Capturing and Reconstructing the Appearance of Complex 3D Scenes

Christian Fuchs

Max-Planck-Institut für Informatik
Saarbrücken

Dissertation zur Erlangung des Grades
Doktors der Ingenieurwissenschaften (Dr.-Ing.)
der Naturwissenschaftlich-Technischen Fakultät I
der Universität des Saarlandes

Eingereicht am 16. März 2009 in Saarbrücken.

Betreuende Hochschullehrer
Prof. Dr. Hendrik P. A. Lensch,
Institut für Medieninformatik, Universität Ulm

Prof. Dr. Hans-Peter Seidel,
Max-Planck-Institut für Informatik, Saarbrücken

Gutachter
Prof. Dr. Hans-Peter Seidel,
Max-Planck-Institut für Informatik, Saarbrücken

Prof. Dr. Hendrik P. A. Lensch,
Institut für Medieninformatik, Universität Ulm

Prof. Dr. Michael Goesele,
GRIS, Technische Universität Darmstadt

Dekan
Prof. Dr. Joachim Weickert,
Universität des Saarlandes, Saarbrücken

Mitglieder des Prüfungsausschusses
Prof. Dr. Philipp Slusallek,
Universität des Saarlandes, Saarbrücken
Vorsitzender des Prüfungsausschusses

Prof. Dr. Hans-Peter Seidel,
Max-Planck-Institut für Informatik, Saarbrücken

Prof. Dr. Hendrik P. A. Lensch,
Institut für Medieninformatik, Universität Ulm

Prof. Dr. Michael Goesele,
GRIS, Technische Universität Darmstadt

Dr.-Ing. Thorsten Thormählen,
Max-Planck-Institut für Informatik, Saarbrücken

Tag des Kolloquiums
29. Mai 2009

Abstract

In this thesis, we present our research on new acquisition methods for reflectance properties of real-world objects. Specifically, we first show a method for acquiring spatially varying densities in volumes of translucent, gaseous material with just a single image. This makes the method applicable to constantly changing phenomena like smoke without the use of high-speed camera equipment.

Furthermore, we investigated how two well known techniques – synthetic aperture confocal imaging and algorithmic descattering – can be combined to help looking through a translucent medium like fog or murky water. We show that the depth at which we can still see an object embedded in the scattering medium is increased. In a related publication, we show how polarization and descattering based on phase-shifting can be combined for efficient 3D scanning of translucent objects. Normally, subsurface scattering hinders the range estimation by offsetting the peak intensity beneath the surface away from the point of incidence. With our method, the subsurface scattering is reduced to a minimum and therefore reliable 3D scanning is made possible.

Finally, we present a system which recovers surface geometry, reflectance properties of opaque objects, and prevailing lighting conditions at the time of image capture from just a small number of input photographs. While there exist previous approaches to recover reflectance properties, our system is the first to work on images taken under almost arbitrary, changing lighting conditions. This enables us to use images we took from a community photo collection website.

Kurzzusammenfassung

In dieser Dissertation präsentieren wir unsere Forschungsergebnisse über neue Methoden zur Akquisition von Reflektionseigenschaften von real existierenden Objekten. Im Einzelnen beschreiben wir zuerst eine Methode, um die Dichte eines durchscheinenden Materials in einem Volumen mit nur einem einzigen Bild komplett aufzunehmen. Dadurch ist das Verfahren besonders gut anwendbar auf sich ständig ändernde Objekte wie z.B. Rauch, ohne auf Hochgeschwindigkeitskameras zurückgreifen zu müssen.

Desweiteren haben wir analysiert, wie zwei sehr bekannte Verfahren – "Synthetic Aperture Confocal Imaging" und "Algorithmic Descattering" – kombiniert werden können um besser durch Nebel oder trübes Wasser hindurchsehen zu können. Wir zeigen, daß durch unser Verfahren die Entfernung in der ein Objekt im Nebel noch wahrgenommen werden kann, vergrößert wird. In diesem Zusammenhang zeigen wir auch wie auf Polarisation und auf Phasenverschiebung basierendes *Descattering* zum 3D-Scannen von durchscheinenden Objekten verwendet werden kann. Normalerweise verhindert die Lichtstreuung unter der Oberfläche eine genaue Entfernungsmessung, weil sich die Position des hellsten erleuchteten Punktes leicht unter die Oberfläche verschiebt. Mit unserem Verfahren wird die Lichtstreuung auf ein Minimum reduziert und somit das 3D-Scannen ermöglicht.

Im Anschluß präsentieren wir ein System, welches mit nur wenigen Eingabebildern die Geometrie und die Reflektionseigenschaften von undurchsichtigen Objekten bestimmt und außerdem die Lichtverhältnisse zum Aufnahmezeitpunkt rekonstruiert. Obwohl schon existierende Verfahren auch in der Lage sind, Reflektionseigenschaften zu bestimmen, ist unser Verfahren das Erste, welches auf Bildern arbeitet, die unter fast beliebigen, wechselnden Lichtverhältnissen aufgenommen wurden. Dies ermöglicht es uns, Bilder aus einer Internet Fotoplattform als Eingabe zu verwenden.

Acknowledgments

First of all, I would like to thank my supervisor Prof. Dr. Hendrik Lensch for his excellent supervision during my PhD studies, for his advice, his patience, and his support. I thank Prof. Dr. Hans-Peter Seidel for providing me with the opportunity to pursue a PhD at the MPI and an excellent working environment.

I am also very grateful to Prof. Dr. Michael Goesele, who supervised my PhD studies in the beginning, not only for introducing me into computer graphics in the first place, or for teaching me how to do research, but for providing me with the opportunity to go through some life-changing experiences.

The research in this dissertation would not have been possible without my collaborators and co-workers. Most importantly, I would like to thank Tongbo Chen and Tom Haber for being such great collaborators. I also thank all members of the computer graphics group at the MPI for their ideas, fruitful discussions and for being wonderful colleagues. I cannot name them all here, but I would especially like to thank the following people (in alphabetical order): Boris Ajdin, Tom Annen, Martin Fuchs, Miguel Granados, Michael Heinz, Matthias Hullin, Grzegorz Krawczyk, Andrei Lintu, Rafal Mantiuk, Karol Myszkowski, Oliver Schall, Kristina Scherbaum, Kaleigh Smith, Carsten Stoll, Robert Strzodka, Holger Theisel, Michael Wand, Oliver Wang, and Gernot Ziegler.

Special thanks go to Svetlana Borodina, Sabine Budde, Ellen Fries, Conny Liegl, and Sonja Lienard for all their patient help with all administrative tasks.

Maintaining a measurement lab and building new acquisition rigs would not have been possible without the help of our technical staff at the MPI. I therefore would like to thank Axel Köppel, Michael Laise, and Uwe Meiser for their great help.

The IST department is responsible for design, implementation and maintenance of the IT infrastructure at the MPI. During my PhD-studies they provided a stable, coherent, and scalable working environment with negligible downtimes. Needs for new hardware, sometimes custom built, were always fulfilled as fast as possible, especially before upcoming conference deadlines. I greatly appreciate their work.

Last, but not least, I would like to thank my family for their never-ending support.

Contents

Chapter 1
Introduction

Gathering appearance data about complex real-world objects has always been of great interest in computer graphics. The main reason for this is that accurate appearance data of an object or a scene is necessary for all kinds of manipulation, for example changing the object's appearance, relighting it in a novel illumination environment, viewing the object from novel viewpoints, or for simply preserving a changing object the way it appeared at a given instant in time. Additionally, the more we learn about an object and how its image is formed, the more accurately we can decompose the appearance into the contributing factors, a process known as modeling.

Models can be obtained from a variety of sources. They can be synthesized manually by artists, or they can be derived from the data that was used to construct the object in the first place. Rapidly changing objects such as smoke or natural phenomena can be simulated in a physically meaningful way and the result can then be used for rendering.

Another approach for obtaining a model of any given complex real-world object is to acquire its properties directly from the object itself. These properties also define the complexity of an object; just capturing the reflected color from an object is easily done with a camera, whereas for example the geometry takes significantly more effort to acquire. In this thesis, we will introduce acquisition methods for very complex object appearance data such as volumetric descriptions of constantly moving translucent smoke.

1.1 Problem Statement

The first step in any acquisition project is to define what properties should be acquired. With this information at hand, firstly a suitable measurement setup can

be designed and built. Secondly, one normally creates a model suitable for representing the desired appearance properties with enough accuracy for subsequent rendering.

In most cases, the captured data is however still a very complex, non-linear function of the actual model parameters. Recovering them requires an additional optimization or interpolation step, where the parameters of the underlying model are fit to the data. In this case, we call the problem of determining the model parameters from the observations an *inverse problem*. Solving an inverse problem allows us to gain insight into an object's behavior just through observation.

Inverse problems are typically ill-posed. To solve the problem anyway, the solution space can be limited to a small number of parameters with additional constraints. In this case *regularization* is normally employed to enforce smoothness in the solution and prevent overfitting.

This thesis introduces new acquistion techniques for specific problems or object classes. Although some of those were already covered by previous work, we managed to extend the space of what objects can be acquired further with our research. We demonstrate a system that acquires a volume of smoke with just a single image and thus enables acquisition without high-speed cameras and without high-output lasers even though the smoke is constantly moving. Furthermore, we show the first system that extracts geometry, incident lighting and reflectance properties for an object from just a few photographs. Since these photographs can be taken under nearly arbitrary condition our system even works on images taken from a community photo collection. We also further push the limit for seeing through a translucent medium like fog or murky water by successfully analyzing and combining two well-known techniques in computer graphics.

1.2 Contributions

In this thesis, we solve inverse problems with the goal of capturing accurate representations for different classes of real-world objects such as the densities in a dynamically changing volume of smoke, objects that are obscured by a scattering medium, and finally a complete scene description including geometry, environment maps and surface reflectance from a couple of photographs. Parts of this thesis have already been published or are to appear as scientific articles at international conferences or journals [Fuchs et al. 2006, Fuchs et al. 2007a, Chen et al. 2007, Fuchs et al. 2008, Haber et al. 2009].

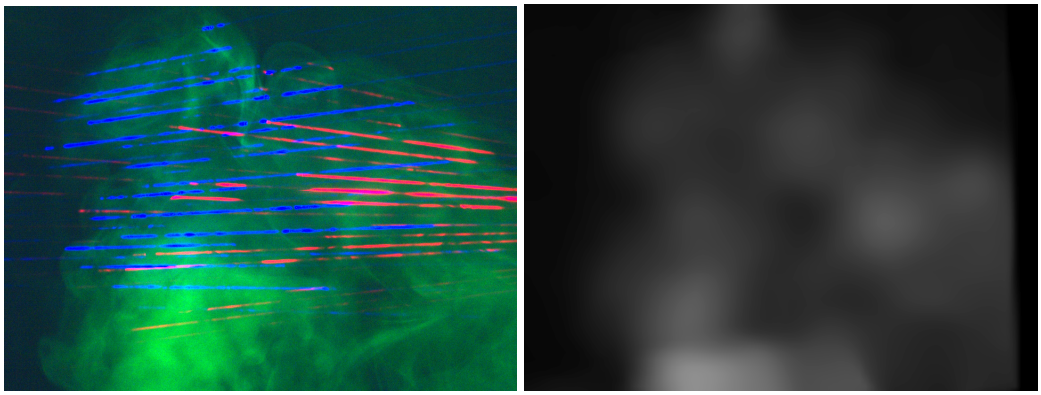


Figure 1.1*: Acquired image of smoke and reconstructed density field.* Left: *Input image with two independent grids of laser lines (red, blue) sampling the volume and diffuse green illumination (image intensity rescaled for display).* Right: *Density field reconstructed from this single image.*

1.2.1 Density Estimation for Dynamic Volumes

The first problem this thesis addresses is capturing a volume of spatially varying, constantly moving smoke. The smoke represents a whole class of dynamic, participating media which exhibit multiple light scattering. Capturing densities in a volume is a difficult task since we cannot observe every point in the volume directly. All points which do not lie on the outside of the volume will show combined effects both from the point itself and all the points on the way to the observer. Therefore, we need to solve this as an inverse problem, i.e., answer the following question: given an image of the volume, what is the most likely density distribution of smoke that created this image? Once we have determined the smoke densities correctly at each point in the volume, we can infer the amount of multiple scattering and create a rendering and compare it to the original image.

Since the smoke is constantly changing, all our measurements need to be done in a single exposure. We do this by only measuring a subset of all points within the volume with a set of laser lines. Using this setup, we achieve a sparse sampling between individual laser lines. On each laser line however, we obtain a very dense sampling of the volumetric densities. In combination, we are able to faithfully interpolate the missing data and reconstruct the whole volume (see also Figure 1.1).

Our contributions to the problem of reconstructing volumetric densities in dynamically changing volumes are:

- **A new approach to sampling the volume:** We illuminate the volume with a grid of laser lines in order to sample the whole volume simultaneously. Placing camera and light sources carefully ensures that all laser lines are

projected onto the image plane without occlusions. We further increase the sampling density by using multiple laser grids of different colors.

- **Decoupling of spatial and temporal sampling:** The proposed system captures density information in the whole volume with a single camera image. Therefore, it allows for a continuous sampling in the time domain.

- **Simulation of the sampling method:** We performed a full simulation of our sampling strategy to asses its capabilities and limitations. The simulated results furthermore allow us to compare our interpolated volumes to ground truth data. We analyze the strengths and the limitations of our sampling approach.

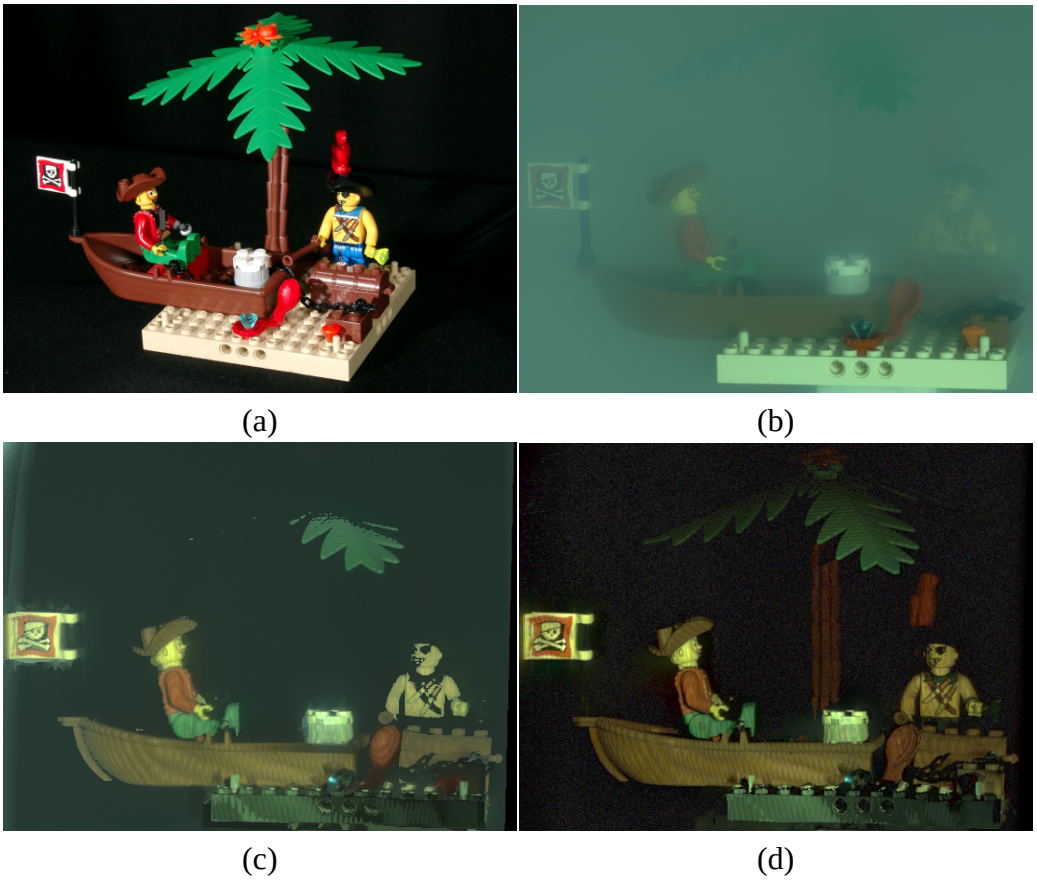
1.2.2 Combining Descattering and Confocal Imaging

Sometimes it may be desirable to not acquire a translucent scattering medium, but rather an object which is obscured by it. A classic application for this is seeing through fog or murky water. In this case, the fog must be separated in such a way that it can be subtracted without destroying the signal arriving from the object itself.

In Chapter 5, we analyze two widespread techniques, *algorithmic descattering* [Nayar et al. 2006] and synthetic aperture *confocal imaging* [Levoy et al. 2004]. We combine both methods with the goal of computing cross-sectional images in translucent media with higher contrast and better resolution than previously possible. In translucent objects, observations along individual light paths are polluted by in- and outscattering as can be seen in Figure 1.2(b). *Confocal imaging* reduces the effects of multiple scattering by carefully focusing illumination and observation onto a specific location within the measurement volume. The light path selected in this way nevertheless contains some global illumination effects. This is also evident in Figure 1.2(c). *Algorithmic descattering* is based on high-frequency illumination. It also aims at reducing multiple scattering effects in translucent media. It works however by removing global illumination effects from any light path. We show that confocal imaging and descattering are orthogonal and that they can be combined efficiently (see Figure 1.2(d)).

In this work, we contributed the following parts to solving the descattering problem:

- **Analysis of confocal imaging and algorithmic descattering:** We analyze both methods in terms of their effects on individual light paths through a translucent medium. From this, we draw the conclusion that both methods



(a) (b)

(c) (d)

Figure 1.2*: Example results from our descattering approach:* (a) *shows a photograph of the object that is immersed in a fishtank filled with dilute milk. After that, an image taken under normal flood-light illumination is shown in* (b)*. The bottom row shows the advantage of combining confocal imaging and descattering. With just confocal imaging* (c) *some details can not be recovered. Combining both methods* (d) *leads to a sharper image with more detail and better color reproduction.*

can be combined with very little extra effort in order to achieve superior results.

- **Validation through simulation:** Using a Monte-Carlo photon simulator, we qualitatively and quantitatively analyze the proposed approach and its limitations.

- **Real-world measurement setup:** Based on our observations, we designed a measurement setup, which acquires volumetric representations of objects embedded in a translucent medium with improved contrast and penetration depth.

1.2.3 Polarization and Phase-shifting for 3D Scanning

Algorithmic descattering is also at the heart of the next contribution. It has been developed in close collaboration with Tongbo Chen [Chen et al. 2007] during his PhD-studies. We therefore only present a short outline of the approach and the underlying algorithms in this thesis and refer the reader to Tongbo Chen's thesis [Chen et al. 2007] or the paper for a detailed description.

3D scanning of translucent objects can be very difficult because of the subsurface light transport. The signal that is projected onto the object becomes weaker since it is spread out beneath the surface. Highly transmissive objects also introduce a systematic bias in the range measurement, because the peak intensity is not observed at the surface but at some point beneath it. Range measurements are furthermore polluted by interreflections and light scattering towards the measurement point.

To enable reliable range measurements on translucent objects, it is necessary to separate the first surface reflection from all the scattering effects mentioned above. One possible way to descattering is to exploit the fact that polarized light becomes depolarized while undergoing multiple scattering. Therefore projecting polarized light patterns and computing the difference between images captured with a polarization filter at orthogonal orientations removes most of the subsurface scattering effects. Another method is algorithmic descattering as described above.

In Chapter 6, we present a method for robust shape acquisition of translucent 3D objects that contributes the following to the field of computer graphics:

- **Analysis of the descattering properties of polarization and phase-shifting:** We show the advantages and disadvantages of both methods, and combine them to obtain better results.

- **Measurement setup:** We show an efficient and practical measurement setup for 3D scanning of highly translucent objects.

1.2.4 Relighting Objects from Image Collections

In this part of the thesis, we do not deal with translucent objects. We restricted ourselves to opaque objects to achieve a rather ambitious goal: In this fourth contribution of the thesis, we use inverse rendering to extract almost all information necessary for relighting a scene from just a small number of photographs. Not only do we estimate scene geometry, but we use this geometry to subsequently recover the prevailing lighting conditions per input image and the surface reflectance properties (BRDF [Nicodemus et al. 1977]) per surface point. We also show that our system is robust enough to use images from a community photo collection site such as Flickr [Flickr 2009]. The resulting data can then for example be used

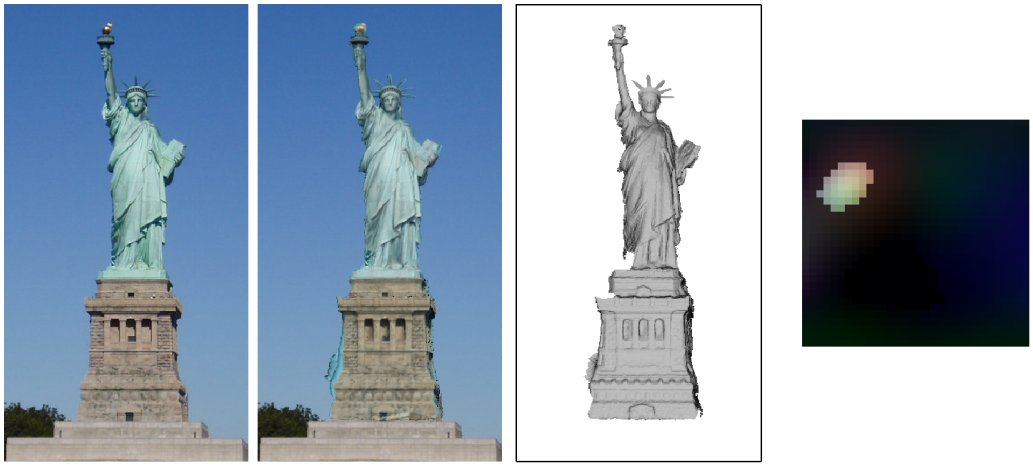


Figure 1.3*: Overview of our reconstruction pipeline. From left to right: an example image taken from Flickr [Flickr 2009], re-rendered model using the recovered reflectance properties, estimated geometry, and the illumination from the estimated environment.*

to render the object in arbitrary environments (*relighting*). The complete pipeline from an example input image to the rerendered result is shown in Figure 1.3.

In summary, the contributions of the system developed in this part of the thesis are:

- **Simultaneously estimate reflectance and illumination:** Our system estimates the reflectance and illumination parameters for a scene captured under varying distant illumination.
- **Using community photo collections:** We demonstrate reconstructions based exclusively on images retrieved from an Internet photo sharing site. Scene geometry, reflectance, and illumination are all estimated from these images.

1.3 Thesis overview

The remainder of this thesis is organized as follows: in the next chapter we are discussing previous work that is relevant for our research. Chapter 2 will summarize the theoretical background behind light transport in translucent media as well as for light interacting with opaque surfaces. The next four chapters (Chapter 4, Chapter 5, Chapter 6, and Chapter 7) are devoted to the contributions outlined above. Finally, the thesis is concluded with an outlook to future work in Chapter 8.

Chapter 2
Background

In this chapter, we will review some of the physical foundations of light transport required in the remaining chapters. We will also look into the theory behind light reflection, BRDFs (Bidirectional Reflectance Distribution Functions), subsurface light transport and the BSSRDF (Bidirectional Scattering Surface Reflectance Distribution Function).

2.1 Radiometry

Light is electromagnetic radiation that shows both properties of particles (photons) and waves carrying energy at different frequencies (wave optics). In particle optics the energy is carried by photons, with each individual photon carrying a certain amount of energy depending on the wavelength. In contrast to that, wave optics describes light as a continuous phenomenon, where the energy is transported by electromagnetic waves with a certain frequency. Light that only consists of waves oscillating in one direction is called *polarized light*.

Radiometry is the science of measuring electromagnetic radiation. In the following, we will give an overview over the most important quantities in radiometry:

Radiant Energy: The energy transported by light of all wavelengths, measured in *Joule* $\left[J = Ws = kg\ \frac{m^2}{s^2}\right]$.

Radiant Flux, Radiant Power Φ: The power (energy per unit time) of the radiation. It is measured in *Watts* $[W]$.

Radiant Intensity I: Radiant Flux per solid angle ω measured in *Watts per unit solid angle* $\left[\frac{W}{sr}\right]$.

$$I := \frac{d\Phi}{d\omega}$$

The intensity I can be used to characterize point light sources. Since a full sphere of directions has a solid angle of $4\pi \cdot sr$, an isotropic point light (a light source that emits the same amount of light in each direction) has an intensity of $\frac{\Phi}{4\pi \cdot sr}$.

Irradiance E and Radiant Exitance B: The irradiance $E := \frac{d\Phi}{dA}$ represents the radiant flux $d\Phi$ *arriving* at a surface with area dA while the radiant exitance B, which is often called *radiosity* in computer graphics, describes the radiant flux per unit area *leaving* a surface. Both are measured in $\left[\frac{W}{m^2}\right]$.

Radiance L: Radiant Flux per unit projected area per unit solid angle arriving at or leaving a point on a surface measured in $\left[\frac{W}{m^2 \cdot sr}\right]$.

$$L(x, \omega) := \frac{d^2\Phi}{cos\theta \cdot d\omega \cdot dA}$$

where x is the point on the surface and θ is the angle between the direction ω and the surface normal.

The relationship between irradiance and incoming radiance is

$$E(x) = \int_{\Omega(\vec{n})} L_i(x_i, \omega_i) cos\theta_i d\omega_i$$

where $\Omega(\vec{n})$ represents the hemisphere of directions around the surface normal $\vec{n}$ and L_i is the radiance arriving at the surface point x.

Radiance is an important quantity in computer graphics since it is constant along a ray in empty space. Therefore it is used by almost all rendering systems including ray-tracing applications.

With these quantities, we are able to formulate mathematical expressions for light arriving at or leaving a surface.

2.2 Rendering Equation

Given a complete description of a scene consisting only of opaque objects, the rendering equation [Kajiya 1986] allows for the simulation of light traveling through the scene in any environment.

$$L_o(x, \omega_o) = L_e(x, \omega_o) + \int_{\Omega^+} S(x, \omega_i, \omega_o) L_i(x, \omega_i)(n \cdot \omega_i) d\omega \qquad (2.1)$$

The right-hand side of the rendering equation contains the integral over the whole hemisphere of directions around the surface normal n. $L_e(x, \omega_o)$ is the radiance

emitted at surface point x in direction ω_o. $S(x, \omega_i, \omega_o)$ describes the reflectance properties for incident light from direction ω_i arriving at x and being reflected in direction ω_o. This special function – the *BRDF* is explained in detail in Section 2.4. $L_i(x, \omega_i)$ denotes the incident radiance arriving at x from direction ω_i.

2.3 Inverse Rendering

The rendering equation is evaluated for image synthesis given a complete scene description, i.e., scene geometry, light sources, reflectance properties of all surfaces, and camera location. However, this process can also be inverted: given a set of correctly rendered images L_{input}, infer the scene description or parts of it. This is in general called *inverse rendering*. It usually involves minimizing the error between the observed images and the re-rendered inferred scene properties along the viewing ray v:

$$E(x, v) = \left(L_{input}(x, v) - \int_{\Omega^+} S(x, \omega_i, v) L_i(x, \omega_i)(n \cdot \omega_i) d\omega \right) \tag{2.2}$$

2.4 Reflectance Functions

The rendering equation describes how an image can be synthesized from a complete description of the scene. This also includes the reflectance properties of all objects included in the scene ($S(x, \omega_i, \omega_o)$). In the following section, we will define this reflectance function for translucent and opaque objects.

How the light interacts with an object depends on several physical properties like the wavelength of the light, object geometry, and the material properties of the object. Translucent objects behave differently from purely opaque objects, they allow the light to enter and distribute within the object. This different behavior requires different formulations. In the following, we will review some basic formulas that describe the interaction of light with different kinds of material.

2.4.1 Reflectance in Translucent Objects

When light hits a translucent object it is partly reflected at the surface and another part is absorbed by the object immediately. The remainder enters the object where it is scattered until it either leaves the object again at some point which is not necessarily the entering location, or until it has been completely absorbed. How much of the incident light actually enters the object is determined by the Fresnel equations and depends on the polarization of the light and the indices of refraction n_1 and n_2. For the situation illustrated in Figure 2.1, if the light is polarized with

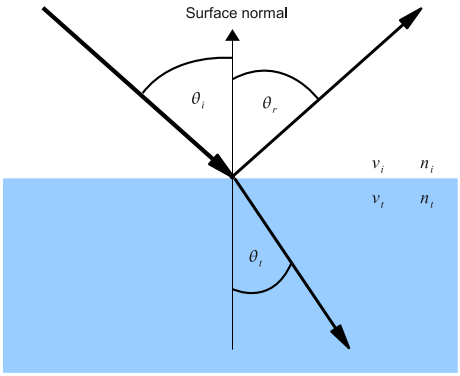


Figure 2.1: *Reflection and refraction at a medium boundary.* n_i *and* n_t *denote the indices of refraction for the different media,* v_i *and* v_t *are their respective light propagation speeds.*

its electric field perpendicular to the object surface, the ratio of light entering the object is given by:

$$\begin{aligned} R_s &= \left[\frac{\sin(\theta_t - \theta_i)}{\sin(\theta_t + \theta_i)}\right]^2 \\ &= \left[\frac{n_1\cos(\theta_i) - n_2\cos(\theta_t)}{n_1\cos(\theta_i) + n_2\cos(\theta_t)}\right]^2 \\ &= \left[\frac{n_1\cos(\theta_i) - n_2\sqrt{1 - \left(\frac{n_1}{n_2}\sin\theta_i\right)^2}}{n_1\cos(\theta_i) + n_2\sqrt{1 - \left(\frac{n_1}{n_2}\sin\theta_i\right)^2}}\right]^2 \end{aligned} \tag{2.3}$$

If the polarization plane coincides with the objects surface, the computation changes slightly:

$$\begin{aligned} R_p &= \left[\frac{\tan(\theta_t - \theta_i)}{\tan(\theta_t + \theta_i)}\right]^2 \\ &= \left[\frac{n_1\cos(\theta_t) - n_2\cos(\theta_i)}{n_1\cos(\theta_t) + n_2\cos(\theta_i)}\right]^2 \\ &= \left[\frac{n_1\sqrt{1 - \left(\frac{n_1}{n_2}\sin\theta_i\right)^2} - n_2\cos(\theta_i)}{n_1\sqrt{1 - \left(\frac{n_1}{n_2}\sin\theta_i\right)^2} + n_2\cos(\theta_i)}\right]^2 \end{aligned} \tag{2.4}$$

In the case of unpolarized light, i.e., consisting of a homogeneous mix of polarizations, the reflection coefficient is simply the average:

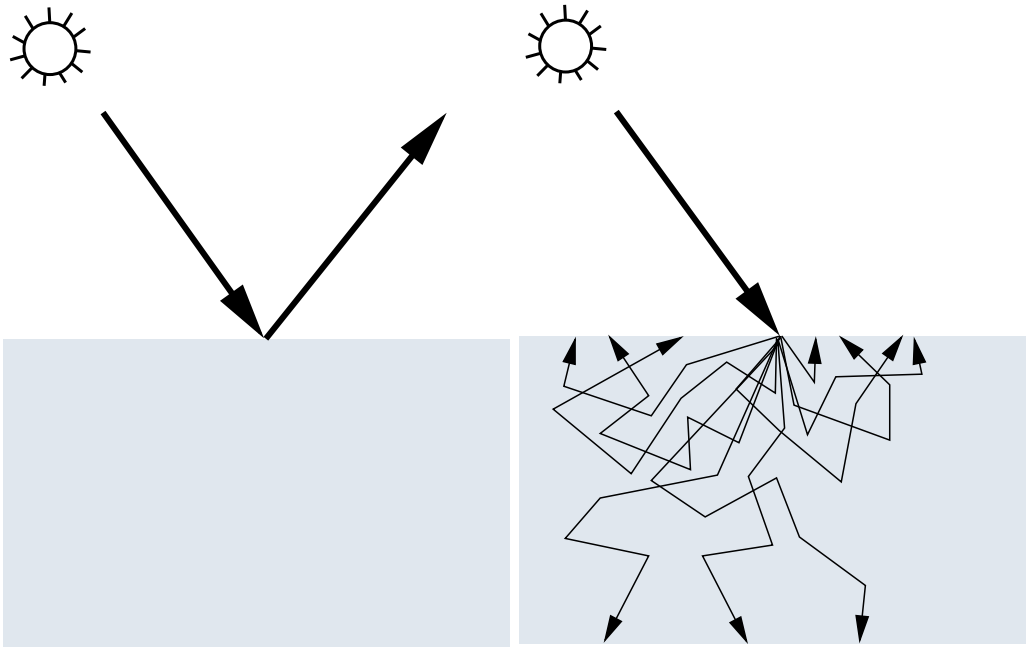


Figure 2.2*: In contrast to just the reflection at the surface as it occurs in opaque objects (*left*), light enters translucent objects, where it is scattered until it is either absorbed or leaves the object at a point that can be different from the point of incidence (*right*).*

$$R = \frac{(R_s + R_p)}{2} \tag{2.5}$$

The relationship between the angles θ_i and θ_t is governed by Snell's law. In the following formula v_i and v_t are the propagation speed of the light before and after entering the medium respectively. Their ratio is equivalent to the opposite ratio of the indices of refraction n_t and n_i.

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{v_i}{v_t} = \frac{n_t}{n_i}, \tag{2.6}$$

BSSRDF

The ratio between the outgoing radiance $L_o(x_o, \omega_o)$ leaving the surface at x_o in direction ω_o and the radiant flux $\Phi(x_i, \omega_i)$ arriving at x_i from direction ω_i is described by the 8D *BSSRDF* (Bidirectional Scattering Surface Reflectance Distribution Function):

$$S(x_i, \omega_i, x_o, \omega_o) = \frac{dL_o(x_o, \omega_o)}{d\Phi_i(x_i, \omega_i)} \tag{2.7}$$

The unit of the BSSRDF S is $\left[\frac{1}{m^2 \cdot sr}\right]$.

Neglecting spectral effects, fluorescence, or polarization, the outgoing radiance $L_o(x_o, \omega_o)$ can be computed given the incoming radiance at all surface points from all directions $L_i(x_i, \omega_i)$ by integrating over the whole surface A and the complete hemisphere of directions Ω_+ at x_i:

$$L_o(x_o, \omega_o) = \int_A \int_{\Omega_+(x_i)} L_i(x_i, \omega_i) \cdot S(x_i, \omega_i, x_o, \omega_o) d\omega_i dx_i. \tag{2.8}$$

A physically motivated BSSRDF should also have the following properties:
Helmholtz reciprocity [von Helmholtz 1856]:

$$S(x_i, \omega_i, x_o, \omega_o) = S(x_o, \omega_o, x_i, \omega_i), \tag{2.9}$$

Conservation of energy:

$$\int_A \int_\Omega L_o(x_o, \omega_o) d\omega_o dx_o \leq \int_A \int_\Omega L_i(x_i, \omega_i) d\omega_i dx_i \tag{2.10}$$

2.4.2 Reflection at the Surface

When light hits an opaque object (for example metal), part of its energy is absorbed in the material and converted to heat. The remaining energy is reflected according to the BRDF, as shown in Equation 2.11. Figure 2.3 shows what the BRDF looks like for a number of selected materials.

BRDF

The BRDF is a special case of the more general BSSRDF [Nicodemus et al. 1977] in which the incoming and outgoing surface point must be the same. This is a valid assumption for most opaque materials as well as metals. The 8D BSSRDF therefore reduces to the 6D *BRDF* (Bidirectional Reflectance Distribution Function):

$$S(x, \omega_i, \omega_o) = \frac{dL_o(x, \omega_o)}{dE_i(x, \omega_i)}, \tag{2.11}$$

where $E_i(x, \omega_i)$ is the irradiance at surface point x arriving from direction ω_i.

To be physically plausible, the BRDF should also obey Helmholtz reciprocity and conservation of energy:

$$S(x, \omega_i, \omega_o) = S(x, \omega_o, \omega_i) \tag{2.12}$$

$$\int_\Omega S(x, \omega_i, \omega_o) d\omega_o \leq 1 \tag{2.13}$$

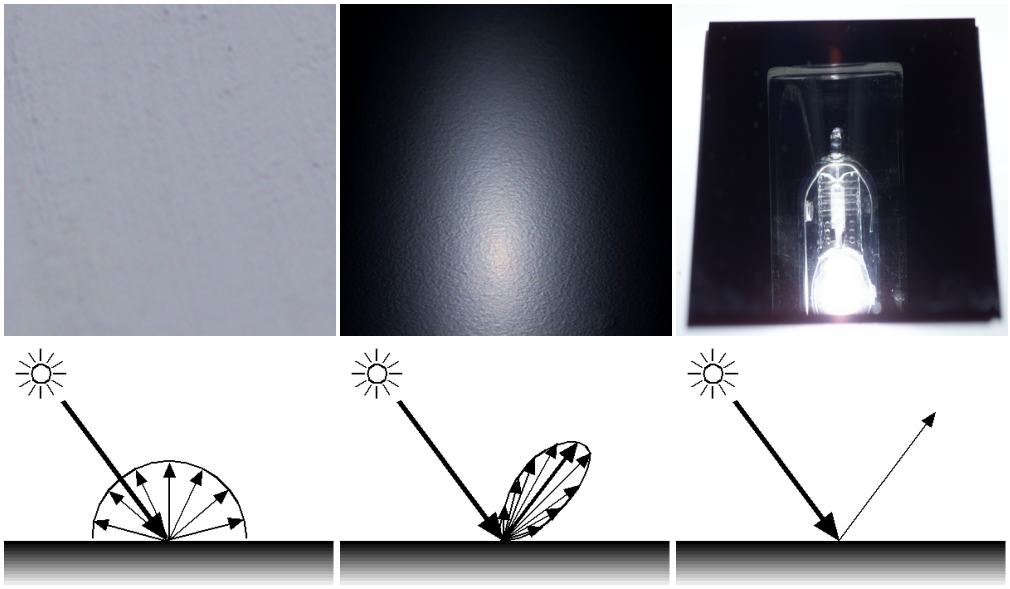


Figure 2.3*: Different materials and their corresponding BRDFs.*

2.5 Reflectance Models

How a surface interacts with light can be described in different forms. In the following, we will list different approaches to representing BRDFs and BSSRDFs ranging from raw point-to-point light transport matrices to analytical models for specific effects. But first, we will see how light transport in a scene can be formulated.

2.5.1 Light Fields and Reflectance Fields

The *light field* [Gershun and Pole 1939, Levoy and Hanrahan 1996, Levoy et al. 2006, Levoy 2006] is a 4D structure which describes the distribution of light within unoccluded space. For each surface point of a scene, it stores how much light leaves this point for every possible outgoing direction. Capturing the light field of a scene means that this scene can later be viewed from arbitrary view points under the illumination conditions which prevailed during acquisition.

The *reflectance field* [Debevec et al. 2000, Wenger et al. 2005, Garg et al. 2006, Fuchs et al. 2007b] accounts for varying illumination conditions. It is a 6D function similar to the light field. It represents the surface light transport at each surface point in the scene for every incoming and outgoing angle. A reflectance field therefore enables rendering of the scene from any

viewpoint under arbitrary lighting conditions. The BRDFs which are present in the scene are included in the light transport described by the reflectance field.

2.5.2 Transport Matrices

The easiest and most straightforward way to store a BRDF is to store its value for a discrete set of incoming and outgoing directions in a table [DeYoung and Fournier 1997]. During rendering, intermediate values can be interpolated. Tabulated BRDFs are however quite memory intensive. BSSRDFs can of course also be stored in point-to-point transfer matrices. Storing the whole 8D function is impractical in practice, though. Goesele et al. [Goesele et al. 2004] therefore neglected the angular dependency of the incoming and outgoing light and thus reduced the BSSRDF to a 4D function. Measuring this 4D transport matrix point by point still required hundreds of thousands of images.

2.5.3 Factored Representations

To overcome the problems associated with storing large tables for reflectance functions, the data is usually fitted to a set of lower-dimensional basis functions which can be evaluated to obtain the original data, or an approximation thereof. Kautz and McCool [Kautz and McCool 1999] factorized BRDFs for example into a sum of 1D or 2D textures. Latta and Kolb also include the incident lighting into the factorization [Latta and Kolb 2002]. Recently, Lawrence et al. [Lawrence et al. 2006] used constrained matrix factorization to represent spatially varying BRDFs. Peers et al. [Peers et al. 2006] use a similar approach for factorization of a BSSRDF after re-parametrization of the measured transport matrix.

Spherical harmonics are a low-dimensional basis function suitable to represent smoothly varying data. They have been applied to BRDFs in [Cabral et al. 1987, Ramamoorthi and Hanrahan 2001]. The big advantage of spherical harmonics is that in this basis the convolution of BRDF with the incident lighting reduces to a simple dot product which makes them especially desirable for interactive applications.

Later in this thesis (Chapter 7), we use Haar wavelets to represent BRDFs. Compared to spherical harmonics, Haar wavelets are better suited to representing high frequencies. Rendering can still be efficiently done by representing occlusion and lighting also in the wavelet domain and computing triple product wavelet integrals [Ng et al. 2004].

2.5.4 Analytical Models

Analytical models represent a BRDF or a BSSRDF with a mathematical function that depends only on a small number of parameters. Evaluation of that function yields the value of the reflectance function. The parameters are usually determined by fitting against measured data. This was done for BRDFs in [Lensch et al. 2001] and for BSSRDFs in [Fuchs et al. 2005].

Phong [Phong 1975] proposed a BRDF model which consists of a diffuse part and one specular lobe. The Blinn-Phong Model [Blinn 1977] provides more realistic reflections. It consists of a diffuse component, a specular lobe, and an exponent for the specular lobe. The model proposed by Lafortune et al. [Lafortune et al. 1997] expresses the specular part of the reflection as a combination of several specular lobes and can also handle anisotropic reflections. In the Torrance-Sparrow model [Torrance and Sparrow 1967] the specular component also considers the distribution of micro-facets on the surface.

For BSSRDFs, Jensen et al. [Jensen et al. 2001] approximate the light distribution within a volume of translucent medium with two light sources, a positive light source (i.e. emitting light) is placed beneath the surface to illuminate the volume, while a negative one (absorbing light) is placed above the surface.

2.6 Summary

In this chapter, we reviewed the basic terms in radiometry, which are needed to formulate light transport, image formation, etc. We have introduced the basic concepts of image synthesis and inverse rendering and have set up the theoretical background of reflectance functions and introduced various approaches to modeling of reflectance functions. These basic concepts are repeatedly used throughout the remainder of this thesis. Reflectance fields and light fields are important concepts used in Chapters 5 and 7. In the same chapter we will also use a Haar wavelet representation for the BRDFs that we recover.

Equipped with this knowledge, we will now look into the previous work that has been carried out in the context of appearance capture and reconstruction.

Chapter 3

Related Work

In the following chapter, we will first look at the related work that was carried out in the context of translucent objects. More specifically, we will discuss previous work covering the complete range from the acquisition of real-world translucent objects, modeling their behavior when interacting with light to realistic rendering. Secondly, we will discuss previous work related to the acquisition of reflectance properties and illumination conditions for complex opaque objects.

3.1 Translucent Objects

In contrast to opaque objects, where light interacts with the surface locally only at the point of incidence, translucent objects allow the light to enter the material and exit again at a possibly different surface point, as illustrated in Figure 3.1. Therefore, the light transport inside the object makes for example determining the exact point of incidence very difficult. There are two common approaches to deal with this. On one hand, research concentrated on techniques for disabling the subsurface light transport altogether, ranging from simply spray painting the object with diffuse paint through the use of polarization to algorithmic *descattering*. In Chapter 6, we show how polarization can be combined with descattering for simpler shape acquisition. On the other hand, for faithfully reproducing the appearance of a translucent object, the global light transport that is going on inside the object must be known. Translucency mainly depends on some physical properties of the material, like the scattering coefficient (σ_s) and the extinction coefficient (σ_t). For a detailed explanation of these and other properties, see [Jensen et al. 2001] or [Ishimaru 1978].

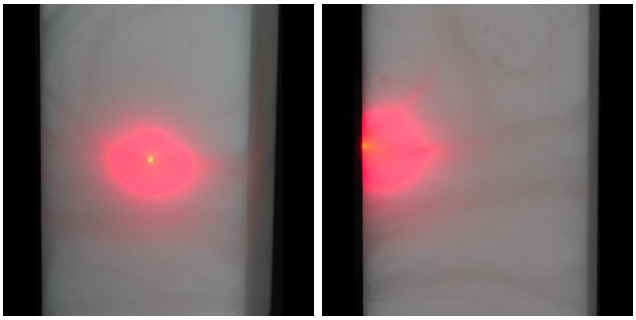


Figure 3.1*: A single laser beam hitting a translucent object. At the front, the single laser dot is spread out over a large surface area. From the side view, we can see, how deep the light penetrates the object and that the light is transported further along the veins in the marble.*

3.1.1 Acquisition of the Light Transport in Translucent Objects

In recent years, substantial amounts of research have been carried out in acquiring the global light transport properties of a translucent object. The light fields captured by [Levoy and Hanrahan 1996] and the Lumigraph by [Gortler et al. 1996] both represent light rays propagating through unobstructed space. They are both able to capture the appearance of a subsurface scattering object faithfully, but only for a fixed set of viewing and illumination conditions. Surface light fields [Miller et al. 1998, Wood et al. 2000] assign color values to each ray originating at a surface point. They are able to represent surface texture, specularity and global effects faithfully for arbitrary viewpoints in a fixed environment. Reflectance fields [Debevec et al. 2000] and polynomial texture maps [Malzbender et al. 2001] are captured from a fixed viewpoint under varying point light illumination. They enable relighting of the object with any linear combination of the captured point light sources.

Environment matting [Zongker et al. 1999, Chuang et al. 2000] techniques acquire reflection and refraction properties of strongly transparent or translucent objects such as glass. After that, the object can be put into an arbitrary new environment, where it will reflect and refract light from the new scene correctly.

[Sen et al. 2005] measured a 6D slice out of the reflectance field that allows for relighting with arbitrary 4D light fields. Furthermore, they exploited Helmholtz reciprocity to computationally exchange the illumination and the camera. They were thus able to record the 6D slice of the reflectance field using an array of cameras instead of an array of projectors. Since the cameras do not interfere with each other, they can capture images in parallel which massively speeds up recording the 6D datastructure. Garg et al. [Garg et al. 2006] measured 8D reflectance

fields as a transport matrix between incident and outgoing 4D light fields. To make measurement of 8D reflectance fields feasible, they exploit the fact that the transport matrix is symmetric due to Helmholtz reciprocity. They also exploit its data-sparseness using a hierarchical measurement approach.

Goesele et al. [Goesele et al. 2004] built the first system to measure a 4D slice out of the 8D BSSRDF (neglecting incident and outgoing angles) of a heterogeneous translucent object. They acquired a full vertex-to-vertex light transport matrix under the assumption that it is locally smooth. Later Peers et al. [Peers et al. 2006] used a slightly modified setup with a projector instead of lasers to reduce measurement time by acquiring multiple samples in each image.

3.1.2 Modeling Translucent Objects

Acquiring the subsurface light transport usually yields huge amounts of data. Therefore, the data is in general fit against a subsurface scattering model. A wide variety of different representations for subsurface scattering objects have been proposed in recent years. While all of these representations share the common goal of accurately portraying translucent objects, there are quite some differences between them, both in methodology and in the resulting models.

Most image-based methods (e.g., [Zongker et al. 1999, Chuang et al. 2000, Goesele et al. 2004, Sen et al. 2005]) acquire the scattering properties for one individual object at a time. Using these *object model*-based techniques one ends up with an accurate description of the captured object that can later be used for relighting. However, editing the appearance properties (such as the scattering parameters, material properties, etc.) is usually not easily possible.

In contrast to that, other methods strive to recover a mathematical description to resynthesize the captured object or even a different object using the acquired data. Depending on the model used, these *material models* also allow for changing the objects appearance by altering the parameters of the model.

Jensen et al. [Jensen et al. 2001] approximated the volumetric light distribution in a translucent medium by two point sources, i.e., a dipole like it is illustrated in Figure 3.2. The light intensity away from the dipole sources is then estimated with a diffusion process. The resulting images look convincing, although in practice, some assumptions of the diffusion process are violated. Furthermore, this technique is only applicable to homogeneous materials. Tong et al. [Tong et al. 2005] later extended this approach to *quasi-homogeneous* materials, i.e., objects that consist of a mix of several homogeneous materials. Peers et al. [Peers et al. 2006] measure the point-to-point light transport of a heterogeneous object in a fashion similar to [Goesele et al. 2004]. They use clustering and factorization of the resulting transport matrix to obtain a compact description of the scattering behavior for each material involved. This model can be applied to

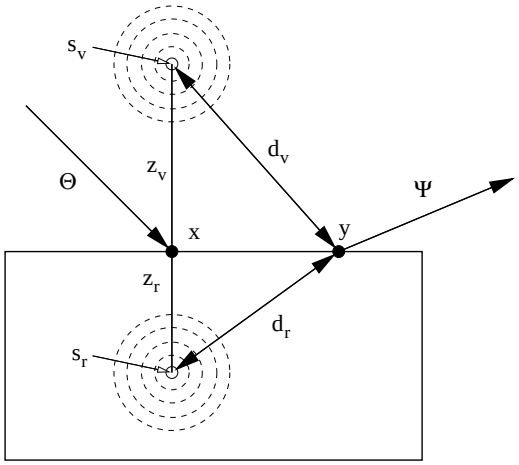


Figure 3.2*: The BSSRDF model of Jensen et al. [Jensen et al. 2001] based on a dipole source approximation. A pair of imaginary point light sources is placed above and below the surface point x. The distances z_r and z_v are calculated from the reduced scattering coefficient σ_s' and the absorption coefficient σ_a of the material. The light source within the translucent volume emits light, while the (negative) light source above the surface absorbs light.*

novel geometry. The Shell Texture Functions (STF) model by [Chen et al. 2004] was designed for efficient rendering under varying illumination conditions. In order to achieve this, they represent an object as a homogeneous inner core surrounded by a spatially varying volumetric outer shell. The STFs, which represent the irradiance of each shell voxel with respect to incident illumination are then precomputed.

Narasimhan et al. [Narasimhan et al. 2006] dilute liquids to a level where multiple scattering almost completely disappears. This allows them to estimate the scattering parameters of the liquid from the single scattering only. Recently, [Wang et al. 2008] have proposed a method which uses inverse rendering together with conjugate gradient optimization to recover physically meaningful scattering parameters for a flat material sample of a translucent object. The recovered material model can be used for rendering immediately or after editing. Their approach is however not suitable for acquiring complex 3D objects.

3.1.3 Shape Acquisition for Translucent Objects

Shape acquisition is in general hindered by the subsurface light transport. [Godin et al. 2001] analyzed the systematic bias that occurs during laser range measurements of translucent objects. In practice, the object is therefore usually covered with a fine layer of lambertian spray paint. We were able to reproduce

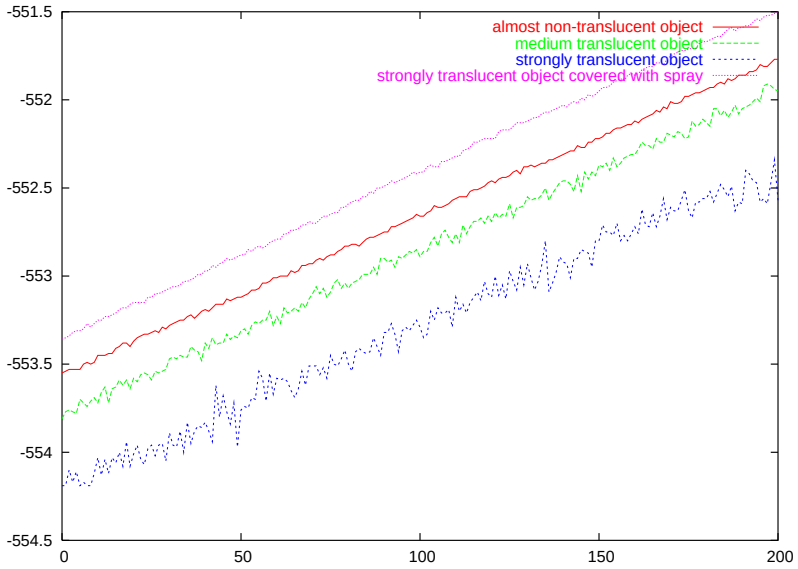


Figure 3.3*: This plot shows the effect of translucency on the depth measurement (y-axis in mm). The measurement offset and noise increase with increasing translucency. Covering the object with a lambertian spray reduces the bias and the measurement noise. The range difference between the sprayed object and the almost lambertian object can be partly due to the thickness of the paint layer.*

the results of [Godin et al. 2001] and show the effect that spraying the object has in Figure 3.3.

Whenever an object is too delicate or too valuable to be spray painted, there are other options for shape acquisition. Tactile scanners sparsely sample the surface at specified points. This approach however is very time consuming, especially for complex objects. Computer tomography scanners or MRI (Magnetic Resonance Imaging) devices produce volumetric representations of the object regardless of surface appearance [Lensch et al. 2003], but are generally not readily available.

3.1.4 Descattering

So far, we have seen a number of approaches that acquire and model subsurface scattering in translucent objects. Sometimes, the translucency can be an unwanted phenomenon. Translucency poses for example a big problem for shape acquisition (see Section 1.2.3). In Section 3.1.3, we have seen previous work that analyzes how translucency hinders 3D shape acquisition and possible ways to overcome this problem. Another example is looking through fog, where the visibility of objects is reduced by the scattering fog in front of it. There are a number of approaches that try to remove as much of the multiple scatter-

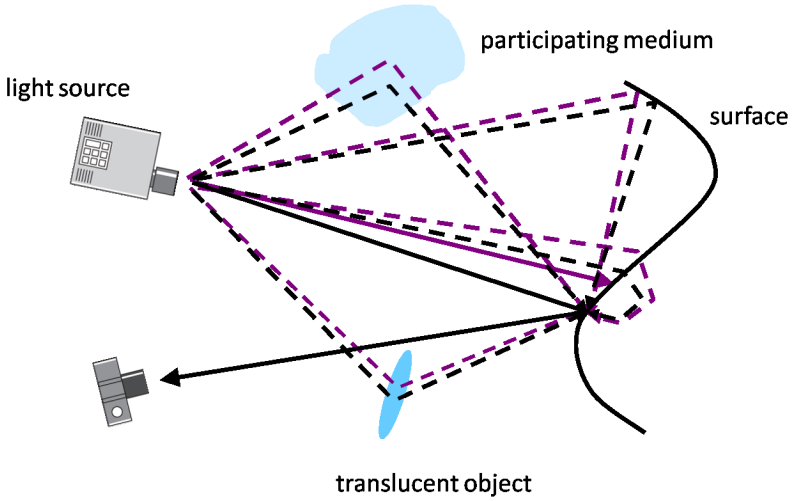


Figure 3.4*: Algorithmic descattering as proposed in [Nayar et al. 2006]. A number of global illumination effects influence the light paths in this setting: there is scattering in a participating medium, interreflection at the surface, scattering in a translucent object, and subsurface scattering. While the two shifted rays illuminate totally different surface points, the global effects remain roughly the same. This observation allows efficient separation of the direct and global components of the scene.*

ing effects as possible, a process which we refer to in the following as *descattering*. Using polarized light and computing the difference of images taken with polarization filters at orthogonal orientations removes most of the multiple scattering [Wolff 1994, Rowe and Pugh Jr. 1995, Schechner and Karpel 2005, Treibitz and Schechner 2006], a fact which we will exploit later in this thesis.

Nayar et al. [Nayar et al. 2006] proposed a simple but efficient method for separating the direct reflection and the global illumination effects due to scattering in a scene. The key idea to their approach is the observation that high frequencies are attenuated much stronger by multiple scattering than lower frequencies. They therefore illuminate the scene with shifted high-frequency patterns. The high-frequency parts of the illumination will only be seen for the direct reflection component, while the global component consists of a low-pass filtered version of the illumination, which roughly stays constant even when the illumination is shifted. This is illustrated in Figure 3.4. By subtracting the constant parts from the input images, they obtain the direct contribution. We employ similar ideas in Chapter 4 and furthermore use their technique in Chapters 5 and 6 of this thesis.

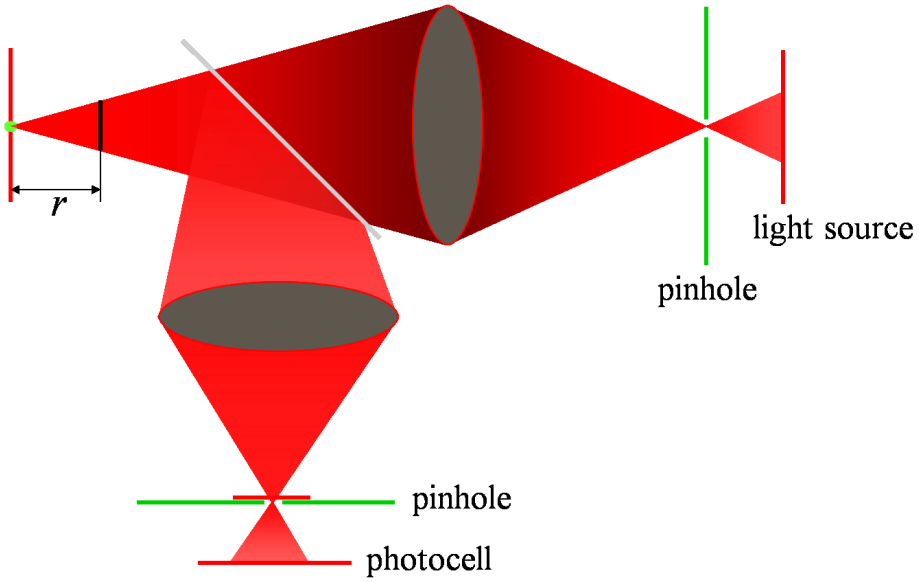


Figure 3.5*: Image formation in confocal imaging. The light source is focused through a pinhole onto a specific point on the objects surface. At the same time, the observation from the camera is focused onto the same point. Any point at a distance r from the common focal plane will receive only $1/r^2$ times the illumination. Accordingly, its image will be spread out over a bigger surface, reducing its intensity on the image sensor by a factor of $1/r^2$. In total, this amounts to a reduction in intensity by a factor of $1/r^4$.*

3.1.5 Volumetric Measurements in the Presence of Multiple Scattering

Most of the acquisition concepts introduced so far consider the object's appearance only as surface-to-surface light transport [Goesele et al. 2004, Peers et al. 2006]. Some of them even model the inner parts of the object [Chen et al. 2004] but they are never explicitly measured. In Chapter 4, we describe a method for measuring a full volume containing a participating medium.

The usual way for sampling a 3D volume is to use multiple input views. The recorded images can then be used to recover the volume in a variety of ways. Most often, when dealing with participating media, the reconstruction process involves some kind of tomographic reconstruction [Ihrke and Magnor 2004, Ihrke and Magnor 2005, Trifonov et al. 2006].

Confocal imaging [Corle and Kino 1996, Wilson et al. 1996, Neil et al. 1997, Levoy et al. 2004] is also suitable for creating volumetric measurements. Since it focuses the illumination and/or the observation to a specific focal plane in space, a whole volume can be sampled by moving that plane. Contributions from objects

before or behind the common focal plane are greatly attenuated. This situation is illustrated in Figure 3.5.

Usually, volumetric measurements are done by combining information from multiple views. For some methods however, a single view is enough for reconstruction. In [Hawkins et al. 2005] the volume is sampled with a laser plane at high speeds. Imaging is performed using a high-speed camera in order to achieve interactive frame rates. Hullin et al. [Hullin et al. 2008] use a similar sampling approach in a different context. In order to scan glass objects, they immerse them in a fluorescent liquid with matched index of refraction. This effectively removes multiple scattering because light that underwent one fluorescent scattering event does not contain enough energy anymore to excite another fluorescent molecule. This allows for separating the parts where the liquid has been displaced by the object.

3.2 Relighting of Opaque Objects

So far, this chapter dealt exclusively with objects that exhibit subsurface light transport. In the following, we will now look at reflectance properties of opaque objects where we can assume that light leaves the object from the same point where it arrived. This simplification allows us to build a system that estimates geometry, incident lighting and the object's BRDF with the goal of enabling relighting of the object.

In this context, the term relighting refers to photo-consistently rendering the target object using novel illumination conditions. In general, this requires the BRDF of each surface point of the object to be known. The incident illumination is hereby often assumed to be distant, i.e., it is the same for all surface points and depends only on surface orientation, and is usually given as an environment map as shown in Figure 3.6.

3.2.1 Reflectance from Known Illumination

A common strategy for relighting is to sample the incident illumination densely using a controlled acquisition setup [Debevec et al. 2000]. Other reconstruction approaches use a sparse set of images captured under known, point light illumination and combine them with additional assumptions. Assuming *spatial coherence*, Marschner et al. [Marschner et al. 2000] estimate a single BRDF from a set of images of an object while Zickler et al. [Zickler et al. 2006] reconstruct a relightable model from a single image. Lensch et al. [Lensch et al. 2003] capture multiple images of a heterogeneous object. They cluster the surface into regions with similar appearance and determine a BRDF for each cluster which serve as *basis BRDFs*.

Figure 3.6*: Example environment maps. The* top *row shows the images as they were acquired. The mirroring sphere shows the surrounding environment for the full hemisphere directed at the camera. The* bottom *row shows the environment maps in the* hemi-octahedral representation *where the sphere is unrolled onto a flat surface with simple boundary extension and without distortion.*

The BRDF of each surface point is then determined as linear combination of these basis BRDFs. Weistroffer et al. [Weistroffer et al. 2007] extend this approach and determine the reflectance of a surface point as a weighted sum of materials, each consisting of several basis BRDFs.

Due to the lack of control, reflectance recovery for outdoor scenes generally operates under complex illumination. Yu and Malik [Yu and Malik 1998] recovered the photometric properties of architectural scenes from photographs using an explicit outdoor illumination model. Debevec et al. [Debevec et al. 2004] first captured the scene geometry and reflectance samples. They then used an inverse rendering approach to create a relightable model of the Pantheon from images with known, distant illumination. Most recently, Romeiro et al. [Romeiro et al. 2008] introduced a passive reflectometry approach that estimates a bi-variate representation of an isotropic BRDF from a single image of a curved surface captured under known distant illumination. All of these systems require either known or controlled incident illumination and often also a detailed, captured or manually created scene model and are thus unable to reconstruct a scene's reflectance from existing imagery.

3.2.2 Illumination Estimation

Distant illumination in a scene can be directly measured using a light probe [Debevec 1998]. Illumination can also be estimated from other features such as shadows in the scene. Sato et al. [Sato et al. 2003] estimate the illumination distribution of a scene with known geometry from shadows cast by an occluder. They define an adaptive refinement criterion based on the intensity distribution of direct light sources. Later, Okabe et al. [Okabe et al. 2004] also investigated illumination estimation from cast shadows. For the case of lighting expressed in a Haar wavelet basis, they start with a coarse representation and iteratively add basis functions with increasing resolution.

3.2.3 Estimating Reflectance and Illumination

Ramamoorthi and Hanrahan [Ramamoorthi and Hanrahan 2001] give an algorithm to factor the reflected light field of a scene captured under unknown illumination into the reflectance and illumination components. They also provide a theoretical framework that predicts under which conditions this can be achieved and how the recoverable detail of the reflectance functions is determined by the frequency content of the incident illumination. Their analysis applies directly to the system we will propose in Chapter 7. Based on a sparse set of images, Nishino et al. [Nishino et al. 2001, Nishino et al. 2005] first separate the diffuse texture from the specular reflectance of an object captured under fixed, unknown lighting. Given the object's geometry, they then compute an initial estimate for the illumination and determine both the illumination and the parameters of a simplified Torrance-Sparrow model. Both of these methods assume known scene geometry and constant specular reflectance over the surface. Georghiades [Georghiades 2003] uses generalized photometric stereo to estimate geometry, positions of a single point light source, and the parameters of a single reflectance model for a sparse set of images with fixed viewpoint.

Finally, Yu et al. [Yu et al. 2006] recover both lighting and surface reflectance from a small set of images captured under constant distant illumination. Given the geometry of the scene, they compute the diffuse albedo and determine the parameters of a Gaussian filtered mirror BRDF model for each surface point. Illumination is modelled in a spherical harmonics framework. Our system is closely related to this work but has two important advantages. First, we allow varying illumination between the input views. This simplifies acquisition and makes our method applicable to existing datasets captured under arbitrary distant lighting conditions. Second, we operate in an all-frequency framework which simplifies modelling of complex illumination and allows the use of arbitrary BRDFs.

3.3 Summary

All the previous publications listed in this chapter have their specific advantages and disadvantages. In some places, however, there was still room for improvement. In the following, we would like to shortly summarize these missing spots in the previous work and show how the research carried out in the context of this thesis fills the gaps.

Acquisition of translucent media in a 3D volume has been done in various ways. The most common method is to use multiple views distributed around the volume in question. In this setting, the participating medium is best recovered using a tomographic reconstruction approach as presented in [Ihrke and Magnor 2004, Ihrke and Magnor 2005, Trifonov et al. 2006]. If measurements from multiple views are not possible or undesirable, the common approach is sampling the volume with active illumination. While the approach of Hawkins et al. [Hawkins et al. 2005] produces high-quality results, it requires high-speed camera equipment and a high-output laser to do so. Acquisition with a high-speed camera generates large amounts of data to be processed and ultimately limits acquisition time. The approach we present in Chapter 4 trades spatial resolution for speed so that acquisition becomes possible with a single image. It is therefore especially well suited for long term observation or real-time processing of the input data.

Synthetic aperture confocal imaging [Levoy et al. 2004] and algorithmic descattering [Nayar et al. 2006] are two well-known methods in computer graphics. Nevertheless, the approach we present in Chapter 5 is the first that analyzes both methods in terms of their effect on the reflectance field of a scene. Our analysis leads to the conclusion that both methods are suitable for removing multiple scattering effects within a volume of a homogeneous translucent medium. But they do so with different effect on the reflectance field: while confocal imaging aims at removing unwanted illumination and observation paths through the medium, algorithmic descattering removes global illumination effects from every single light path. Our work shows that the two methods can therefore be efficiently combined and confirms this by showing that visibility through a translucent medium is improved.

In related research, we analyzed methods for the acquisition of the 3D shape of highly translucent objects. Previous approaches to prevent subsurface scattering range from simply spray painting the object with matte paint to using polarization filters [Wolff 1994, Rowe and Pugh Jr. 1995, Schechner and Karpel 2005, Treibitz and Schechner 2006]. Spray painting the object is usually a tedious work or it is impossible due to the delicate nature of the object. Polarization alone is also usually not sufficient to remove all subsurface scattering, leading to noisy and biased results. In Chapter 6 we show a method that combines the polariza-

tion already used in previous methods with the algorithmic descattering proposed in [Nayar et al. 2006]. We show that compared to descattering or polarization alone, our method produces high-quality geometry models for a variety of highly translucent, "uncooperative" objects.

Recovering geometry, reflectance properties, and illumination conditions simultaneously for a scene from images has long been considered to be one of the hardest problems in computer graphics. Earlier, we presented a number of approaches that recover surface reflectance alone. Most of them require the illumination conditions to be either controllable [Debevec et al. 2000] or least known [Marschner et al. 2000, Zickler et al. 2006, Lensch et al. 2003, Weistroffer et al. 2007], for example by measuring them with a light probe [Debevec 1998]. Even though Yu et al. [Yu et al. 2006] recover both lighting and surface reflectance, they require the illumination to stay constant. All previous approaches that estimate both lighting and reflectance require the geometry of the object to be known in advance. Our work estimates the geometry from the input images and actually requires the illumination to vary between input views to estimate surface reflectance robustly. This has a number of implications: firstly, our method is not confined to any laboratory setting, since we do not need to control illumination. Secondly, estimating the geometry directly from the input images means that we do not need physical access to the object to acquire the geometry separately. All this results in our approach being able to work on images which we took from a community photo collection website like Flickr [Flickr 2009].

In the last two chapters of this thesis, we have established the theoretical background for our work and have shown where this dissertation is situated within previously published research. We have identified where problems were not tackled by previous work and how this thesis fills in these gaps. In the next chapter we are now going to introduce our first acquisition method.

Chapter 4

Density Estimation for Dynamic Volumes

The first acquisition method, we would like to present is an approach for capturing the density distribution of a translucent medium within a volume. The presented method has been published in [Fuchs et al. 2006] and [Fuchs et al. 2007a].

4.1 Introduction

Sampling a volume usually requires information from several viewpoints around the volume. Hasinoff and Kutulakos [Hasinoff and Kutulakos 2003] presented a multi-view method to volumetrically reconstruct flames using a photo-consistency approach. They assume however, that the flames have negligible scattering and therefore their emission dominates radiance. Ihrke and Magnor [Ihrke and Magnor 2004, Ihrke and Magnor 2005] used sparse view tomography to reconstruct flames and optically thin smoke from a small set of camera views.

All these approaches have in common, that they require multiple viewpoints. For reconstructing the full volume from only a single view, approaches are usually employed which scan the volume slice by slice. They suffer however from the problem, that scanning takes time, smoke or flames are constantly in motion and cannot be stopped easily. Hawkins et al. [Hawkins et al. 2005] sampled time-varying smoke density anyway by rapidly scanning a high-power laser plane through the smoke volume and imaging the light scattered by individual smoke particles from a lateral view with a high-speed camera (see Figure 4.1, left). This allows them to sample locations in the moving light plane with high spatial resolution yielding high quality renderings of the captured model. Physical mea-

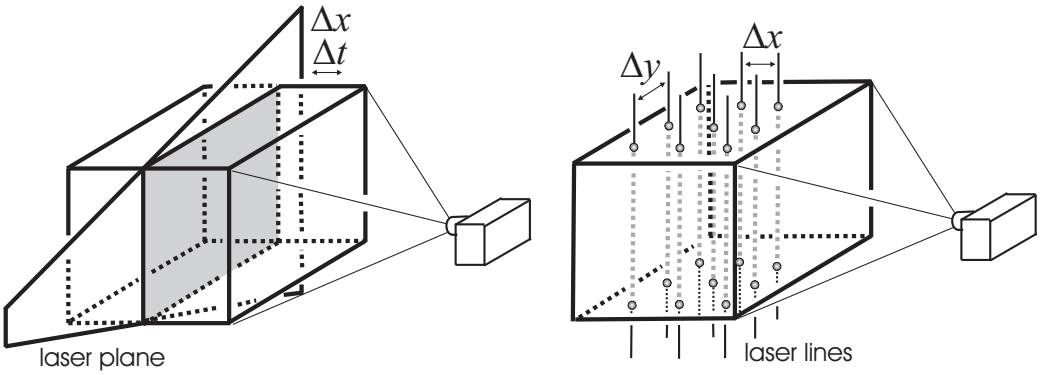


Figure 4.1*: Different acquisition approaches.* Left: *Sequential scanning of the volume with a plane of light as implemented by Hawkins et al. [Hawkins et al. 2005] with discrete sampling in time (Δt) and one spatial dimension (Δx).* Right: *New approach with simultaneous capture of volume density using line grids as illumination and discrete sampling in two spatial dimensions ($\Delta x, \Delta y$).*

surement systems such as Yip et al. [Yip et al. 1987] or laser induced fluorescence (LIF) [Deutsch and Dracos 2001] follow a similar approach and capture the whole volume *sequentially* slice by slice, from a single view.

Our method is inspired by single view techniques but takes a fundamentally different sampling approach: The volume is illuminated with a grid of laser lines in order to sample the whole volume *simultaneously.* Essentially, the 2D laser plane is discretized and spread out to discrete locations in space. The volume is captured with a standard camera. As illustrated in Figure 4.1, this trades (potentially) continuous sampling in one spatial domain (Δy) against continuous sampling in the time domain (Δt). Careful placement of camera and light sources avoids occlusions when the laser lines are projected onto the image plane and the full but sparser sampled 3D information is captured with a single image. The sampling density can be increased by projecting multiple grids of differently colored illumination into the volume.

This new sampling paradigm has several consequences:

Decoupling of spatial and temporal sampling: The system captures the volume with a single camera image. It therefore enables continuous sampling in the time domain allowing both integration over long time intervals for weak signals and extremely short acquisition times for fast-changing datasets.

Increased time resolution: Using the same camera hardware, frame rates can be increased by 1–2 orders of magnitude. Furthermore, the system contains no mechanically moving parts, i.e., the achievable time resolution is only limited by the imaging hardware and the illumination intensity.

Limited spatial resolution: Spatial resolution is fundamentally limited by the number of laser lines that can be resolved by the imaging system. Sampling is denser along the direction of the laser lines but spatial detail between sampling lines is lost, resulting in potential aliasing.
Density instead of photo-consistency: The actual density of scatterers is reconstructed (up to a scale factor) instead of a photo-consistent representation of the volume as in Hasinoff and Kutulakos [Hasinoff and Kutulakos 2003].

In the remainder of the chapter we first describe the basic concepts underlying our capture approach (Section 4.2). We then analyze its properties and study the resulting errors using a synthetic smoke dataset (Section 4.3). We describe our prototypical acquisition system and show results for real captured datasets (Section 4.4). The chapter concludes with a discussion (Section 4.5) and an outlook to future work (Section 4.6).

4.2 System Description

Figure 4.1, right, gives an overview over the measurement setup. We assume that the measurement volume contains a spatially varying density of scattering particles that we would like to measure. Apart from their density, the scattering properties of the individual particles should be constant (or be well approximated by a constant). Depending on the size of the particles, scattering will either be dominated by Mie scattering (for larger particles such as typically found in smoke) or by Raleigh scattering [Bohren and Huffman 1983]. In the remainder of this section, we first describe the principles of radiative transfer (Section 4.2.1) and develop our image formation model (Section 4.2.2). Section 4.2.3 shows how we can recover the density of scatterers along the laser lines illuminating the volume. We finally describe in Section 4.2.4 how we can recover the full density field from this information.

4.2.1 Radiative Transfer

We start our analysis with the equation of radiative transfer [Ishimaru 1978] which describes the change in specific intensity $I(\mathbf{r}, \hat{\mathbf{s}})$[1] for a small volume element ds at a position $\mathbf{r}$ in space and in a direction $\hat{\mathbf{s}}$:

$$\frac{dI(\mathbf{r}, \hat{\mathbf{s}})}{ds} = -\rho\sigma_t I(\mathbf{r}, \hat{\mathbf{s}}) \quad + \quad \frac{\rho\sigma_t}{4\pi} \int_{4\pi} p(\hat{\mathbf{s}}, \hat{\mathbf{s}}') I(\mathbf{r}, \hat{\mathbf{s}}') d\omega' \quad + \quad \epsilon(\mathbf{r}, \hat{\mathbf{s}}). \tag{4.1}$$

[1] Note that specific intensity can be converted into radiance by integrating over the spectrum of the radiation.

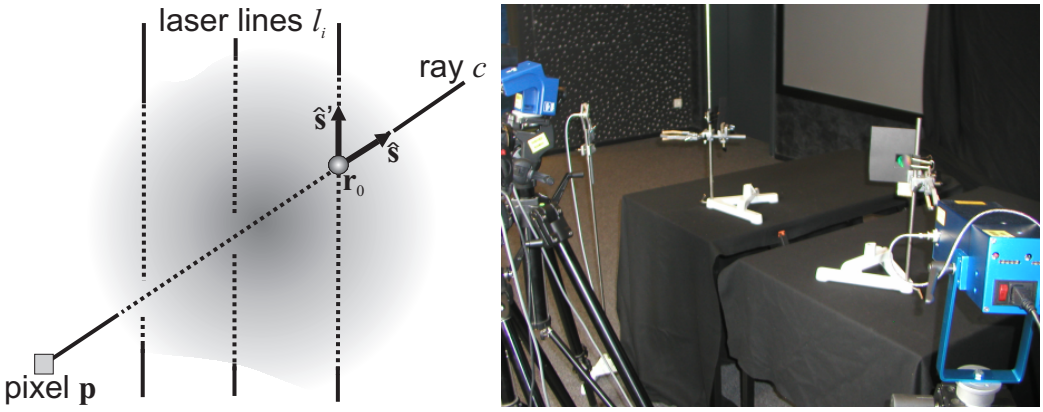


Figure 4.2*:* Left: *Situation for an individual ray c.* Right: *Image of the actual acquisition setup with camera on the left and two laser sources (left near camera and blue box on the right). When active, smoke is generated between the two tables.*

The first term models a decrease in specific intensity due to absorption and outscattering which is proportional to the number of particles per unit volume ρ and the extinction cross section σ_t. The second term describes the increase in specific intensity due to inscattering which depends additionally on the phase function $p(\hat{\mathbf{s}}, \hat{\mathbf{s}}')$. $\epsilon(\mathbf{r}, \hat{\mathbf{s}})$ is the emission from within ds.

In our measurement system (see Figure 4.1, right, and Figure 4.2 for the principle of the approach and the notation) we assume that light is scattered from homogeneous particles inside the volume. The extinction cross section σ_t and the phase function $p(\hat{\mathbf{s}}, \hat{\mathbf{s}}')$ are constant throughout the volume but the number of particles per unit volume $\rho = \rho(\mathbf{r})$ varies. We furthermore assume that the laser lines l_i are generated outside the observation volume so that there is no emission inside, i.e., $\epsilon(\mathbf{r}, \hat{\mathbf{s}}) = 0$. The goal of the measurements is to recover $\rho(\mathbf{r})$ up to a scale factor, i.e., it is sufficient to recover $D(\mathbf{r}) = \sigma_t \rho(\mathbf{r})$. Equation 4.1 simplifies then to the first order differential equation

$$\begin{aligned} \frac{dI(\mathbf{r}, \hat{\mathbf{s}})}{ds} = & - D(\mathbf{r}) I(\mathbf{r}, \hat{\mathbf{s}}) \\ & + \frac{D(\mathbf{r})}{4\pi} \int_{4\pi} p(\hat{\mathbf{s}}, \hat{\mathbf{s}}') I(\mathbf{r}, \hat{\mathbf{s}}') d\omega'. \end{aligned} \tag{4.2}$$

which we would like to solve under the boundary conditions given by our setup and assumptions.

4.2.2 Image Formation Model

We now take the specific situation of our measurement setup into account: The incident intensity I_i at a position $\mathbf{r}_0$ where the backprojected ray c of a pixel $\mathbf{p}$ with direction $-\hat{\mathbf{s}}$ intersects with a laser line l_i with direction $\hat{\mathbf{s}}_{l_i}$ can be computed as

$$I_i(\mathbf{r}_0, \hat{\mathbf{s}}) = \frac{D(\mathbf{r}_0)}{4\pi} I_{l_i}(\mathbf{r}_0, \hat{\mathbf{s}}_{l_i}) p(\hat{\mathbf{s}}, \hat{\mathbf{s}}_{l_i}). \tag{4.3}$$

This is, however, only valid if $\mathbf{p}$ covers the full width of the laser line l_i. We will show in Section 4.4.1 how we lift this restriction in practice.

We assume in the following that $I_{l_i}(\mathbf{r}_0, \hat{\mathbf{s}}_{l_i}) = I_{l_i}$ is constant along each laser line l_i and can be calibrated in a separate step. Following Ishimaru [Ishimaru 1978] we can now split the intensity along the ray c into two terms:

$$I(\mathbf{r}, \hat{\mathbf{s}}) = I_{\text{ri}}(\mathbf{r}, \hat{\mathbf{s}}) + I_{\text{d}}(\mathbf{r}, \hat{\mathbf{s}}). \tag{4.4}$$

The reduced intensity $I_{\text{ri}}(\mathbf{r}, \hat{\mathbf{s}})$ models the decay of I_i due to absorption and outscattering along c according to the first term in Equation 4.2:

$$I_{\text{ri}}(s) = I_i(\mathbf{r}_0, \hat{\mathbf{s}}) e^{-\tau} = I_i(\mathbf{r}_0, \hat{\mathbf{s}}) e^{-\int_0^s D(\mathbf{r}) ds}. \tag{4.5}$$

s measures the distance from $\mathbf{r}_0$ along the ray c to the pixel $\mathbf{p}$. Note that unlike [Ishimaru 1978], we treat $I_i(\mathbf{r}_0, \hat{\mathbf{s}})$ as a radiation source inside the volume. The remaining contributions caused by inscattering from the volume are accumulated in the diffuse intensity $I_{\text{d}}(\mathbf{r}, \hat{\mathbf{s}})$. The specific intensity that reaches pixel $\mathbf{p}$ can therefore be described as:

$$I_{\mathbf{p}} = I_{\text{ri}}(s) + I_{\text{d}}(s). \tag{4.6}$$

4.2.3 Recovering $D(\mathbf{r})$ along Laser Lines l_i

The goal of this section is to recover the scaled density values $D(\mathbf{r})$ along the laser lines l_i. We rewrite Equation 4.6 using Equations 4.3 and 4.5:

$$I_{\mathbf{p}} = \frac{D(\mathbf{r}_0)}{4\pi} I_{l_i} p(\hat{\mathbf{s}}, \hat{\mathbf{s}}_{l_i}) e^{-\int_0^s D(\mathbf{r}) ds} + I_{\text{d}}(s). \tag{4.7}$$

This can be rewritten as

$$D(\mathbf{r}_0) = 4\pi \frac{I_{\mathbf{p}} - I_{\text{d}}(s)}{I_{l_i} p(\hat{\mathbf{s}}, \hat{\mathbf{s}}_{l_i})} \cdot e^{\int_0^s D(\mathbf{r}) ds}. \tag{4.8}$$

The phase function $p(\hat{\mathbf{s}}, \hat{\mathbf{s}}')$ can either be assumed to be isotropic, theoretically derived from the properties of the scattering media [Bohren and Huffman 1983],

or be measured experimentally [Hawkins et al. 2005]. Note, however, that our acquisition setup evaluates $p(\hat{\mathbf{s}}, \hat{\mathbf{s}}')$ only over a small range of angles (all possible intersection angles of a laser line l_i and a ray c in Figure 4.2). Locally approximating $p(\hat{\mathbf{s}}, \hat{\mathbf{s}}')$ with a constant yields therefore only a small error. The error can be minimized by placing the camera far away from the scene in order to approximate an orthographic projection. I_{l_i} can be determined in a calibration step for each laser line l_i.

The diffuse intensity $I_{\mathrm{d}}(s)$ can be estimated from the set of camera pixels $N(\mathbf{p})$ in the spatial neighborhood of $\mathbf{p}$ whose backprojected ray does not intersect with any laser line l_i:

$$I_{\mathrm{d}}(s) \approx \tilde{I}_{\mathrm{d}}(s) = \frac{1}{||N(\mathbf{p})||} \sum_{\mathbf{p}' \in N(\mathbf{p})} I_{\mathbf{p}'}. \tag{4.9}$$

This assumes that $I_{\mathrm{d}}(s)$ varies smoothly over the image plane and corresponds to the removal of multiple scattering in the calibration part of Hawkins et al. [Hawkins et al. 2005]. We furthermore need to set $\tilde{I}_{\mathrm{d}}(s) = I_{\mathbf{p}}$ if $\tilde{I}_{\mathrm{d}}(s) > I_{\mathbf{p}}$ to avoid physically implausible results, e.g., due to noise in the acquisition. Under the assumption of optically thin scattering material, we can furthermore set $\int_0^s D(\mathbf{r})ds = 0$ without introducing a too large error. Equation 4.8 then becomes

$$D(\mathbf{r}_0) \approx 4\pi \frac{I_{\mathbf{p}} - \tilde{I}_{\mathrm{d}}(s)}{I_{l_i} p(\hat{\mathbf{s}}, \hat{\mathbf{s}}_{l_i})} \tag{4.10}$$

which allows us to recover the scaled density values along all laser lines.

4.2.4 Reconstruction of the Entire Density Field $D(\mathbf{r})$

The previous section introduced an approach to capture D along the laser lines. We will now discuss both interpolation and approximation approaches to recover $D(\mathbf{r})$ from this information.

Interpolation

Given the density values D along the laser lines, we can employ an interpolation technique to interpolate $D(\mathbf{r})$. The push-pull algorithm [Gortler et al. 1996, Drori et al. 2003], for example, fills in missing parts in an image by iterative downsampling and upsampling of the image using a filter kernel at multiple resolutions. It can be easily generalized from 2D image data to 3D volume data.

The push-pull algorithm is able to fill in large missing parts at coarse resolution levels while it approximates the high-frequency detail at the finer levels. Before each iteration, the known values are reintroduced into the image resp. volume

thus making the push-pull algorithm an interpolation method that preserves the original, measured values. Any such interpolation technique will however also preserve the non-uniform sampling inherent in our data structure. High frequency details will only be available along the laser lines and yield noticeable artifacts. Figure 4.4 (b) shows a reconstruction result of the push-pull algorithm. The lines on which the original densities have been sampled can still be seen going from left to right.

Approximation

We therefore opted for a more general reconstruction approach that approximates the sample values and yields a smoother reconstruction. We formulate our approximation problem as follows: Given are n discrete samples of measured density values $D(\mathbf{p}_i)$ at locations $\mathbf{p}_i$ ($i = 0, \ldots, n-1$) on the laser lines. We then approximate the field at a position $\mathbf{r}$ as

$$\widetilde{D}(\mathbf{r}) = \frac{\sum_{i=0}^{n-1} D(\mathbf{p}_i) \cdot w(\|\mathbf{r} - \mathbf{p}_i\|)}{\sum_{i=0}^{n-1} w(\|\mathbf{r} - \mathbf{p}_i\|)}. \tag{4.11}$$

The weighting function w is defined as

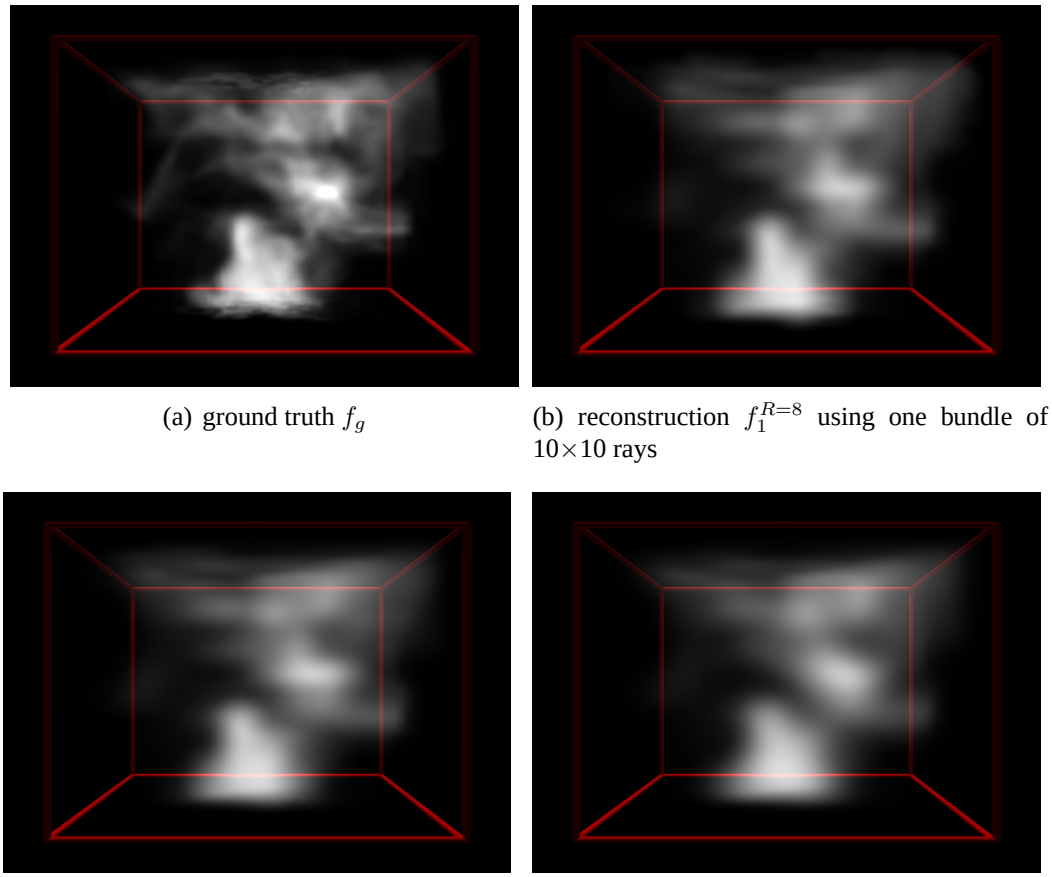
$$w(x) = \begin{cases} 0.5 \cdot \cos(\frac{x \cdot \pi}{R}) + 0.5 & \text{for } x < R \\ 0 & \text{else} \end{cases}. \tag{4.12}$$

The parameter R which determines the width of the reconstruction kernel needs to be manually selected for a given sampling configuration. Note that R has to be chosen such that in the R-neighborhood of every point $\mathbf{r}$ of the domain there is at least one sample point $\mathbf{p}_i$. Figure 4.4 (c) shows a result of this reconstruction approach. It looks much more smooth and lacks the artifacts due to the irregular sampling of the original volume.

4.3 Simulation

To perform an analysis of a dataset with ground truth we used a 100 frame simulation of smoke emitted from the border of a volume using the technique of Treuille et al. [Treuille et al. 2006]. The data was stored as a 64^3 voxel density field. Assuming that each dataset is defined over the domain $[0, 1]^3$ and assuming a trilinear interpolation between the grid points, we have continuous scalar fields f_g for each time frame which act as ground truth.

In order to reconstruct one scalar field, we consider a bundle of 100 rays starting from $(\frac{1}{2}, \frac{1}{2}, -3)$ and passing through $(\frac{i}{9}, \frac{j}{9}, 1)$ for $i, j = 0, \ldots, 9$. Then we



(a) ground truth f_g

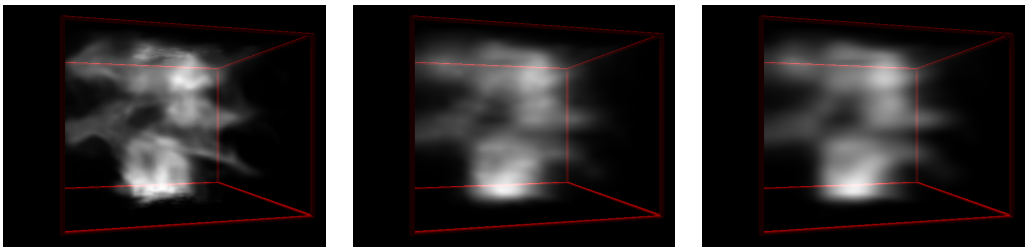
(b) reconstruction $f_1^{R=8}$ using one bundle of 10×10 rays

(c) reconstruction $f_2^{R=8}$ using two bundles of 10×10 rays

(d) smoothed ground truth $f_g^{R=8}$

Figure 4.3*: Result for the simulated smoke dataset.* (a): *Rendering of Frame 60 of the synthetic smoke dataset.* (b) and (c): *Reconstruction using one bundle and two bundles of 10×10 rays, respectively, and* $R = 8$*.* (d): *Rendering of the smoothed original volume with* $R = 8$*.*

sampled f_g in its domain along the rays and apply our approximation technique with parameter $R = 8$ to get a reconstructed field $f_1^{R=8}$. In a second test, we add a second bundle of 100 rays starting at $(-5, \frac{1}{2}, \frac{1}{2})$ and passing through $(1, \frac{i}{9}, \frac{j}{9})$ for $i, j = 0, \ldots, 9$. The field reconstructed from these 200 rays with $R = 8$ is $f_2^{R=8}$. Figure 4.3 shows f_g, $f_1^{R=8}$, and $f_2^{R=8}$ for Frame 60 of the dataset. While both reconstructions faithfully represent the overall structure of the field, it is clear that many high-frequency details are lost. We therefore computed a smoothed version of the ground truth field by convolving it with a normalized version of the reconstruction kernel w (Equation 4.12). Figure 4.3(d) depicts the smoothed ground truth field $f_g^{R=8}$ for a kernel radius $R = 8$ which is well approximated by both,



(a) ground truth f_g (b) interpolation f_2^{pp} using the push-pull algorithm with 2 bundles of 10×10 rays (c) approximation $f_2^{R=8}$ using 2 bundles of 10×10 rays

Figure 4.4*: Frame 60 of the synthetic smoke dataset, rotated by 75 degrees.* (a): *Ground truth.* (b) and (c): *Reconstruction using the push-pull algorithm and our approximation technique. Note the line artifacts in the interpolated result* f_2^{pp}.

$f_1^{R=8}$ and $f_2^{R=8}$.

For comparison, we also used the push-pull algorithm described in Section 4.2.4 on the same dataset. For the reconstruction, we sampled each volume with 2 bundles of 100 rays as before so that both reconstruction algorithms work on the same input data. We show the interpolation result using the push-pull algorithm for Frame 60 of the dataset (f_2^{pp}) in Figure 4.4(b). The result is not so smooth as $f_2^{R=8}$ for the same frame shown in Figure 4.4(c). The lines on which the volume has been sampled can be still be seen as horizontal artifacts.

4.3.1 Error Analysis

We define the RMS error between two scalar fields f and f' stored as a voxel densities sampled at a set of identical locations $\mathbf{V}$ as

$$\mathrm{RMS}(f, f') = \sqrt{\frac{\sum_{v_i \in \mathbf{V}} (f(v_i) - f'(v_i))^2}{\|\mathbf{V}\|}}. \tag{4.13}$$

In practice, $\mathbf{V}$ corresponds to the set of 64^3 voxels defining our field. Figure 4.5 depicts the RMS error between various versions of the field for all 100 frames in the dataset. The density values in the original dataset vary between 0 and 1.6.

The RMS errors between the ground truth f_g and the reconstructions from one and two ray bundles ($f_1^{R=8}$ and $f_2^{R=8}$) are almost identical. The RMS error decreases drastically when it is computed against the smoothed version of the ground truth $f_g^{R=8}$. Furthermore, the reconstruction from two ray bundles $f_2^{R=8}$ compares now much better than the reconstruction from a single ray bundle $f_1^{R=8}$. This suggests that most of the error in the reconstructions is due to the sparse sampling and smooth approximation that suppresses high frequency detail. To

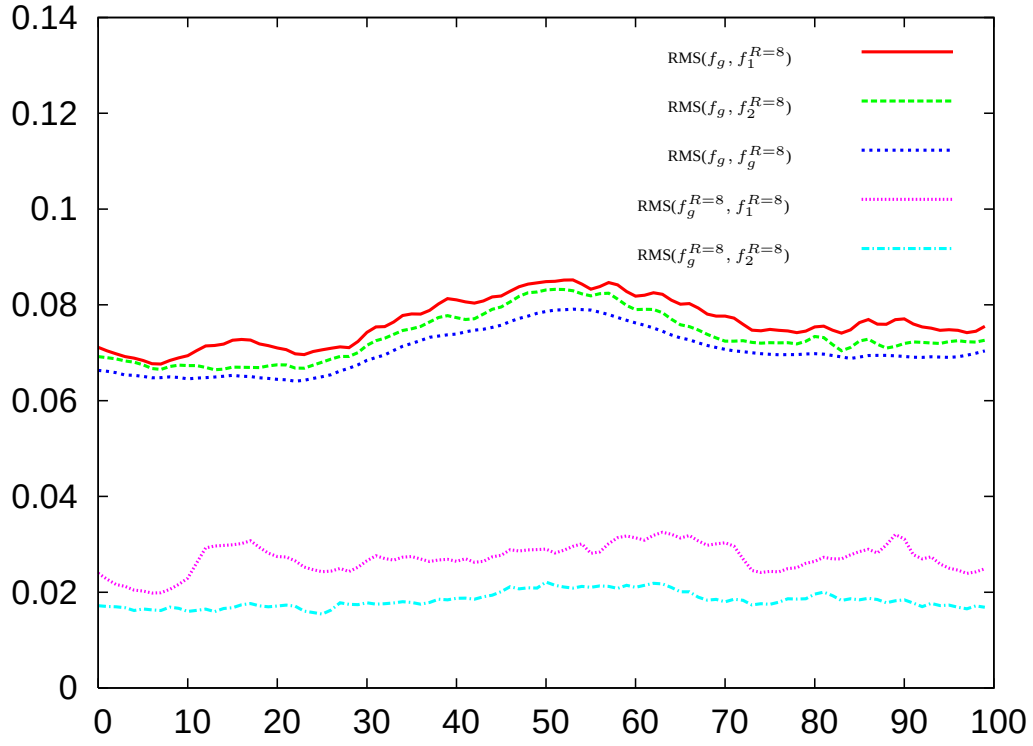


Figure 4.5: *Various RMS errors for the 100 frames contained in the simulated dataset. Horizontal axis: frame number. Vertical axis: RMS error. The density values in the original dataset vary between 0 and 1.6. The first two lines show the RMS error between ground truth f_g and the reconstruction using 1 and 2 bundles of laser rays, RMS($f_g, f_1^{R=8}$) and RMS($f_g, f_2^{R=8}$), respectively. The third line represents the RMS error between ground truth and a version of the ground truth which has been smoothed using the same kernel that is used in the reconstruction (RMS($f_g, f_g^{R=8}$), Equation 4.12). The two lines on the bottom of the graph show the RMS error between the smoothed ground truth and the reconstruction using 1 and 2 bundles of laser rays RMS($f_g^{R=8}, f_1^{R=8}$) and RMS($f_g^{R=8}, f_2^{R=8}$), respectively.*

verify this, we computed the RMS error RMS$(f_g, f_g^{R=8})$ between the ground truth field and the smoothed version of the ground truth field. Figure 4.5 shows that this error is only slightly lower than RMS$(f_g, f_2^{R=8})$, i.e., most of the reconstruction error seems to be due to the loss of high frequency detail.

Figure 4.6 shows the RMS error for the approximation technique and the interpolation using the push-pull algorithm. The error produced by the push-pull interpolation is lower than that of the reconstruction throughout the whole dataset. This might be an indication that the push-pull algorithm does a better job reconstructing the high-frequency detail.

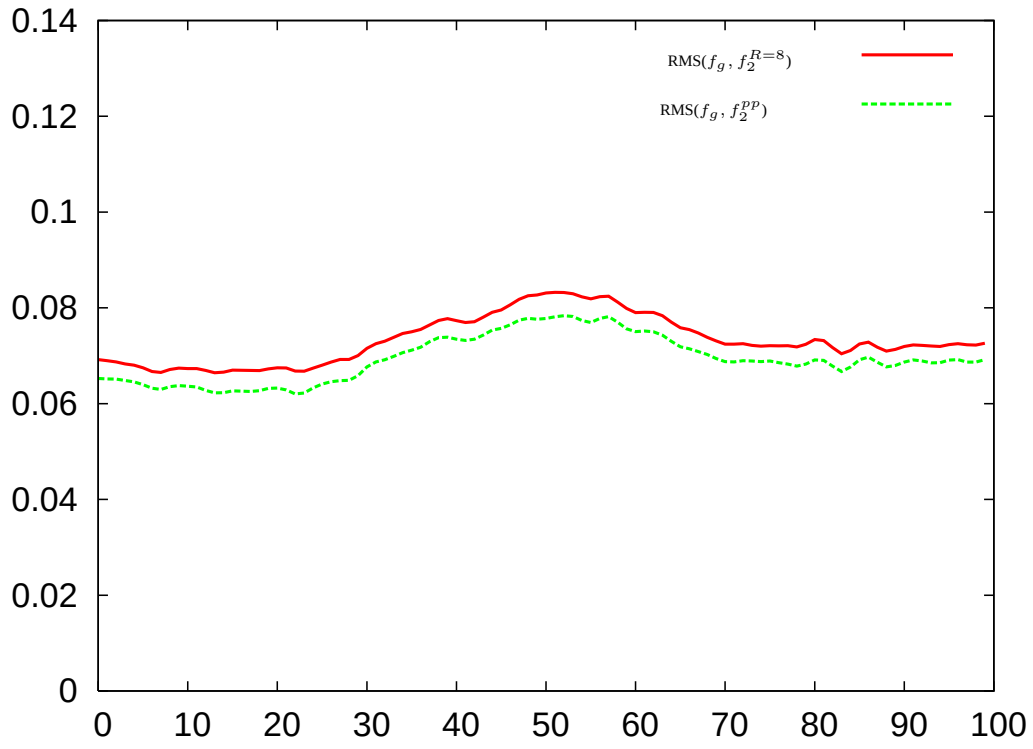


Figure 4.6: *Comparison of the RMS errors from different reconstruction techniques for the 100 frames contained in the simulated dataset. The approximation was computed using 2 bundles of laser rays (RMS($f_2^{R=8}$)). The error for the push-pull algorithm is shown by RMS(f_2^{pp}). Horizontal axis: frame number. Vertical axis: RMS error.*

4.4 Acquisition System

We built a prototype of an acquisition system to test our ideas in practice. The setup contains two collimated laser sources – a 130 mW red laser and a 30 mW blue laser. Each collimated laser beam is transformed into a bundle of perspective rays using a burst grating. A black mask limits the number of rays in each bundle to 5×5 rays for each laser. The generated ray bundles are roughly perpendicular to each other; their intersection defines a measurement volume of about $50\times24\times50$ cm^3 (see Figure 4.2, right, for an image of the measurement setup). We use a smoke machine to create a spatially varying density of scattering media.

A high quality color CCD camera is used to capture images of the measurement volume. Its placement ensures that no two rays of the same color project to the same location on the image plane. We are thus able to capture the two bundles independently using the camera's red and blue channel. In addition, we illuminate the measurement volume diffusely with green LEDs and use the green channel of

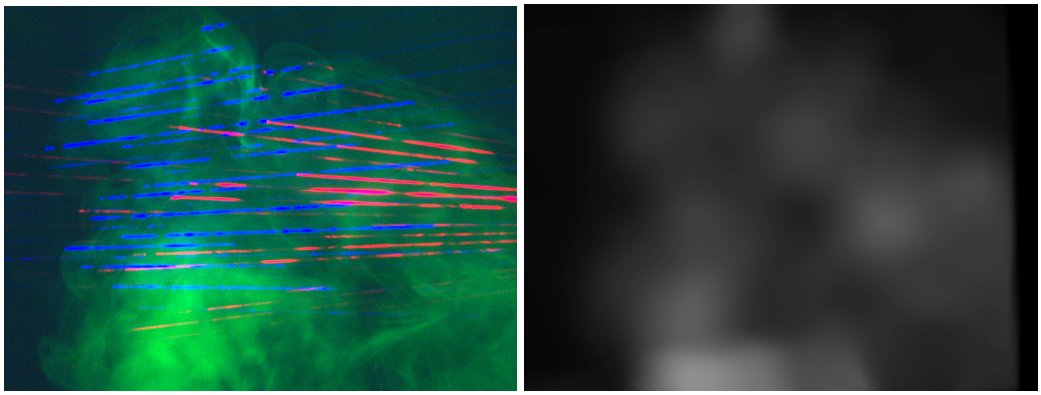


Figure 4.7*: Acquired image of smoke and reconstructed density field.* Left: *Input image with two independent grids of laser lines (red, blue) sampling the volume and diffuse green illumination (image intensity rescaled for display).* Right: *Density field reconstructed from this single image.*

the camera to capture simultaneously a ground truth image of the acquired dataset. Figure 4.7, left, shows an example input image of the system. Although all three light sources have a narrow spectral response, we observe crosstalk between the two lasers which we remove using standard image processing techniques.

4.4.1 Calibration and Capture

For geometric calibration of the camera and the lasers we capture several high-dynamic range images [Debevec and Malik 1997] of each bundle of laser rays illuminating a yellow and white checkerboard pattern. This allows us to reliably detect the corners of the calibration pattern as well as the centers of the laser spots. After computing the 3D location of these spots using a calibration toolbox [Bouguet 2006] we can estimate the position of the laser lines in space relative to the camera. Figure 4.8 (left) shows a visualization of the spatial sampling in which the two bundles of laser rays are clearly visible. Using the calibration data, we can now extract samples $I_{\mathbf{p}}$ from the camera images by marching densely along the projections of the rays and taking a sample at each step. As noted in Section 4.2.2, we need to ensure that the full width of the projected laser line is captured. We therefore integrate the contributions to $I_{\mathbf{p}}$ along a small line segment perpendicular to the projected ray direction. We then low-pass filter and downsample the obtained intensity samples along each ray in order to reduce noise and facilitate further processing, yielding about 150 samples per ray.

The burst gratings create bundles of rays where each laser line can have a different intensity. We therefore need to recover the relative intensity I_{l_i} of all laser lines in each bundle in the calibration phase. To this end, we capture a

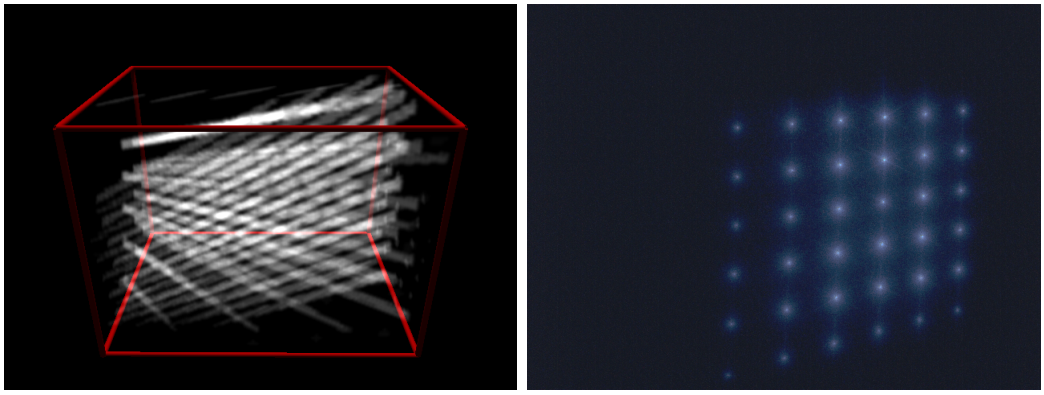


Figure 4.8: *Geometric and photometric calibration of the setup.* Left: *Visualization of the spatial sampling. The two bundles of laser rays are clearly visible.* Right: *Calibration image for the blue laser to recover the relative intensities I_{l_i} for the individual rays (tone-mapped).*

high-dynamic range image per laser of a sheet of subsurface scattering plastics[2] illuminated by the corresponding ray bundle (see Figure 4.8 right). We integrate the laser intensity over a constant area around each laser spot to get I_{l_i}. The two lasers might still behave differently, e.g., due to different scattering properties for different wavelengths. In a subsequent calibration step, we therefore capture images of smoke and recover separate density fields D_{red} and D_{blue} for the red resp. the blue laser bundle. We then determine a scale factor k that minimizes the RMS error between D_{red} and $k \cdot D_{\text{blue}}$ and scale the input data for the blue channel accordingly before reconstructing the density field.

4.4.2 Acquisition Results

We captured several datasets with the acquisition system described in Section 4.4 and depicted in Figure 4.2 (right). All images were taken with 0.25 s exposure time. Captured results are shown in Figures 4.7 and 4.9. The results are rendered using a raytracing-based direct volume rendering approach.

Figure 4.7 shows an input image capturing smoke from a smoke machine. The intensity variations along the laser lines are clearly visible. In addition, the blue laser lines are brighter and tighter focused than the red lines. This is corrected by integrating over the width of the laser line (Section 4.4.1 and by the intensity scaling described in Section 4.4). In the result image on the right, note, how much detail could be reconstructed from this single input view.

[2]The subsurface scattering material spreads out the laser intensity over a greater area. Thereby the dynamic range of intensity that has to be measured is reduced.

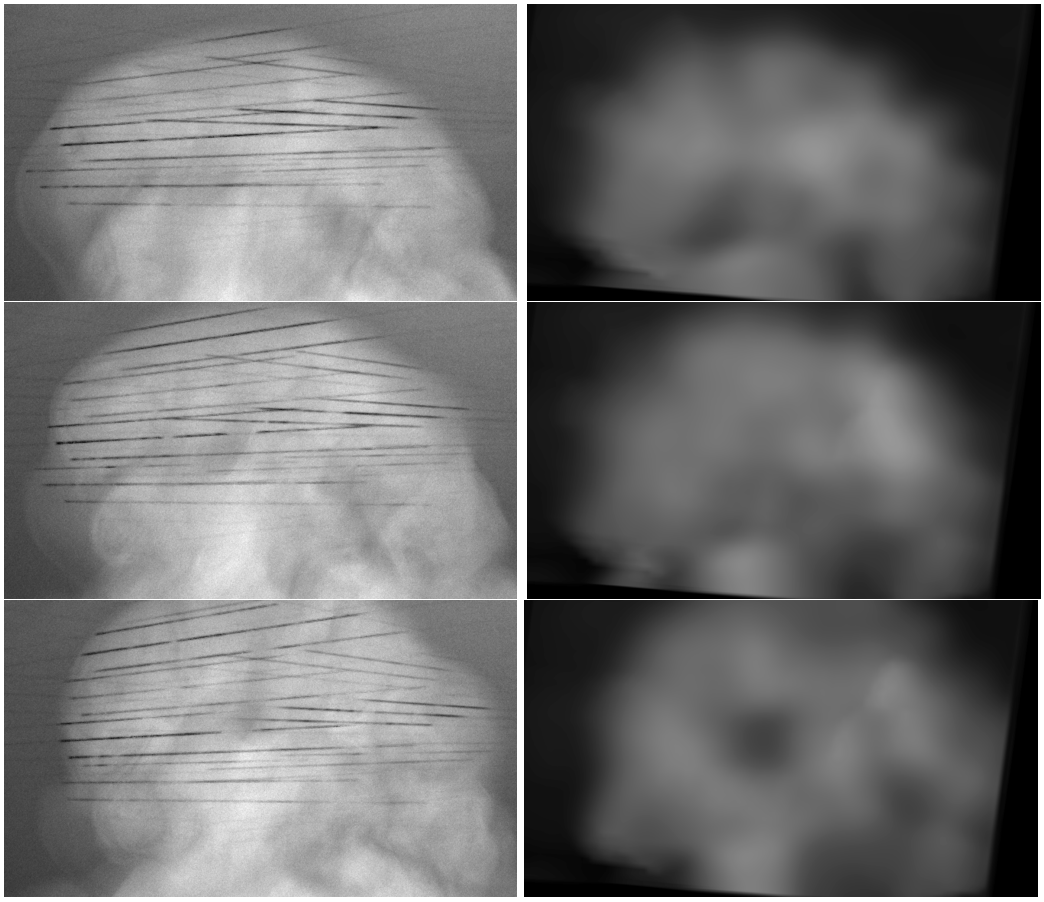


Figure 4.9*: Results for smoke from a smoke machine. The three images were taken in sequence from top to bottom. The images on the left show ground truth photographs of the scene extracted from the green channel. The images on the right show renderings of our reconstruction.*

Figure 4.9 shows comparisons between ground truth photographs and renderings for three images of dataset consisting of 50 images captured at approximately 3 fps. Note that the speed is mainly constrained by the frame rate of the camera and not by the measurement principle. The ground truth photographs were extracted from the camera's green channel as described in Section 4.4. The reconstructed dataset captured the overall shape of the smoke as well as prominent features. Its resolution is, however, limited due to the sparse sampling using only 5×5 grids of laser lines.

4.5 Discussion

We presented a new approach to sample dynamic volumetric density fields using grids of laser lines as illumination. This allows us to continuously sample in the time domain at the cost of sparser sampling on the spatial domain.

Like other measurement systems such as Hawkins et al. [Hawkins et al. 2005] or Narasimhan et al. [Narasimhan et al. 2005], we are assuming that the scattering behavior inside the measurement volume is dominated by single scattering. This limits both, the size of the measurement volume and the density of the scatterer inside the volume, due to two effects: First, the intensity of the laser rays I_{l_i} and the scattered intensity I_{p} decrease inside the measurement volume due to outscattering and absorption yielding a systematic bias in the reconstructed field $D(\mathbf{r})$. We also currently neglect the absorption of the laser ray along its path from the source to the camera, which causes errors if the density distribution is not sufficiently homogeneous. Modeling and inverting this effect for spatially varying densities is difficult even if all scattering parameters are known.

Second, as the diffuse intensity I_{d} increases, the signal-to-noise ratio for the measurement of the directly scattered intensity decreases significantly. This is a fundamental problem to all approaches of this kind which can be only partially compensated for, e.g., by subtracting the diffuse illumination [Hawkins et al. 2005]. Narasimhan et al. [Narasimhan et al. 2006] recognized this problem and solve it by avoiding multiple scattering in the first place. They dilute their working medium until single scattering clearly dominates its scattering behavior.

4.6 Future Work

There are several directions for further research improving both the acquisition setup and the reconstruction of the complete density field.

A different camera system would allow us to operate the whole setup at a higher frame rate removing motion blur from captured images.

The properties of the gratings that are used to generate the ray bundles have a great impact on the size of the volume that can be effectively measured and on the number of lines that can be used. The rapid decrease in intensity for the outer lines of the grid imposes a strong limit on the number of useful rays. The angle between individual lines of the grid limits the size of the volume which can be sampled at a certain spatial sampling rate. All these properties do not only depend on the grating itself, but also on the wavelength of the laser source. Thus, the gratings have to be carefully chosen to match an existing laser.

Due to all these problems it might be preferable to use individual laser sources

(e.g., a set of laser pointers) to generate the lines independently instead of a single laser and a grating. This would allow a much denser sampling grid. The number of laser lines is then only limited by their projected width in the captured image. It could also greatly facilitate the setup of the whole system, since an occluded line could be moved independently until it is visible. Furthermore it would minimize the need for a calibration of the intensities.

The current data processing approach makes no assumption about the structure of the density field $D(\mathbf{r})$ and yields therefore a smooth reconstruction of the smoke volume (especially in sparsely sampled dimensions). Reconstruction algorithms that make use of prior knowledge of the structure of the data (e.g., [Owada et al. 2004]) can improve the visual quality of the reconstructed density field. The structure could also be inferred by analyzing the frequency content in denser sampled dimensions in the spirit of Dischler et al. [Dischler et al. 1998]. Alternatively, we could try to synthesize the missing data while forcing the statistics of the completed volume to match the input or statistics from the smoke simulation, similar to the texture synthesis algorithm proposed in [Kopf et al. 2007].

Chapter 5

Combining Confocal Imaging and Descattering

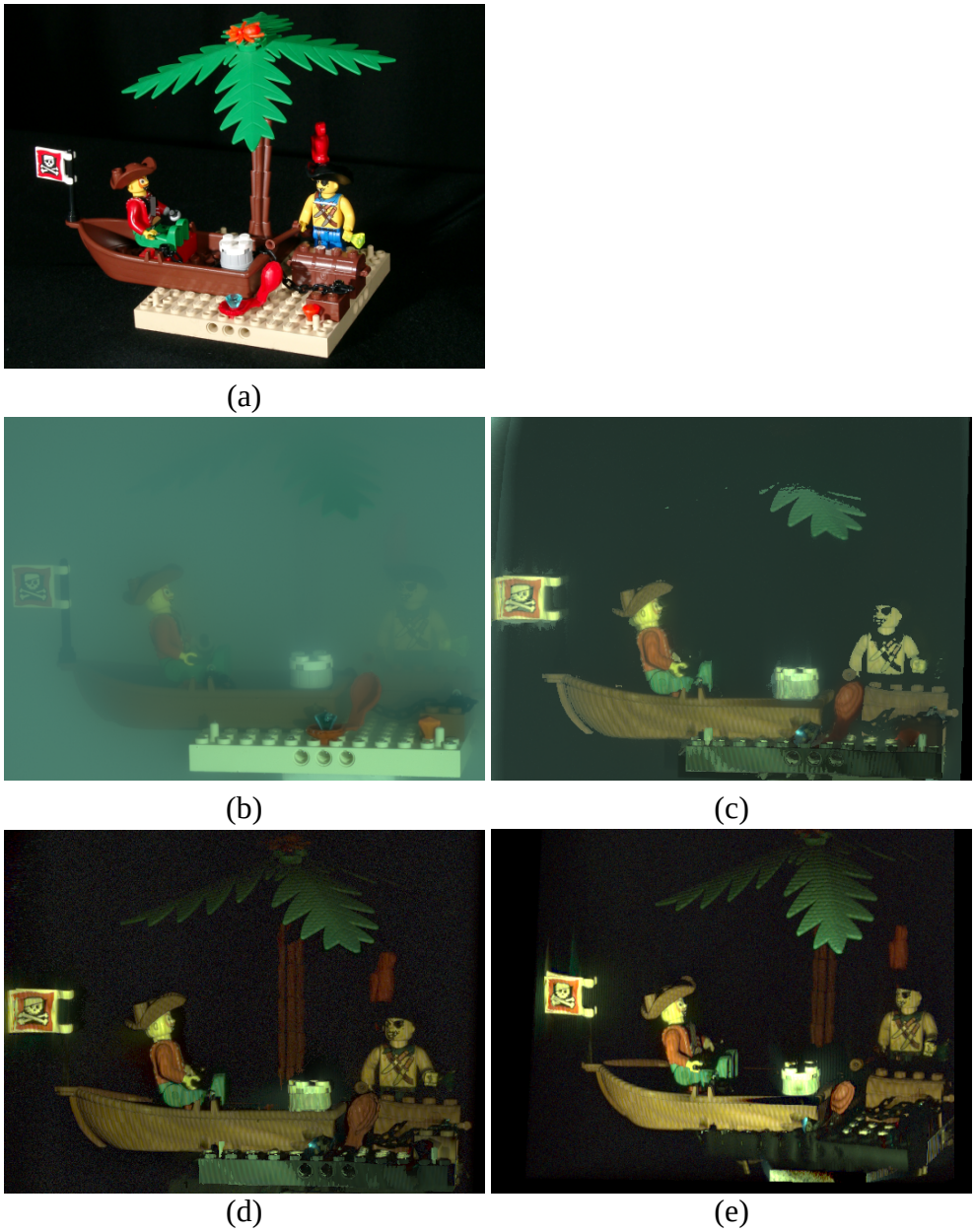
In the previous chapter, we focused on acquiring spatially varying densities in a full volume with just a single image. In the following chapter, we change our sampling paradigm and therefore drop the ability to image a specific instant in time. Furthermore, we employ multiple cameras to observe the now static scene and use projectors instead of lasers to provide active illumination.

In the following, we propose a technique for computing cross-sectional images of translucent objects or scenes for visible wavelengths by combining *confocal imaging* and algorithmic *descattering* [Fuchs et al. 2008].

5.1 Introduction

Confocal imaging [Corle and Kino 1996, Wilson et al. 1996, Neil et al. 1997] allows for computing optical sections of partially *transparent* volumes. Arbitrary slices through the volume are assembled by focusing the illumination and the observation rays from a large aperture onto individual voxels. Due to the large aperture, all contributions of points off the selected plane are significantly blurred and darkened, and their influence on the confocal image is drastically reduced. In our approach, we perform synthetic aperture confocal imaging [Levoy et al. 2004], where the aperture is sampled by a small set of cameras and projectors. Individual voxels are extracted by combining the viewing and illumination rays that intersect at the voxel.

As can be seen in Figure 5.1(c), confocal imaging works to some extent even in translucent media where the light transport is governed by multiple scattering. Confocal imaging can be seen as extracting a particular region, a hypersurface,



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

Figure 5.1: *Reconstructed volume of a 3D scene in a fish tank with diluted milk. While in the floodlit image (b) only the front objects are recognizable, confocal imaging (c) partially removes the haze. Using our novel descattering algorithm combined with confocal imaging more global scattering is removed and objects at even larger distances become visible (d). The image is sharper and features more saturated colors. In (e), the recovered 3D structure is visualized for a different view. Images (c)-(e) have been created by computing a confocal or descattered confocal volume followed by maximum intensity projection.*

from the scene's reflectance field. Hereby, most of the scattering that degrades the floodlit image (Figure 5.1(b)) is avoided. The individual confocal samples however still contain global illumination effects due to in- and out-scattering along the selected paths. The recorded photons, instead of following the direct path given by the illumination and the viewing ray, travel through a larger volume, the so-called *photon banana* [Feng et al. 1993] (see Section 5.3).

The second component, algorithmic *descattering* aims at separating the reflections of an object observed for individual rays into the component, which is due to direct reflection and the contribution due to global illumination effects. In algorithmic descattering, high frequency illumination patterns are often applied [Nayar et al. 2006].

In this part of the thesis, we provide the following contributions:

- We develop a novel framework, where separation is performed on similar input data, by computing the difference of reflectance samples measured for adjacent illumination samples or light paths.
- A qualitative analysis based on Monte-Carlo photon simulation is performed which indicates that our approach corresponds to reducing the original photon banana to a small channel around the direct path.
- We show that confocal imaging and descattering are similar but orthogonal concepts. Our novel descattering algorithm can be easily combined with confocal imaging, exploiting exactly the same illumination patterns for both approaches.
- The performance of our combined approach is demonstrated by computing significantly improved views through scattering fluids and into translucent objects.

5.2 Related Work

5.2.1 Separation of Reflection Components

The separation of measurements into different reflection components such as specular, diffuse, subsurface, or interreflections has so far been addressed by a number of different techniques. Using images captured with a polarization filter at different orientations, one can for example separate diffuse from specular reflections [Rowe and Pugh Jr. 1995, Tyo et al. 1996, Nayar et al. 1997] or remove global illumination effects in participating media [Schechner et al. 2003, Schechner and Karpel 2005, Treibitz and Schechner 2006] since multiple scattering tends to depolarize the incoming light.

Separation can also be performed using high frequency illumination patterns, based on the observation that only direct reflections will propagate high frequencies while global effects drastically dampen them. Nayar et al. [Nayar et al. 2006] proposed computing the difference of measurements obtained from shifted high frequency patterns. By subtracting two measurements, the global component is removed. We compare our results to this method in detail later in this chapter.

Narasimhan and Nayar [Narasimhan et al. 2005] used swept line patterns for 3D-scanning through participating media. Based on the estimated distance they compute a ”clear-air” view. They estimate the scattering parameters of a *single* scattering model and correct only for extinction along the illumination and viewing ray given the estimated depth. We make use of the geometric calibration technique proposed in this paper. In addition, we will also use swept line patterns for acquisition. In contrast to their method, we combine descattering with confocal imaging recovering a volume rather than single surfaces. Furthermore, we do not assume an explicit scattering model. Rather than just accounting for extinction along a single path, our descattering method corrects for multiple scattering in a photon banana by analyzing a local neighborhood of paths.

In our approach we will analyze the task of descattering in the context of reflectance fields [Debevec et al. 2000]. Reflectance fields can be captured by scanning over all illumination rays as in [Goesele et al. 2004] or structured light patterns [Seberry and Yamada 1992]. Sen et al. [Sen et al. 2005] presented an adaptive algorithm that can parallelize the acquisition of multiple rays if they do not affect the same camera pixels. Since translucent objects result in rather dense reflectance fields almost no parallelization will be achieved. Garg et al. [Garg et al. 2006] developed a technique that determines reflectance samples of multiple illumination rays in parallel even for dense reflectance fields. The acquisition of a reflectance field in itself does not yet provide any means for descattering.

Seitz et al. [Seitz et al. 2005] presented a theory for inverse light transport computation based on measured reflectance fields. They propose a cancellation operator to remove the multiple scattering events from the reflectance fields, leaving the desired direct reflectance. This operator is computed using the inverse of the reflectance field, which does not always exist. In our framework we consider descattering as a local operation applied to a small neighborhood around each entry of the reflectance field; no inversion is necessary in our approach.

5.2.2 Volumetric Reconstruction

Confocal imaging, which we will explain in more detail in the next section, is related to other volumetric reconstruction techniques operating in the visible light range. Techniques for tomographic reconstruction of transparent objects

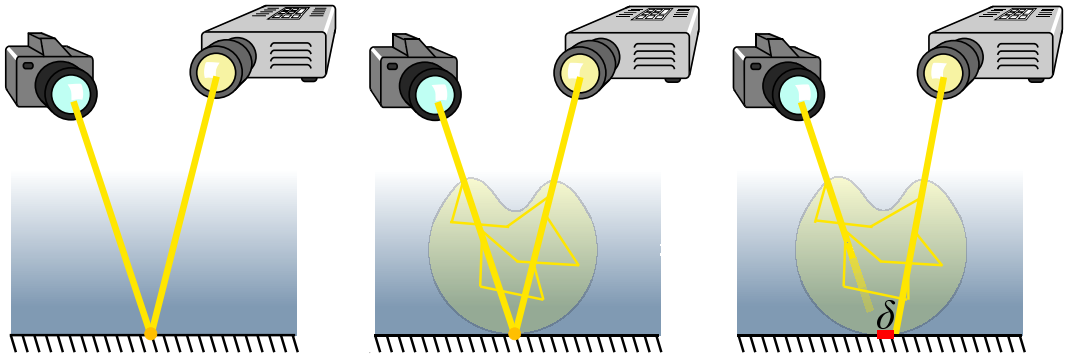


Figure 5.2*: In translucent objects a reflectance sample for one pair of intersecting rays will always contain a direct contribution from the scattering at the intersection point (a) plus some contribution due to multiple scattering (b), the so-called photon-banana [Feng et al. 1993]. (c) For a pair of adjacent but non-intersecting rays offset by δ, there will be no direct component but approximately the same global component. Combining both measurements, the direct component in (b) can be estimated.*

have been proposed by Sharpe et al. and Trifonov et al. [Sharpe et al. 2002, Trifonov et al. 2006]. In the first step, they determine the geometry of rays passing through the object by placing the possibly refracting object into some liquid with the same index of refraction. They measure the absorption along rays through the object from different directions and then perform algorithmic reconstruction techniques [Gordon et al. 1970]. Their technique only works for transparent objects since it is not resistant to multiple scattering.

Diffuse optical tomography [Arridge 1999] tries to perform volumetric reconstruction of translucent objects but considers only the diffuse light transport. It must solve an ill-posed inversion problem to obtain any localized information.

5.3 Confocal Imaging

With confocal imaging one can perform optical sectioning, i.e. capturing images of slices through a given volumetric object, or one can capture the full confocal volume and then apply volume rendering techniques. In confocal imaging the incident light and the observation rays from a large aperture are focused to a specific location within the object [Corle and Kino 1996]. The shallow depth of field of a large aperture blurs the contribution of any point not in focus, and at the same time darkens it relative to the illuminated focused point, since any point out of focus will receive light only from a small fraction of the aperture. In mostly transparent

scenes the influence of any point out of focus in the final image decays with r^4 where r is the distance to the focal plane.

Sectioned slices can be generated by illuminating individual points (for example by scanning or by means of a spinning Nipkow disk [Egner et al. 2002]), or by illuminating and measuring multiple points at once with varying patterns. In the latter case the confocal image is computed after decorrelating the measurements for individual voxels, which can be done using pseudo random noise patterns [Wilson et al. 1996, Heintzmann et al. 2001] or periodic patterns [Wilson et al. 1998, Neil et al. 1997].

In the scanning configuration the illumination is directed to a single point v from an illumination aperture Ω_i. The computed confocal irradiance $L_{conf}(v)$ observed with aperture Ω_o is the integral over all rays intersecting at v:

$$L_{conf}(v) = \int_{\Omega_o} \int_{\Omega_i} L(\omega_i, v, \omega_o) d\omega_i d\omega_o. \tag{5.1}$$

5.3.1 Synthetic Aperture Confocal Imaging

Levoy et al. [Levoy et al. 2004] performed confocal imaging with a synthetic aperture setup where the large aperture is sampled at a sparse and discrete set of directions from a set of cameras and projectors. In order to extract the information about a given voxel v, one selects the pair of rays $(\omega_o(v), \omega_i(v))$ that intersects at v. Given a set of cameras C and projectors P the confocal radiance is computed as:

$$L_{conf}(v) = \sum_{c \in C} \sum_{p \in P} L(\omega^p(v), \omega^c(v)). \tag{5.2}$$

Note, that the pair of rays $(\omega^p(v), \omega^c(v))$ that intersect at v will be different for each camera/projector pair due to the geometric setup. They are determined using the calibration data described in Section 5.6.1.

5.3.2 Separation Due to Confocal Imaging

The configuration for one pair of rays is depicted in Figure 5.2. In a purely *transparent* medium, the light transport will be governed by the reflection at v and the extinction along the incident light and the viewing ray (Figure 5.2(a)). Since each camera/projector pair observes v along a different path, the influence of the extinction along the paths is averaged in the confocal image, minimizing the influence of voxels other than v.

In the case of *translucent* media (Figure 5.2(b)), the light transport will include in- and outscattering along the path. The measurements will be affected by multiple scattering, effectively incorporating contributions from voxels within the

resulting photon banana [Feng et al. 1993]. The photon banana is formed by all photon traces that finally reach the camera pixel. It is clearly visible that the photon banana integrates the contributions of a much larger volume than the original direct path. At the same time, only a small fraction of the observed intensity will depend on the reflection at v.

When the observations of different aperture samples are combined, there is a large overlap in the photon bananas. The voxel v is no longer singled out but instead the net effect of the intersection of photon bananas will contribute to the confocal image. The influence of defocused voxels will no longer decay with r^4.

5.3.3 Descattering for Individual Rays

Besides the direct observation of the reflection at v and the extinction along the path, the reconstructed confocal image will contain multiple scattering events. We search for a separation into a non-local, or global, term L_g and the remaining local, or direct component L_d, which would correspond to the observation in transparent media:

$$L(\omega_o, \omega_i) = L_d(\omega_o, \omega_i) + L_g(\omega_o, \omega_i) \tag{5.3}$$

The key observation by Nayar et al. [Nayar et al. 2006] is that the global component acts as a low-pass filter on high-frequencies in the incident illumination, while only the direct component will keep them. Posed differently, the global component L_g will be approximately the same in the vicinity of paths around (ω_o, ω_i) while the direct component L_d will be different.

Let us assume two neighboring rays ω_i and $\omega_i + \delta$ with the following two properties:

1. ω_o and ω_i intersect at v while ω_o and $\omega_i + \delta$ do not. It follows that $L_d(\omega_o, \omega_i + \delta) = 0$ (Figure 5.2(c)).

2. If the scene is homogeneous in the δ-neighborhood around the path (ω_o, ω_i) one can state that their global component is the same, $L_g(\omega_o, \omega_i) = L_g(\omega_o, \omega_i + \delta)$.

The direct component can in theory be determined as:

$$L_d(\omega_o, \omega_i) = L(\omega_o, \omega_i) - L(\omega_o, \omega_i + \delta), \quad \text{or} \tag{5.4}$$

$$= L(\omega_o, \omega_i) - L(\omega_o + \delta, \omega_i). \tag{5.5}$$

Due to the duality of the light transport [von Helmholtz 1856, Sen et al. 2005] the role of viewing and illumination rays can be interchanged such that the difference can be computed for neighboring viewing rays as well.

The assumption of a homogeneous medium is often violated, especially at the places of interest, where there are volumetric features. In practice, we estimate the global component more robustly by incorporating a larger neighborhood. In the algorithm in Section 5.5 we construct a smooth neighborhood by weighted averaging.

5.4 Simulation

We use a Monte-Carlo photon simulator to qualitatively and quantitatively analyze the proposed descattering approach of Equation 5.4. The simulated scene represents the fish tank setup for which real measurements are demonstrated in Section 5.7.1. Photons enter a homogeneous scattering medium in the direction of a diffuse reflector which is embedded at the center of the volume (Figure 5.3). The volume extends over $20 \times 20 \times 20\ cm^3$, while the diffuse target has a size of $2 \times 2 \times 2\ cm^3$. The distance from the front plane of the volume to the target is $9\ cm$. We only consider photons that leave the volume at a predefined exit point, the camera pixel. Any other photon is discarded since it does not contribute to the measurement. For the selected photons, we record every scattering location and the energy along its trace.

5.4.1 From Photon Bananas to Photon Channels

In the first experiment (Figure 5.3), we simulated three neighboring incident light rays, marked red, green, and blue (c). The photon banana of the central illumination ray is shown in (b). By subtracting the averaged contribution of red and blue from the green photon banana, we reduce the green photon banana to a photon channel that encloses the original direct path (Figure 5.3(d)). Instead of the large volume of the full photon banana, only the much smaller volume covered by the photon channel will contribute to the final descattered image. In the confocal setting, the overlap of neighboring photon channels is significantly smaller than that of neighboring photon bananas.

While all photons within the channel might undergo multiple scattering events, they have a similar net effect as simple extinction in a transparent medium and therefore closely resemble the direct component.

5.4.2 Dependence on Scattering Parameters

In a second experiment, we investigate the dependence of the signal strength on varying scattering parameters of the medium: extinction coefficient

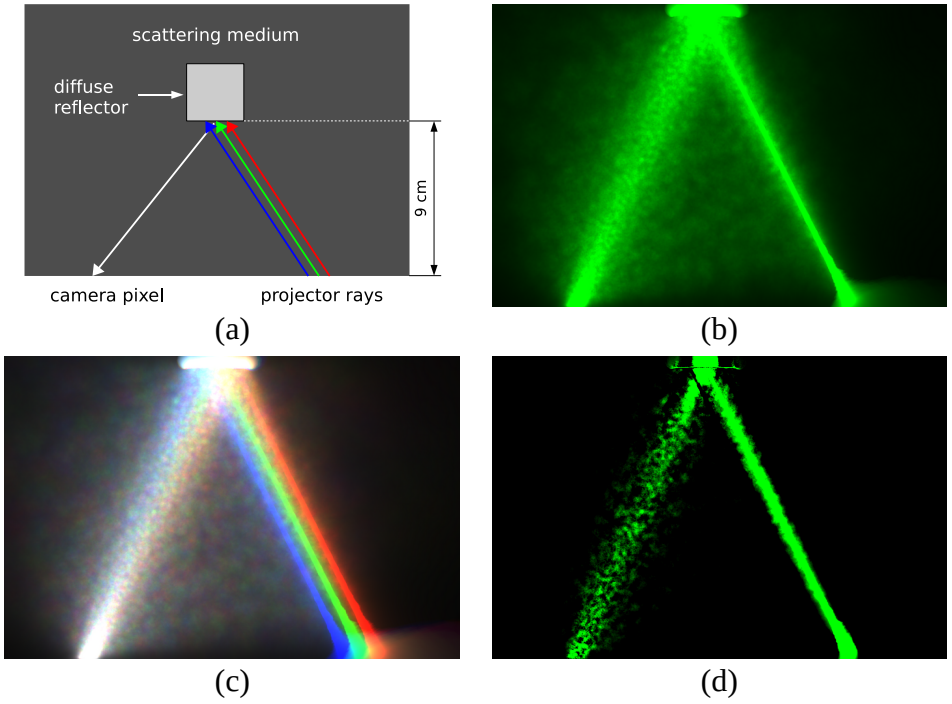
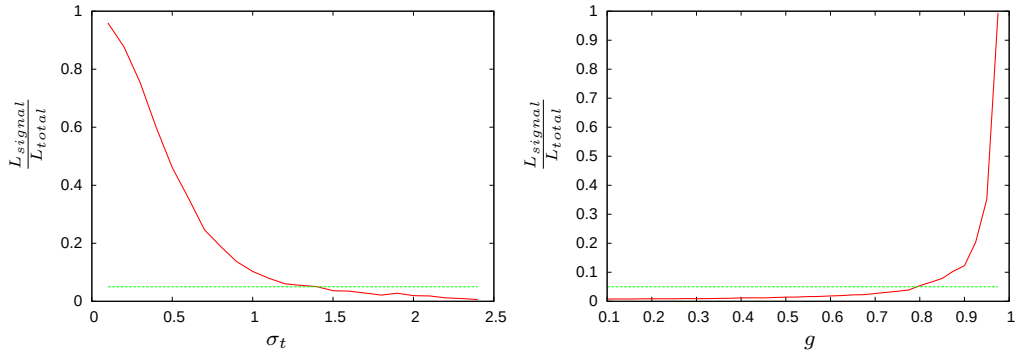


Figure 5.3: *(a) Simulation of photon traces from projector rays to a camera pixel in scattering media ($\sigma_t = 0.1$, $g = 0.9$, $\alpha = 0.9$). (b) Multiple scattering widens the direct path into the shape of a photon banana. (c) Neighboring illumination rays (red and blue) result in similar photon distributions. (d) Subtracting the averaged neighboring distributions extracts a photon channel which closely resembles the direct path.*



(a) (b)

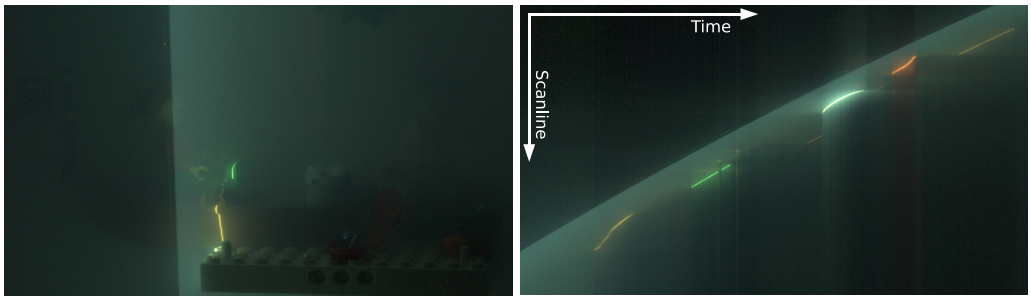
Figure 5.4*: Simulation results for varying scattering parameters. (a)* L_{signal}/L_{total} *falls off exponentially with increasing absorption (*$\alpha = 0.9$*,*$g = 0.9$*). A meaningful signal (above 5%) can be obtained up to an extinction coefficient of about* $1.4\ cm^{-1}$*. (b) SNR for varying forward scattering coefficient (*$\sigma_t = 0.1\ cm^{-1}$*).*

(σ_t $[cm^{-1}]$) and the average scattering cosine (g) represented by the Henyey-Greenstein [Henyey and Greenstein 1941] model. The range of values covers most of the materials measured in [Narasimhan et al. 2006]. As we focus at a point on the diffuse reflector we distinguish the total energy L_{signal} of photons arriving at the camera after interacting with the diffuse reflector from the total energy carried by all photons L_{total} [1].

Figure 5.4(a) shows how the ratio L_{signal}/L_{total} decreases with increasing σ_t. For an albedo of $\alpha = 0.9$ and a scattering angle of $g = 0.9$, the signal to noise ratio drops exponentially with σ_t. Around $\sigma_t = 1.4\ cm^{-1}$ the signal will become indistinguishable from the typical noise floor of the cameras (5%). A similar curve will be observed for the relative drop in signal with regard to the distance of the object to the projector and camera.

The signal-to-noise ratio furthermore depends significantly on the selected average scattering cosine g. Given the number of scattering events, photons are likely to reach the diffuse reflector only if the material is mostly forward scattering.

[1] This is a slightly different definition than used before, where L_d incorporates only photons whose traces stays within the photon channel.



(a) (b)

Figure 5.5*: (a) A single vertical line projected into the island scene of Figure 5.1. (b) A horizontal slice of the recorded reflectance field. The opaque surfaces are clearly visible as well as the haze due to the scattering fluid.*

5.5 Analysis in the Context of Reflectance Fields

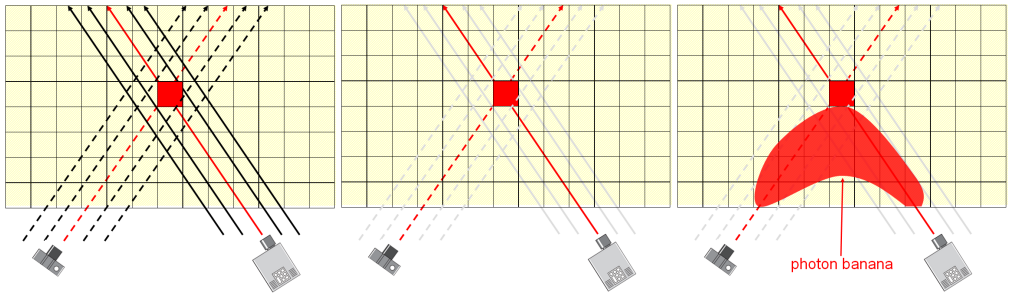
We will now analyze confocal imaging and descattering in the context of reflectance fields between pairs of cameras and projectors. The reflectance field T [Debevec et al. 2000, Masselus et al. 2003, Sen et al. 2005] describes how an incident light field L_i will be reflected by the scene, forming an outgoing light field $L_o = TL_i$. While the reflectance field for arbitrary light fields is an 8D function, the reflectance field between a single camera and one projector is only four-dimensional. Representing T as a tensor, each of its entries $T_{(\omega_o,\omega_i)}$ is the transport coefficient for one pair of camera and projector pixels or rays.

Figure 5.5(b) shows a 2D slice of the reflectance field of the island scene (Figure 5.1). The rows show always the same scan line of the camera while a vertical projector line (a) sweeps through the scene. The various object surfaces light up for different projector planes, in different rows. At the same time, multiple scattering in the surrounding fluid is clearly visible as it affects all entries in the reflectance field.

Under floodlit illumination (see Figure 5.1(b)), the contribution of all projector pixels, i.e. every line, integrates to the final image. Besides the actual surface, all entries that represent reflections by the fluid do contribute to the floodlit image. The interesting signal will be masked by the multiple scattering.

5.5.1 Confocal Imaging in Reflectance Fields

In confocal imaging, focusing to a specific voxel v corresponds to selecting the single entry in the reflectance field that is due to the particular selection of the view and the illumination ray. Extracting a volume slice means extracting all en-



(a) (b) (c)

Figure 5.6*: Image formation in confocal imaging. In* (a) *the volume is illuminated and observed along multiple rays in parallel. In* (b) *observation and illumination are concentrated along one pair of rays through confocal imaging. However, the photon banana due to multiple scattering along this pair of rays remains* (c).

tries on a specific hyperplane whose geometry is defined by the camera/projector configuration and the selected volume slice.

The improved contrast in confocal images is mainly achieved by focusing on the entries of interest, where the viewing and the light ray do intersect, while discarding all those entries in the reflectance field that are just due to multiple scattering. We have illustrated this effect in Figure 5.6(a) and 5.6(b). Thus, compared to a floodlit image, in a confocal slice most of the global effects are separated out. However, the net reflectance recorded for a single intersecting pair of viewing and illumination rays still contains some global portion (Figure 5.6(c)).

5.5.2 Descattered Confocal Imaging

As explained in the previous section, the global component L_g will be similar for neighboring, non-intersecting paths, as illustrated in Figure 5.7. In reflectance fields, neighboring paths correspond to neighboring entries. Thus, one can interpret descattering as a local difference operator applied to the reflectance field of the scene.

Since in practice the homogeneity assumption (condition (2), Section 5.3.3) will be violated, a good approximation is found by averaging the difference over multiple samples in the neighborhood of $T_{(\omega_o,\omega_i)}$ as long as condition (1) (non-intersecting rays) is met. We compute the averaged difference by applying a Laplacian of Gaussian (LoG) in the neighborhood.

For simplicity, we only apply a 1D LoG kernel in the dimension of the illumination, i.e. for the camera pixels' recordings at different illumination.

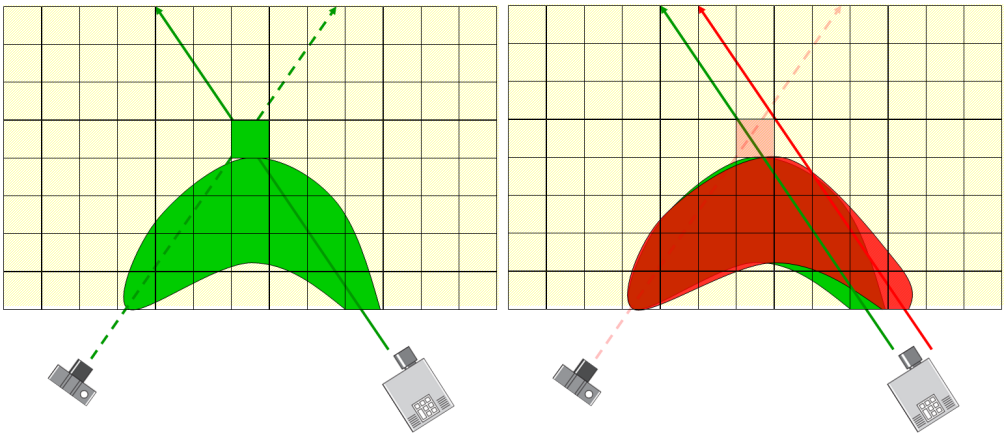


Figure 5.7*: Image formation in algorithmic descattering. The* left *image shows the situation for a single pair of rays.* Right: *shifting the illumination cause the center voxel to change its intensity drastically while the photon banana due to multiple scattering stays approximately the same.*

5.6 Acquisition Approach

As outlined in Section 5.5, we first record a reflectance field T^{cp} for every camera projector pair (c, p) and then compute the descattered confocal volume by applying a local LoG kernel. In order to produce the result images we either directly render the recovered volume or extract a single slice.

5.6.1 Setup

Our measurement setup (Figure 5.8 is similar to the one proposed by Levoy et al. [Levoy et al. 2004]. We employ three cameras (Jenoptic CFcool) with about 1.3 megapixel resolution and three Mitsubishi 490 DLP projectors. The devices are placed at a distance of $60\ cm$ away from the scene. The largest angle between the cameras and the projectors is about 65 degrees.

In order to obtain a pixel-precise alignment between a camera and a projector we perform calibration as described by Narsimhan and Nayar [Narasimhan et al. 2005]. A planar calibration target with a printed checker board is placed at three different known distances, moving it perpendicular to the calibration plane. The recorded images allow for precise recovery of the view-ray-to-voxel mapping $(c2v)$. Similarly, the three projectors are calibrated using projected checkerboard patterns, resulting in the illumination-ray-to-voxel mapping $(p2v)$. Assuming a rather smooth mapping we can easily compute the inverse mappings $(v2c)$ and $(v2p)$.

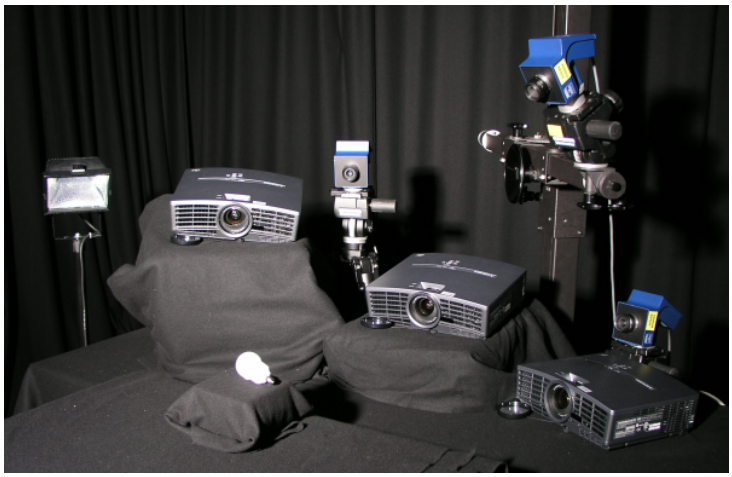


Figure 5.8*: Our synthetic aperture confocal imaging setup consists of three Jenoptik CFcool cameras and three Mitsubishi 490 DLP projectors.*

5.6.2 Planes of Light

Instead of recording a reflectance field with individual light rays, we sweep a plane of light through the volume generated by a line of one pixel width. Compared to the ray-based scanning, this reduces the number of required images from $O(N^2)$ to $O(N)$ for N projector rows. As Masselus et al. [Masselus et al. 2003] demonstrated, one can approximate a reflectance field from observations of swept horizontal and vertical lines.

5.6.3 Image Acquisition

In order to maximize contrast, every projector is turned on individually and its sweep is recorded while the two other projectors are turned off. Because every projector emits some black-level even for off pixels, we first record a black frame for every camera c and every projector p. We then sweep a single pixel line horizontally and vertically and record images I_x^{cp} and subtract the black frames. Here, x is the projector coordinate of the line.

In order to produce a confocal volume, we apply Equation 5.2 for every voxel, extracting the reflectance field samples for intersecting pairs of rays. Let $(s,t)^c$ be the projection location of voxel v in camera c determined by the calibrated $v2c$ map, and x be the x-coordinate of v being projected back to the projector image using $v2p$. We determine the confocal irradiance by averaging over the tri-linear interpolated reflectance field entries, first interpolating in the camera image and

(a) (b)

Figure 5.9*: (a) A bright illumination plane in front of the considered confocal plane will darken the image region if a symmetric descattering kernel is applied. (b) The one-sided kernel neglects those planes, rendering a slightly brighter image of the internal structures and resulting in more recovered detail.*

then between the selected slices with weight $\xi = x - \lfloor x \rfloor$:

$$\begin{aligned} L_{conf}(v) &= \sum_{c \in C} \sum_{p \in P} (1-\xi) \cdot I^{cp}_{(\lfloor x^p \rfloor)}(s,t)^c + \xi \cdot I^{cp}_{(\lceil x^p \rceil)}(s,t)^c \\ &= \sum_{c \in C} \sum_{p \in P} T^{cp}((s,t)^c, x^p). \end{aligned} \tag{5.6}$$

In order to compute a descattered confocal image this step is augmented by weighting the local illumination neighborhood with a LoG kernel of size $2K$:

$$L_d(v) = \sum_{c \in C} \sum_{p \in P} \sum_{k=-K}^{K} w(k) T^{cp}((s,t)^c, x^p + k). \tag{5.7}$$

We chose $K = 3$ and $w[-3,3] = \{-3, -10, 1, 24, 1, -10, -3\}$. The results for horizontal and vertical sweeps are simply averaged.

Within a translucent object, for a non-coaxial camera and projector setup there will typically be one intersection of the ray (s,t) with every plane in the neighborhood. If there would be a strong scattering event in front of voxel v along (s,t) the computed $L_d(v)$ will be too small. Without loss of generality, let $x + k$, with $k < 0$, be the planes where the intersection is in front of the voxel v. In this case, the direct component is better approximated by $L_d(v) = \sum_{c \in C} \sum_{p \in P} \sum_{k=0}^{K} w'(k) T^{cp}((s,t)^c, x^p + k)$ with a one-sided kernel $w'[0,k] = \{24, 2, -20, -6\}$. The difference is shown in Figure 5.9.

5.7 Results

Our proposed combined descattered confocal imaging method is demonstrated in two different kinds of experiments. First, we demonstrate the superior result of our method in imaging through murky water, after which we acquire a partially translucent object.

5.7.1 Looking through Murky Water

Computing clear images in participating media has been addressed in quite a number of previous applications, e.g. [Schechner et al. 2003]. In this section we demonstrated the effect of performing descattering on top of confocal imaging in a fish tank filled with diluted milk and compare our results against those obtained with descattering or confocal imaging alone. We try to estimate clear-air views into the tank for varying milk concentration. The tank has a dimension of $39 \times 25 \times 24\ cm^3$. We have acquired a resolution test chart placed at a distance of $9\ cm$ from the front plane (see Figure 5.10) and show a scene with higher depth complexity in Figure 5.1.

We swept about 200 horizontal and 240 vertical lines from three projectors as shown in Figure 5.8 and recorded HDR images of the scene with the three cameras. The acquisition took roughly 1 hour followed by 15 minutes of further processing, most of which is due to I/O.

In Figure 5.10, the performance of various algorithms is compared for two different concentrations. The first row (a) shows the tank in floodlit illumination. In the next two rows, we applied the descattering algorithm proposed by Nayar et al. [Nayar et al. 2006]. As input the technique requires a set of images recorded with periodic illumination pattern shifted in phase over time. Similar global patterns have been used even in the context of confocal imaging [Wilson et al. 1998, Neil et al. 1997]. A simple calculation approximates the direct component by $L_d = \frac{1}{1-b}(I_{\max} - I_{\min})$, where $I_{\min}$, $I_{\max}$ are the minimum/maximum of each pixel in the sequence, and b is the relative black-level of the projector. We have calculated the direct component on two different input sets. In column (b), the illumination pattern showed alternating on/off stripes of five pixels width, repeated over the entire projector pattern. The shifted sequence has been captured five times to reduce camera noise. In column (c), the same algorithm is applied to a synthetic input sequence generated from the reflectance field recorded by sweeping a single pixel line. The input patterns are computed by adding the captured images according to the previously described periodic patterns.

The next column (d) shows a single slice of the captured synthetic aperture confocal volume [Levoy et al. 2004] as computed using Equation 5.6. Finally, we

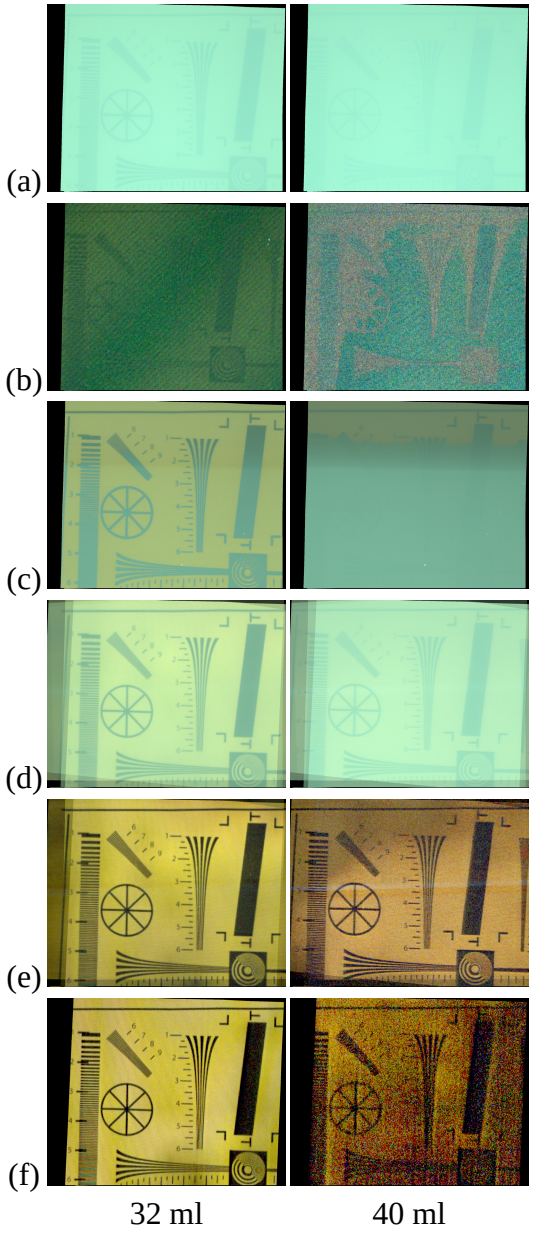


Figure 5.10: *Comparison for different milk concentrations in a* $23\ l$ *water tank: (a) white frame, (b) fast separation [Nayar et al. 2006] with periodic full-frame patterns, (c) fast separation with patterns synthesized from line sweeps, (d) a single confocal slice, (e) a descattered confocal slice incorporating all cameras and projectors, (f) a descattered slice from a single camera/projector pair. Please zoom in for full resolution.*

show our descattered confocal image in column (e) using three cameras and three projectors and a single camera/projector pair in (f) (c.f. Equation 5.7).

The pattern is barely visible in the floodlit configuration (a). In (b), descattering using images acquired with a periodic illumination pattern increases the contrast very little but increases the noise significantly. It produces even partially inverted patterns. This can be explained by the fact that half the projector pixels are on at any time. The tiny direct component, which is due to individual light rays, can hardly be separated from the camera noise because the added global component of half a million illumination rays is so much stronger in each input image. In the input images for column (c), the direct component is observed together with the global component of one line only. The signal is recorded much stronger. However, during the synthesis of the input sequence, half of all reflectance field entries are again combined. Recovering the direct component works better for the $32\ ml$ case but breaks down at $40\ ml$ milk in $23\ l$ water, due to the exponential fall-off of signal-to-noise ratio with increasing concentration (cf. Figure 5.4).

Combining the recordings from three cameras and three projectors, confocal imaging (d) succeeds in producing a much clearer image of the resolution chart since it considers only those entries of the reflectance field that contain a direct contribution, while the remainder of the entries is neglected. The pattern is visible in both concentrations indicating a precise calibration. Close to the image boundary the overlap of the swept projector volumes ends, resulting in intensity differences. While of good quality, the confocal images are still affected by multiple scattering. The pattern gets blurred more with increasing milk concentration. Since the target has a constant depth the method proposed by Narasimhan et al. [Narasimhan et al. 2005] would just produce a scaled version of the confocal image.

Applying our proposed descattering algorithm (column (e)) removes the remaining global effects yielding higher contrast. It furthermore extracts sharper pictures of the direct component with higher resolution than any other method. Since only a small neighborhood is used for computing the direct component, the contrast is significantly increased compared to methods using global periodic kernels.

Note, that rows (c), (d) and (e) are computed from the same set of camera images. Only the processing is different.

In Figure 5.10(e), one can furthermore observe a color shift from 32 to $40\ ml$ in our results. This indicates that we are indeed extracting the direct component. Increasing the concentration increases the out-scattering along the path. Since milk scatters blue more strongly than red, the blue part of the direct component is filtered out faster than the red one. The difference between our descattered confocal imaging approach and only applying local descattering becomes obvious when comparing Figure 5.10(e) with (f) which is much more noisy.

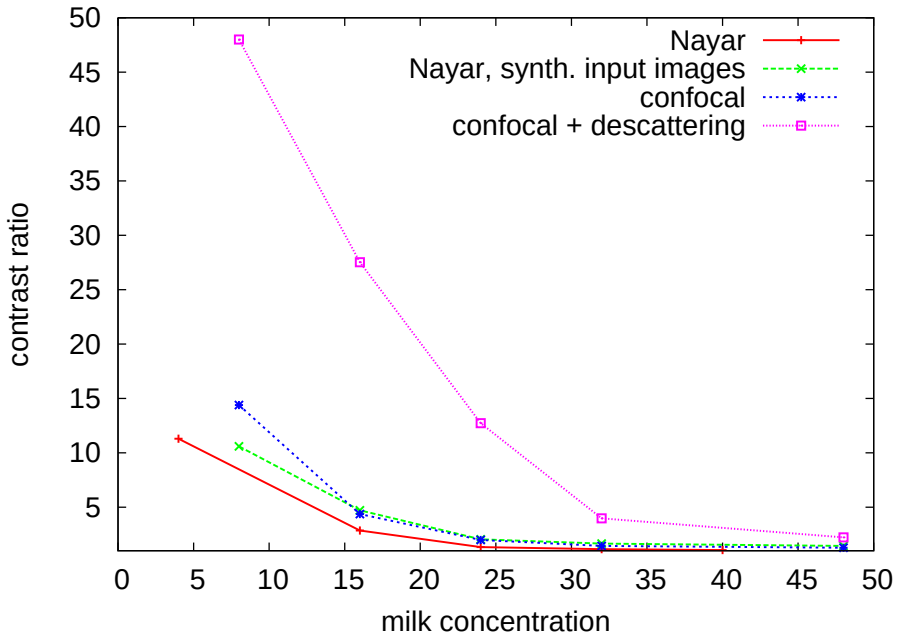


Figure 5.11*: The maximum contrast between a white and a dark patch in the result images of the approaches in Figure 5.10 decreases with the scattering density (*ml *milk in* $23l$ *water).*

The achieved contrast of the various methods for different milk concentrations is plotted in Figure 5.11. As expected, the signal drops exponentially with increasing concentration, as also predicted by the simulations (c.f. Figure 5.4). Descattering with large kernels and pure confocal imaging consistently produce a smaller contrast, rendering the image unrecognizable much earlier than with our descattered confocal imaging technique.

Island Scene

While the previous scene consisted of a single plane at fixed depth, the second experiment captures a scene with some depth complexity (Figure 5.1). It touches the front of the fish tank and has a depth of approximately $9.5\ cm$.

We have captured the full confocal and descattered confocal volume to recover the volumetric 3D structure of the scene as can be seen in the offset view in Figure 5.1(e). We render orthographic views of the scene using maximum intensity projection and increase the reconstructed intensity linearly with depth in order to counteract the extinction.

Compared to pure confocal imaging our descattered confocal imaging technique can look much deeper into the volume. The pirate on the right and the palm

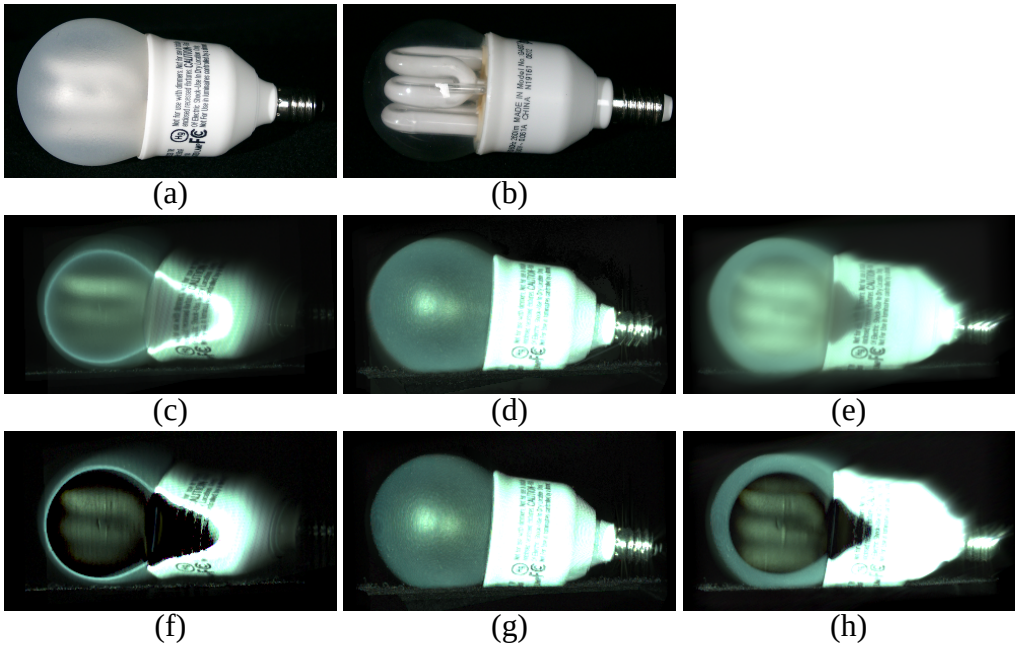


Figure 5.12*: Confocal (middle row) and descattered confocal images (bottom row) of an energy saving light bulb with frosted coating (a). In (b), the coating has been removed for demonstration. The depth discrimination for a single confocal slice (c) increases using descattering (f). (d) and (g) show the full integrated volume, from which in (e) and (h) the front slices have been removed.*

tree in the back are recovered. As for the resolution chart, our results are sharper and the colors are more saturated.

In the offset view one can observe some artifacts around depth discontinuities where the assumption of a homogeneous medium around the path is violated. For rather bright objects such as the flag, the silhouette is partially extruded in depth due to small errors in the camera calibration and the low aperture sampling rate.

The ground plane is not rendered white in the confocal images because a couple of volume slices have been removed from the front in order to exclude any contribution from the fish tank's front glass.

5.7.2 Looking into Translucent Objects

We further investigate the internal structure of a light bulb with frosted coating (Figure 5.12). Here, partially translucent layers are separated by air.

The first column shows one slice of the reconstructed volumes. While one sees a blurry image of the internal tubes in the confocal image (middle row), applying descattering (bottom row) reveals sharper detail. In addition, the considered plane is much more focused in the descattered version. The second column shows the

integrated volume, i.e. all slices added up. The outer surface is not affected by our descattering algorithm. In the third column, the front most volume slices have been clipped away, revealing the internal structure of the fluorescent tubes. While being slightly more noisy the descattered image provides a significantly higher contrast. Upon further inspection (see video), one can even make out the traces of a tiny, completely transparent glass pipe in front of the white fluorescent tubes. It is hardly visible in the pure confocal volume.

5.8 Discussion

As indicated in Figure 5.4 the performance of active light volumetric acquisition methods, such as ours, with regard to contrast or penetration depth will be limited in principle by the scattering density σ_t, since the ratio of the direct to the global contribution will decrease exponentially. A low noise camera and a high-contrast projector can help increasing the maximum recoverable depth.

Increasing the number of aperture samples can further help both increasing the contrast, as well as minimizing the aliasing artifacts which are currently visible in the island scene when solid surfaces come out of focus.

Currently, we calibrate the paths for the illumination and viewing rays in clear air for the bulb example or with water in the fishtank experiments. This approach assumes that the index of refraction remains constant in order to predict the intersection of viewing and illumination rays. We are therefore limited to test objects which have the same index of refraction as used during calibration. For the light bulb this assumption does not hold, since its housing has a different index of refraction than air. As a result, the internal structures of the light bulb appear slightly distorted.

The performance of confocal imaging in translucent media can be significantly improved by combining it with our novel descattering procedure based on a local descattering kernel. The same input data is used for both steps. While confocal imaging reduces the effect of multiple scattering by selecting individual light paths, our descattering operator further removes global effects from the selected paths by analyzing the vicinity of paths: The photon banana corresponding to a confocal sample is effectively reduced to photon channels around the direct path.

For descattering, we have so far applied a local LoG kernel in the illumination domain only. In the future, one might obtain even better descattering by computing spatially varying deconvolution of the volume around the confocal entries in the reflectance field, or by further investigating the duality of light transport.

Our approach for combined descattering and confocal imaging currently operates at macroscopic scales. Using a confocal microscope, or by augmenting a

light field microscope [Levoy et al. 2006] by a light field projector, migration to microscopic scales should be straightforward.

Chapter 6

Combining Polarization and Algorithmic Descattering

The work outlined in the following chapter was done in close collaboration with Tongbo Chen [Chen et al. 2007] during their PhD studies.

Besides combining confocal imaging and algorithmic descattering as described in the previous chapter, we also investigated towards another approach to remove multiple scattering effects.

6.1 Introduction

We make use of the fact that polarized light becomes depolarized while undergoing multiple scattering. With this idea in mind, we developed a method which is designed to facilitate 3D scanning of translucent objects. As already outlined in Section 3.1.3, for scattering objects, the signal observed for a particular surface point may be polluted by the subsurface light transport. Of course, for 3D scanning the translucency can be dealt with by painting the object's surface with matte paint. This is however tedious and for some objects simply not possible. In this chapter of the dissertation, we therefore propose several methods for 3D scanning in the presence of translucency.

Projecting polarized light and computing the difference of images captured with a polarization filter at two orthogonal orientations thus removes most of the multiple scattering contribution [Wolff 1994, Rowe and Pugh Jr. 1995, Schechner and Karpel 2005, Treibitz and Schechner 2006]. Since polarization alone is usually not sufficient to enable reliable range scanning of highly translucent objects, we combine it with the algorithmic descattering proposed by Nayar et al. [Nayar et al. 2006] which was already discussed in Section 5.3.3 of this the-

Figure 6.1*: By combining phase-shifting and polarization our method faithfully captures the 3D geometry of very translucent objects such as this alabaster Venus figurine (height ≈ 19cm).*

sis, albeit in slightly different form. In the same paper Nayar et al. also mention that phase-shifting [Srinivasan et al. 1985, Zhang and Yau 2006] can perform the separation and 3D scanning at the same time. For these reasons, our 3D scanning approach for translucent objects is based on phase-shifting.

To summarize, in this chapter of this dissertation we contribute the following:

- We demonstrate and analyze why descattering based on either polarization or structured light alone is not sufficient to obtain high quality depth maps of heterogeneous translucent objects.
- We combine phase-shifting and polarization and show that the accuracy and reliability of the 3D scanning process is greatly improved on a variety of different objects.

In the following, we will analyze the problems associated with 3D scanning of translucent objects.

6.2 Direct Reflection vs. Multiple Scattering

In this section we will discuss the relevant effects of direct reflection, multiple scattering or interreflections on projected polarized or non-polarized structured light patterns.

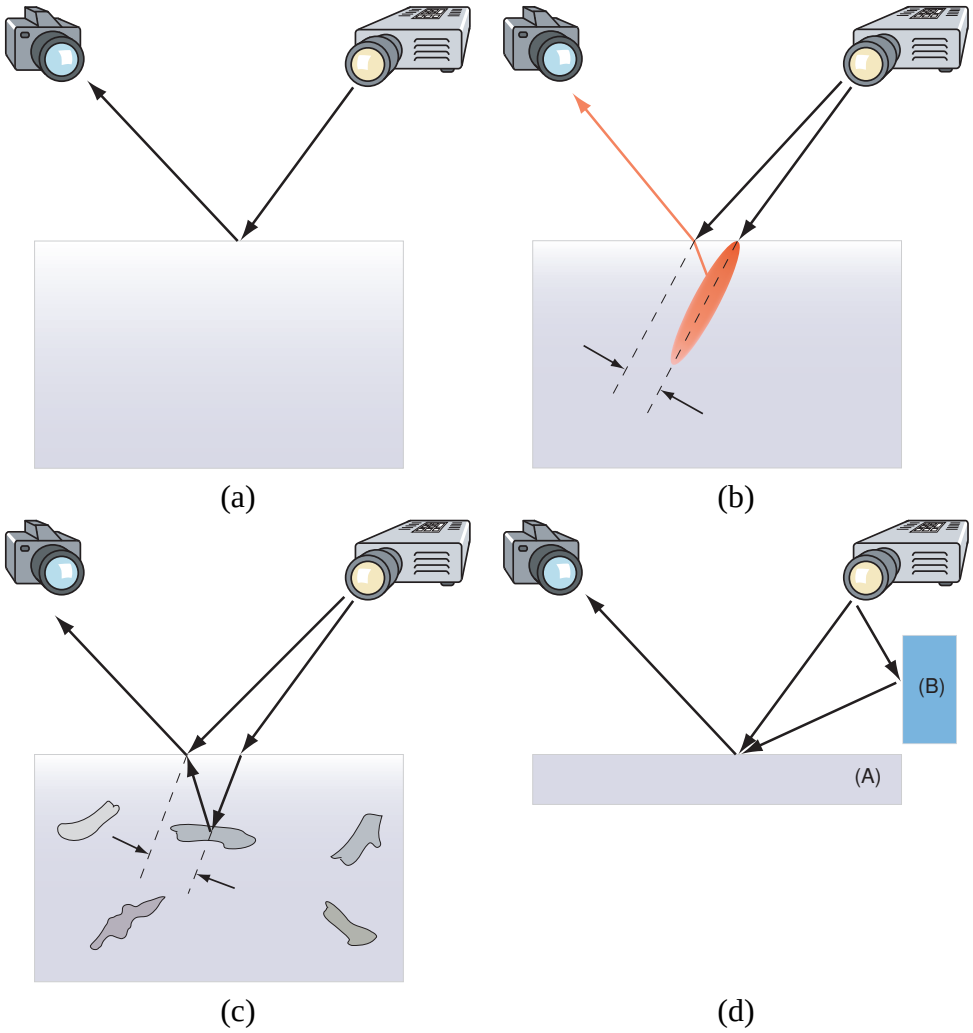


Figure 6.2: *Rays to consider in 3D scanning. (a) 3D geometry can be estimated reliably only from the direct reflection off the surface. (b) The subsurface scattering in translucent objects can shift the observed intensity peak away from the point of incidence. (c) Opaque structures beneath the surface pollute the range estimate. (d) The signals of different projector rays are overlaid due to interreflection from another surface.*

6.2.1 Direct Reflection

The light reflected at the surface towards the camera consists of different components [Jensen et al. 2001, Nicodemus et al. 1977, Ishimaru 1978]: direct reflection off the surface, subsurface scattering or interreflections (see Figure 6.2). The signal we are most interested in for 3D scanning is the directly reflected light (Figure 6.2a). The amount of directly reflected light depends on the surface properties such as color, roughness, etc. which can be summarized in the BRDF [Nicodemus et al. 1977]. In addition, light that is reflected from a smooth surface of a dielectric (or insulating) material is partially polarized. The amount of polarization of the reflected light can be computed according to Fresnel's formulae [Born et al. 1999] and depends on the material properties and the orientation of surface with regard to the incident and reflected ray directions. A comprehensive polarization reflectance model can be found in Wolff and Boult [Wolff and E.Boult 1991].

6.2.2 Bias due to Multiple Scattering

The prominent effect that distinguishes translucent from opaque materials is that some light penetrates the surface and is scattered multiple times inside the object before it finally leaves the surface at some other location. Determining the 3D shape of a translucent object requires detecting the first surface intersection of the incoming light ray, i.e. to observe the pure direct reflection (Figure 6.2a). Unfortunately, the signal of the direct reflection will be rather weak since some fraction of the incident light will penetrate the surface instead of being reflected. The reflected signal will furthermore be heavily polluted by single or multiple scattering created by light incident on other scene points. As pointed out by Godin et al. [Godin et al. 2001] multiple scattering results in a measurable bias in the depth estimate since the location of the observed intensity peak is shifted away from the point of incidence (Figure 6.2b). Multiple scattering can be approximated by a diffusion process [Jensen et al. 2001] and leads to a significant damping of the high frequencies in the incident illumination. Projecting shifted high frequency patterns, the global component will remain the same while changes can be observed in the direct reflection only. This can be used to remove this global effect algorithmically [Nayar et al. 2006]. We will further investigate this approach in Section 6.3.

Multiple scattering further influences the state of polarization. While single scattering polarizes light according to the size and shape of a particle and the reference plane spanned by the direction of the incoming light and the scattered direction, multiple scattering due to the random orientation of particles to some degree depolarizes the incident light [van de Hulst 1981, Schechner et al. 2003,

Schechner and Karpel 2005]. In Section 6.4 we make use of the depolarization properties to remove multiple scattering effects from the measurements.

Another important source of error is depicted in Figure 6.2c. Here, some structure beneath the surface actually reflects more light than the direct reflection at the surface leading to wrong depth estimates (compare Figures 6.6 and 6.4). While light reflected by those structures keeps the high frequencies of the incident light pattern we show in our experiments that it undergoes some degree of depolarization, which can be utilized.

6.2.3 Interreflections

Similar effects are introduced by interreflections due to nearby surfaces (Figure 6.2d). The signal of the direct reflection off an arbitrary surface (not necessarily translucent) is disturbed by the indirect reflection from another surface. The resulting artefacts might range from a small bias added to the depth estimate of the original surface (A) to wrongly detecting the depth of the mirror image of the other surface (B).

Depending on the reflection properties of the other surface (B) the high frequencies of the original pattern will typically be significantly reduced in the indirect reflection; for a glossy or diffuse BRDF, the illumination of a single point on surface (B) will indirectly illuminate a larger region on surface (A), hereby spreading out the signal. For second and higher order interreflections the loss of high frequencies is even more prominent.

Note, however, that interreflections might still result in linearly polarized light depending on the arrangement of surfaces (A) and (B). As a result, polarization is not always suitable for separating the direct component from interreflections.

6.3 Phase-Shifting for 3D Scanning and Reflection Separation

In Chapter 5, we used the algorithmic descattering approach of Nayar et al. [Nayar et al. 2006] in combination with synthetic aperture confocal imaging to improve visibility through translucent media. In this chapter, we will use the descattering properties of that algorithm to eliminate the light transport beneath the surface and thus facilitate 3D scanning.

Algorithmic descattering is based on the insight that global effects significantly dampen high frequencies (compare Section 6.2). Illuminating the scene with shifted high frequency patterns therefore will result in high frequencies observable in the direct reflection part only.

6.3.1 Descattering Properties of Phase-Shifting

Various patterns have been proposed by Nayar et al. [Nayar et al. 2006] to perform the separation ranging from checker board and simple stripe patterns to sinusoids. As pointed out by the authors shifted sinusoids can be used simultaneously for 3D scanning since the patterns allow for deriving the phase of the sinusoid function. We implemented this technique in our 3D scanning approach as well. Assuming that the output of the projector is linear, which we establish through photometric calibration, a set of sine patterns is generated as $L_i(m,n) = 0.5\cos(\lambda m + \delta_i) + 0.5$, where λ is the frequency for all patterns and δ_i is the phase-shift for each individual pattern. Given a sufficiently high λ, the observed intensity I reflected of a scene point at camera pixel (x,y) will be

$$I_i(x,y) = \frac{1}{2}[L_d(x,y)cos(\Phi(x,y) + \delta_i) + L_d(x,y) + L_g(x,y)]. \tag{6.1}$$

Note that only the direct reflection L_d will depend on the phase Φ of the surface point while the global part L_g will not. The observed phase $\Phi(x,y)$ is correlated to the depth or disparity of the surface point and depends on the specific camera and projector parameters. From a set of at least three different phase shifts, e.g. $\delta_i \in -2\pi/3, 0, 2\pi/3$, one can separate the global and the direct components as

$$L_g = \frac{2}{3}(I_0 + I_1 + I_2) - L_d \quad \text{and} \tag{6.2}$$

$$L_d = \frac{2}{3}\sqrt{3(I_0 - I_2)^2 + (2I_1 - I_0 - I_2)^2}. \tag{6.3}$$

L_g is supposed to be in low frequency and leads to the fact that it can be cancelled out implicitly in the least square evaluation of Φ. For N evenly spaced phase shifts in one cycle the following equation computes the phase at pixel(x,y):

$$\Phi(x,y) = tan^{-1}\left(\frac{-\sum I_i \sin(\delta_i)}{\sum I_i \cos(\delta_i)}\right), \tag{6.4}$$

where all sums are over the N measurements, a result which has also been observed in communication theory when detecting noise corrupted signals using synchronous detection [Bruning et al. 1974]. At the same time we can use the ratio γ of the observed amplitude over the observed bias as a measure for the reliability of the phase estimation:

$$\gamma = \frac{L_d}{L_d + L_g} = \frac{N\sqrt{(\sum I_i \sin(\delta_i))^2 + (\sum I_i \cos(\delta_i))^2}}{\sum I_i \sum \sin^2(\delta_i)}. \tag{6.5}$$

6.3.2 Temporal Phase-Unwrapping

Using shifted patterns with a single frequency we can detect the phase within one period of the selected frequency ($\Phi \in [0, 2\pi]$). The period however might be repeated multiple times over the entire scene. The problem is to locate the absolute unwrapped phase Ψ that uniquely identifies the pixel's phase. A number of different methods have been proposed to obtain an unwrapped phase map [Ghiglia and Pritt 1998]. If the scene contains depth discontinuities the exact phase and period can be obtained by repeating the phase extraction for multiple (lower) frequencies [Huntley and Saldner 1993, Huntley and Saldner 1997]. Possible approaches are choosing frequencies such that the greatest common divisor of the periods is larger than the number of columns in the projector image [Tarini et al. 2005]. More robust unwrapping is obtained by creating a series of frequencies $\lambda_j = 0.5\lambda_{j-1}$ until one period spans the projector image width s resulting roughly in $F = \log_2(s)$ frequencies [Huntley and Saldner 1997]. Given the unwrapped phase at one frequency $j+1$, the unwrapping algorithm iteratively locates the phase at step j, the next higher frequency. Starting with $j = F-1$ and $\Psi_F = \Phi_F$ one computes the unwrapped phase at the next higher frequency by

$$\Psi_j = \Phi_j - 2\pi \text{NINT}\left(\frac{\Phi_j - 2\Psi_{j+1}}{2\pi}\right), \tag{6.6}$$

where NINT rounds to the nearest integer. The unwrapping itself is to some extent similar to decoding binary encoded structured light patterns [Salvi et al. 2004], but more robust.

In our experiments (see Section 6.5) we used periods of $8, 16, 32, 64, 128, 256, 512$ and 1024 pixels. At the higher frequencies (8 and 16 pixels) we use 8 and 16 phase-shifts to obtain the best precision while the lower frequencies are only used for disambiguating the period (making rough and stable binary decisions), therefore six phase shifts turned out to be sufficient. Overall, 60 images are captured for each range scan but the number could be further reduced if necessary.

Using multiple frequencies poses the problem that Equation 6.1 only holds for high frequencies. For low frequencies, the global component will also vary with the phase-shift and thus the phase and depth estimates will be biased. Based on the scanning of the planar block of alabaster shown in Figure 6.6 we demonstrate this effect in Figure 6.3a by comparing the phase estimates for different frequencies. The lower the frequency, the larger the deviation of the estimated phase for the individual frequency. A small or moderate drift at a low frequency has typically only very little effect on the combined result since the lower frequencies are just used for estimating the 2π modulo jumps. Figure 6.4c shows an example where

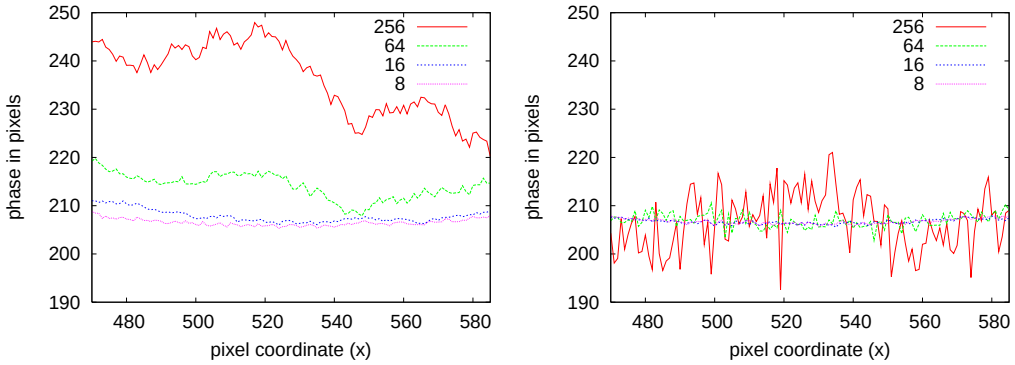


Figure 6.3*: Biased phase reconstruction for low frequency patterns. (a) Phase profiles of individual frequencies for one line on the planar alabaster block. (b) After polarization-difference imaging (Section 6.4) even the lower frequencies result in correct depth estimates. Note that the curves are tilted by the same factor for illustration.*

the deviation on a lower frequency is larger than one period and thus introduces a major offset in the 3D scan.

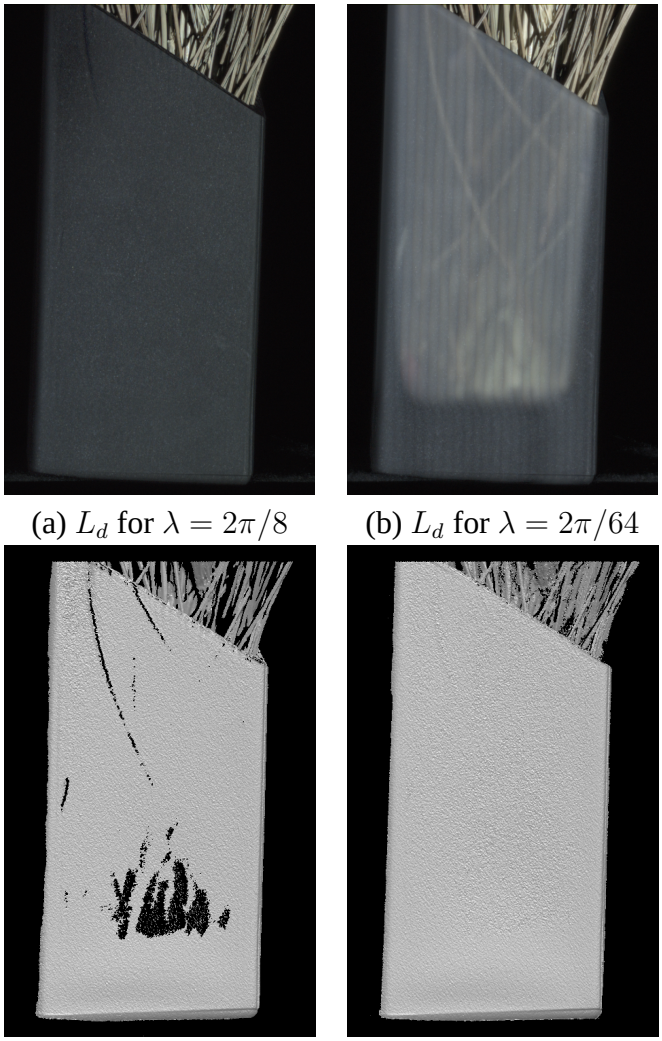
In the next section we demonstrate how reliable depth profiles can be computed even for low frequencies when polarization is used in addition to phase-shifting to separate out the global component (Figure 6.3b and Figure 6.4d).

6.4 Polarization-Difference Imaging for Descattering

As discussed in Section 6.2 multiple scattering depolarizes the incoming light. Schechner et al. [Schechner et al. 2003, Schechner and Karpel 2004, Schechner and Karpel 2005, Treibitz and Schechner 2006] have made extensive use of this phenomena to compute clear pictures through haze or murky water by taking several polarized images from which the depolarized part can be removed afterwards. Based on the estimated signal loss induced by the participating media the authors further compute rough depth maps of the underlying scene.

In our setup depicted in Figure 6.5 linear polarizers are put in front of the camera and the projector. We then capture the phase-shift image sequence twice, once when the camera's polarizer axis is oriented parallel to the projector's polarizer axis, yielding $I_j^{\parallel}$, and a second time using cross-polarization, $I_j^{\perp}$. A polarization difference image [Rowe and Pugh Jr. 1995, Tyo et al. 1996] is then computed as

$$I_j^{\Delta} = |I_j^{\parallel} - I_j^{\perp}|. \tag{6.7}$$



(a) L_d for $\lambda = 2\pi/8$ (b) L_d for $\lambda = 2\pi/64$

(c) phase-shifting (d) PDI

Figure 6.4*: For this translucent vase filled with lavender, the reconstructed direct reflection is dependent on the frequency of the illumination pattern. (a) Most subsurface scattering is removed using the highest frequency. (b) At lower frequencies structures beneath the surface contribute to the direct component polluting the phase-unwrapping results in (c). (d) Using PDI the influence of subsurface structures is largely reduced and the desired shape is captured.*

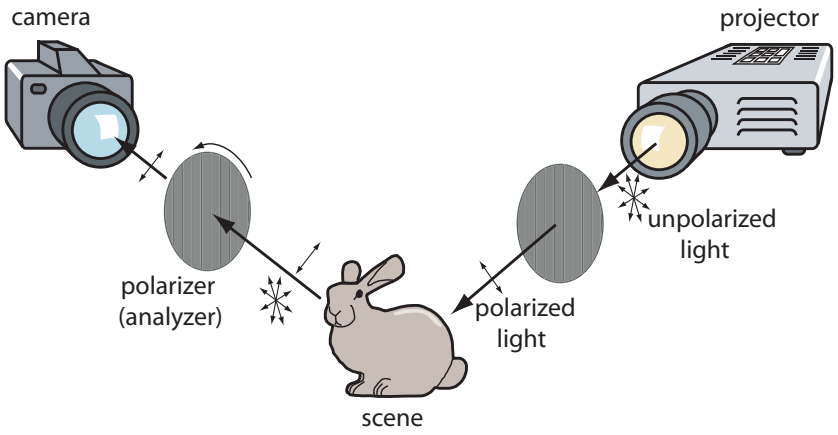


Figure 6.5*: Our setup for polarization-difference imaging (PDI). The projector is equipped with a linear polarization filter at fixed orientation. The camera captures two image sequences with parallel and with perpendicular orientation of the polarization filters.*

The idea is that depolarized light will add exactly the same contribution to both image sequences, independent of the camera's filter orientation, and thus will be completely removed in I_j^{Δ}. Based on I_j^{Δ} we then perform the 3D reconstruction.

Figure 6.6 demonstrates the effect of polarization filtering on the quality of the 3D reconstruction of a quite planar block of alabaster. The most important difference between 3D reconstruction by phase-shifting without polarization filtering and with polarization-difference imaging applied is that scattering events beneath the surface are much better removed in the latter case. Using PDI, the contrast and thus the signal of the input images is largely improved (bottom row of Figure 6.6). However, as will be discussed in the result section, there are some scenes where the PDI approach filters out too much of the direct reflection. In these cases using the image sequence with parallel orientation of the polarization filters provides a good trade-off between no polarization and PDI.

6.5 Results

In the following section we assess the descattering capabilities of phase-shifting with and without polarization on a set of translucent scenes: a highly translucent, almost homogeneous alabaster figurine (Figure 6.1 and 6.7), a filled, translucent vase (Figure 6.4), a heterogeneous planar slab of alabaster (Figure 6.6), some grapes and a starfruit (Figure 6.7). Except for removing spurious background pixels and pixels having a too weak signal no further processing, i.e., noise removal or smoothing has been applied to the results.

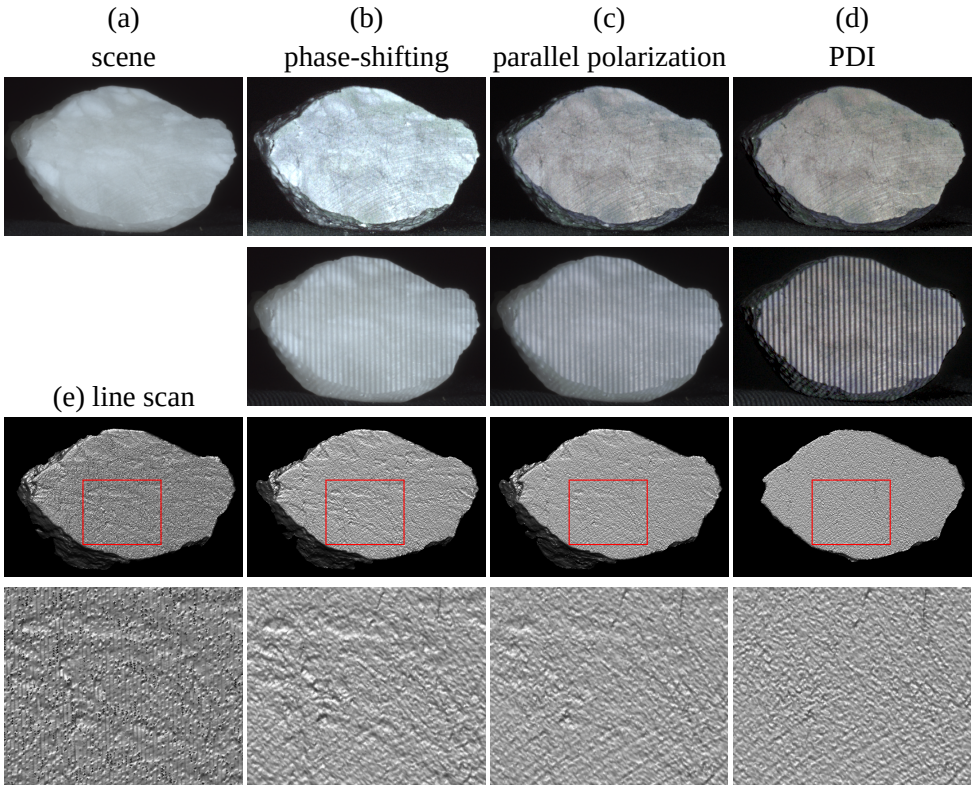


Figure 6.6: *Reconstruction results for a planar surface of heterogeneous alabaster. First row: (a) Photograph. While the direct component L_d (b) extracted using no polarization filters clearly contains some subsurface structures they are partially removed by parallel polarization $L_d^{\parallel}$ (c) and not present after applying PDI L_d^{Δ} (d). Second row: The contrast in the high frequency input images is improved by parallel polarization and further by PDI. Third and fourth row (magnified region): Geometry reconstruction results for (e) line sweeping, (b) phase-shifting without polarization, (c) with parallel polarization, and (d) with PDI. The influence of the subsurface structures on the final 3D geometry has been completely removed by PDI.*

6.5.1 Setup

All images in this section have been acquired with a 14-bit 1360×1024-pixel Jenoptik ProgRes CFcool CCD camera and a Mitsubishi XD490U XGA DLP Projector whose native resolution is 1024×768. We performed a photometric calibration for both devices and captured HDR images [Robertson et al. 2003] using four different exposures. The measured maximum simultaneous contrast of a sine pattern with a period of 8 pixels reflected by a gray card is 180:1 (max/min). We performed geometric calibration between the camera and the projector [Zhang 2000]. Linear polarization filters have been placed in front of the projector and the camera

to acquire the PDI image sequences.

6.5.2 Structured Light Results

Descattering based on phase-shifting without polarization can deal pretty well with translucent objects and clearly removes some amount of the subsurface scattering (Figure 6.7), as predicted by Nayar et al. [Nayar et al. 2006]. The phase-unwrapping, however, relies on low frequency patterns which clearly suffer from global effects (see Section 6.4). Furthermore, structures beneath but close to the surface will have some influence on the estimated direct component which is unwanted in the context of 3D scanning (Figure 6.4 and 6.6).

Sweeping a single line is an alternative to phase-shifting and performs surprisingly similar. As can be seen in Figure 6.6 even for highly translucent objects one obtains a reasonable 3D scan if high quality equipment and HDR sequences are used. The noise in the figure indicates that the SNR of line sweeping compared to phase-shifting is considerably lower. Although line sweeping is still sensitive to the bias introduced by subsurface scattering or subsurface structures (Section 6.2) global effects are minimized by the comparably small amount of incident light concentrated on a small region.

6.5.3 Polarization Results

Polarization-difference imaging also separates the direct from the global component very well. It faithfully removes all traces of subsurface structures. At grazing angles PDI however filters out too much of the direct reflection (see Figure 6.7). It is worthwhile to note that depending on the surface properties also some fraction of the direct reflection might be depolarized. This fraction will also be removed in the polarization difference image. For some scenes, we actually observed a better contrast of direct vs. global reflection in the parallel polarization setting $I_j^{\parallel}$ producing smoother 3D scans (see Figures 6.7, bottom row). Even though parallel polarization in theory only removes some fraction of multiple scattering effects (compare second row in Figure 6.6), combining it with phase-shifting adds the descattering capabilities of both techniques. Figure 6.6 further shows that parallel polarization also renders phase-shifting slightly more robust against subsurface structures, though not as robust as PDI. On the other hand parallel polarization is much easier to acquire since it requires only half the amount of images and a fixed orientation of the filters.

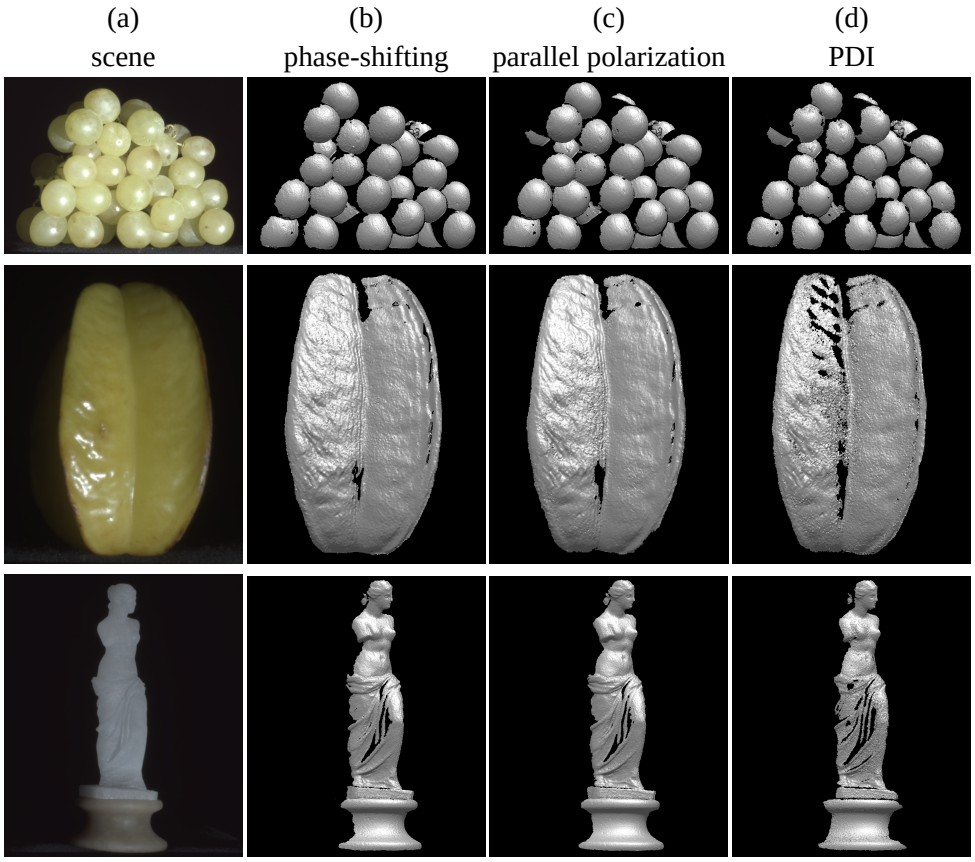


Figure 6.7*: Reconstruction results for a selection of translucent objects. For these scenes the reconstruction results of phase-shifting without polarization (b) and with parallel polarization (c) are of comparable quality since the objects are mostly homogeneous. At grazing angles too much signal from the direct reflection is filtered out by PDI (d) resulting in more holes and noisier depth maps.*

6.6 Discussion

In the previous chapter, we have already seen that Nayar's descattering method alone is not sufficient to remove all of the translucency in subsurface scattering objects. Therefore, we combined it with synthetic aperture confocal imaging to reconstruct 3D structures in a translucent medium.

In this chapter, we have shown that descattering by phase-shifting alone is also not sufficient. Usually, enough scattering remains to disturb the 3D scanning process. We therefore combined phase-shifting with polarization filtering, to provide a robust 3D scanning technique for translucent objects. A careful analysis of phase-shifting without polarization, combined with parallel polarization and with polarization-difference imaging has shown that some of the shortcomings of pure phase-shifting such as its sensitivity to subsurface structures can be over-

come. Even though PDI has the ability to robustly remove all global effects due to subsurface scattering which otherwise renders accurate 3D scanning a hard problem, parallel polarization sometimes provides a better SNR resulting in less noisy range maps. Depending on the richness of subsurface structure, the planarity of the object and the sensitivity of the camera one might choose one method over the other.

Chapter 7

Relighting from Image Collections

This chapter contains joint work which was developed together with Tom Haber [Haber et al. 2009]. In contrast to the previous chapters of this thesis, which all dealt with translucent objects, we will consider here only opaque objects.

7.1 Introduction

Estimating scene reflectance using image-based techniques is a well-studied topic in the graphics and vision literature. Common approaches are based on controlled, active illumination [Debevec et al. 2000] or special setups to record the illumination during scene capture [Yu and Malik 1998, Lensch et al. 2003]. In specific cases it is also possible to estimate illumination from shadows [Sato et al. 2003] or from other properties of the scene. An alternative line of work assumes that multiple images of the scene are captured under identical illumination, which is then estimated together with the surface reflectance [Nishino et al. 2001, Yu et al. 2006]. In contrast, our goal is to estimate the reflectance of each scene point from a set of photographs captured under unknown, constant or varying distant illumination.

This has two striking implications: First, we do not require special lighting or capture equipment to shoot our images. Instead, we can just casually acquire a set of photographs provided that the distant lighting assumption holds. Second, it is not even necessary to capture the images ourselves. Instead, we are able to make use of existing photographs, even images taken from online community photo collections (CPC). Combined with a robust multi-view stereo approach for CPCs [Goesele et al. 2007], we can estimate the scene geometry from these images as

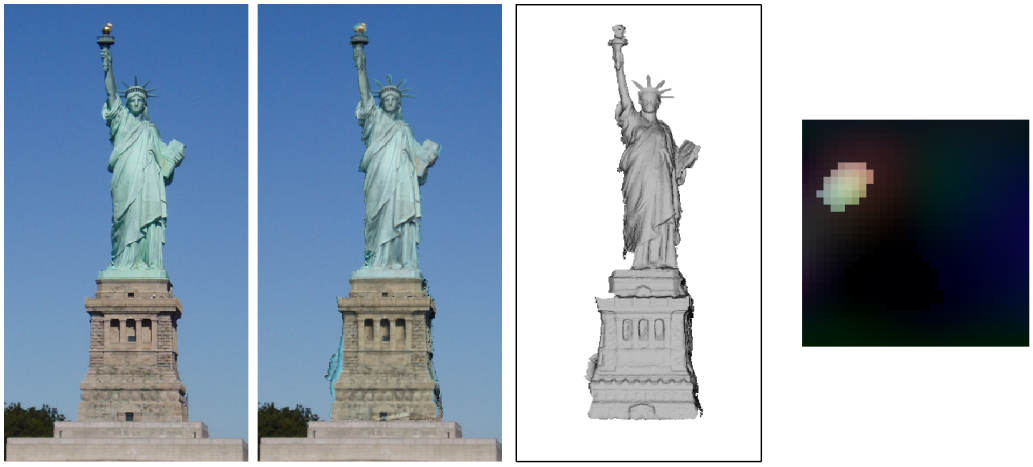


Figure 7.1*: Overview of our reconstruction pipeline. From left to right: an example image taken from Flickr [Flickr 2009], re-rendered model using the recovered reflectance properties, estimated geometry, and the illumination from the estimated environment.*

well. This is the first system that recovers a complete scene model from images acquired under general conditions.

Our system consists of two main components: We use an *all frequency relighting framework* based on a wavelet representation of the local visibility and the current estimate of illumination and scene reflectance to render the scene efficiently using the triple-wavelet product integral [Ng et al. 2004]. Factoring illumination and scene reflectance requires *solving a bilinear system of equations*. We therefore employ an iterative optimization to estimate illumination given scene reflectance and vice versa. To summarize, our contributions are as follows:

- We simultanously estimate the reflectance and illumination parameters for a scene captured under varying distant illumination. This simplifies the capture process and makes our system applicable to a larger range of scenes and existing datasets.
- We show reconstructions solely based on images captured from an Internet photo sharing site. Scene geometry, reflectance, and distant illumination are all estimated from these images.

The achievable quality of our system is bound by fundamental limitations such as the bandwidth of the reflectance and the frequency content of the illumination [Ramamoorthi and Hanrahan 2001, Romeiro et al. 2008]. We nevertheless achieve good quality results for the datasets presented in this chapter.

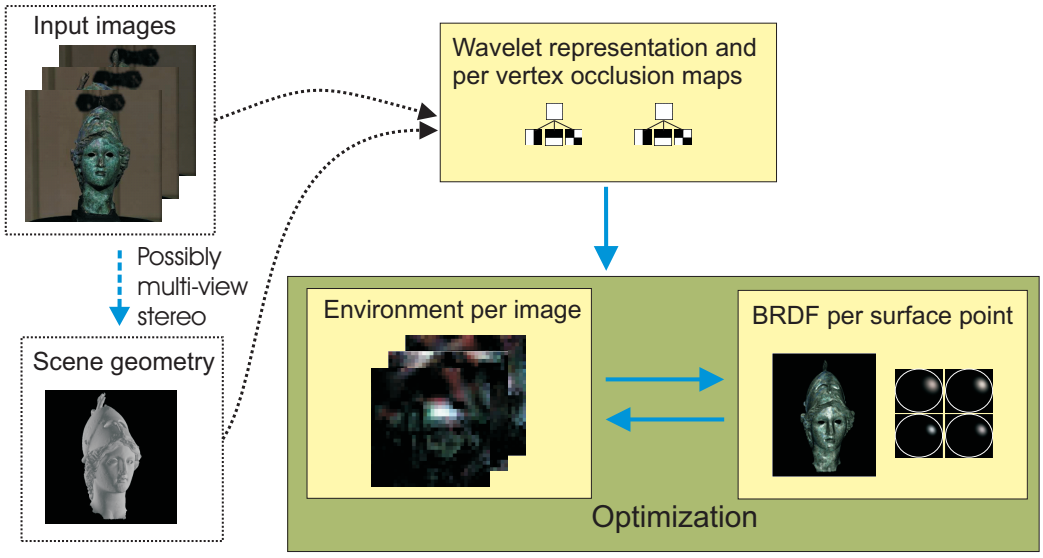


Figure 7.2*: System overview.*

7.2 Overview and Assumptions

Our system takes as input a set of photographs of an object or a scene captured under unknown and possibly varying distant illumination (see Figure 7.2). We assume that these images are already in a photometrically linear space. Overexposed pixels cannot be trusted and receive a weight of zero in the reconstruction process.

A triangle mesh representing the scene geometry is either given (e.g. manually modeled or captured in a separate acquisition process), or reconstructed using a multi-view stereo approach [Goesele et al. 2007]. The mesh is manually cleaned and cropped from spurious or inaccurate geometry. We assume that the mesh models the complete relevant scene geometry including blockers that cast shadows on the object.

The system operates within an all-frequency relighting framework [Ng et al. 2004, Yu et al. 2006] based on Haar wavelets. In a preprocessing step, we compute an occlusion map per mesh vertex, i.e. determine which part of the hemisphere of illumination directions is occluded, and transform the occlusion map into the wavelet basis. We then set up an efficient rendering pipeline based on triple-wavelet products, which allows us to easily update the current reflectance and illumination estimates. Local surface reflectance is expressed as a weighted sum of basis materials. We assume that interreflections in the scene can be neglected. The relighting task leads to a set of bilinear equations, which is solved iteratively yielding a reflectance estimate per surface point and an illumination estimate per input image. Provided that the illumina-

tion used for image capture and the BRDF contains sufficient high frequencies [Ramamoorthi and Hanrahan 2001, Romeiro et al. 2008], we are able to reconstruct a detailed model of the scene, the reflection properties, and of the illumination.

7.3 Rendering Pipeline

A central part in our estimation process is the simulation of light propagation in the scene based on the reflectance equation (see Section 2.2), which is evaluated for every rendered pixel:

$$L(\vec{x}, \omega_o) = \int_\Omega \rho(\vec{x}, \omega_o, \omega_i) V(\vec{x}, \omega_i) \tilde{L}(\vec{x}, \omega_i) d\omega_i, \tag{7.1}$$

where ρ is the BRDF including the cosine term for the incident illumination from direction ω_i at position $\vec{x}$ and the outgoing direction ω_o. $\tilde{L}$ is the incident illumination and V represents the occlusion at $\vec{x}$. Since we iteratively optimize for ρ and $\tilde{L}$, we need an efficient evaluation scheme.

7.3.1 All Frequency Relighting

Efficient techniques for solving this rendering problem under the distant illumination assumption have been investigated in the field of precomputed radiance transfer [Sloan et al. 2002], where the incident illumination and the reflectance functions are represented using spherical basis functions such as spherical harmonics (SH) or wavelet bases. In these bases, the otherwise costly integral reduces to a dot product. In order to not limit the resolution capabilities of our reconstruction we follow the all-frequency relighting approach by Ng et al. [Ng et al. 2004].

The local visibility, the local environment map, and the slice of the BRDF that corresponds to the viewing direction ω_o are expressed in the Haar wavelet basis Ψ:

$$V(\omega) = \sum_l V_l \Psi_l(\omega) \tag{7.2}$$

$$\tilde{L}(\omega) = \sum_m \tilde{L}_m \Psi_m(\omega) \tag{7.3}$$

$$\rho(\vec{x}, \omega) = \sum_k \rho_k(\vec{x}) \Psi_k(\omega) \tag{7.4}$$

At each point, the directions are defined over the local hemisphere using the hemi-octrahedral parameterization introduced by Praun and Hoppe [Praun and Hoppe 2003]. Their parametrization is illustrated in Figure 7.3.

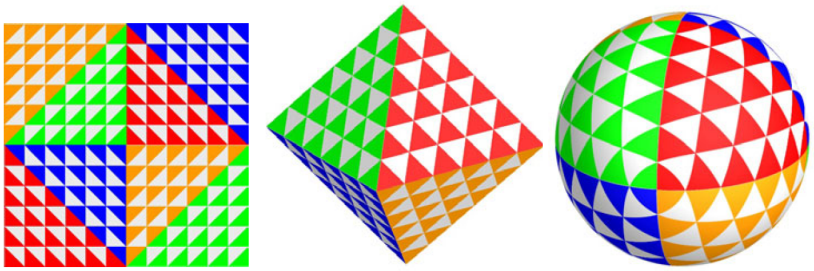


Figure 7.3: *Overview over the parametrization. The planar 2D surface is first mapped to an octahedron, which is then reprojected onto a sphere. For a hemisphere, only a hemi-octahedron and only the inner part of the 2D surface is used. (Image courtesy of Hugues Hoppe.)*

In practice, we compute local visibility only per mesh vertex and interpolate it for any point in a triangle. Note that representing the BRDF in the wavelet basis (Equation 7.2) allows us to use arbitrary BRDFs (including isotropic and anisotropic, measured and analytically represented BRDFs) in the rendering pipeline.

Once all three components are represented in the same Haar wavelet basis the integral over the hemisphere is evaluated as a triple-wavelet product integral [Ng et al. 2004]

$$L(\vec{x}, \omega_o) = \sum_k \sum_l \sum_m C_{klm} \rho_k V_l \tilde{L}_m \tag{7.5}$$

with tripling coefficients $C_{klm} = \int_\Omega \Psi_k \Psi_l \Psi_m d\omega$. Exploiting the hierarchical nature and sparseness of the Haar wavelet representation drastically reduces the required computation effort.

7.3.2 Lighting Rotation

One problem of this approach is that for every pixel each slice of the BRDF needs to be represented in the wavelet basis Ψ. While Ng et al. [Ng et al. 2004] used precomputed 6D BRDF data sets, parameterized by ω_i, ω_o, and the surface normal $\vec{n}$, we employ the Efficient Wavelet Rotation algorithm [Wang et al. 2006] to transform the environment map $\tilde{L}$ given in an octahedral parameterization into the local coordinate frame of each pixel and the corresponding hemi-octahedral representation. This step reduces the dimensionality of the BRDF per pixel to 4D resulting in reduced memory overhead at the cost of additional computation. In a preprocessing step, we discretize the sphere into 32×32 normal directions $\{\vec{\nu}\}$ and compute a rotation matrix $\mathbf{R}_{\vec{\nu}}$ for each of them. $\mathbf{R}_{\vec{\nu}}$ encodes the influence of

every wavelet basis coefficient in the global coordinate frame onto the wavelet coefficients in the rotated space. A rotated environment map $\tilde{L}(\vec{n})$ for an in-between normal is constructed by interpolation from the rotated maps of the four nearest normals with interpolation weights $\phi_{\vec{\nu}}(\vec{n})$:

$$\tilde{L}(\vec{n}) = \sum_{\vec{\nu} \in \text{Neighborhood}(\vec{n})} \phi_{\vec{\nu}}(\vec{n}) \cdot \mathbf{R}_{\vec{\nu}} \tilde{L}. \tag{7.6}$$

7.3.3 Bilinear Model

Given the object geometry and a set of images $\{I_i\}$ captured under distant illumination, we can extract the set of measurements $\{y_p\}$ that show the response of the scene to an environment map $\tilde{L}_{I(y_p)}$ recorded in image $I(y_p) \in \{I_i\}$ measured at surface point $\vec{x}(y_p)$ on the object's surface S.

The visibility term V_l in Equation 7.5 is constant for a given surface point $\vec{x}$ as it encodes the portion of the hemisphere that is blocked by the object geometry. It can therefore conceptually be combined with the tripling coefficients to a new set of coefficients

$$T_{\vec{x},km} = \sum_l C_{klm} V_l. \tag{7.7}$$

Using matrix notation, Equation 7.5 simplifies then for a particular y_p to a bilinear system in L and ρ

$$L_p \;\; = \;\; \rho^T_{\vec{x}(y_p)} \mathbf{T}_{\vec{x}(y_p)} \tilde{L}_{I(y_p)}. \tag{7.8}$$

7.4 Optimization

We can now define our inverse rendering task. From a set of images $\{I_i\}$ we seek to determine the reflectance functions $\rho(\vec{x})$ and the incident illumination $\tilde{L}_I$ for each I_i. This is equivalent to minimizing the following objective function in a least squares sense with respect to $\tilde{L}$ and ρ:

$$\mathcal{O} = \sum_{p=1}^{N} \alpha_p (y_p - \rho^T_{\vec{x}(y_p)} \mathbf{T}_{\vec{x}(y_p)} \tilde{L}_{I(y_p)})^2 \tag{7.9}$$

with $\tilde{L} = \{\tilde{L}_0, \ldots, \tilde{L}_{Q-1}\}$ and $\rho = \{\rho_{\vec{x}} | \vec{x} \in S\}$. y_p is a measurement of surface point $\vec{x}_p$ seen in an input image. We additionally introduce confidence values α_p per observation to reduce the effect of less reliable data, such as measurements at grazing angles and overexposed pixels.

Since our system depends linearly on both ρ and $\tilde{L}$ we have to solve a bilinear system of equations. This can be performed iteratively by solving linear problems once for $\tilde{L}$ and once for ρ [Cohen and Tomasi 1997, Bai and Liu 2006]. Note that solving for both simultaneously would yield clear advantages [Buchanan and Fitzgibbon 2005] but is impractical due to the size of our problem. The solution can only be unique up to a global scale factor. In addition, we constrain both $\tilde{L}$ and ρ to be strictly positive in order to ensure physically plausible results. Each subproblem can now be efficiently solved with arbitrary linear equality and inequality constraints by posing it as a convex Quadratic Programming problem. We apply an implementation of the primal-dual interior point algorithm [Gertz and Wright 2003] for this task.

7.4.1 Optimization of Environment Maps

Given an initial guess of the reflectance ρ we first solve for the environment $\tilde{L}$ minimizing the linear system

$$Y_I = \mathbf{M}\tilde{L}_I \tag{7.10}$$

where we combined the constant terms into a single matrix $\mathbf{M}_p = \rho^T_{\vec{x}(y_p)}\mathbf{T}_{\vec{x}(y_p)}$. This inverse lighting problem is in general ill-posed [Marschner 1998, Ramamoorthi and Hanrahan 2001]. The recoverable information of $\tilde{L}$ depends on the bandwidth of the reflectance function ρ and on the geometry of the object, as shadows carry additional information about the illumination [Sato et al. 2003]. In general, it will not be possible to recover all details of the illumination.

To obtain a sensible solution despite the ill-posedness of the problem, we apply a small amount of regularization by imposing smoothness on the environment maps. This prevents the optimization from introducing unnecessary high-frequency details, reducing the amount of overfitting while hardly destroying any useful information since the recoverable frequencies are anyway limited. We improve robustness to noise, inaccurate geometry, interreflection and other unmodeled effects by appyling an iteratively reweighted least squares approach [Björck 1996] which effectively reduces the contribution of pixels with high reconstruction error.

7.4.2 BRDF Optimization within a Linear Basis

In order to obtain a linear system for the reflectance $\rho(\vec{x})$, we follow the approach of Weistroffer et al. [Weistroffer et al. 2007] and represent the reflectance at each surface point as a linear combination of materials. Each material in turn consists

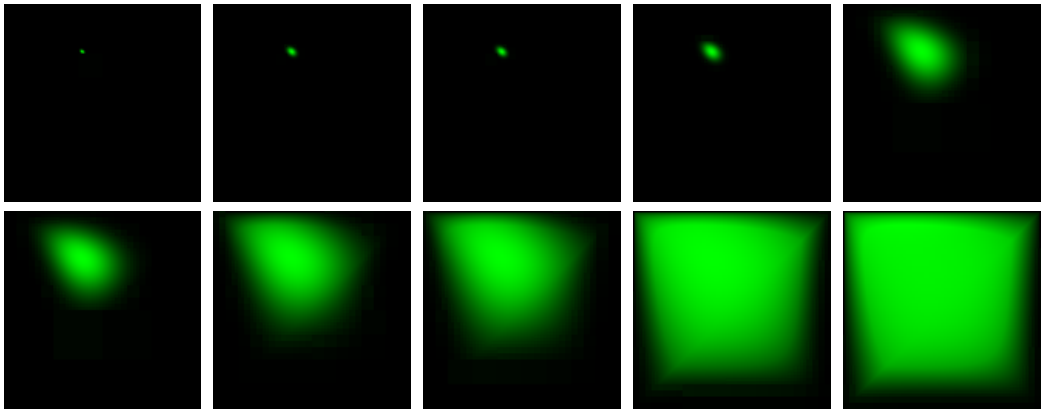


Figure 7.4*: Slice through the specular basis BRDFs selected from the MERL database (only green channel).*

of a linear combination of basis BRDFs

$$\rho(\vec{x}) = \sum_k (w_k(\vec{x}) \sum_b \lambda_{kb} \rho_b). \tag{7.11}$$

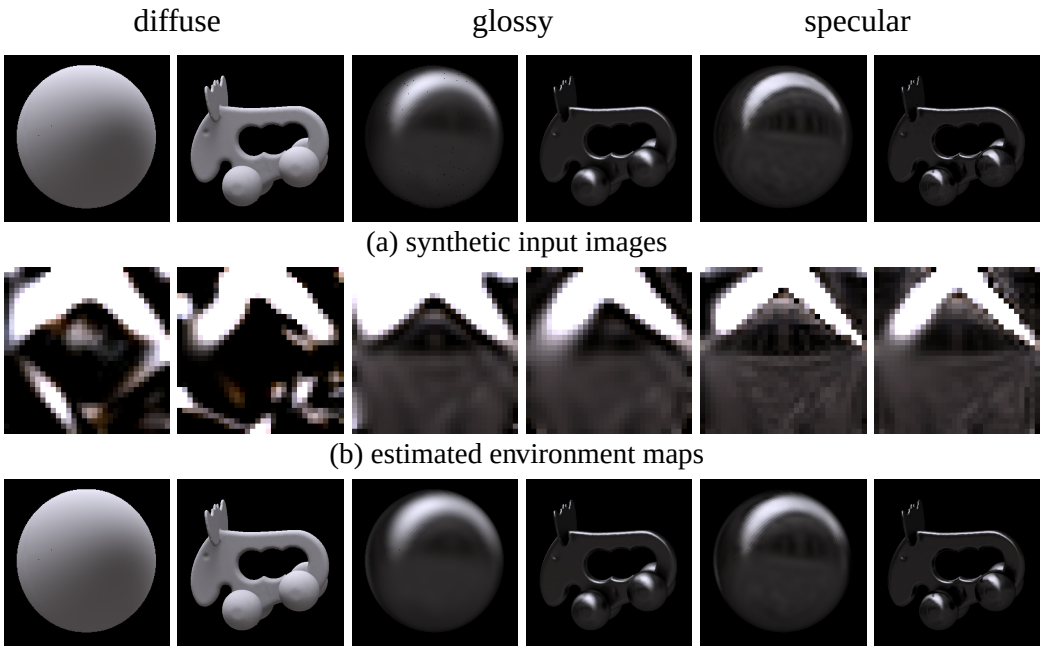
The optimization for each λ_{kb} uses all available samples that are part of material k (i.e. $w_k > 0$). This two-level approach improves the stability for sparse and noisy data. We again solve this bilinear problem by alternating between the estimation of material weights λ_{kb} and blending weights $w_k(\vec{x})$ per surface point

$$\mathcal{O} = \sum_{\{p|\vec{x}(y_p)=\vec{x}\}} \alpha_p \left(y_p - \rho(\vec{x})^T \mathbf{T}_{\vec{x}(y_p)} \tilde{L}_{I(y_p)} \right)^2 \tag{7.12}$$

for each surface point $\vec{x}$. We again reduce the effect of less reliable data using the confidence value α_p. All coefficients $w_k(\vec{x})$ and λ_{kb} are constrained to be strictly positive. To improve numerical stability while minimizing the number of used basis BRDFs, we modify this system by additionally minimizing $\beta \sum_k \lambda_{kb}^2$.

To decrease computation times, we define this system using mesh vertices and the blending weights are then later bilinearly interpolated for every surface point. Using a sufficiently dense mesh, the additional smoothing introduced is barely noticable.

One of the advantages of our framework is that we can use arbitrary BRDFs (including measured or analytical BRDFs) as basis provided they can be expressed in the wavelet basis. For the examples shown in this chapter, we performed k-means clustering on the fitted Cook-Torrance parameters for the MERL BRDF database [Ngan et al. 2005] and selected the specular lobes from representatives of the ten largest clusters (see Figure 7.4). The actual basis $\{\rho_b\}$ for a reconstructed scene then consists of a subset of these lobes plus a Lambertian BRDF to model the diffuse reflectance.



(a) synthetic input images

(b) estimated environment maps

(c) object rendered under estimated illumination

Figure 7.5*: Estimated environment maps for synthetic diffuse, glossy, and specular test cases with a fixed BRDF. The amount of detail recoverable in the environment maps depends on both the bandwidth of the BRDF as well as the scene's geometry.*

To compute an initial estimate of the material and blending weights, we transform the averaged per-vertex samples into HSV colorspace and cluster the vertices using the parametrization $(\cos(2\pi \cdot H), \sin(2\pi \cdot H), S)$. This transformation reduces the influence of shadows and specular highlights. We use k-means clustering to get the separation into regions. For each region, a different material is constructed by computing a diffuse color based on the median of the per-vertex averages, the other BRDF weights are set to zero. The blending weights can then be obtained from the k-means algorithm.

7.5 Results

We present results for three different types of input data: synthetic datasets, datasets captured under controlled conditions, and a dataset solely based on image collections from Internet photo sharing sites (see also Table 7.1).

(a) (b) (c)

Figure 7.6: *a) The* Uffizi *environment map [Debevec 1998] in the octahedral representation as used in our experiments. b) The acquisition setup to capture the* Minerva *dataset. The two main light sources are facing towards the head. c) The ground truth environment map for the* Minerva *dataset.*

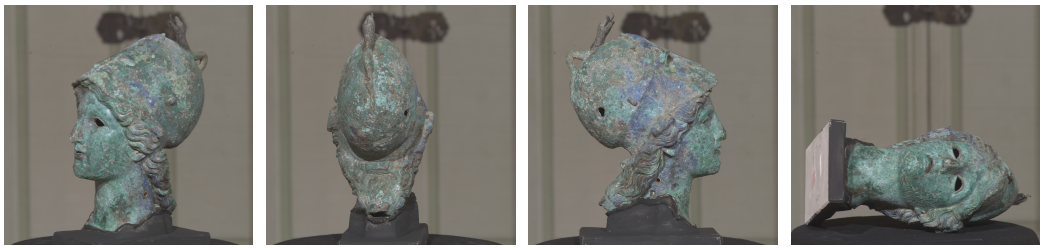
7.5.1 Synthetic Data

We render two objects (a sphere and an elk, see Figure 7.5a) with varying surface properties (diffuse, glossy with Lafortune exponent $N = 50$, and specular with $N = 500$ [Lafortune et al. 1997]) using the *Uffizi* environment map (Figure 7.6a). We then discard the lighting information and reconstruct it from the rendered image, known object geometry, and BRDF (Section 7.4.1). The resulting environment maps (Figure 7.5b) clearly show that the lighting is recovered faithfully. Note how remarkably close the sky region reconstructed from the diffuse sphere is to the original. As predicted by Ramamoorthi and Hanrahan [Ramamoorthi and Hanrahan 2001], the amount of detail in the lighting increases with the specularity and thus with the bandwidth of the BRDF. In all cases, illuminating the model with the recovered lighting yields results hardly distinguishable from the input (Figure 7.5c).

7.5.2 Captured Data

Figure 7.7 gives an overview over the *Minerva* dataset. The dataset consists of a detailed geometry model captured with a 3D scanning system and a set of high-dynamic range (HDR) images captured under distant illumination in a church-like environment. Lighting and camera were fixed while the object was rotated in front of the camera (see Figure 7.6 b) and c)). This dataset is therefore an ideal test case to compare the environment maps reconstructed for different input images.

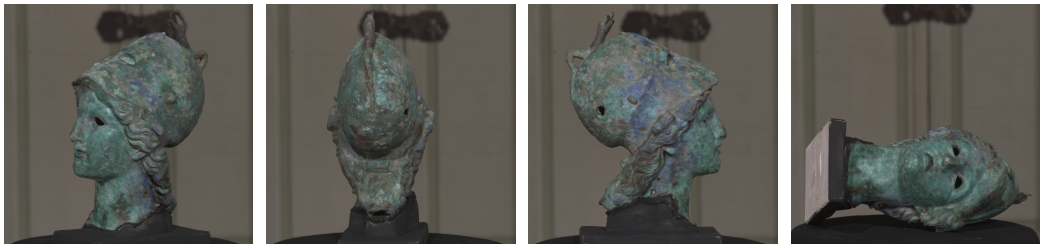
We used a collection of 13 images captured under general lighting (Figure 7.7a) shows 4 of these images) and reconstructed the per-view illumination and the per-pixel BRDF using 6 basis materials each composed of 7 BRDFs. Figure 7.7d) clearly demonstrates that the resulting model rendered with the esti-



(a) input images for the *Minerva* dataset

(b) rotated ground truth environment map

(c) estimated environment maps



(d) object rendered under estimated illumination

Figure 7.7*: Overview of the* Minerva *dataset showing four of the eight input images used in the estimation.*

mated illumination matches the input images very well. Even small highlights on the temple and chin are faithfully recovered. The estimated environment maps Figure 7.7c) locate the main light sources correctly and otherwise approximate roughly the ground truth illumination since the head features limited specularity. Nevertheless, the materials are well estimated which is also shown in the albedo map (Figure 7.12) which reveals hardly any geometric feature or shadow, as expected but show some color artifacts which cause a color shift when the model is relit. The reddish color in the hair region can be explained by the fact that the floor of the room is red. The algorithm contributed this to the BRDF instead of the environment map.

As second example, we selected 30 images from the *van Gogh* dataset [Chen et al. 2002] (see Figure 7.8a). This dataset was captured under static illumination using a moving camera. Although we reconstruct an individual environment map per view, Figure 7.8c) and the relighting of a novel view with the average of all environment maps (Figure 7.10) clearly show how consistent our reconstructions are even without assuming a constant environment as in Yu et al. [Yu et al. 2006]. Note that the estimated environment maps are not identical but consistent for the different views. Missing geometric detail in the input mesh prevents us from reconstructing a high resolution environment map and also reduces the quality of the material estimation especially in the hair region. As we reconstruct materials per vertex, the signature on the bust is highly undersampled.

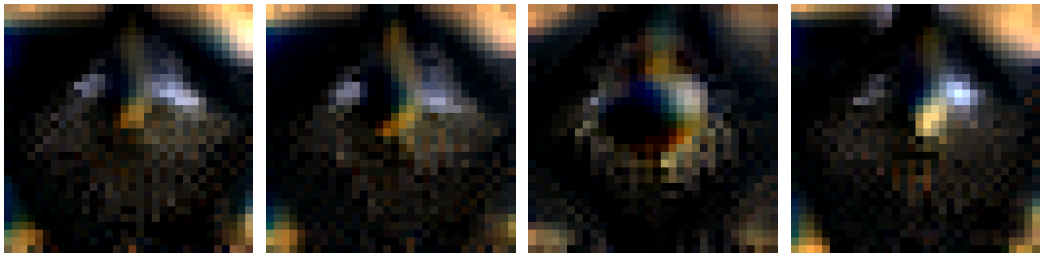
The relit models shown illuminated by the *Uffizi* environment (Figure 7.9a) and under point light illumination (Figure 7.9b) are of high quality. They show clearly that we were able to reconstruct more high frequency details than Yu et al. [Yu et al. 2006]. However, our color estimates are slightly desaturated. Here we observe the fact that the inverse rendering problem can only be solved up to scale. The red channel is slightly underestimated in the BRDF while it is slightly overestimated in the environment.

7.5.3 Internet Data

We downloaded an image dataset of the Statue of Liberty from the Internet photo sharing site Flickr [Flickr 2009] and reconstructed the scene geometry using a multi-view stereo approach [Goesele et al. 2007]. The resulting mesh was manually cleaned and simplified to reduce its complexity. Images found on the Internet are generally neither HDR nor photometrically calibrated. We therefore assume that they are encoded according to the sRGB standard and convert them into photometrically linear space by inverse gamma mapping.

For the lighting and reflectance estimation only six images have been used, they are shown in Figures 7.11 and 7.1. The estimated environment maps clearly locate the sun direction correctly. Note that the light source might wrap around

(a) input images for the *van Gogh* dataset



(b) estimated environment maps

(c) object rendered under estimated illumination

Figure 7.8*: Overview of the* van Gogh *dataset showing 4 out of the 30 input images.*

(a) (b)

Figure 7.9*: The* van Gogh *model relit with the* Uffizi *environment map* (a) *and under point light illumination* (b)*.*

model	#vertices	#mat	#brdfs	#images	time
Minerva	100K	6	7	13	7h
Liberty	100K	2	4	6	3h
van Gogh	32K	3	4	30	14h

Table 7.1*: Overview of the different models. "#mat", "#brdfs", and "#images" refer to the number of materials used, number of basis brdfs incorporated, and number of input images, respectively.*

borders due to the parametrization of the environment maps. Its narrow size can only be achieved by including the visibility information and the use of an all frequency framework. Other regions in the environment map are however reconstructed with low frequency due to their smaller intensity. Again, the color separation between materials and environment is not perfect. This could also be due to the different white balancing of images found on the Internet. Nevertheless, the estimated materials are consistent and of high quality as evident in Figures 7.11 c) and d).

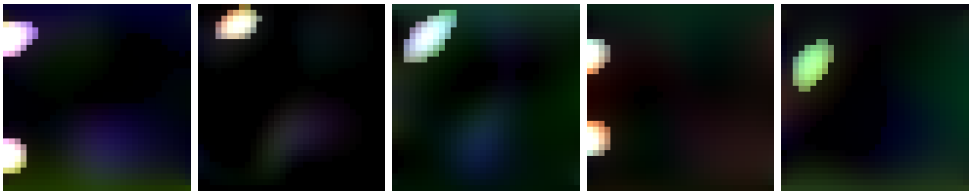
Figure 7.10*: An image* (left) *from the* van Gogh *dataset that was not used in the reconstruction process is regenerated using the average environment map* (right)*. The difference image shows that the largest error occurs in the specular highlights and in places where geometric detail is missing.*

7.6 Discussion

In the presented inverse rendering framework we estimate both the reflection properties of an object and the incident illumination. We have demonstrated that it is in principle possible to recover the environment map for each individual input view. The reconstruction is however limited by the bandwidth of the BRDFs. Furthermore, the quality of the results is to some extent influenced by the precision of the input geometry. Another strength of our approach is that no assumptions are placed on the image set. We flexibly can incorporate image collections gathered from various sources. In particular, this will allow the exploitation of Internet photo sharing communities for reconstructing relightable scene models.

A problem that remains is the ambiguity between illumination and surface color. Currently, this can only be resolved by having sufficiently different input environments. A couple of extensions of the current framework would be interesting. So far, our system does not take interreflections into account. An additional geometry optimization step could help with the reconstruction, since more precise geometry and normals will improve the estimation. This is supported by the synthetic test cases where we see a close to perfect reconstruction. Furthermore, it would be interesting to incorporate other ways to stabilize the estimation. For this, techniques such as reflectance sharing [Zickler et al. 2006] could be incorporated. Kuthirummal et al. [Kuthirummal et al. 2008] recently published a method that could help improve the quality of the camera calibration.

(a) input images for the *Liberty* dataset



(b) estimated environment maps

(c) object rendered under estimated illumination

(d) object rendered under point light illumination and under the *Uffizi* environment map.

Figure 7.11*: Overview over the* Liberty *dataset.*

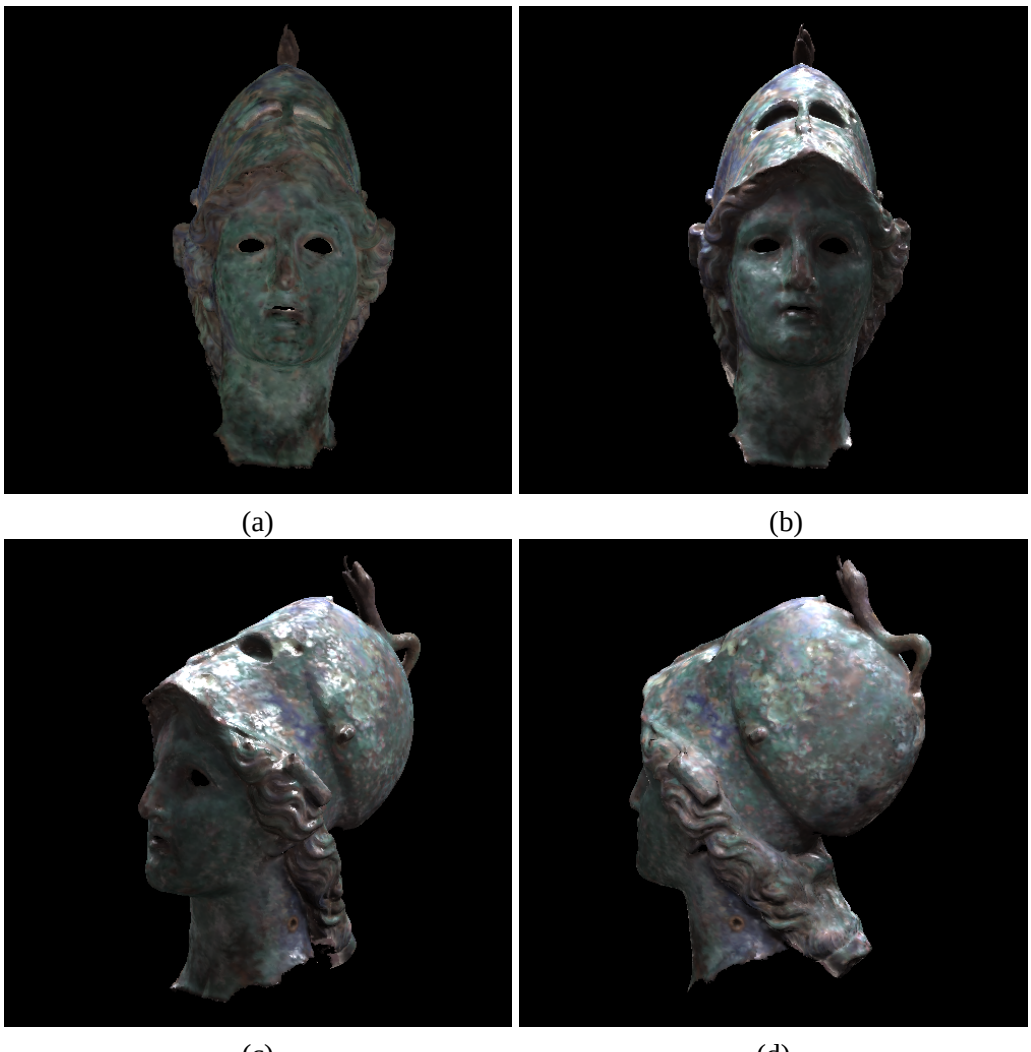


Figure 7.12*: (a) shows the recovered albedo map for the* Minerva*, while (b)-(d) show the relit model using the* Uffizi *environment.*

Chapter 8

Conclusion and Future Work

The focus of this thesis lies on the acquisition and modeling of different phenomena. The resulting models contain lots of information about the object in question, in all cases sufficient information for rerendering the object from novel viewpoints. However, more information can be obtained from the models: in the case of the smoke acquisition system it is knowledge about the different smoke densities in the complete volume. For the relighting part, it is knowledge about the environment each input image was taken in plus the reflectance properties of each surface point of the object. This information allows for a rerendering of the object not only from novel viewpoints, but also under novel illumination conditions.

This thesis proposes acquisition techniques for different classes of objects: in the first part of the thesis, we put emphasis on translucent objects. We show how to acquire certain classes of subsurface scattering objects and how to remove scattering to look further into a translucent medium. In the second part, we limit ourselves to opaque objects and drop the more general BSSRDF in favor of the BRDF. This simplification allows us to recover lighting and reflectance information from just a small number of input images.

8.1 Contributions of this Thesis

In the following, we summarize our individual contributions and discuss their advantages and disadvantages and show their advance over previously existing methods.

8.1.1 Acquisition of Volumetric Densities

In Chapter 4 we propose a novel technique for the acquisition of full volumes of participating media like smoke. We sacrifice some spatial resolution to gain improved temporal resolution, which enables us to acquire all necessary data within a single image. Missing data due to the limited spatial resolution is filled in smoothly using the push-pull algorithm.

This approach has the advantage that it is able to recover volumetric densities at interactive framerates without the use of high-speed camera equipment or high-output lasers. Since we enforce a smooth interpolation of the missing data, we can only recover details up to a certain frequency which depends on the density of the employed laser grids.

8.1.2 Combining Confocal Imaging and Descattering

With this work, which we present in Chapter 5, we analyzed two well-known techniques in computer graphics with respect to their effect on a reflectance field. It turned out, that both methods are suitable for removing multiple scattering effects within a homogeneous translucent medium like fog or murky water. Furthermore, confocal imaging also allows for the whole 3D volume medium to be scanned and therefore also yields volumetric measurements.

Although both methods achieve a similar goal, we found that the two methods are actually orthogonal and can therefore be effectively combined to improve the results. However, we still need to capture thousands of images for good quality results and therefore our approach is currently only applicable to static scenes.

8.1.3 Descattering using Polarization

In a related publication, we analyzed how polarization difference imaging combined with algorithmic descattering using phase-shifting facilitate 3D scanning of translucent objects. Normally any light-based approach to range sensing fails when dealing with subsurface scattering objects, because of the shift in peak intensity caused by the subsurface light transport.

With our approach, we are able to perform 3D scanning on a variety of difficult objects at the cost of a small number of additional images. The descattering property of polarization is not limited to phase-shifting. It can also be used with other structured light 3D scanning techniques like gray code.

8.1.4 Relighting from Image Collections

We presented the first system capable of determining object geometry, incident illumination for every input image and reflectance properties from a small number of input images. Our inverse rendering framework also works on images that were not taken under laboratory conditions. This is demonstrated by our reconstruction results using images obtained from a photo community collection website like Flickr.

We demonstrated that it is in principle possible to determine the incident lighting for each of the input images. The reconstruction is however band-limited by the bandwidth of the BRDF of the object. A problem that remains to be solved is the ambiguity between incident light and object color. Currently, the only way to resolve this issue is by having a sufficient number of input images with sufficiently different illumination colors. The quality of our results also depends greatly on the quality of the recovered surface mesh. This could be helped by more robust camera calibration.

8.2 Outlook

Although we developed new and interesting solutions for some important acquisition problems in computer graphics, many new questions and possibilities for further investigation arose during our research.

Combining confocal imaging and descattering the way we proposed it in Chapter 5 applies a local LoG filter kernel in the illumination domain only – a rather simple and straightforward approach. We strongly believe that we could exploit additional information by computing a spatially varying deconvolution on the space-time volume of the measured reflectance field.

Confocal microscopy is a standard technique among biologists and chemists. In the future, we would like to analyze further related methods used within these communities with respect to their effect on reflectance fields. At the same time, we would like to introduce the combination of confocal imaging and algorithmic descattering in biology and chemistry communities.

Another very interesting direction of research involves trying to replace the complex camera setup with a human observer. If we could get the human visual system to do the necessary calculations, we could build an illumination setup which provides instant descattering to the human observer. One possible application for this would be a novel kind of fog-lamp which enables humans to see further through fog without the need for a camera setup and displays.

The volumetric density measurements described in Chapter 4 would definitively benefit from higher resolution sampling and faster acquisition devices to

eliminate some of the motion blur. One possible direction for future research in this topic would be to combine a smoke simulation with the measurement of the real volume to obtain better interpolations for the missing parts.

Our reflectance and illumination estimation from Chapter 7 could be made more reliable by improving the underlying geometry estimation. The most important direction for future research however is solving the ambiguity between the illumination and the surface reflectance.

Bibliography

[Arridge 1999] ARRIDGE, S. 1999. Optical tomography in medical imaging. *Inverse Problems 15*, R41–R93. 51

[Bai and Liu 2006] BAI, E.-W., AND LIU, Y. 2006. Least squares solutions of bilinear equations. *Systems & Control Letters 55*, 6 (June), 466–472. 89

[Björck 1996] BJÖRCK, Å. 1996. *Numerical Methods for Least Squares Problems*. SIAM. 89

[Blinn 1977] BLINN, J. 1977. Models of Light Reflection For Computer Synthesized Pictures. 192–198. 17

[Bohren and Huffman 1983] BOHREN, C. F., AND HUFFMAN, D. R. 1983. *Absorption and Scattering of Light by Small Particles*. Wiley & Sons. 33, 35

[Born et al. 1999] BORN, M., WOLF, E., AND BHATIA, A. B. 1999. *Principles of Optics*. Cambridge University Press. 72

[Bouguet 2006] BOUGUET, J.-Y., 2006. Camera Calibration Toolbox for MATLAB. http://www.vision.caltech.edu/bouguetj/calib_doc/. 42

[Bruning et al. 1974] BRUNING, J. H., HERRIOTT, D. R., GALLAGHER, J. E., ROSENFELD, D. P., WHITE, A. D., AND BRANGACCIO, D. J. 1974. Digital wavefront measuring interferometer for testing optical surfaces and lenses. *Appl. Opt. 13*, 2693–2703. 74

[Buchanan and Fitzgibbon 2005] BUCHANAN, A. M., AND FITZGIBBON, A. W. 2005. Damped newton algorithms for matrix factorization with missing data. In *CVPR*, 316–322. 89

[Cabral et al. 1987] CABRAL, B., MAX, N., AND SPRINGMEYER, R. 1987. Bidirectional Reflection Functions From Surface Bump Maps. In *Proceedings SIGGRAPH*, 273–281. 16

[Chen et al. 2002] CHEN, W.-C., BOUGUET, J.-Y., CHU, M. H., AND GRZESZCZUK, R. 2002. Light field mapping: efficient representation and hardware rendering of surface light fields. In *SIGGRAPH*. 94

[Chen et al. 2004] CHEN, Y., TONG, X., WANG, J., LIN, S., GUO, B., AND SHUM, H.-Y. 2004. Shell Texture Functions. *ACM Trans. Graph. (SIGGRAPH 2004) 23*, 3, 343–353. 22, 25

[Chen et al. 2007] CHEN, T., LENSCH, H., FUCHS, C., AND SEIDEL, H.-P. 2007. Polarization and phase-shifting for 3d scanning of translucent objects. *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on* (June), 1–8. 2, 6, 69

[Chuang et al. 2000] CHUANG, Y.-Y., ZONGKER, D. E., HINDORFF, J., CURLESS, B., SALESIN, D. H., AND SZELISKI, R. 2000. Environment Matting Extensions: Towards Higher Accuracy and Real-time Capture. In *SIGGRAPH 2000*, 121–130. 20, 21

[Cohen and Tomasi 1997] COHEN, S., AND TOMASI, C. 1997. Systems of bilinear equations. Tech. Rep. CS-TR-97-1588, Stanford University. 89

[Corle and Kino 1996] CORLE, T. R., AND KINO, G. S. 1996. *Confocal Scanning Optical Microscopy and Related Imaging Systems*. Academic Press. 25, 47, 51

[Debevec and Malik 1997] DEBEVEC, P., AND MALIK, J. 1997. Recovering High Dynamic Range Radiance Maps from Photographs. In *SIGGRAPH 1997*, 369–378. 42

[Debevec et al. 2000] DEBEVEC, P., HAWKINS, T., TCHOU, C., DUIKER, H.-P., SAROKIN, W., AND SAGAR, M. 2000. Acquiring the reflectance field of a human face. In *Proceedings of ACM SIGGRAPH 2000*, Computer Graph-

ics Proceedings, Annual Conference Series, 145–156. 15, 20, 26, 30, 50, 57, 83

[Debevec et al. 2004] DEBEVEC, P., TCHOU, C., GARDNER, A., HAWKINS, T., POULLIS, C., STUMPFEL, J., JONES, A., YUN, N., EINARSSON, P., LUNDGREN, T., FAJARDO, M., AND MARTINEZ, P. 2004. Estimating surface reflectance properties of a complex scene under captured natural illumination. Tech. Rep. ICT-TR-06.2004, USC ICT. 27

[Debevec 1998] DEBEVEC, P. 1998. Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *SIGGRAPH*, 189–198. 28, 30, 92

[Deutsch and Dracos 2001] DEUTSCH, S., AND DRACOS, T. 2001. Time Resolved 3D Passive Scalar Concentration-Field Imaging by Laser Induced Fluorescence (LIF) in Moving Liquids. *Measurement Science and Technology 12*, 2, 188–200. 32

[DeYoung and Fournier 1997] DEYOUNG, J., AND FOURNIER, A. 1997. Properties of Tabulated Bidirectional Reflectance Distribution Functions. In *Graphics Interface*, 47–55. 16

[Dischler et al. 1998] DISCHLER, J.-M., GHAZANFARPOUR, D., AND FREYDIER, R. 1998. Anisotropic Solid Texture Synthesis Using Orthogonal 2D Views. *Computer Graphics Forum 17*, 3, 87–96. 46

[Drori et al. 2003] DRORI, I., COHEN-OR, D., AND YESHURUN, H. 2003. Fragment-Based Image Completion. *ACM Transactions on Graphics 22*, 3, 303–312. 36

[Egner et al. 2002] EGNER, A., ANDRESEN, V., AND HELL, S. W. 2002. Comparison of the axial resolution of practical nipkow-disk confocal fluorescence microscopy with that of multifocal multiphoton microscopy: theor y and experiment. *Journal of Microscopy 206*, 1, 24–32. 52

[Feng et al. 1993] FENG, S., ZENG, F., AND CHANCE, B. 1993. Monte Carlo Simulations of Photon Migration Path Distributions in Multiple Scattering Media. In *Photon Migration and Imaging in Random Media and Tissues, Proc. of SPIE Vol. 1888*, 78–89. 49, 51, 53

[Flickr 2009] FLICKR, 2009. `http://www.flickr.com`. 6, 7, 30, 84, 94

[Fuchs et al. 2005] FUCHS, C., GOESELE, M., CHEN, T., AND SEIDEL, H.-P. 2005. An empirical model for heterogeneous translucent objects. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Sketches*, ACM, New York, NY, USA, 24. 17

[Fuchs et al. 2006] FUCHS, C., CHEN, T., GOESELE, M., THEISEL, H., AND SEIDEL, H.-P. 2006. Volumetric density capture from a single image. In *Eurographics / IEEE VGTC Workshop on Volume Graphics*, 17–22. 2, 31

[Fuchs et al. 2007a] FUCHS, C., CHEN, T., GOESELE, M., THEISEL, H., AND SEIDEL, H.-P. 2007. Density estimation for dynamic volumes. *Comput. Graph. 31*, 2, 205–211. 2, 31

[Fuchs et al. 2007b] FUCHS, M., BLANZ, V., LENSCH, H. P., AND SEIDEL, H.-P. 2007. Adaptive sampling of reflectance fields. *ACM Trans. Graph. 26*, 2, 10. 15

[Fuchs et al. 2008] FUCHS, C., HEINZ, M., LEVOY, M., SEIDEL, H.-P., AND LENSCH, H. P. A. 2008. Combining confocal imaging and descattering. *Computer Graphics Forum, Special Issue for the Eurographics Symposium on Rendering 2008 27*, 4, 1245–1253. 2, 47

[Garg et al. 2006] GARG, G., TALVALA, E.-V., LEVOY, M., AND LENSCH, H. P. A. 2006. Symmetric photography: Exploiting data-sparseness in reflectance fields. In *Proceedings of Eurographics Symposium on Rendering*, 251–262. 15, 20, 50

[Georghiades 2003] GEORGHIADES, A. S. 2003. Recovering 3-d shape and reflectance from a small number of photographs. In *EGSR*, 230–240. 28

[Gershun and Pole 1939] GERSHUN, A., AND POLE, S. 1939. The light field. *Journal of Mathematics and Physics XVIII*, 51–151. 15

[Gertz and Wright 2003] GERTZ, E. M., AND WRIGHT, S. J. 2003. Object-oriented software for quadratic programming. *ACM Trans. Math. Softw. 29*, 1, 58–81. 89

[Ghiglia and Pritt 1998] GHIGLIA, D. C., AND PRITT, M. D. 1998. *Two-dimensional phase unwrapping: theory, algorithms, and software*. Wiley, New York. 75

[Godin et al. 2001] GODIN, G., RIOUX, M., BERALDIN, J.-A., LEVOY, M., COURNOYER, L., AND BLAIS, F. 2001. An Assessment of Laser Range Measurement on Marble Surfaces. In *5th Conf. on Optical 3D Measurement Techniques*. 22, 23, 72

[Goesele et al. 2004] GOESELE, M., LENSCH, H. P. A., LANG, J., FUCHS, C., AND SEIDEL, H.-P. 2004. Disco: acquisition of translucent objects. In *ACM SIGGRAPH 2004*, 835–844. 16, 21, 25, 50

[Goesele et al. 2007] GOESELE, M., SNAVELY, N., CURLESS, B., HOPPE, H., AND SEITZ, S. M. 2007. Multi-view stereo for community photo collections. In *ICCV*. 83, 85, 94

[Gordon et al. 1970] GORDON, R., BENDER, R., AND HERMAN, G. 1970. Algebraic Reconstruction Techniques (ART) for Three-dimensional Electron Microscopy and X-ray Phtography. *Journal of Theoretical Biology 29*, 471–481. 51

[Gortler et al. 1996] GORTLER, S. J., GRZESZCZUK, R., SZELINSKI, R., AND COHEN, M. F. 1996. The Lumigraph. In *Computer Graphics Proceedings (ACM SIGGRAPH 96)*, 43–54. 20, 36

[Haber et al. 2009] HABER, T., FUCHS, C., GOESELE, M., BEKAERT, P., SEIDEL, H.-P., AND LENSCH, H. P. A. 2009. Relighting objects from image collections. *to appear in: Computer Vision and Pattern Recognition, 2009. CVPR '09.* (June). 2, 83

[Hasinoff and Kutulakos 2003] HASINOFF, S. W., AND KUTULAKOS, K. N. 2003. Photo-Consistent 3D Fire by Flame-Sheet Decomposition. In *International Conference on Computer Vision*, 1184–1191. 31, 33

[Hawkins et al. 2005] HAWKINS, T., EINARSSON, P., AND DEBEVEC, P. 2005. Acquisition of Time-Varying Participating Media. *ACM Transactions on Graphics 24*, 3, 812–815. 26, 29, 31, 32, 36, 45

[Heintzmann et al. 2001] HEINTZMANN, R., HANLEY, Q. S., ANRDT-JOVIN, D., AND JOVIN, T. M. 2001. A dual path programmable array microscope (pam): simultaneous acquisition of conjugate and non-conjugate images. *Journal of Microscopy 204*, 2, 119–137. 52

[Henyey and Greenstein 1941] HENYEY, L. G., AND GREENSTEIN, J. L. 1941. Diffuse radiation in the Galaxy. *Astrophys. J 93* (Jan.), 70–83. 56

[Hullin et al. 2008] HULLIN, M. B., FUCHS, M., IHRKE, I., SEIDEL, H.-P., AND LENSCH, H. P. A. 2008. Fluorescent immersion range scanning. *ACM Transactions on Graphics 27*, 3 (Aug.), 87:1–87:10. 26

[Huntley and Saldner 1993] HUNTLEY, J. M., AND SALDNER, H. O. 1993. Temporal phase unwrapping algorithm for automated interferogram analysis. *Appl. Opt. 32*, 3047–3052. 75

[Huntley and Saldner 1997] HUNTLEY, J. M., AND SALDNER, H. O. 1997. Phase unwrapping: application to surface profiling of discontinuous objects. *Appl. Opt. 36*, 2770–2775. 75

[Ihrke and Magnor 2004] IHRKE, I., AND MAGNOR, M. 2004. Image-Based Tomographic Reconstruction of Flames. In *ACM Siggraph / Eurographics Symposium on Computer Animation*, 367–375. 25, 29, 31

[Ihrke and Magnor 2005] IHRKE, I., AND MAGNOR, M. 2005. Adaptive Grid Optical Tomography. In *Vision, Video, and Graphics*, 141–148. 25, 29, 31

[Ishimaru 1978] ISHIMARU, A. 1978. *Wave Propagation and Scattering in Random Media*. Academic Press. 19, 33, 35, 72

[Jensen et al. 2001] JENSEN, H. W., MARSCHNER, S. R., LEVOY, M., AND HANRAHAN, P. 2001. A Practical Model for Subsurface Light Transport. In *SIGGRAPH 2001*, 511–518. 17, 19, 21, 22, 72

[Kajiya 1986] KAJIYA, J. T. 1986. The rendering equation. In *Computer Graphics (SIGGRAPH '86 Proceedings)*, 20(4):143-150. 10

[Kautz and McCool 1999] KAUTZ, J., AND MCCOOL, M. 1999. Interactive Rendering with Arbitrary BRDFs using Separable Approximations. In *Tenth Eurographics Workshop on Rendering*, 281–292. 16

[Kopf et al. 2007] KOPF, J., FU, C.-W., COHEN-OR, D., DEUSSEN, O., LISCHINSKI, D., AND WONG, T.-T. 2007. Solid texture synthesis from 2d exemplars. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2007) 26*, 3, 2:1–2:9. 46

[Kuthirummal et al. 2008] KUTHIRUMMAL, S., AGARWALA, A., GOLDMAN, D. B., AND NAYAR, S. K. 2008. Priors for Large Photo Collections and What They Reveal about Cameras. In *European Conference on Computer Vision*, 74–87. 97

[Lafortune et al. 1997] LAFORTUNE, E. P. F., FOO, S.-C., TORRANCE, K. E., AND GREENBERG, D. P. 1997. Non-linear approximation of reflectance functions. In *SIGGRAPH*, 117–126. 17, 92

[Latta and Kolb 2002] LATTA, L., AND KOLB, A. 2002. Homomorphic Factorization of BRDF-based Lighting Computation. In *Proceedings SIGGRAPH*, 509–516. 16

[Lawrence et al. 2006] LAWRENCE, J., BEN-ARTZI, A., DECORO, C., MATUSIK, W., PFISTER, H., RAMAMOORTHI, R., AND RUSINKIEWICZ, S. 2006. Inverse shade trees for non-parametric material representation and editing. In *SIGGRAPH*, 735–745. 16

[Lensch et al. 2001] LENSCH, H., KAUTZ, J., GOESELE, M., HEIDRICH, W., AND SEIDEL, H.-P. 2001. Image-Based Reconstruction of Spatially Varying Materials. In *Rendering Workshop 2001*, 104–115. 17

[Lensch et al. 2003] LENSCH, H. P. A., KAUTZ, J., GOESELE, M., HEIDRICH, W., AND SEIDEL, H.-P. 2003. Image-based Reconstruction of Spatial Appearance and Geometric Detail. *ACM Trans. Graph. 22*, 2, 234–257. 23, 26, 30, 83

[Levoy and Hanrahan 1996] LEVOY, M., AND HANRAHAN, P. 1996. Light Field Rendering. In *Computer Graphics Proceedings (ACM SIGGRAPH 96)*, 31–42. 15, 20

[Levoy et al. 2004] LEVOY, M., CHEN, B., VAISH, V., HOROWITZ, M., MCDOWALL, I., AND BOLAS, M. 2004. Synthetic aperture confocal imaging. *ACM TOG (Proc. ACM SIGGRAPH 2004) 23*, 3, 825–834. 4, 25, 29, 47, 52, 59, 62

[Levoy et al. 2006] LEVOY, M., NG, R., ADAMS, A., FOOTER, M., AND HOROWITZ, M. 2006. Light field microscopy. *ACM Trans. Graph. 25*, 3, 924–934. 15, 68

[Levoy 2006] LEVOY, M. 2006. Light fields and computational imaging. *Computer 39*, 8, 46–55. 15

[Malzbender et al. 2001] MALZBENDER, T., GELB, D., AND WOLTERS, H. 2001. Polynomial Texture Maps. In *SIGGRAPH 2001*, 519–528. 20

[Marschner et al. 2000] MARSCHNER, S., WESTIN, S., LAFORTUNE, E., TORRANCE, K., AND GREENBERG, D. 2000. Image-based BRDF measurement including human skin. In *EGWR*, 139–152. 26, 30

[Marschner 1998] MARSCHNER, S. 1998. *Inverse rendering for computer graphics*. PhD thesis, Cornell University. 89

[Masselus et al. 2003] MASSELUS, V., PEERS, P., DUTRÉ, P., AND WILLEMS, Y. D. 2003. Relighting with 4D incident light fields. *ACM Transactions on Graphics 22*, 3, 613–620. 57, 60

[Miller et al. 1998] MILLER, G. S. P., RUBIN, S., AND PONCELEON, D. 1998. Lazy Decompression of Surface Light Fields for Precomputed Global Illumination. In *Rendering Workshop 1998*, 281–292. 20

[Narasimhan et al. 2005] NARASIMHAN, S., NAYAR, S., SUN, B., AND KOPPAL, S. 2005. Structured light in scattering media. *IEEE International Conference on Computer Vision (ICCV) I*, 420–427. 45, 50, 59, 64

[Narasimhan et al. 2006] NARASIMHAN, S. G., GUPTA, M., DONNER, C., RAMAMOORTHI, R., NAYAR, S. K., AND JENSEN, H. W. 2006. Acquiring Scattering Properties of Participating Media by Dilution. *ACM Transactions on Graphics 25*, 3, 1003–1012. 22, 45, 56

[Nayar et al. 1997] NAYAR, S. K., FANG, X. S., AND BOULT, T. 1997. Separation of reflection components using color and polarization. *IJCV 21*, 3, 163–186. 49

[Nayar et al. 2006] NAYAR, S. K., KRISHNAN, G., GROSSBERG, M. D., AND RASKAR, R. 2006. Fast separation of direct and global components of a scene using high frequency illumination. *ACM TOG (Proc. ACM SIGGRAPH 2006) 25*, 3, 935–944. 4, 24, 29, 30, 49, 50, 53, 62, 63, 69, 72, 73, 74, 80

[Neil et al. 1997] NEIL, M. A. A., JUSKAITIS, R., AND WILSON, T. 1997. Method of obtaining optical sectioning by using structured light in a conventional microscope. *Optics Letters 22*, 24. 25, 47, 52, 62

[Ng et al. 2004] NG, R., RAMAMOORTHI, R., AND HANRAHAN, P. 2004. Triple product wavelet integrals for all-frequency relighting. In *SIGGRAPH*, 477–487. 16, 84, 85, 86, 87

[Ngan et al. 2005] NGAN, A., DURAND, F., AND MATUSIK, W. 2005. Experimental analysis of BRDF models. In *EGSR*, 117–126. 90

[Nicodemus et al. 1977] NICODEMUS, F. E., RICHMOND, J. C., HSIA, J. J., GINSBERG, I. W., AND LIMPERIS, T. 1977. Ge-

ometrical Considerations and Nomenclature for Reflectance. NBS. 6, 14, 72

[Nishino et al. 2001] NISHINO, K., ZHANG, Z., AND IKEUCHI, K. 2001. Determining reflectance parameters and illumination distribution from a sparse set of images for view-dependent image synthesis. In *ICCV*, 599–606. 28, 83

[Nishino et al. 2005] NISHINO, K., IKEUCHI, K., AND ZHANG, Z. 2005. Re-rendering form a sparse set of images. Tech. Rep. DU-CS-05-12, Drexel University. 28

[Okabe et al. 2004] OKABE, T., SATO, I., AND SATO, Y. 2004. Spherical harmonics vs. haar wavelets: Basis for recovering illumination from cast shadows. In *CVPR*, 50–57. 28

[Owada et al. 2004] OWADA, S., NIELSEN, F., OKABE, M., AND IGARASHI, T. 2004. Volumetric Illustration: Designing 3D Models with Internal Textures. *ACM Transactions on Graphics 23*, 3, 322–328. 46

[Peers et al. 2006] PEERS, P., VOM BERGE, K., MATUSIK, W., RAMAMOORTHI, R., LAWRENCE, J., RUSINKIEWICZ, S., AND DUTRÉ, P. 2006. A compact factored representation of heterogeneous subsurface scattering. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, ACM, New York, NY, USA, 746–753. 16, 21, 25

[Phong 1975] PHONG, B.-T. 1975. Illumination for Computer Generated Pictures. *Comm. of the ACM 18*, 6, 311–317. 17

[Praun and Hoppe 2003] PRAUN, E., AND HOPPE, H. 2003. Spherical parametrization and remeshing. *ACM TOG 22*, 3, 340–349. 86

[Ramamoorthi and Hanrahan 2001] RAMAMOORTHI, R., AND HANRAHAN, P. 2001. A signal-processing framework for inverse rendering. In *SIGGRAPH*, 117–128. 16, 28, 84, 86, 89, 92

[Robertson et al. 2003] ROBERTSON, M., BORMAN, S., AND STEVENSON, R. 2003. Estimation-theoretic approach to dynamic

range improvement using multiple exposures. *Journal of Electronic Imaging 12*, 2, 219–228. 79

[Romeiro et al. 2008] ROMEIRO, F., VASILYEV, Y., AND ZICKLER, T. 2008. Passive reflectometry. In *ECCV*. 27, 84, 86

[Rowe and Pugh Jr. 1995] ROWE, M. P., AND PUGH JR., E. N. 1995. Polarization-difference imaging: a biologically inspired technique for observation through scattering media. *Optics Letters 20*, 6, 608–610. 24, 29, 49, 69, 76

[Salvi et al. 2004] SALVI, J., PAGES, J., AND BATLLE, J. 2004. Pattern codification strategies in structured light systems. *Pattern Recognition 37*, 4, 827–849. 75

[Sato et al. 2003] SATO, I., SATO, Y., AND IKEUCHI, K. 2003. Illumination from shadows. *Trans. PAMI 25*, 3, 290–300. 28, 83, 89

[Schechner and Karpel 2004] SCHECHNER, Y. Y., AND KARPEL, N. 2004. Clear under water vision. In *Proceeding of CVPR*, 536–543. 76

[Schechner and Karpel 2005] SCHECHNER, Y. Y., AND KARPEL, N. 2005. Recovery of underwater visibility and structure by polarization analysis. *IEEE Journal of Oceanic Engineering 30*, 3, 570–587. 24, 29, 49, 69, 73, 76

[Schechner et al. 2003] SCHECHNER, Y. Y., NARSIMHAN, S. G., AND NAYAR, S. K. 2003. Polarization-based vision through haze. *Applied Optics 42*, 3, 511–525. 49, 62, 73, 76

[Seberry and Yamada 1992] SEBERRY, J., AND YAMADA, M. 1992. Hadamard matrices, sequences, and block designs. In *Dinitz, J. H. and Stinson, D. R., editors (1992), Contemporary Design Theory: A Collection of Essays*, 431–560. 50

[Seitz et al. 2005] SEITZ, S., MATSUSHITA, Y., AND KUTULAKOS, K. 2005. A theory of inverse light transport. In *ICCV*, 1440–1447. 50

[Sen et al. 2005] SEN, P., CHEN, B., GARG, G., MARSCHNER, S., HOROWITZ, M., LEVOY, M., AND LENSCH, H.

P. A. 2005. Dual photography. *ACM Trans. on Graph 24*, 3, 745–755. 20, 21, 50, 53, 57

[Sharpe et al. 2002] SHARPE, J., AHLGREN, U., PERRY, P., HILL, B., ROSS, A., HECKSHER-SORENSEN, J., BALDOCK, R., AND DAVIDSON, D. 2002. Optical Projection Tomography as a Tool for 3D Microscopy and Gene Expression Studies. *Science 296*, 19. 51

[Sloan et al. 2002] SLOAN, P.-P., KAUTZ, J., AND SNYDER, J. 2002. Precomputed Radiance Transfer for Real-time Rendering in Dynamic, Low-frequency Lighting Environments. In *SIGGRAPH 2002*, 527–536. 86

[Srinivasan et al. 1985] SRINIVASAN, V., LIU, H.-C., AND HALIOUA, M. 1985. Automated phase-measuring profilometry: A phase mapping approach. *Appl. Opt. 24*, 185–188. 70

[Tarini et al. 2005] TARINI, M., LENSCH, H. P. A., GOESELE, M., AND SEIDEL, H.-P. 2005. 3d acquisition of mirroring objects. *Graphical Models 67*, 4 (July), 233–259. 75

[Tong et al. 2005] TONG, X., WANG, J., LIN, S., GUO, B., AND SHUM, H.-Y. 2005. Modeling and rendering of quasi-homogeneous materials. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, ACM, New York, NY, USA, 1054–1061. 21

[Torrance and Sparrow 1967] TORRANCE, K., AND SPARROW, E. 1967. Theory for off-specular reflection from roughened surfaces. *Journal of Optical Society of America 57*, 9, 1105 1114. 17

[Treibitz and Schechner 2006] TREIBITZ, T., AND SCHECHNER, Y. Y. 2006. Instant 3Descatter. In *Proceedings of CVPR*, 1861–1868. 24, 29, 49, 69, 76

[Treuille et al. 2006] TREUILLE, A., LEWIS, A., AND POPOVIĆ, Z. 2006. Model Reduction for Real-Time Fluids. *ACM Transactions on Graphics 25*, 3, 826–834. 37

[Trifonov et al. 2006] TRIFONOV, B., BRADLEY, D., AND HEIDRICH, W. 2006. Tomographic reconstruction of transparent objects. In *Rendering Techniques 2006: 17th Eurographics Workshop on Rendering*, 51–60. 25, 29, 51

[Tyo et al. 1996] TYO, J. S., ROWE, M. P., PUGH, E. N., AND ENGHETA, N. 1996. Target detection in optically scattering media by polarization difference imaging. *App. Opt. 35*, 639–647. 49, 76

[van de Hulst 1981] VAN DE HULST, H. C. 1981. *Light scattering by small particles*. Courier Dover Publications, New York, Dover. 73

[von Helmholtz 1856] VON HELMHOLTZ, H., 1856. Treatise on physiological optics. The Optical Society of America. Electronic edition(2001):University of Pennsylvania http://psych.upenn.edu/backuslab/helmholtz. 14, 53

[Wang et al. 2006] WANG, R., NG, R., LUEBKE, D., AND HUMPHREYS, G. 2006. Efficient wavelet rotation for environment map rendering. In *EGSR*, 173–182. 87

[Wang et al. 2008] WANG, J., ZHAO, S., TONG, X., LIN, S., LIN, Z., DONG, Y., GUO, B., AND SHUM, H.-Y. 2008. Modeling and rendering of heterogeneous translucent materials using the diffusion equation. *ACM Trans. Graph. 27*, 1, 1–18. 22

[Weistroffer et al. 2007] WEISTROFFER, R. P., WALCOTT, K. R., HUMPHREYS, G., AND LAWRENCE, J. 2007. Efficient Basis Decomposition for Scattered Reflectance Data . In *EGSR*, 207–218. 27, 30, 89

[Wenger et al. 2005] WENGER, A., GARDNER, A., TCHOU, C., UNGER, J., HAWKINS, T., AND DEBEVEC, P. 2005. Performance relighting and reflectance transformation with time-multiplexed illumination. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, ACM, New York, NY, USA, 756–764. 15

[Wilson et al. 1996] WILSON, T., JUSKAITIS, R., NEIL, M., AND KOZUBEK, M. 1996. Confocal microscopy by aperture correlation. *Optics Letters 21*, 3. 25, 47, 52

[Wilson et al. 1998] WILSON, T., NEIL, M. A. A., AND JUSKAITIS, R. 1998. Real-time three-dimensional imaging of macro-

scopic structures. *Journal of Microscopy 191*, 2, 113–220. 52, 62

[Wolff and E.Boult 1991] WOLFF, L. B., AND E.BOULT, T. 1991. Constraining object features using a polarization reflectance model. *PAMI 13*, 7, 635–657. 72

[Wolff 1994] WOLFF, L. B. 1994. Polarization camera for computer vision with a beam splitter. *Journal of the Optical Society of America 11*, 2935–2945. 24, 29, 69

[Wood et al. 2000] WOOD, D. N., AZUMA, D. I., ALDINGER, K., CURLESS, B., DUCHAMP, T., SALESIN, D. H., AND STUETZLE, W. 2000. Surface Light Fields for 3D Photography. In *SIGGRAPH 2000*, 287–296. 20

[Yip et al. 1987] YIP, B., LAM, J. K., WINTER, M., AND LONG, M. B. 1987. Time-Resolved Three-Dimensional Concentration Measurements in a Gas Jet. *Science 235*, 1209–1211. 32

[Yu and Malik 1998] YU, Y., AND MALIK, J. 1998. Recovering photometric properties of architectural scenes from photographs. In *SIGGRAPH*, 207–217. 27, 83

[Yu et al. 2006] YU, T., WANG, H., AHUJA, N., AND CHEN, W.-C. 2006. Sparse lumigraph relighting by illumination and reflectance estimation from multi-view images. In *EGSR*, 41–50. 28, 30, 83, 85, 94

[Zhang and Yau 2006] ZHANG, S., AND YAU, S.-T. 2006. High-resolution, real-time absolute 3-D coordinate measurement based on the phase shifting method. *Opt. Express 14*, 2644–2649. 70

[Zhang 2000] ZHANG, Z. 2000. A flexible new technique for camera calibration. *PAMI 22*, 11, 1330–1334. 79

[Zickler et al. 2006] ZICKLER, T., RAMAMOORTHI, R., ENRIQUE, S., AND BELHUMEUR, P. 2006. Reflectance sharing: predicting appearance from a sparse set of images of a known shape. *Trans. PAMI 28*, 8, 1287–1302. 26, 30, 97

[Zongker et al. 1999] ZONGKER, D. E., WERNER, D. M., CURLESS, B., AND SALESIN, D. H. 1999. Environment Matting and Compositing. In *SIGGRAPH 1999*, 205–214. 20, 21