCG*, 2009.

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

[132] S. Zhang and S. Negahdaripour. 3D shape recovery of planar and curved surfaces from shading cues in underwater images. *IEEE Journal of Oceanic Engineering*, 27:100–116, 2002.

[133] S. Zhang, D. V. D. Weide, and J. Oliver. Superfast phase-shifting method for 3-D shape measurement. *Optics Express*, 18(9), 2010.

[134] T. Zickler, P. N. Belhumeur, and D. J. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *IJCV*, 49(2/3).