THE UNIVERSITY OF CALGARY

Biologically and Physically-Based Rendering of Natural Scenes

by

Gladimir V. Guimarães Baranoski

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE

CALGARY, ALBERTA

NOVEMBER, 1998

THE UNIVERSITY OF CALGARY
FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a dissertation entitled "Biologically and Physically-Based Rendering of Natural Scenes" submitted by Gladimir V. Guimarães Baranoski in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Supervisor, Jon G. Rokne
Department of Computer Science

Brian Wyvill
Department of Computer Science

Przemyslaw Prusinkiewicz
Department of Computer Science

Peter Ehlers
Department of Mathematics and Statistics

External Examiner, Alain Fournier
Department of Computer Science
The University of British Columbia

Date

Abstract

Physically-based rendering methods represent the core of current realistic image synthesis frameworks. These methods, through a plausible simulation of the processes of light propagation and interaction with objects, have contributed considerably to the improvement of photorealistic rendering. The state of art research in this area includes the simulation of natural phenomena and the incorporation of biological aspects affecting light propagation in natural environments. The search for more efficient rendering solutions is also of major interest for the rendering community.

In this dissertation biologically and physically-based models for light interaction with plant leaves are presented. Moreover, since the light that reach a plant leaf may be propagated directly from a light source or indirectly, due to multiple interactions with other objects in the environment, global illumination issues are also addressed, more specifically related to the radiosity method. This method is commonly used to account for the indirect mechanisms of light transfer. For this reason, faster solutions for radiosity systems are also investigated in this dissertation.

The fundamental aspects of physically-based rendering as well as the main factors affecting the light propagation by plant leaves are first presented to form a technical background for the research that follows. An algorithmic reflectance and transmittance model for plant tissue is presented next. Its accuracy is evaluated through comparisons with real measured data. For applications involving several foliar primitives a nondeterministic algorithm for the reconstruction of isotropic reflectance and transmittance functions is presented as well as a simplified model for light interaction with foliar tissues. Both models and the reconstruction algorithm have a formulation based on standard Monte Carlo methods.

The study of the spectral properties of the radiosity systems has not been extensively explored by computer graphics and illuminating engineering researchers despite its importance. These properties are investigated with the purpose of finding faster so-

lutions for radiosity problems. Besides this theoretical study, the Chebyshev method for solving systems of linear equations is introduced to the rendering literature. This method is particularly suitable for environments with high average reflectance and occlusion commonly occurring in environmental design and remote sensing applications. Tests show that this method outperforms the iterative methods usually applied to solve the radiosity systems under these conditions.

Finally, this dissertation closes with an outline of aspects that are worthy of further investigation or extension. Among these aspects, the investigation of the physical meaning of eigenvalues and eigenvectors in the global illumination context is highlighted. Preliminary experiments are presented in the appendices.

Acknowledgements

First of all, I would like to give my warmest thanks to two very special women. My mother, Eledi, use to say that she gave me "wings". She's right, and I am deeply indebted to her because of this. My wife, Adriane, knows better than anybody else how tough this pursuit of a Ph.D. degree was. Her patience, confidence and love were crucial in key moments of my program. This degree belongs to her too. Thanks a lot!

My main achievements during this period in North America were not academic. I am talking about my kids, Patricia and Gabriel. So, I want to give my sincere thanks to the friends who looked after them many times. Without their help in some difficult moments, this journey would have been much tougher.

I am very grateful to Jon Rokne, my supervisor, for his unconditional support and friendship that made this journey much more enjoyable and fruitful. In fact, I am not sure if I would have made to the other end of the tunnel without his help. I consider myself very fortunate for having worked, during my graduate studies, with supervisors who always consider the best interest of their students.

I also enjoyed working with Randall Bramley and Peter Shirley at Indiana University. They provided a valuable background for my doctorate research. While at Indiana, I also enjoyed the friendship and support of Younghee Lee, Chih-Yi Chen, Alan Keahey, Shankar Swamy, Tom Loos, Rajesh Kamath ("Romário") and Paulo Maciel.

I am also grateful to all graphics jungle colleagues. My special thanks to Daniel Dimian, whose friendly conversations about sports, soccer, movies, and even computer graphics, were much appreciated. My special thanks also to Radomir Mech and Dale Brisinda, who many times listened to my complaints and gave me the opportunity to release some steam.

I want to thank my defense committee: Peter Ehlers, Przemyslaw Prusinkiewicz, Brian Wyvill and Alain Fournier. Their comments and suggestions helped me to improve the quality of this dissertation.

I also want to show my appreciation to my "mock defense" committee: Robson Lemos, Mark Tigges, Dave Wilson, Jalal Kawash, Guangwu Xu, Radek Karwowski and Xikun Liang. Their confidence on the "eve of the battle" had a positive effect on my own attitude.

During graduate studies one often needs something to keep one's sanity. Soccer and running were my main "therapeutic" activities. I therefore thank my soccer buddies and the Human Performance Laboratory (HPL) guys. My special thanks to Marco Aurélio Vaz, who provided valuable logistics support during the early stages of my program and introduced me to the HPL relay team, and to the team's captain, Joanne Archambault.

Many people have provided valuable insights and feedback for the research presented in this dissertation: James Arvo, Wayne Eberly, Richard Levy, Laslo Neumann, Fred Baret, Michael Chelle, Stephane Jacquemoud, Rui Bastos, Ian Ashdown, Manuel Menezes de Oliveira Neto, Shirley Bray ... I thank all of them for their contributions. My special thanks to Ken Girard of the Department of Biology for being so generous with his time growing the soybean plants used in this research. I am also grateful to the Joint Research Center of the European Commission and the Advanced Techniques Unit of the Space Applications Institute (Italy) for granting me access to the *LOPEX* data set used in my experiments.

I could not have performed my doctorate program without financial support. Therefore, I want to express my sincere gratitude for the financial support provided by the Brazilian National Council of Research (CNPq), the Department of Computer Science and the Faculty of Graduate Studies of the University of Calgary.

To my family.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

Computer graphics involves the creation, storage, and manipulation of models and synthetic images of objects. These images are used in areas as diverse as education, science, engineering, medicine and entertainment [75]. Initially the main goal was mostly to generate visually pleasing images. As the field developed this goal was modified to include physical realism. Although there are applications, such as imagery for the entertainment industry, for which the aesthetic aspects are important, there are also applications, such as illuminating engineering, for which there is an increasing demand for realistic images that can provide a physically truthful representation of an object or an environment.

In order to add physical realism to a synthetic image the physical environment has to be investigated. A substantial amount of computer graphics research has been devoted to the simulation of light transport, *i.e.* the large-scale interaction of light and matter [7], through rendering algorithms. In this context, the problem of determining the appearance of an environment by simulating the transport of light within it, which includes the problems of light emission, propagation, scattering and absorption, is known as global illumination [7, 211].

This dissertation deals primarily with the research of physical models and numerical methods for the simulation of light transport aiming at realistic image synthesis applications. In the following section a historical survey of the global illumination methods used in computer graphics is presented. Then the interplay between the local and global aspects of light transport is briefly discussed. Finally this chapter closes with an outline of the theoretical and practical contributions and with an overview of the contents of this dissertation.

1.1 Historical Survey

The roots of the global illumination methods used in computer graphics can be traced back to the turn of the last century. In 1900, Planck introduced the idea that energy comes only in discrete quantities, which is considered the beginning of quantum mechanics, a revolutionary theory for submicroscopic phenomena [113]. In 1905, building on Planck's idea, Einstein postulated that energy along an incident beam is quantized into small packets or "particles", later called *photons* [144], whose individual energy is given in terms of their wavelength and Planck's constant [86, 113]. The concept of photons is fundamental for geometrical optics [113, 149], also called ray optics, which involves the study of the particle nature of light. Also in 1905, Schuster [192] published a paper on radiation through a foggy atmosphere, and later, in 1906, Schwarzschild published a paper on the equilibrium of the Sun's atmosphere [159]. These two astrophysical works are considered the beginning of the radiative transfer theory [7, 45, 179]

In geometrical optics the large-scale behavior of light, such as reflection and refraction, is described by assuming light to be composed of non-interacting rays, each of them carrying a certain amount of energy. The first algorithm for ray tracing was presented by Appel [6]. Whitted [236] extended this algorithm by incorporating many of the principles of geometrical optics. This improved algorithm became the substrate for a group of global illumination methods such as path tracing. This stochastic ray tracing based method was introduced in the graphics literature by Kajiya [131] along with a general physical framework for the global illumination problem known as the rendering equation (further discussed in Section 2.4).

The radiative transfer theory combines principles of geometrical optics and thermodynamics to characterize the flow of radiant energy at scales large compared with its wavelength and during time intervals large compared to its frequency [7]. This theory is the substrate of another group of global illumination methods, namely the radiosity based methods (further discussed in Chapters 7 and 8), introduced in the computer graphics literature by Goral *et al.* [94]. The introduction of these methods is considered the first step towards a more accurate approach to the global illumination problem, known as physically-based rendering [136].

There are phenomena at the level of electromagnetism, such as interference, diffraction and polarization [113, 198], that cannot be explained by geometrical optics or radiative transfer theory. These phenomena are addressed by physical optics [113, 149], which involves the study of the wave nature of light. Despite this limitation, geometrical optics and radiative transfer theory are the two levels of physical description largely used in the simulation of visual phenomena aimed at image synthesis applications. As mentioned by Glassner [86], this is a pragmatic choice, since the simulation of phenomena such as interference, diffraction and polarization usually requires enormous computational resources to produce results of significantly higher error fidelity and complexity than those attainable by geometrical optics. Models derived from physical optics theory [112] can be incorporated into a global illumination algorithm operating in the radiative transfer level [207]. When physical optics phenomena are relevant for a given image synthesis application they are usually handled as special cases [164, 214] or modeled using rendering algorithms based on geometrical optics [219].

Ray tracing and radiosity based methods have been extended and combined to solve different parts of the global illumination problem since their introduction in computer graphics. Ray tracing based methods are particularly suitable for applications involving specular (mirror-like) reflections and refractions of light, but present a high computational cost for applications involving multiple diffuse scatterings of light. Radiosity methods, on the other hand, are particularly suitable to handle multiple diffuse interactions of light, but present a high computational cost for handling the high frequency signals associated with specular reflections and refractions of light.

The different approaches proposed to combine ray tracing and radiosity methods, known as hybrid or multipass methods, take advantage of the strengths of ray tracing and radiosity while insuring that a single type of light transport is not accounted for more than once [86]. For a comprehensive review of these methods the reader is referred to the reference works by Cohen and Wallace [57], Sillion and Puech [211] and Glassner [86]. Although these methods are responsible for the generation of very realistic images, they usually do not accommodate materials that present properties intermediate to those expected from ideal specular and ideal diffuse materials.

It may be noteworthy to point out that, despite the latest achievements in rendering, it is not feasible, with the current technology, to give a viewer the same impression from a synthetic image as from viewing a real scene. Among the factors that preclude this are aspects beyond the scope of this dissertation such as hardware limitations, characteristics of the human visual system and cognitive aspects of perception. As pointed out by Shirley [198], these aspects have forced computer graphics researchers to aim considerably lower than completely duplicating the perceptual response.

Nevertheless, the physical accuracy and the computational trade-offs involved in the processes of image synthesis are the dominant aspects of the current physically-based rendering research. Physical accuracy not only ensures the correctness of the resulting images, but also makes the images predictive [7], a key requisite for practical applications such as architectural and environmental design. Furthermore, the use of physical principles makes the rendering processes more intuitive, since, as pointed out by Arvo [7], it is far easier to control a simulation using familiar physical concepts than through arcane parameters. However, an accurate simulation of the behavior of light in an environment, using current technology, is still computationally expensive. Rather than sacrificing the quality of the synthetic images or simply increasing the complexity of the rendering algorithms, the current trend in this field is to push the boundaries of both accuracy and efficiency, and to extend the application of physically-based rendering methods to materials with complex modes of light interaction.

1.2 Interplay between Local and Global Illumination

In studying the core problem of physically-based rendering, global illumination, it is often useful to distinguish its two components, direct and indirect illumination. By direct illumination we mean the light from a light source[1] that arrives at an object without interacting with any other object along its path [86]. By indirect illumination we mean the light that undergoes at least one and possibly multiple interactions with other objects before arriving at a given object. The rendering algorithms that

[1]Typical light sources or luminaires include thermal radiators, such as flames or incandescent filaments, fluorescent bulbs and decaying phosphors [86].

explicitly gather illumination information only from direct light sources are usually called local illumination models and the rendering algorithms that include estimates of indirect illumination are usually called global illumination models [86].

Although these two terms local and global describe the extremes of a continuum, one can not overlook the interplay between them. As pointed out by Fournier [76], the only way the light acts at the rendering level is locally, through intricate phenomena, such as reflection, transmission and absorption of the incident light, which are physically represented by reflectance and transmittance models (discussed further in Section 2.6). In addition, considering that well-separated objects in an environment can influence one another's appearance, either by blocking light in the case of direct illumination, or by scattering light in the case of indirect illumination [7], direct and indirect illumination are globally interconnected.

1.3 Outline of Contributions

In this dissertation original contributions to the study of local and global aspects of light interaction with biological materials are made. These contributions range from the development of reflectance and transmittance models for plants to the introduction of faster solutions for radiosity systems.

The propagation[2] to of light by most natural materials cannot be described as being purely specular or perfectly diffuse. Plants form a large class of natural materials that fit this characteristic and for which there is a lack of physically and biologically-based reflectance models aiming at rendering applications. Rather than proposing a general purpose reflectance model, we intend to present reflectance models specifically designed to take into account the biological factors that affect light propagation by plants.

The proposed models present different levels of complexity to increase their efficiency according to the rendering problem at hand. We consider this modeling approach compatible with the current trends of seeking for practical solutions that

[2]Instead of explicitly mention reflection and transmission we will use the term propagation when we refer to both processes in this dissertation.

allow us to render materials with complex modes of reflection faster without undermining the accuracy of the simulations. In addition, it is also consistent with the radiative transfer modeling approaches used in the study of vegetation as stated by Asrar and Mynemi [12]:

> We view the two types of modeling activities (*i.e.* simple *vs.* detailed models) as complementary. Development of simple models could be pursued either in parallel or as a follow-on to the development of more sophisticated radiative transfer models.

The contributions to the global aspects of the simulation of light transport are theoretical and practical. As mentioned by Arvo [7], every global illumination problem involves the solution of some linear system at some level. Moreover, several fundamental operators that arise in global illumination can be uniformly approximated by matrices, which suggests that functional analysis might be a useful tool for providing a better theoretical foundation for global illumination [9]. This suggestion is followed in this dissertation in the form of an investigation of the spectral properties of the radiosity matrix. We combine the results of this investigation with the introduction of a numerical method which provides faster solutions for radiosity systems associated with high average reflectance environments. We also discuss a more accurate approach to evaluate the performance of the iterative solvers used in the radiosity context.

The connection between the contributions of this dissertation to the local and global aspects of the simulation of light transport resides in the importance of radiosity algorithms for radiative transfer in vegetation, which was appropriately expressed by Asrar and Mynemi [12]:

> The radiative transfer models that are currently in use for the study of vegetative surfaces account only for the direct and diffuse components of the incident solar energy, and the diffuse components that may emanate from the vegetation and underlying soil surface. ... The concept of radiosity would permit accounting for all sources of radiation that may arrive at a surface independent of sensor position. ... The frontiers of research in

this area are wide open and deserve the attention of the radiative transfer modeling community concerned with the subject of vegetation canopies.

Moreover, although the development and improvement of radiosity algorithms for rendering applications was not primarily aimed at organic materials, their use in the field of radiative transfer in vegetation has grown rapidly, which is exemplified by related works not only in computer graphics [155, 158], but also in remote sensing [35, 84, 89]. As a result of this trend, the development of faster solutions for radiosity systems has also become relevant to applications involving plant canopies, as described in a recent paper by Chelle *et al.* [48].

1.4 Dissertation Overview

In this dissertation we will focus on the physical aspects of the light transport and we will describe light propagation in terms of geometrical optics. As mentioned in Section 1.1, geometrical optics is usually used in computer graphics instead of physical optics, since, from a practical point of view, it is more efficient to model light as rays rather than waves. We can think of a wave as just a ray with an energy, and the wavelength of light, a physical optics parameter important for rendering applications, can be included in geometrical optics by associating a wavelength with each ray [198]. Furthermore, as pointed out by Shirley [198] and Arvo [7], in many situations physical optics effects are not visually important, and do not dominate the scenes that we commonly wish to simulate. For instance, the light sources commonly used in rendering applications are usually incoherent, and effects related to phase, such as interference, are usually masked [7]. Also, diffraction phenomena are noticeable for long wavelength radiation, but have a fairly small effect for visible light [198].

In addition, in the experiments presented in this dissertation we will assume that the energies of different wavelengths are decoupled. In other words, the energy associated with some region of the space, or surface, at wavelength λ_1 is independent of the energy at λ_2 [86]. We will also assume that objects within an environment exchange energy directly with no atmospheric attenuation. As mentioned by Arvo [7], atmospheric effects over tens of meters are insignificant under normal circumstances.

Finally, participating media such as smoke, fog or water will not be addressed in this dissertation.

Many of the techniques and results reported in this dissertation have been presented at computer graphics conferences [14, 15, 17, 18, 19]. The main body of this document is organized as follows:

- Chapter 2 introduces relevant aspects of physically-based rendering that are used throughout this dissertation, including the formulation of the light transport yield by the rendering equation and the definition of important reflectance and transmittance terms. A brief survey of reflectance models used in computer graphics is also presented along with an overview of global illumination applications aimed at vegetation.

- Chapter 3 presents the main factors affecting the interaction of light with plant tissues, with a particular emphasis to leaves since they are the most important plant surface interacting with light. An overview of the most important reflectance models for plants, mostly oriented to remote sensing applications, is also presented. Although the requirements are different in computer graphics, some of the intuitive concepts used in remote sensing models can be incorporated into models aimed at image synthesis applications.

- Chapter 4 introduces a new algorithmic reflectance and transmittance model for plant tissue aiming at image synthesis applications. The model accounts for the three components of light propagation in plant tissues, namely surface reflectance, subsurface reflectance and transmittance, and mechanisms of light absorption by pigments present in these tissues. Its design is based on the available biological information. The model is controlled by a small number of biologically meaningful parameters, and its formulation is based on standard Monte Carlo techniques. The spectral curves of reflectance and transmittance computed using the model are compared with measured curves from actual experiments for real specimens. The formulation and techniques used to simulate the virtual measurement devices used in the testing of the models proposed in Chapters 4 to 6 are also presented in this chapter.

- Chapter 5 introduces a nondeterministic reconstruction algorithm for isotropic reflectance and transmittance functions that can be used as an alternative for deterministic approaches, such as spherical harmonics [44, 207] or wavelets [51, 139, 140, 141, 235], in applications involving isotropic materials. This algorithm, also based on standard Monte Carlo methods, is general enough to be not only applied to the rendering of plants, but also to access arbitrary isotropic reflectance and transmittance functions. The reconstructed spectral curves obtained through this algorithm are compared with spectral curves provided by the model introduced in Chapter 4 in order to determine their accuracy.

- Chapter 6 introduces a simplified model for light interaction with plant tissue aiming at applications demanding higher performance. It accounts for the main biological characteristics of these materials needed to preserve their rendering quality, while avoiding undue complexity in order to increase their rendering efficiency. The model's formulation is also based on Monte Carlo techniques, and it can be incorporated into global illumination frameworks without a significant computational overhead. Its accuracy and performance is also examined through comparisons with the model introduced in Chapter 4.

- Chapter 7 discusses some important characteristics of the radiosity system of linear equations and presents tools used in the investigation of its spectral properties. Particular emphasis is given to the definition and basic properties of the eigenvalues and eigenvectors of the system along with their relationship to the iterative methods used to solve the radiosity system. The major original contribution introduced in this chapter is the analytical proof that all of the eigenvalues of the radiosity matrix are real and positive. This fact, among other implications, permits the computation of tight estimates for the eigenvalues of the radiosity matrix, which are used to obtain faster solutions for radiosity systems.

- Chapter 8 introduces a numerical method to the graphics literature that converges faster than the methods usually used in the radiosity problems associated with environments with high average reflectance. We compare the performance

of this method to other iterative solvers, and show that the performance evaluation approach based on a small number of iterations, as opposed to small amounts of time, commonly used in this area of research, can lead to incorrect conclusions. We also show numerically the important relationship between the characteristics of a given environment, such as average reflectance and level of occlusion, and the spectral characteristics of the corresponding radiosity matrix. We focus this discussion on the impact of this relationship on the convergence of the iterative methods used to solve the radiosity system.

- Finally, Chapter 9 summarizes the results and main contributions of this dissertation. A list of open questions is also presented along with possible directions for future research.

The references and appendices follow the main body of this document. The appendices are organized as follows:

- Appendix A presents some fundamental Monte Carlo definitions and techniques used in this research as well as the concise derivation of functions used to generate scattered directions in rendering applications.

- Appendix B compares a deterministic method for form factor computation with a nondeterministic one. This comparison is focused on their numerical accuracy regarding selected geometries, using the corresponding analytical form factors as accuracy standards.

- Appendix C shows some interesting features of using the eigenvectors of the radiosity matrix as solution vectors in graphics settings. The information gathered from these experiments is expected to contribute to the understanding of physical meaning of the eigenvectors in the radiative transfer context, which, in turn, may lead to faster solutions for global illumination problems.

- Finally, Appendix D proposes a parameter to scale the eigenvector corresponding to the largest eigenvalue of the symmetric radiosity matrix. The purpose of this scaling operation is to use the resulting scaled eigenvector as an initial

approximation for the solution vector of a radiosity system. The use of such an initial approximation may also result in faster solutions for global illumination problems, particularly those associated with radiosity systems.

Chapter 2
Fundamentals of Physically-Based Rendering

In this chapter we will concentrate on the fundamental aspects of physically-based rendering that will be used throughout this dissertation. Physically-based rendering involves simulating the behavior of light, starting from luminaires, *i.e.* area light sources, traveling through the environment, interacting with different objects, and finally reaching the viewer. Reflectance and transmittance models are used to describe how the light interacts with different objects. These models must satisfy certain physical requirements to avoid excluding important physical effects and to maintain the energy consistency needed for global illumination calculations [235]. These issues will also be examined in this chapter.

The sections of this chapter are organized as follows:

- Section 2.1 presents basic laws of optics aiming at rendering applications.
- Section 2.2 describes important spectral radiometric terms along with their properties.
- Section 2.3 presents the laws that describe the absorption of transmitted light in homogeneous materials.
- Section 2.4 outlines the basic formulation of light transport yield by the rendering equation.
- Section 2.5 presents the basic mechanisms of light transfer, the terminology used to describe the reflective and refractive behavior of the surfaces and a notation proposed by Heckbert [116] to represent these mechanisms.

- Section 2.6 introduces a concise overview of reflectance and transmittance models used in computer graphics.

- Finally, Section 2.7 summarizes the contents of this chapter.

2.1 Basic Optics

Throughout this dissertation we will use the following terminology suggested by Meyer-Arendt [161]. Terms ending in *-ion*, such as reflection, transmission and absorption, describe a process. Terms ending in *-ivity*, such as reflectivity, transmissivity and absorptivity, refer to a general property of a material. Terms ending in *-ance*, such as reflectance, transmittance or absorptance, refer to properties of a given object or surface.

Reflection is the process in which light at a specific wavelength incident on a material is propagated outward by the material without a change in wavelength. Similarly, transmission is the process in which light at a specific wavelength incident on the interface between materials passes through the interface and into the other material without a change in wavelength [86].

Hall [102] suggested that reflection and transmission can be broken into two components, a coherent component and an incoherent or scattered component [23]. The coherent component is reflected using the law of reflection and transmitted using the law of refraction, which are described later. The incoherent component is reflected and transmitted in all directions based upon a statistical probability function associated with surface properties (Appendix A).

Absorption is a general term for the process by which the light incident on a material is converted to another form of energy, usually to heat. All of the incident light is accounted for by the processes of reflection, transmission and absorption [5].

The reflection and transmission (refraction[1]) of light at the smooth surfaces of pure materials is described by the Fresnel equations [86, 113, 198]. Before getting to

[1]Refraction, or the coherent component of transmission, can be defined as the bending or the change in the direction of the light rays as they pass from one medium to another [122]. This bending is determined by the change in the velocity of propagation associated with the different indexes of refraction of the media [71].

the specifics of the Fresnel equations, however, we shall review some relevant physical parameters, definitions and laws.

Materials such as conductors (metals), semi-conductors and dielectrics are characterized by their complex index of refraction, $N(\lambda)$, which is composed of a real and an imaginary term. The real term corresponds to the real index of refraction (refractive index, for short), which measures how much an electromagnetic wave slows down relative to its speed in vacuum [86]. The imaginary term corresponds to the extinction coefficient, which represents how easily an electromagnetic wave can penetrate into the medium [86]. The resulting expression for $N(\lambda)$ is given by:

$$N(\lambda) = \eta(\lambda) + j\mu(\lambda) \tag{2.1}$$

where:

λ	=	wavelength,
η	=	real index of refraction as a function of λ,
μ	=	extinction coefficient as a function of λ,
j	=	imaginary unit ($j = \sqrt{-1}$).

Semi-conductors are conductors with a small extinction coefficient. Dielectrics are essentially non-conductors whose extinction coefficient is by definition zero [198]. For notational simplicity, we will remove the explicit dependency on λ in the remaining equations presented in this section.

The reflection direction, represented by the vector $\vec{r}$ (Figure 2.1), for light incident at an interface is obtained using the law of reflection [113]. This states that the angle of the reflection direction, θ_r, is equal to the angle of incidence, θ_i, and will be in the same plane as the incident direction, represented by the vector $\vec{i}$, and the surface normal, represented by the vector $\vec{n}$:

$$\theta_i = \theta_r \tag{2.2}$$

Considering the geometry described in Figure 2.1 and applying the law of reflection stated above, the reflection direction, $\vec{r}$, is given by:

$$\vec{r} = \vec{i} + 2\vec{n}\cos\theta_i = \vec{i} - 2\vec{n}(\vec{i}\cdot\vec{n}) \tag{2.3}$$

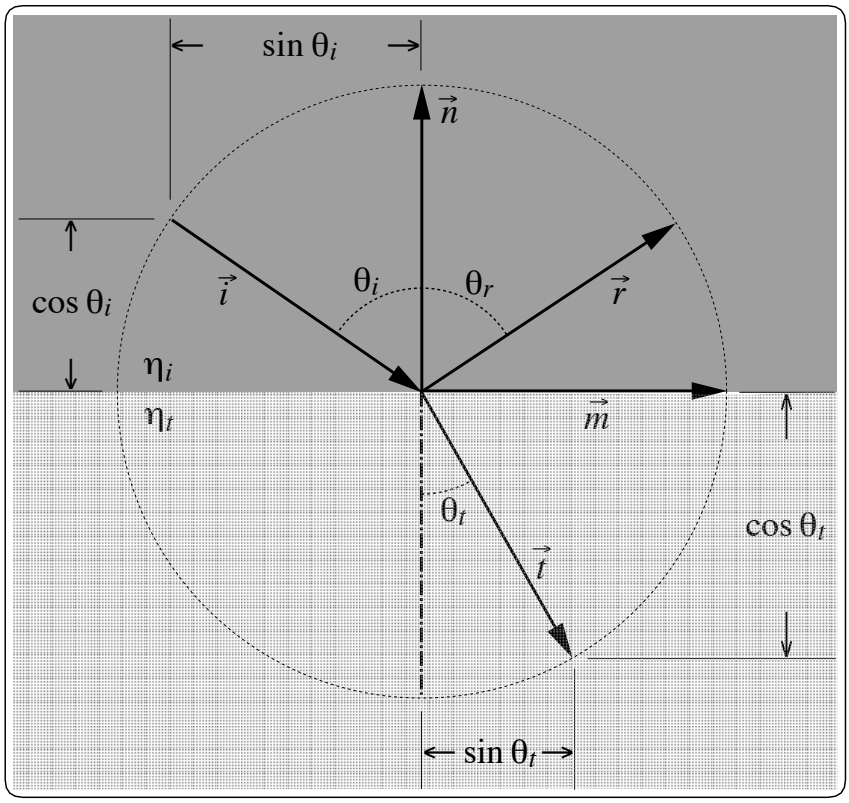

Figure 2.1: Geometry for light incident at an interface between different materials.

The transmission (refraction) direction, represented by the vector $\vec{t}$ (Figure 2.1), is obtained using the law of refraction, also known as Snell's law [113]:

$$\eta_i \sin\theta_i = \eta_t \sin\theta_t \tag{2.4}$$

where:

η_i = refractive index of the incidence material (medium),

η_t = refractive index of the transmission material (medium).

More specifically, the refraction direction $\vec{t}$ is given by:

$$\vec{t} = -\vec{n}\cos\theta_t + \vec{m}\sin\theta_t \tag{2.5}$$

where:

$\vec{m}$ = vector perpendicular to $\vec{n}$ and in the same plane as $\vec{i}$ and $\vec{n}$.

Equation 2.5 can be expanded to yield the expression presented by Heckbert [114]:

$$\vec{t} = \left[-\frac{\eta_i}{\eta_t}(\vec{i} \cdot \vec{n}) - \sqrt{1 - \left(\frac{\eta_i}{\eta_t}\right)^2 \left(1 - (\vec{i} \cdot \vec{n})^2\right)} \right] \vec{n} + \frac{\eta_i}{\eta_t}\vec{i} \tag{2.6}$$

The incident rays are not only reflected and/or transmitted (refracted) at an interface between dielectrics, but also attenuated. This attenuation is given by the Fresnel coefficients for reflection and transmission (refraction), which are computed through the Fresnel equations. For a complete derivation of these equations, the reader is referred to the texts by Hecht and Zajac [113] and Glassner [86]. The Fresnel equations for a smooth interface between two dielectrics ($\mu = 0$) can be simplified to the following expressions collected by Kajiya [130]:

$$R_\perp = \frac{\eta_i cos\theta_i - \eta_t cos\theta_t}{\eta_i cos\theta_i + \eta_t cos\theta_t} \tag{2.7}$$

$$R_{||} = \frac{\eta_t cos\theta_i - \eta_i cos\theta_t}{\eta_i cos\theta_t + \eta_t cos\theta_i} \tag{2.8}$$

$$T_\perp = \frac{2\eta_i cos\theta_i}{\eta_i cos\theta_i + \eta_t cos\theta_t} \tag{2.9}$$

$$T_{||} = \frac{2\eta_i cos\theta_i}{\eta_t cos\theta_i + \eta_i cos\theta_t} \tag{2.10}$$

In the previous equations $R_\perp$ and $R_{||}$ are the Fresnel coefficients for reflection of light polarized in directions perpendicular to ($\perp$) and parallel ($||$) to the interface. Similarly, $T_\perp$ and $T_{||}$ are the Fresnel coefficients for transmission (refraction) of light polarized in directions perpendicular to ($\perp$) and parallel ($||$) to the interface.

The Fresnel coefficient for reflection, or reflectivity [161], R, for polarized light is the weighted sum of the polarized components, in which the weights must to sum to unity [86]. In this dissertation we are interested in the Fresnel coefficients for unpolarized light. In this case, the Fresnel coefficient for reflection is simply the average of the two coefficients $R_\perp$ and $R_{||}$. Then, the equation used to compute this

coefficient reduces to following expression used by Shirley *et al.* [202]:

$$R = \frac{(\eta_i^2 - \eta_t^2)^2 c_{it}^2 + (\cos\theta_i^2 - \cos\theta_t^2)^2 n_{it}^2}{(c_{it}(\eta_i^2 + \eta_t^2) + n_{it}(\cos\theta_i^2 + \cos\theta_t^2))^2} \tag{2.11}$$

where:

$$\begin{aligned} c_{it} &= \cos\theta_i \cos\theta_t, \\ n_{it} &= \eta_i \eta_t. \end{aligned}$$

An important property of these equations is that they can be applied without regard to the direction of propagation [198], which becomes relevant when one applies ray tracing methods. To find the Fresnel coefficient for transmission (refraction), or transmissivity [161], T, we observe that there is no absorption at an interface between dielectrics. Thus, T can be easily obtained from R through a simpler relation: $T = 1 - R$, rather than resorting to an expression similar to Equation 2.11. Incidentally, absorption may occur once light is transmitted into a medium. Absorption of the transmitted light is discussed in Section 2.3.

2.2 Spectral Radiometric Terms and Properties

Radiometric terms describe measures of light integrated over all wavelengths. These measures may be also evaluated at a specific wavelength λ. When a radiometric term is written making this dependency on wavelength explicit, it is called a spectral radiometric term [86].

Radiant energy, denoted by Q (measured in joules, J), represents the energy of a packet of rays. In computer graphics we are interested in the amount of light hitting a surface or film plane during a set period of time. Radiant power or flux, denoted by Φ (measured in Watts, W, or $\frac{J}{s}$), is, therefore, often used as pointed out by Shirley [198]. Moreover, Shirley notes that radiant power is convenient to work with because it allows energy balance constraints to be applied, assuming that either the solution is steady state or the speed of light is infinite. The later assumption is usually appropriate, since the time it takes light to travel across a typical scene is very small compared to a camera shutter speed or the human temporal visual threshold [198]. The amount of radiant power traveling from a source in a certain direction,

per unit of solid angle[2], is called the radiant intensity and denoted by I (measured in $\frac{W}{sr}$).

The underlying purpose of the rendering process is to determine the colors of the surfaces within an environment. The color of a given surface will depend on how much light is emitted, reflected, absorbed and transmitted by the surface. Since radiant intensity depends on the area of the light source, it is not convenient to approximate color, which is independent of surface area. As pointed out by Shirley [198], the radiometric quantity that more closely approximates the color of a surface, through an indication of its brightness[3], is the radiance, denoted by L (measured in $\frac{W}{sr.m^2}$), which is not dependent on the size of the object being viewed, or the distance to the viewer. The spectral radiance at a point x of a surface and in a direction ψ (usually represented by a pair of spherical coordinates) is denoted by $L(x, \psi, \lambda)$ and can be expressed as:

$$L(x, \psi, \lambda) = \frac{dI(x, \psi, \lambda)}{dA\, cos\theta} = \frac{d^2\Phi(x, \psi, \lambda)}{d\vec{\omega}\, dA\, cos\theta} = \frac{dM(x, \psi, \lambda)}{d\vec{\omega}\, cos\theta} \tag{2.12}$$

where:

$dI(x, \psi, \lambda)$	=	spectral radiant intensity at x and in a direction ψ,
$d\Phi(x, \psi, \lambda)$	=	spectral radiant power at x and in a direction ψ,
$dM(x, \psi, \lambda)$	=	spectral radiant exitance at x and in a direction ψ,
θ	=	angle between the surface normal and the direction ψ,
dA	=	differential area surrounding x,
$d\vec{\omega}$	=	differential solid angle at which $d\Phi$ arrives at or leaves from x.

The fraction of light at wavelength λ incident from a direction ψ_i at a point x that is neither absorbed into nor transmitted through a given surface is called the reflectance, $\rho(x, \psi_i, \lambda)$, of the surface. Similarly, the fraction of light transmitted through the surface is called the transmittance, $\tau(x, \psi_i, \lambda)$. The light that is neither

[2]A solid angle is the three-dimensional analog to the two-dimensional concept of angle [86]. For example, the solid angle subtended by an area A on a sphere with radius $\mathcal{R}$ is equal to $\frac{A}{\mathcal{R}^2}$. This quantity is the measure of the angle in steradians (radians squared), denoted by sr [57].

[3]Brightness can be defined as the attribute by which an area of color of finite size is perceived to emit, transmit, or reflect a greater or lesser amount of light [5].

reflected nor transmitted by the surface is absorbed. The parameter that describes the amount of absorbed light is absorptance [5]. The sum of the reflectance, transmittance and absorptance is one.

The reflectance and the transmittance do not describe the distribution of the reflected and transmitted light. The *bidirectional reflectance distribution function* (BRDF), f_r, and the *bidirectional transmittance function* (BTDF), f_t, are used to overcome this limitation. As suggested by Glassner [86], these functions can be combined into the *bidirectional surface-scattering distribution function* (BSSDF, or simply BDF [86]). The BDF, f, can be expressed in terms of the ratio between the spectral radiance propagated at a point x of a surface in the direction ψ and the spectral radiant energy (per unit of area and per unit of time) incident from a direction ψ_i at the point x of the surface:

$$f(x, \psi_i, \psi, \lambda) = \frac{dL(x, \psi, \lambda)}{L_i(x, \psi_i, \lambda)\, d\vec{\omega_i}\, cos\theta_i} \tag{2.13}$$

where:

$f(x, \psi_i, \psi, \lambda)$	=	BDF of the surface at x,
$dL(x, \psi, \lambda)$	=	spectral radiance propagated at x and in a direction ψ,
$L_i(x, \psi_i, \lambda)$	=	spectral incident radiance at x and in a direction ψ_i,
θ_i	=	angle between the surface normal at x_i and the direction ψ_i,
$d\vec{\omega_i}$	=	differential solid angle at which L_i arrives at x.

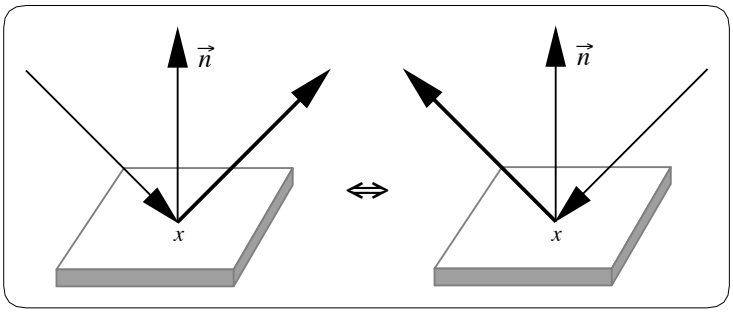

Figure 2.2: Reciprocity of the BDF.

An important property of the BDF is its symmetry or reciprocity condition, which is based on *Helmholtz Reciprocity Rule*[4] [52]. This condition states that the BDF for a particular point and incoming and outgoing directions remains the same if these directions are exchanged (Figure 2.2). It allows, for instance, the "forward" simulation of light rays traveling from a viewer to a light source, which is used by global illumination methods such as path tracing [131, 198]. Quantitatively, this condition can be expressed as:

$$f(x, \psi_i, \psi, \lambda) = f(x, \psi, \psi_i, \lambda) \tag{2.14}$$

Another important property of the BDFs is that they must be normalized, *i.e.* conserve energy. This means that the total energy propagated in response to some irradiation must be no more than the energy received [86]. In other words, for any incoming direction the radiant power propagated over the hemisphere can never be more than the incident radiant power [136]. Any radiant power that is not propagated is absorbed. Formally, in the case of reflection of light, the so-called directional-hemispherical reflectance [5] should therefore be less than or at most equal to 1:

$$\rho_(x, \psi_i, 2\pi, \lambda) = \int_{\text{outgoing } \psi} f_r(x, \psi_i, \psi, \lambda) \cos\theta d\vec{\omega} \leq 1, \quad \forall \psi_i \in \text{incoming directions} \tag{2.15}$$

where:

$f_r(x, \psi_i, \psi, \lambda)$	=	BRDF of the surface at x,
θ	=	angle between the surface normal and the outgoing direction ψ,
$d\vec{\omega}$	=	differential solid angle at which the radiance is reflected.

A similar relation given in terms of the directional-hemispherical transmittance [5] and the BTDF is used for the transmission of light. Reflectance and transmittance models, or simply BDF models, that are energy-conserving and reciprocal are considered *physically plausible*[5]. This is a crucial requirement for physically-based rendering

[4]The original statement of *Helmholtz Reciprocity Rule* does not include non-specular reflection of any sort [52, 225]. Recently Veah [225] derived a reciprocity condition for general BDFs using Kirchhoff's laws regarding radiative transfer [205].

[5]Lewis [145, 146] uses the term "plausible" to describe BDF models whose existence does not violate the laws of physics.

frameworks aimed at global illumination applications, and it will be further addressed in Section 2.6.

Sometimes, when energy transport or energy balance is of concern as opposed to lighting at a point, it is more convenient to work with the radiant power (radiant flux) [5] than with the radiance [198]. Under these circumstances, it is more natural to describe the surface reflection and transmission properties in terms of the probability distribution of the reflected and transmitted light. This term is called the *scattering probability function* (SPF) [197, 198]. It describes the amount of energy scattered in each direction ψ, at a point x of a surface and at wavelength λ as:

$$s(x, \psi_i, \psi, \lambda) = \frac{dI(x, \psi, \lambda)}{\rho(x, \psi_i, \lambda) d\Phi(x, \psi_i, \lambda)} \tag{2.16}$$

where:

$dI(x, \psi, \lambda)$ = spectral radiant intensity reflected at x and in a direction ψ,
$\rho(x, \psi_i, \lambda)$ = reflectance of the surface at x,
$d\Phi(x, \psi_i, \lambda)$ = spectral radiant power incident at x and in a direction ψ_i.

The term $\rho(x, \psi_i, \lambda)$ appears in the numerator when we are dealing with reflection of light. It scales the function to a valid *probability density function* (PDF) (see Appendix A) over the solid angle through which the reflected light leaves the surface [197, 198]. In the case of transmission of light, a similar expression is used, in which $\rho(x, \psi_i, \lambda)$ is replaced by $\tau(x, \psi_i, \lambda)$.

2.3 Absorption in a Homogeneous Medium

In this section we will focus on the losses affecting the transmittance in a homogeneous medium, *i.e.* a material in which the physical properties that affect light propagation are assumed to be identical everywhere. The losses affecting the transmittance in a inhomogeneous medium can be simulated through successive application of the laws for homogeneous medium [4]. Another alternative is to think of an inhomogeneous material as a structure composed of two or more homogeneous layers [188]. The reader interested in the spectrophotometry regarding the transmittance in inhomogeneous materials is referred to the text by MacAdam [151].

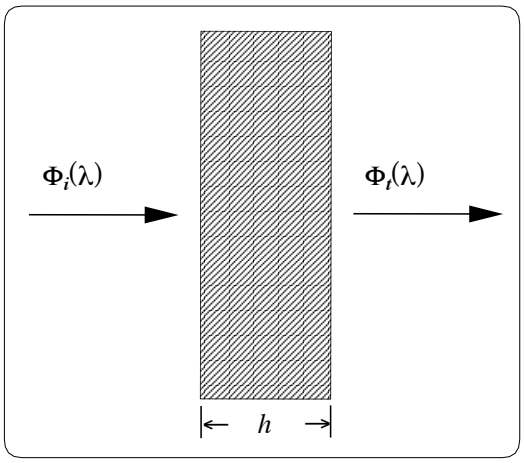

Figure 2.3: Loss of light at wavelength λ in a medium of thickness h.

The transmittance of a homogeneous material, after correction for surface losses, varies in accordance with Bouguer's law (Figure 2.3), also called Lambert's law of absorption [151]. This law states that the loss due to the process of absorption is proportional to the power of the light incident on the medium, to the thickness of the medium (or the distance traveled by the light in the medium) and to a constant of proportionality called absorptivity [161]. This constant, also known as absorption coefficient, is a characteristic of the medium and a function of wavelength. A complete derivation of this law is presented by Meyer-Arendt [161]. It is usually written as follows:

$$\tau(\lambda) = \frac{\Phi_t(\lambda)}{\Phi_i(\lambda)} = e^{-a(\lambda)\,h} \tag{2.17}$$

where:

$\Phi_t(\lambda)$	=	spectral power after being transmitted through the medium,
$\Phi_i(\lambda)$	=	spectral power incident on the medium,
$a(\lambda)$	=	absorption coefficient of the medium at wavelength λ,
h	=	thickness of the medium.
e	=	Euler's number ($2.718281\cdots$).

Beer's law [149] states that for a dye solution, the absorption coefficient of the solution is directly proportional to its concentration. Combining Beer's law with

Bouguer's law [151] for samples of thickness h and concentration c results in the following expression for the transmittance of a homogeneous material:

$$\tau(\lambda) = e^{-a(\lambda)\, c\, h} \tag{2.18}$$

where:

$a(\lambda)$	=	absorption coefficient of the medium at wavelength λ,
c	=	concentration of the solution,
h	=	thickness of the medium,
e	=	Euler's number ($2.718281\cdots$).

Sometimes it is more convenient to specify the absorption of a medium by means of the extinction coefficient [149], μ, which is given by:

$$\mu = \frac{a(\lambda)\lambda_m}{4\pi} \tag{2.19}$$

where:

$a(\lambda)$	=	absorption coefficient of the medium at wavelength λ,
λ_m	=	wavelength of light in the medium.

2.4 Rendering Equation

Three major global illumination approaches have been used in rendering to simulate the light transfer mechanisms: ray tracing, radiosity and hybrid methods. Kajiya [131] unified the discussion of global illumination methods with the rendering equation. This equation, also known as transport equation, can be expressed in terms of radiances (Equation 2.20) on the basis of the ray law (the radiance is constant along a line of sight between objects [198]), and the definition of the BDF. In a simplified form it is given by:

$$\underbrace{L(x,\psi,\lambda)}_{total} = \underbrace{L_e(x,\psi,\lambda)}_{emitted} + \underbrace{L_p(x,\psi,\lambda)}_{propagated} \tag{2.20}$$

Equation 2.20 states that the radiance of a point x on a surface, in a direction ψ and at wavelength λ is given by the sum of the emitted radiance component, L_e, and

the propagated radiance component, L_p. Usually L_e is known from the input data, and the computation of L_p constitutes the major computational problem.

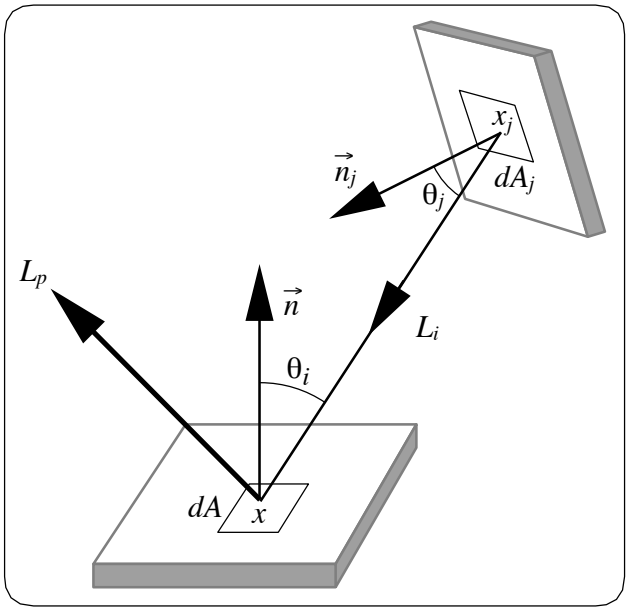

Figure 2.4: Geometry for computing L_p as an integral over all the surfaces within the environment.

The term L_p can be written as an integral over all the surfaces within the environment (Figure 2.4), resulting in the formulation presented in the following equation:

$$L_p(x,\psi,\lambda) = \int_{\text{all xj}} f(x,\psi,\psi_i,\lambda) L_i(x,\psi_i,\lambda)\, \cos\theta_i\, V(x,x_j)\, \frac{cos\theta_j dA_j}{||x_j - x||^2} \qquad (2.21)$$

where:

$f(x,\psi,\psi_i,\lambda)$	=	BDF of the surface at x,
$L_i(x,\psi_i,\lambda)$	=	spectral incident radiance at x and in a direction ψ_i,
θ_i	=	angle between the surface normal at x and the direction ψ_i,
θ_j	=	angle between the surface normal at x_j and the direction ψ_i,
dA_j	=	differential area surrounding x_j,
$V(x,x_j)$	=	visibility term.

The visibility term $V(x, x_j)$ used in the Equation 2.21 is one if a point x_j of a certain surface can "see" a point x of another surface, and zero otherwise. This equation is commonly used by deterministic rendering methods based on standard numerical techniques [136]. In the context of global illumination these techniques are used to solve the high dimensional integrals and large systems of equations usually associated with the application of the radiosity method (Chapters 7 and 8).

Alternatively, L_p can also be expressed in terms of all directions visible to x (Figure 2.5) [198]. This expression for L_p is more suitable to nondeterministic rendering methods based on Monte Carlo techniques [107], and it is given by:

$$L_p(x, \psi, \lambda) = \int_{\text{incoming } \psi_i} f(x, \psi, \psi_i, \lambda) L_i(x, \psi_i, \lambda) \cos\theta_i d\vec{\omega}_i \tag{2.22}$$

where:

$f(x, \psi, \psi_i, \lambda)$	=	BDF of the surface at x,
$L_i(x, \psi_i, \lambda)$	=	spectral incident radiance at x and in a direction ψ_i,
θ_i	=	angle between the surface normal at x and the direction ψ_i,
$d\vec{\omega}_i$	=	differential solid angle where L_i arrives.

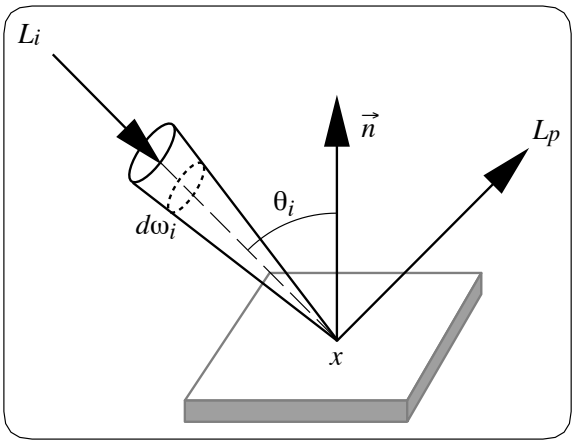

Figure 2.5: Geometry for computing L_p in terms of all directions visible to a point x.

2.5 Light Transfer Mechanisms

The reflection and transmission of light from several materials used in computer graphics are usually approximated by the behavior of two simple types of materials namely, specular and diffuse (Figure 2.6a), and their combination. The resulting basic mechanisms of light transfer identified by Wallace [231] are the following:

- diffuse to diffuse,
- diffuse to specular,
- specular to diffuse and
- specular to specular.

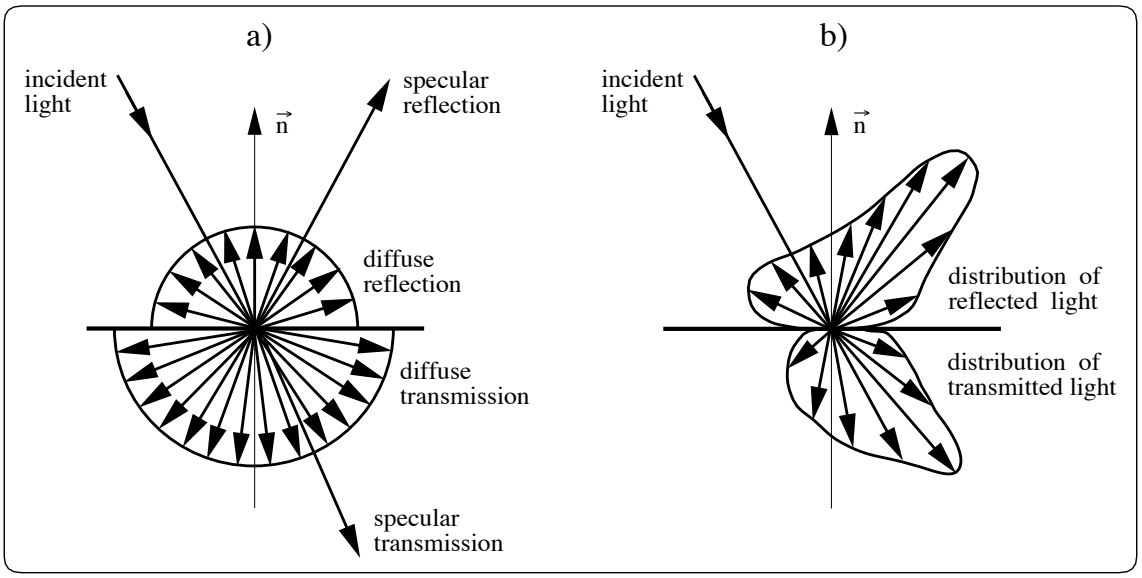

Figure 2.6: a) Ideal diffuse and ideal specular distribution of reflected and transmitted light. b) General distributions of reflected and transmitted light.

Ideal diffuse or Lambertian surfaces (see Section 2.6), such as chalk, appear equally bright from all viewing angles because they have a constant spectral radiance at all viewing angles under steady lighting conditions [198]. For these surfaces, according

to Lambert's cosine law of reflection [75], the amount of radiant intensity seen by the viewer is directly proportional only to the cosine of the angle of incidence of the light. Specular surfaces, however, reflect light unequally in different directions. A perfect specular surface, such as a mirror, reflects light only in the direction of reflection, which corresponds to the viewer's direction mirrored about the surface normal.

Actual surfaces, however, present reflectance and transmittance characteristics which are quite varied and have a complex directional pattern (Figure 2.6b). Usually these surfaces are not adequately approximated by diffuse, specular, or combined models [86]. Surfaces that approach closely the conditions of coherent (specular-like) reflection are ordinarily said to be glossy, and those that approach closely the conditions of incoherent (diffuse-like) reflection are said to be matte [151]. Surfaces that allow some light to pass through are called transparent or translucent. The term transparent is commonly used for surfaces that exhibit propagation characteristics that can be better approximated by coherent transmission. The term translucent is commonly used for surfaces whose propagation characteristics are better approximated by incoherent transmission.

The spectral energy distribution of the propagated light, measured in terms of reflectance and transmittance, and the spatial distribution of that light, measured in terms of BRDF and BTDF, work together to establish the *measurement of appearance* of an object [93]. This term is used to identify the family of measurements that are necessary to characterize both the color and the surface finish of an object [122]. The variations in the spectral energy distribution affect appearance characteristics such as hue, lightness and saturation, which are described by Hunter and Harold [122] as follows:

- hue - attribute of color perception by means which an object is judged to be red, yellow, green, blue, purple and so forth,
- lightness - perception by which white objects are distinguished from gray objects and light from dark colored objects,
- saturation - attribute of color perception that expresses the degree of departure from the gray of the same lightness.

The changes in the spatial distribution affect appearance characteristics such as gloss, reflection haze, transmission haze, luster and translucency. The reflection haze corresponds to the scattering of reflected light in directions near that of specular reflection by a specimen having a glossy surface [122]. The transmission haze corresponds to the scattering of light within or at the surface of a nearly clear specimen, responsible for cloudy appearance seen by transmission [122]. Finally, the luster, or contrast gloss, as described by Hunter and Harold [122], corresponds to the gloss associated with contrasts of bright and less bright adjacent areas of the surface of an object. Luster increases with increased ratio between light reflected in the specular direction and that reflected in diffuse direction which are adjacent to the specular direction.

Besides the interactions of direct light with individual objects, in order to perform a general examination of light transfer mechanisms it is also necessary to account for the indirect lighting. In this context, a notation suggested by Heckbert [116] can be used to describe the direct and indirect light interactions with different objects in an environment. He uses symbols to represent the various events experienced by a photon. To avoid confusion with the radiometric terms used in this dissertation, we will use boldface fonts to represent these symbols:

- **L** - emission of a photon from a light source,
- **E** - photon arriving at the eye (intersection with the image plane),
- **V** - photon interacting with a volume,
- **S** - photon specularly reflected and/or transmitted,
- **D** - photon diffusively reflected and/or transmitted.

The sequence is written left to right over time, in such a way that **L** is the first character and **E** is the last. Glassner [86] suggested the use of elements of the standard computer-science regular expression symbology [119] to represent combined expressions. As pointed out by Sillion and Puech [207], this notation approach can also be used to characterize the shortcomings of the rendering methods [211]. For example,

distributed ray tracing [60] can in theory capture all possible paths followed by a photon, which are represented by $\mathbf{L}(\mathbf{S}|\mathbf{D})^*\mathbf{E}$ (where the verical bar | denotes a selection among members, and the asterisk superscript $*$ represents 0 or more repetitions). However, the diffuse information is rarely carried out explicitly by this method, due to its high computational cost. Moreover, since this method is particularly suitable to handle near-specular reflection and transmission, the term $\mathbf{S}$ can be replaced by the glossy term $\mathbf{G}$. Thus, its characteristic expression is more appropriately represented by $\mathbf{L}[\mathbf{D}]\mathbf{G}^*\mathbf{E}$ (where the square brackets indicate that the term $\mathbf{D}$ may be included or not) [86].

In order to simulate various physical effects in the same scene, such as *highlights* (view-dependent concentrations of light on a surface caused by non-diffuse reflectors), *color-bleeding* (diffuse reflection of light of one colored surface affecting the appearance of another surface) and *caustics* (view-independent concentrations of light on a surface caused by non-diffuse reflectors), one has to resort to multipass or hybrid algorithms [50, 115, 199, 210].

The main idea behind these algorithms is that all the different paths of light transfer that will be handled are determined beforehand, and each type of path is handled only once [86]. For example, most of these algorithms start with a pass performed by a radiosity method (Chapters 7 and 8) to determine the multiple diffuse interreflections in the environment. Additional passes are performed by other rendering methods to handle the other mechanisms of light transfer. Usually these passes are performed by ray tracing methods such as visibility tracing [236], path tracing [131] and photon tracing [174, 175]. The radiosity method takes into account the direct light transfer paths from the light source, which are also accounted for by other rendering methods. Then, it is necessary to make sure that these direct paths are only computed once. Usually their contribution is removed from the radiosity solution, and computed again in one of the other passes, *e.g.* using photon tracing [198].

When applying the notation described earlier to these algorithms, usually each type of light transfer path or mechanism is represented by a specific characteristic expression [86]. For example, Chen *et al.* [50] developed a hybrid algorithm in which they considered four types of light transfer paths. The first type contains no diffuse

elements, and it is represented by $\mathbf{LS^*E}$. The second corresponds to a single diffuse element with no specular surfaces between itself and the light, and it is represented by $\mathbf{LDS^*E}$. The third is similar to the previous, but with one or more specular elements between the light and the diffuse element, and it is represented by $\mathbf{LS^+DS^*E}$ (where the plus-sign superscript + represents one or more repetitions). The fourth type corresponds to a general case, with two or more diffuse elements, and it is represented by $\mathbf{L(S|D)^*DS^*DS^*E}$.

2.6 Overview of Reflectance Models

The expression reflectance model is restrictive, since there are reflectance models that handle transmittance as well. We will retain this term in this section for consistency with graphics literature, explicitly mentioning the transmittance components when they become relevant. Moreover, since the reflectance models are in fact used to determine the BRDF of the surfaces, we can often find the term BRDF model in the literature. We suggest the term BDF model, which encapsulates the BTDF of the surfaces as well.

Usually reflectance models are divided into two families: empirical and theoretical models [102, 189, 233]. The term empirical is somewhat misleading, since usually these models are not based on experimental data. In addition, they usually include *ad hoc* terms in their formulation. Because of these aspects, we will henceforth refer to these models as *ad hoc* models. These models, although computationally simple, are not physically plausible (Section 2.2) and lack the requirements needed for global illumination calculations, namely energy conservation and reciprocity.

Theoretical models, on the other hand, are designed to attempt the fulfillment of these requirements. These models may, in turn, be divided into two groups: analytical and non-analytical models [189]. Analytical models provide an analytical formulation to compute the BDF. Non-analytical models, on the other hand, use an algorithmic process to compute the BDF for a selected number of incoming and outgoing directions.

The simplest and oldest reflectance model is the Lambert model [86]. It is named after Lambert, who in 1760 introduced the cosine law of reflection (Section 2.5).

In computer graphics and computer vision Lambert's cosine law of reflection has very often been assumed for diffuse-reflecting dielectric surfaces [239]. Recently, a diffuse-reflective model for smooth dielectric surfaces was developed and used in experimental measurements by Wolff [239] to determine the range of applications of Lambert's cosine law of reflection. These measurements showed that this law is a good approximation (within approximately 5% accuracy) across illuminated smooth dielectric objects when the angle of incidence and the viewing angle are simultaneously less than 50°. These measurements also showed that outside of this range large deviations start to occur.

After the Lambert model, several reflectance models have been proposed by researchers from different fields, such as physics, engineering, remote sensing (Chapter 3) and computer graphics. In this section we will focus the discussion on reflectance models aimed at rendering applications, with a particular emphasis on works that are considered milestones in this area. More comprehensive surveys on this topic can be found in the papers by Schlick [188, 189] and in the book by Glassner [86].

2.6.1 *Ad Hoc* and Transitional Models

In 1975 Phong introduced a model [176], known as the Phong model, in which an *ad hoc* exponentiated cosine is used to simulate specular highlights. Afterwards, Blinn [29] proposed a variant of the Phong model based on similar approximations, but slightly faster. While the Phong model does not have a physical basis, it can be restricted to conserve energy. The BRDF resulting from these restrictions has been used in global illumination calculations [123]. Lewis [145, 146] has shown, however, that this "Phong-like" BRDF does not satisfy the reciprocity condition.

In order to obtain a plausible BRDF, Lewis suggested modifications to the original Phong model to make it reciprocal and constrained to conserve energy. These changes originated a theoretical reflectance model known as modified Phong model [138, 145]. The plausible BRDF described by the modified Phong model has been used to derive approximated PDFs for Monte Carlo integration through importance sampling (see Appendix A). The scattering distributions derived from these PDFs, in turn, have being used by nondeterministic physically-based rendering algorithms [136] to simulate non-Lambertian phenomena [8, 7, 198, 199, 203].

Some of the *ad hoc* models are called transitional models [102], since they began a transition towards the energy equilibrium in rendering applications. The models proposed by Kay and Greenberg [133] and Hall and Greenberg [104] are examples of such models. The first model addresses attenuation of light passing through a material while traveling between intersections. The second introduces an incoherent transmission term to provide better representation of light sources seen through objects.

2.6.2 Analytical Models

Torrance and Sparrow [222], working in the optical physics area, developed a geometric reflectance model for specular reflection based on a previous work by Beckmann and Spizzichino [23]. Their model assumes that the surface is made of highly reflective microscopic facets distributed in v-grooves. The random distribution of the facets allows the modeling of isotropic surfaces. A surface is qualified as isotropic when the reflected light intensity at a given point is independent of the surface orientation along this normal at this point. In other words, an isotropic surface has a BRDF that is independent of rotation about the normal [233].

The normal vector to one of the two flat sides of a given groove is used, together with the lighting direction and the viewing direction, to compute the fraction of incoming light which is neither shadowed on the way in (shadowing), nor obscured in the way out (masking). If the groves are randomly distributed over the surface, shadowing and masking functions can be statistically estimated. Figure 2.7 shows the effects of self-shadowing and self-masking by a rough surface. Therefore, using these functions, the BRDF can be determined for a given distribution of slopes of the facets.

The model proposed by Torrance and Sparrow was applied to computer graphics by Blinn [29] and modified by Cook and Torrance [61] to take better account of colored light. The Cook and Torrance model has been used to create very realistic pictures of metallic and plastic objects. Besides having a physical basis, these models have in common the handling of isotropy. Several natural surfaces, such as forest canopies, oceans, snow and sand, however, exhibit anisotropy. This means that if an element

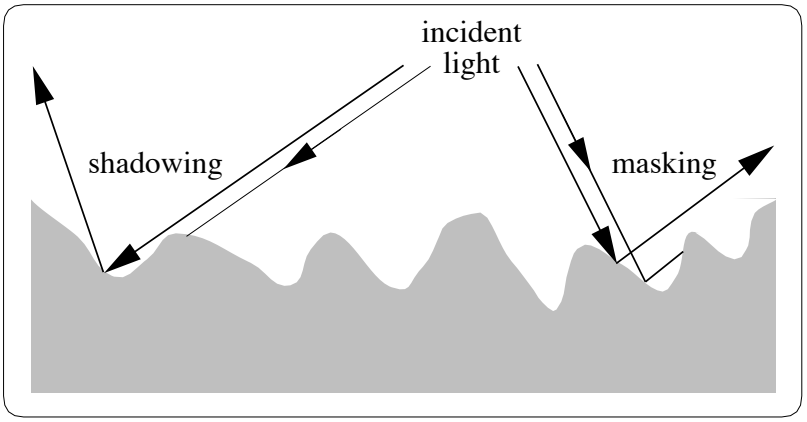

Figure 2.7: Shadowing and masking.

of such an anisotropic surface is rotated around its normal while the light and viewer directions remain unchanged, the light intensity reflected to the viewer will vary.

Poulin and Fournier [178] introduced a reflectance model for anisotropic surfaces based on covering the surface with groups of small cylinders. Different levels of anisotropy are obtained by varying the distance between each cylinder and/or elevating the cylinders more or less above the surface. Multidirectional anisotropy is modeled by orienting groups of cylinders in different directions. The intensity of the reflected light is computed by determining the visible and illuminated portion of the cylinders, taking self-blocking into account. An analytic solution to the diffuse intensity over the cylinders was developed, as well as an approximation to the specular intensity using Chebyshev polynomials (see Section 8.3). The computation of the surface BRDF is performed implicitly within the computation of the intensity of the reflected light.

He *et al.* [112] presented a comprehensive reflectance model, known as *HTSG* model (named after its developers), to improve the generality of the reflectance models. Their model is based on the wave transport theory of light, using the Kirchhoff approximation for the light reflected from a rough surface [23]. Through this theory a rough surface is approximated by its tangent plane at the point considered. The *HTSG* model takes into account the interreflection and self-shadowing/masking factors derived from statistical techniques (the surface height is assumed to follow a

Gaussian distribution). They propose a BRDF consisting of three components: specular, directional-diffuse and uniform-diffuse. The first two components result from the surface reflection process, due to specular reflection by the mean surface and diffraction scattering by the surface roughness respectively. The third component is attributed to multiple surface and/or subsurface reflections.

The *HTSG* model compares favorably with experimental measurements of reflected radiation for metals, non-metals, and plastics, with smooth and rough surfaces. The normalization was, however, not adequately treated in this model [233]. Furthermore, comparisons between the reflectance model and absolute BRDF measurements to demonstrate normalization empirically were not made either. The authors claimed that the data for the BRDF of real surfaces is hard to come by, and they are supported in their claim by Shirley [198] and Ward [233] among other computer graphics researchers.

As pointed out by Ward [233], a reflectance model without symmetry (reciprocity of the BRDF described in Section 2.2) cannot be physical because the simulated surface reflects light differently in one direction than the other. Also, without proper normalization, a reflectance model does not yield a correct energy balance and, thus, can not produce physically meaningful results. The models for anisotropic surfaces, presented above, either fail to satisfy the one or the other of two physical requirements or are somewhat expensive to evaluate.

Ward [233] presented a mathematical model of anisotropic reflectance which was extended from an isotropic Gaussian reflectance model similar to the model presented by Cook and Torrance [61]. The BRDFs are computed through empirical formulas that may incorporate Fresnel effects. According to Ward, one of the advantages of this model over more complex reflectance models, such as the *HTSG* model, is its simplicity. It adapts to stochastic sampling techniques, which allows its implementation through the Monte Carlo methods [107]. However, although this model is accurate for many surfaces, satisfying both reciprocity and normalization, there are still many materials, such as translucent biological surfaces, that do not fit in the scope of the model.

2.6.3 Non-Analytical Models

The first model using an algorithmic approach was presented by Kajiya [130]. It is also based on the wave transport theory of light, using the Kirchhoff approximation for reflection from a rough surface [23]. A numerical integration of the Kirchhoff equation is done for each pair of directions in order to describe the reflection and refraction of light from surfaces with anisotropic characteristics. This model presents some restrictions related to the self shadowing and multiple scattering, and it is not considered computationally practical. It has the merit, however, of rederiving the general formula for scattering from a rough surface in a form more suitable for computer graphics.

Cabral *et al.* [44] proposed a model to generate anisotropic reflectance functions, where the facets are created from a height field. The reflection of each facet is studied, including the block factor for incident and reflected light, in order to compute the surface BRDF. Cabral et al. claim that the anisotropy can be averaged out if desired for isotropic surfaces. Spherical harmonics series [86] are used to represent the BRDFs. As with Kajiya's, this model is not considered to be computationally practical.

Westin *et al.* [235], extending the idea originally presented by Cabral *et al.* [44], proposed a general approach for approximating BDF for a large class of geometries. While the two non-analytical models presented above can afford, at best, to approximate reciprocity, Westin *et al.* enforce the reciprocity by ensuring the symmetry of the matrix used to store the BDFs. The approach proposed by Westin *et al.* consists of using a Monte Carlo ray tracing technique to simulate optical scattering. Moreover, it provides the means of creating a milliscale BDF from microscale scattering events. BDFs for complex surfaces can be simulated hierarchically by using the results of one simulation in generating the BRDF for the next larger scale. This approach can be applied to both isotropic and anisotropic materials such as brushed metals, velvet and woven textiles. Oftentimes, however, the experimental data needed to describe this hierarchical arrangement is not available, and the results may be difficult to validate [109]. Although only three local scattering modes are described, namely specular reflection, specular transmission and diffuse reflection, Westin *et al.* assert that their approach allows the incorporation of other scattering modes as well.

Hanrahan and Krueger [109] proposed a similar[6] model to simulate subsurface scattering and transmission from layered surfaces, known as the *H-K* multiple-layer model [86]. This intuitive idea of a layered surface model has appeared several times in physics [188], remote sensing (see Chapter 3) and computer graphics [188, 198]. The algorithm to compute the BDF is based on the linear transport theory and uses a Monte Carlo sampling scheme. The model explicitly evaluates the reflection and transmission of light at media boundaries, like in the ocean model proposed by Nishita *et al.* [171]. It assumes that the reflected radiance from a surface has two components. One arises due to surface reflectance and the other due to subsurface volume scattering. It also assumes that the transmitted radiance has two components. One represents the amount of light transmitted through the layer without scattering inside the layers, but accounting for absorption, and the other is due to scattering in the volume. Similarly, the BRDF and BTDF have also two components, and the relative contributions of the surface and subsurface terms are modulated by Fresnel coefficients. The physical plausibility of this approach was not demonstrated either analytically or experimentally in spite of its physical basis.

The *H-K* multiple-layer model has the merit of addressing issues related to both inorganic and organic materials. However, its generality causes the overlook of important specific characteristics and properties of organic materials, such as the absorption of light by pigments. Moreover, the reflectance and transmittance are not computed directly, but implicitly introduced into the model as the albedo. In the context of this model the albebo represents the ratio between the scattering cross section[7] and the sum of the scattering cross section and the absorption cross section. If the albedo is close to zero, absorption is more likely to occur than scattering. In the testing of the model on a plant leaf, Hanrahan and Krueger obtained the albedo values of a leaf from an image acquired from a digital scanner. Although spectral curves of reflectance and transmittance available in the literature could be used instead, for a

[6]We use the term similar in this context to refer to the fact that both algorithmic models, the model proposed by Westin *et al.* and the *H-K* model, address subsurface phenomena and use Monte Carlo techniques.

[7]In this context cross section may be interpreted as the probability per unit length of an interaction of a particular type [109].

large group of natural materials, specially plant leaves, these spectral curves are only available for a few illuminating and viewing angles.

Gondek *et al.* [93] have also proposed a model which attempts to characterize the reflectance functions of a material by modeling its subsurface microstructure. Their model is mainly aimed at paints. Gondek *et al.* define a substrate filled with small pigment-particle microspheres. These particles are modeled as dielectric spectral filters, and the attenuation of the transmitted light is governed by Bouguer's law (Section 2.3). This model serves to demonstrate some of the reflective properties that real paints exhibit and simulate interference effects. More specifically, this model is used to predict thin film amplitude splitting interference. This phenomenon occurs when a wave is divided through reflection and transmission, and later recombines [93]. Gondek *et al.* state that, in order to turn their model into a true design tool, comparisons must be performed against experimental data taken from real surfaces for which the surface and the subsurface geometries are known.

Schramm *et al.* [190] extended the model proposed by Gondek *et al.* to include subsurface refraction and multilayer interference effects. They employed their model to explore surface reflectance properties of an iridescent hummingbird feather and paper. In the last case, measurements were taken to attempt to match actual reflectance properties of the material. The results of these measurements showed that the model captured both the anisotropic reflectance functions caused by fiber orientation, and the reflectance in the incident direction (backscatter or retroreflectance[8]).

2.7 Summary

The rendering equation represents the kernel of global illumination methods. In order to solve this equation it is necessary to determine the reflection/transmission behavior of the surfaces. This behavior is described by the BDF of the surfaces, which is a combination of the BRDF and BTDF, or alternatively by the SPF (Section 2.2). Several models have been proposed to represent these macroscopic reflectance and

[8]Surfaces that present a high degree of retroreflection are sometimes said to have "negative gloss" [128]. The reflectance of highway signs is an extreme example of retroreflection or "negative gloss". Commercial surfaces such as paint, plastics, papers and textiles, no matter matte or glossy, show some degree of retroreflection.

transmittance properties of surfaces. Theoretical models, in particular, are designed to be physically plausible to enable physically-based rendering.

Despite the efforts in this area, the current models still present limitations and, in several cases, a low accuracy/cost ratio. Not all theoretical models fulfill the physical requirements, namely normalization and reciprocity. Very few are controlled by a small number of simple and meaningful parameters, and present a formulation with a low computational cost. Another unsatisfactory point of the current models, as mentioned by Schlick [189], is the formulation of the BDF as a linear combination with constant weights between the diffuse and specular components. For example, to make the modified Phong Model conserve energy and be reciprocal at same time, the BDF is formulated using this approach [145, 146].

The justification for using this linear combination of weights is that, for several materials, the diffuse and the specular components come from different physical phenomena [189]. For example, a polished surface, such as varnished wood, can be thought of as a diffuse reflector covered with a thin dielectric coating. The light is reflected from the coating in a coherent way (specular reflection) and reflected from the substrate in an incoherent way (diffuse reflection). As noticed by Shirley [198], such a linear combination with constant weights is incorrect because proportions of diffuse and specular components are usually not constant, but rather a function of the incident angle.

Finally, very few reflectance models used in rendering address issues specifically associated with organic materials. It is important to note, however, that researchers from other fields have advanced further in this area. In the next chapter we present reflectance models for plants, the organic materials within the scope of this dissertation, proposed by biologists and remote sensing researchers. Some of the concepts and the biological data used in these models are relevant for this dissertation. We believe that the incorporation of this information is essential to develop and validate biologically and physically-based BDF models for rendering applications.

Chapter 3
Interaction of Light with Plant Tissue

One of the major difficulties in biologically and physically-based rendering is the lack of available experimental data to be used as input or for testing parameters. Appropriately Ward [233] pointed out that good science requires both theory and data, and one is of little use without the other. Fortunately, in the case of plants, there is a reasonable amount of information that can be incorporated into rendering applications.

Most of this information consists of experimental data for plant leaves, which are the most important plant surface interacting with light [126]. Although other plant surfaces, such as stems and petals, present similar optical and structural characteristics [80, 134], the ways in which they absorb and propagate light and thus acquire their color have not been fully investigated. Recently Biolley and Jay [26] presented a work on the colorimetry of roses. However, there is still a noticeable lack of understanding of the relationships between petal structure and pigment distribution [134]. For these reasons, we are going to focus on light interaction with foliar tissues in this chapter.

A large amount of research has been performed in botany and remote sensing regarding the reflectance and transmittance characteristics of leaves. Most of this research considers the entire light spectrum. However, since, for rendering purposes, we are mainly interested in the visible region of the light spectrum, we are going to stress the phenomena associated with this region in this chapter.

The remainder of this chapter is organized as follows:

- Section 3.1 presents a structural description of foliar tissues.

- Section 3.2 describes the main factors affecting the propagation and absorption of light by these tissues.
- Section 3.3 introduces a concise survey of reflectance and transmittance models for plant leaves aiming at applications in botany and remote sensing.
- Section 3.4 presents a brief overview of global illumination applications in the simulation of radiative transfer processes in vegetation.
- Finally, Section 3.5 summarizes the contents of this chapter.

3.1 Structural Description of a Plant Leaf

A leaf can be described as a diffusing and pigmented structure (mesophyll) having external plates of epidermal cells with a protective skin (cuticle) [240]. Figure 3.1 shows an idealized leaf cross-section.

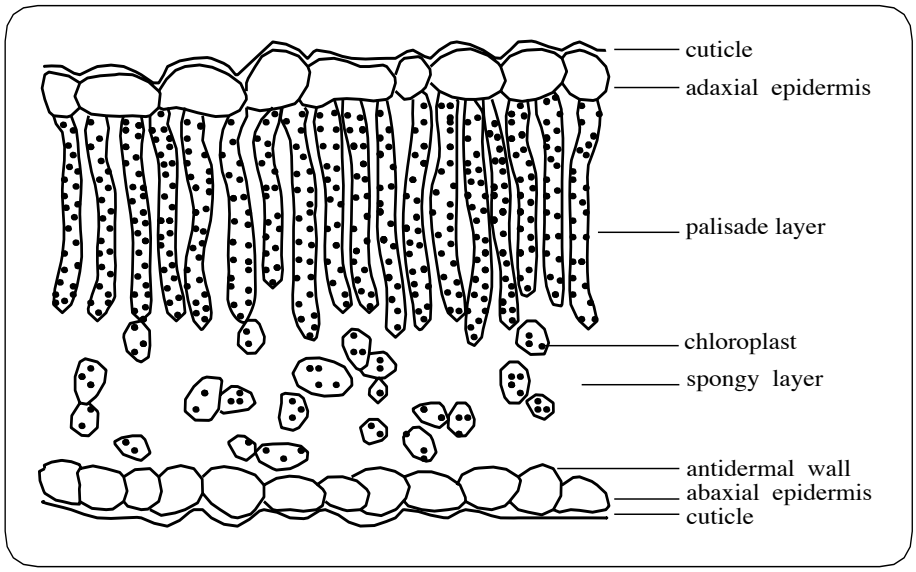

Figure 3.1: Typical cross-section of foliar tissues.

The cuticle is an extracellular, multilayered membrane of pectin, cellulose, cutin, and wax. The outermost portion of the cuticles consists of epiticular wax, which may be extremely thin, or so thick as to be visible to the naked eye [97, 150]. The epiticular

waxes may be amorphous, semicrystalline, or crystalline in form and exhibit a wide range of geometric configurations [98]. They may be orderly oriented vertical to the leaf surface or oriented at varying angles from the cuticle. The surface roughness characteristics and the refraction index of the epiticular wax control the specularly reflected light from the adaxial (front) and abaxial (back) epidermis. No intercellular spaces are normally present in the epidermal tissues, and the cells usually fit one another like the pieces of a jigsaw puzzle (Figure 3.2).

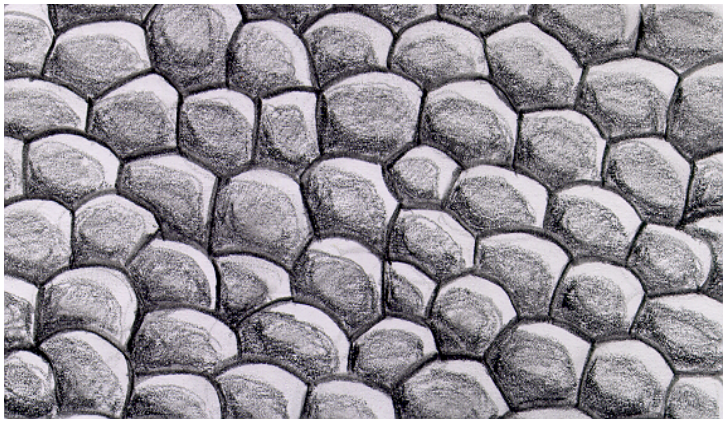

Figure 3.2: Artist's conception of a typical epidermis tissue (Courtesy of Daniel Dimian, The University of Calgary).

The mesophyll of bifacial leaves is usually composed of a densely packed layer of palisade cells, and a loosely packed layer of spongy cells. There are species of plants, however, that present unifacial leaves which either do not have palisade mesophyll [230] or have it attached to both the front and the back surfaces of a leaf [36]. The palisade cells have a cylindrical shape, and 5 to 20% of their volume is air space. These cells present a high concentration of chloroplast which contain pigments, namely chlorophylls (mainly a and b forms) and carotenoids. The concentration and distribution of these pigments control the absorption of the light in the visible region of the light spectrum within the leaf (see Section 3.2). The spongy tissue consists of smaller cells, roughly ovoid to round in shape, that present less densely packed

chloroplasts, with 50 to 80% of their volume occupied by air space [240]. The spaces between the mesophyll cells are also filled with air.

3.2 Factors Affecting the Propagation of Light

The reflection, transmission and absorption of light by leaves is affected by some important leaf characteristics:

- **Internal Structure and Thickness**
  The intra-leaf scattering caused by refractive index differences between cellular organelles, walls, hydrated cells, and adjacent intercellular air spaces is compounded or amplified by the irregular shapes and organization of the cells within the leaf [223]. As pointed out by Grant [97], differences in the reflectance curves among species, as well as changes with maturation and senescence, have also been attributed to differences in the leaf internal structure. These differences directly affect the thickness of the leaves. Experiments by Wooley [240] showed that, although the reflectance of a leaf is not strongly dependent on the leaf thickness, the transmittance is strongly affected by this foliar characteristic. Allen *et al.* [3] assert that the differences in internal structure are substantial between sun and shade leaves of the same plant species. Clearly, even greater differences can be expected between leaves from different species.

- **Pigments Composition, Concentration and Distribution**
  Experiments [223, 240] showed that the reflectance of leaves is relatively low in the visible portion of the light spectrum ($400nm$ to $700nm$), and that absorptance dominates. Most of the absorption in this region is caused by pigments presented in the leaf tissue. Pigments are materials that exhibit selective reflection and selective absorption [86]. The usual pigments found in leaves are the chlorophylls and the caratenoids. As pointed out by Devlin and Baker [66], the chlorophylls are by far the most important and abundant of these pigments, and they are usually concentrated in the palisade mesophyll. There are several forms of chlorophyll [228], and the dominant ones are chlorophylls a and b. Although the relative pigment concentrations vary within the species, most of the plants

contain two to three times more chlorophyll a than b [187]. These two forms of chlorophyll present similar absorption spectra (curve showing the absorption coefficient of the material at various wavelengths), and some researchers combine them into a single curve due to the high correlation between them [125]. The carotenoids are a group of pigments which, in higher plants, are usually red, orange, yellow, or brown and are associated with chlorophyll in the chloroplasts [187]. Their yellow colors are evident in many autumn leaves from which the chlorophyll has disappeared. Senescent leaves present also brown pigments called tannis. However, to the best of our knowledge, there is not yet a method to determine the concentration of brown pigments [125].

- **Water Content**
  The concentration and distribution of water affects the absorption of light in the infrared region, but have no significant effect on the absorption of light in the visible region [223]. However, as water is lost from a fully turgid leaf, the reflectance increases. Different patterns in this increase are seen in different types of leaves. Wooley [240] hypothesized that the reason might be the changes in intercellular air spaces.

- **Surface Features**
  A small amount of incident light is initially reflected from the upper cuticular surface for angles of incidence smaller than 45° [240]. Grant [97] states that the presence of hairs in some species increases the reflectance, by presenting numerous light scattering interfaces which decreased the amount of light entering the leaf, thereby decreasing absorption. The reflectance curves of leaves are also affected by their venation system [240]. It is likely that their increased thickness and lower concentration of pigments affect the reflection and transmission of light [98]. However, biological data regarding these characteristics of the venation systems is scarce.

Grant *et al.* [97] describe leaves as having both specular and diffuse characteristics. The specular (non-Lambertian) character of the leaf reflectance arises at the surface of the leaf. For some viewing directions, the surface reflectance may be so large

that the leaves appear to have the color of the light source instead of the color determined by the foliar pigments, which is usually green. This happens when the reflected light visually overhelms the much smaller amounts of green light scattered by the interior of the leaves (Figure 3.3). The diffuse (Lambertian) character of leaf's reflectance emanates primarily from the mesophyll tissue through multiple scattering, with a small contribution of scattering from rough elements on the leaf surface. The multiple scattering within the foliar tissues also gives the leaf's transmittance a near-Lambertian distribution.

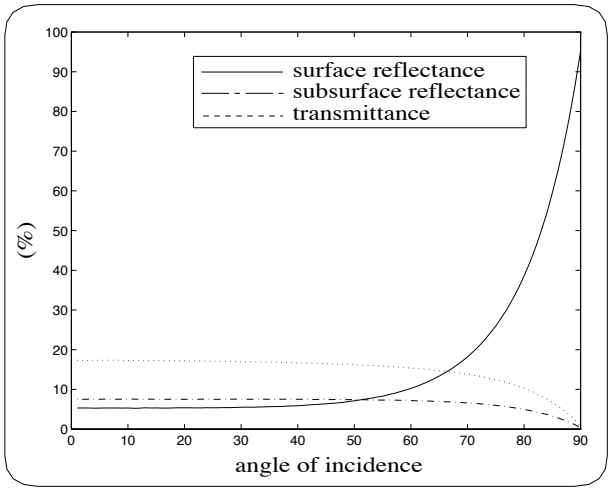

Figure 3.3: Curves of reflectance and transmittance of a soybean leaf obtained using the ABM (Chapter 4) at a wavelength of $550nm$ and considering the front (adaxial epidermis) of the leaf towards the light source.

Three principal types of scattering occur within the foliar tissues: Rayleigh, Mie and refractive-reflective scattering [223]. Rayleigh and Mie scatterings occur for particles of size equal to or less than the wavelength of the incident light respectively [86]. In the case of the foliar tissues these particles correspond to organelles and macromolecules. Although the extent and the exact causes of these forms of scattering in foliar tissues have not yet been completely resolved, their contribution to the total internal scattering is considered to be very low compared with the reflective-refractive scattering [223]. This type of scattering accounts for most of the diffusion or internal

scattering, and it is mainly caused by the arrangement of tissues, and the refractive index differences, which, for the most part, are associated with air-cell wall interfaces regarding cells whose dimensions are quite large compared with the wavelength of light such as the palisade and spongy cells.

In species having bifacial leaves the adaxial epidermis is attached to the mesophyll over most of its inner surface (Figure 3.1), so that once light has passed the outer epidermal surface, it can easily pass into the center of the leaf. The light entering or leaving the abaxial surface of a leaf must, however, pass through two semiplanar interfaces, namely the air-abaxial epidermis and the air-antidermal wall interfaces (Figure 3.1). Thus light from a source on the back side of a leaf can penetrate the back epidermis and be reflected back by the inside surface of the epidermis without encountering the mesophyll. It is largely because of this aspect that the back of the leaves is often pale. Moreover, these structural differences make the abaxial surface of a leaf a greater barrier to the escape of the oblique light than the adaxial surface. These factors explain why leaves that differ markedly in the structure of their two sides show corresponding differences in their reflectance and transmittance curves [240].

Experiments performed by Wooley on maize leaves [240] showed that these leaves reflect light more diffusively in a plane perpendicular to the veins than in a plane parallel to the veins. This aspect suggests that the anisotropy of plant leaves is associated with the venation systems. As described by Esau [72], a leaf may have a single vein or two or more. Single-veined leaves are found among conifers[1], while multiveined leaves are common in higher ferns[2] and angiosperms[3]. Although the major venation system of these species can be divided in two main types of patterns,

[1]Conifers correspond to one of the major groups of gymnosperms (vascular and flowerless plants characterized by the presence of seeds). They typically bear cones and needle-like leaves, *e.g.* pine and redwood [24].

[2]Ferns form the most widely known group of pteridophytes (vascular and flowerless plants which do not produce seeds). They present feathery fronds formed by large leaves, usually deeply parted or divided, *e.g.* wood fern [24].

[3]Angiosperms, or flowering plants, dominate large areas of the land surface and represent the climax of vascular plant evolution. They are very diversified in their form and range in size from few millimiters in diameter (aquatic *Lemna*) to over 90 meters in height (*Eucalyptus*) [36].

namely reticulate (or net) and striate (or parallel), the minor venation system exhibits a wide range of intergrading patterns. Due to the complexity of these patterns, further experiments are required to determine their relationship with the anisotropy of plant leaves. The experiments performed by Wooley also showed that the retroreflection[4] associated with a plant leaf is negligible.

3.3 Reflectance Models for Botany and Remote Sensing Applications

Many researchers from areas like botany and remote sensing have proposed reflectance and transmittance models for leaves where the goal is to understand the physiological processes that relate foliar optical properties to biophysical characteristics [126]. Although we have different goals in realistic image synthesis, some concepts used in those models, such as the intuitive concept of plates or layers, can be useful in computer graphics applications as well.

For most of this century the Willstatter-Stoll (W-S) theory [223] was the accepted explanation for the reflectance from a leaf on the basis of the reflection of light at spongy mesophyll cells wall-air interfaces. Since then, several theories and models have been proposed to describe the mechanisms of light reflection, transmission and absorption by leaves. For a comprehensive literature review on this topic the reader is referred to the texts by Grant [97], Bjorn [27] and Vogelmann [229]. In this section we examine relevant reflectance and transmittance models for leaves, designed mainly for botany and remote sensing applications, and group them according to their similarities.

3.3.1 Plate Models

Allen *et al.* [4] developed the "plate model" where the complex structure of a plant leaf can be simulated by a transparent plate with rough plane-parallel surfaces. The parallel surfaces are assumed to be Lambertian. After its penetration inside the leaf, the light flux is assumed to be diffuse. This model has two optical parameters: an

[4]Usually a "surface" having a pronounced 3-D microstructure, such as that of a forest viewed from an airplane, is likely to show retroreflection, *i.e.* the peak of reflectance along the angle of incidence, instead of near that of mirror reflection [128].

effective[5] index of refraction and an effective coefficient of absorption. Unfortunately, it cannot be applied to leaves which cannot be described as a unique compact layer such as dicotyledons[6] and senescent leaves.

Jacquemoud and Baret [125] have generalized the "plate model" to be used in inversion procedures in remote sensing applications [79]. An inversion procedure is a way to derive leaf biochemical properties from *in situ* and nondestructive optical experiments. Their generalization of the "plate model" resulted in a reflectance and transmittance model for a plant leaf called *Prospect*. This model has three main parameters: pigment concentration, water content and a structure parameter N. A leaf is then assumed to be composed of a pile of N homogeneous plates or layers separated by $N-1$ air spaces. The authors assumed an uniform distribution of water and pigments inside the leaf. The *Prospect* model was recently improved [126] using experimental data [120] to provide specific absorption coefficients for the biochemical constituents.

3.3.2 K-M Theory Based Models

In the beginning of the century, Kubelka and Munk developed a simple relationship between the scattering and absorption coefficients of paint and its overall reflectance. This relationship is known as the K-M theory [86]. It applies the transport theory to describe the radiation transfer in diffuse scattering media with two parameters: the scattering and the absorption coefficients. Paint is physically composed of many small colored pigment particles suspended in some sort of colorless base such as oil [86, 224]. Similarly, a leaf or a petal may be described as a diffuse scattering medium containing pigments.

The original K-M theory is, however, valid only for "macrohomogeneous" layers, *i.e.* layers characterized by invariant scattering and absorption coefficients, with later-

[5]In this context the term *effective* is used to indicate that the parameters refer to the whole leaf instead of any of its constituent materials such as epiticular wax, chlorophyll and water.

[6]The dicotyledons form the larger group of angiosperms bearing two cotyledons (first leaves formed in the embryo), including, for example, the broad-leaved trees, roses and sunflowers The other group of angiosperms is formed by the monocotyledons bearing one cotyledon, including, for example, grasses, lilies, orchids, irises, palms and cannas [24].

ally infinite extensions. As pointed out by Bjorn [27], a leaf or a plant canopy cannot be described as a single macrohomogeneous layer. Moreover, the original K-M theory is strictly valid only if the light is completely diffuse. Duntley [70] has extended it to incorporate, among other factors, specular (directional) light. The extensions proposed by Duntley have been applied to plant canopies by Allen and Richardson [2], Allen *et al.* [1] and Suits [218].

The K-M theory as originally stated is considered to be a two-flux theory, since only two types of radiant flux are involved, namely a diffuse downward flux and a diffuse upward flux. The relations between the fluxes are expressed by two simultaneous linear differential equations [226]. Allen *et al.* [2] proposed a reflectance and transmittance model for leaves based on the K-M-Duntley theory, which later resulted in the *AGR* model [1] (named after its developers). In this model, used to simulate stacked leaves, a direct solar flux is included, making it a three-flux theory with three differential equations and five coefficients.

Suits [218] added another flux type associated with the radiance in the direction of observation, making it a four-flux theory, with four differential equations and nine coefficients. Verhoef [226, 227] improved the angular responses of the model proposed by Suits through a detailed analysis of extinction and scattering of radiant flux by leaf layers. The results of the improvements proposed by Verhoef resulted in the *SAIL* model (Scattering by Arbitrarily Inclined Leaves). The original *AGR* model was later simplified by Baret et al. [20] and applied to wheat leaves.

Yamada and Fujimura [244] proposed a K-M model of reflectance and transmittance for individual leaves aiming at inversion applications to determine chlorophyll content. They considered four macrohomogeneous layers (two cuticles, a palisade layer and a spongy layer), each described by the K-M theory, and assumed that the scattering and absorption coefficients of the K-M theory can be expressed as a linear function of the pigment content of a plant leaf. In the formulation of their model Yamada and Fujimura use several constants which are not biologically meaningful. These constants are determined using a least square method [42]. Their model was tested on different leaves and the direct relationship between the reflectance and

transmittance of a plant leaf and its chlorophyll content was confirmed through experiments and direct measurements.

3.3.3 Ray Tracing Based Models

Ray tracing based approaches were used to simulate the propagation of light within foliar tissues even before being extensively used in computer graphics. Allen *et al.* [3] and Kumar and Silva [135] developed models where they simply traced the passage of a light ray through a drawing of a leaf cross section. Unfortunately, their experiments were not in the visible region of light spectrum and the techniques used were computationally expensive. In spite of that their conclusions form the basis of the current understanding of diffuse reflectance in leaves as pointed out by Grant [97],

Allen *et al.* [3] investigated the W-S theory of leaf reflectance by tracing rays through a two-dimensional (2-D) model of a leaf consisting of a single medium and air, in which the internal cellular structures were approximated by arcs. The optical properties of the medium were specified by a complex index of refraction. For a given incident ray, new reflected and transmitted rays were generated at each interface using Snell's law, Fresnel equations (Section 2.1) and Lambert's law of absorption (Section 2.3). Although their results were not in good agreement with the theory (the predicted reflectance was too high and the predicted transmittance was too low), their experiments suggested that a ray tracing approach might yield more accurate results if it was generalized to a more realistic model of a plant leaf.

In the model proposed by Kumar and Silva the leaf was assumed to consist of homogeneous and isotropic media[7] This assumption was made for mathematical simplicity so that the Fresnel equations could be applied to each interface. The Rayleigh and Mie scattering, as in the previous models, were not considered. Kumar and Silva proposed two different levels of detail to trace the rays within the leaf. In the first level only the two most important interfaces are considered: air to cell wall and cell wall to air. In the second level the four leaf main constituents, namely cell wall,

[7]In an isotropic medium only the angle between the incident and scattered directions matters, not their absolute location [86]. In other words, as described by Glassner [86], the direction of the incident and scattered directions don't matter.

chloroplasts, cell sap and air, give raise to the following eight optical interfaces in the leaf, all of which were considered in the ray tracing:

- air to cell,
- cell sap to cell wall,
- chloroplasts to cell wall,
- cell sap to chloroplasts,
- chloroplasts to to cell sap,
- cell wall to chloroplasts,
- cell wall to cell sap and
- cell wall to air.

Their experiments showed that considering only cell wall to air and air to cell wall interfaces is likely to give less diffuse reflectance and transmittance than that given by considering all the eight interfaces. Kumar and Silva also claimed that their model can be used to simulate the optical phenomena in the visible region of the light spectrum through the use of appropriate indexes of refraction of the leaf constituents in the visible region.

Recently Govaerts *et al.* [96] proposed a Monte Carlo ray tracing based model to simulate the propagation of light in a typical dicotyledon leaf. In their simulation approach the three-dimensional (3-D) internal cellular structure of various leaf tissues is explicitly geometrically modeled. A typical foliar cell is defined by Govaerts *et al.* as a set of concentric objects filled with three different media, namely cell wall material, water and chlorophyll, and the different shapes of these cells are approximated by primitive objects such as spheres, ellipsoids and cylinders. Although the modeling of individual cells may improve the accuracy of the simulations, the complexity of the resulting (3-D) geometrical model is likely to place a substantial demand on computational resources to trace the ray paths.

In their model, each cell constituent is characterized by an index of refraction and an absorption coefficient. The interaction of light with various cell constituents is also simulated using the Fresnel Equations (Section 2.1). The absorption testing assumes a homogeneous distribution of pigments, and it is performed probabilistically, using an expression based on Beer's law (Section 2.3), and considering the incident rays perpendicular to the interfaces. Although the results presented by the model proposed by Govaerts *et al.* seemed to qualitatively approximate the spectral curves of a plant leaf, direct comparisons with curves obtained experimentally would be needed to determine the quantitative accuracy of their results.

3.3.4 Radiative Transfer Theory Based Models

Tucker and Garret [223] proposed a stochastic leaf radiation model based upon leaf structure, pigment composition and concentration of water content. The radiative interactions within a leaf are considered as a random walk process [107], using the *Markov chain*[8] approach. Their stochastic model compartmentalizes a leaf to represent the radiation states, cell parts, and internal scattering. The radiation states, namely reflected, absorbed, and transmitted, are represented by six compartments:

- solar input;
- reflection from cuticle;
- absorption in the palisade cells;
- diffuse reflected radiation;
- absorption in the spongy cells;
- diffuse transmitted radiation.

Although the model predictions for leaf spectral absorption, reflection, and transmission closely agree with measured values for the non-visible region of the light

[8]A sequence of random variables, $x_0, x_1, x_2, ...$, generated such that at each time $t \geq 0$, the next state x_{t+1} is sampled from a distribution $\mathrm{P}(x_{t+1} \mid x_t)$, which depends only on the current state of the chain, x_t, is called a *Markov chain* [85].

spectrum, the simulation results did not account for the high degree of absorption which occurs in the visible region. Nevertheless, this model provides a technically sound approach for simulating light interactions with plant leaves. Tucker and Garret claim that the success in their modeling approach depends upon the calculation and justification of accurate probabilities to represent the flows between and within states.

Ma *et al.* [150] also used the radiative transfer theory combined with wave optics to develop a model in which a leaf is modeled as a slab of water with a irregular surface and containing randomly distributed scatterers. They assume that the surface roughness is of relatively great size compared with the incident wavelength and it undulates on a large scale. This assumption allowed them to apply the Kirchhoff rough surface scattering theory [23] to describe the leaf surface reflectance. Ma *et al.* assumed that the scatterers inside the leaf are spherical to enable them to apply a scattering function derived from the Mie scattering theory. The optical scattering within a leaf was simulated by combining the multiple scattering with the Kirchhoff rough surface theory.

Ma *et al.* solved the radiative transfer equations with Stokes vectors and boundary conditions [45] numerically using Fourier expansion, a discrete-ordinate technique and an eigenvalue-eigenvector method [124, 162]. Instead of providing spectral curves of reflectance and transmittance, the model proposed by Ma *et al.* determines optical parameters for different plant leaves, such as optical thickness[9], surface roughness and refractive indexes. These parameters can, in turn, be incorporated into standard reflectance and transmittance models. Ma *et al.* also considered the depolarization of the polarized incident beam used in their experiments. Although Ma *et al.* claimed that they have modeled a leaf as a slab of water because the high water content of fresh leaves, as mentioned earlier, the role of the absorption pigments in the optical phenomena occurring within a leaf cannot be overlooked.

[9]The optical thickness, as presented by Ma *et al.* is given by the product of the distance along z axis (thickness) by the particle number density and the total scattering cross section per particle.

3.4 Global Illumination Applications in Vegetation

In the previous sections we addressed issues mainly related to the interaction of light with individual leaves. In this section we provide an overview of global illumination applications for radiative transfer in regions of vegetation such as forest canopies or crops. Radiation transport in plant canopies have been extensively studied for interpreting remote sensing data [163], such as satellite images used to survey land use and resources, and for predicting vegetation productivity, litter decomposition processes and nutrient cycles within an ecosystem [79].

From the point of view of a satellite regions of vegetation are "surfaces" with a characteristic BDF that determines their appearance under various observation conditions [57]. In order to determine the BDF of a region of vegetation, such as a forest canopy, a model of this vegetation canopy can be constructed and evaluated through computer simulations [88]. Simplified models used in these applications usually treat a canopy as a system of homogeneous volume elements with a certain density. A model of radiative transfer is then used to compute the radiation transport between these volumes [57, 88].

For heterogeneous environments the radiative transfer processes are simulated using Monte Carlo methods [181, 182], which were introduced to plant canopy radiative transfer studies by Tanaka [220]. The application of these methods is done through the simulation of solar radiation transfer processes as an aperiodic *Markov chain* of collisions between photons and foliar tissues. Ross and Marshak [182] provide a detailed review of works in this area and their contributions to the development of a plant canopy radiative transfer theory.

As mentioned by Ross and Marshak, besides its suitability to the simulation of radiative transfer in 3-D inhomogeneous canopies, the use of Monte Carlo methods involves additional advantages. It allows the consideration of the detailed architecture of individual plants, the shape and the area of the leaves, the penumbral effects by penetration of direct solar radiation. Moreover, it also permits the calculation of the statistical characteristics of fluxes and intensities. However, to obtain results of acceptable accuracy a great number of trials is needed, which may increase the computational costs considerably.

Recently, remote sensing researchers and plant biologists started to resort to global illumination approaches developed or improved by the computer graphics community as alternative solutions to simulate radiation transport in plant canopies. Among these approaches is the radiosity method (Chapters 7 and 8). Gerst and Borel [84] and Borel *et al.* [35] have applied this method to simulate radiative transfer in canopies with several thousand of leaves, and Goel *et al.* [89] have applied it to simulations of radiative transfer in cornfields.

As mentioned by Borel *et al.* [35], the strength of the radiosity method for these applications is the capability of describing quantitatively and physically correctly the interactions of radiation with many surfaces that may be arranged in a complex structure such as a plant canopy. Among the drawbacks of this method one can highlight the difficulties to handle non-Lambertian phenomena. Moreover, this method may demand large computational resources for certain applications, *e.g.* scenes with millions of primitives. Fortunately, one can take advantage of techniques used in computer graphics, such as the clustering of objects hierarchically into volumes [206] and the application of hierarchical radiosity methods [56, 110, 111, 196], to minimize the computational costs.

Methods combining different global illumination techniques have also been used to perform radiative transfer simulations in plant canopies. Govaerts [95] used a Monte Carlo ray tracing algorithm, and Chelle [47] proposed an hybrid approach in which the radiosity method is used for elements of a region of vegetation close to a point of interest and the *SAIL* model (Section 3.3.2) is used for elements distants from this point of interest. This approach was called nested radiosity for plant canopies [48]. Recently, Mech [158] used a radiosity via ray tracing algorithm [200], also based on Monte Carlo methods, to compute the amount of light reaching plant organs in a given environment.

The Monte Carlo ray tracing based algorithms can be applied to complex environments and are easy to implement. A large number of rays, however, is usually required in order to converge to an accurate solution. Moreover, if there are lighting and reflectance changes in the environment the entire process of "shooting" rays to determine the energy balance in the environment must be repeated. For this case,

the application of the standard radiosity may be more appropriate. We examine this issue in more detail in Chapter 8.

Recently Max *et al.* [155] proposed a plane-parallel radiance transport approach for global illumination in vegetation. In this approach the radiance transport problem is reduced to one dimension (1-D) by assuming that the radiance depends angularly on light flow directly on the unit sphere, but positionally only on z, the height above the ground. The resulting partial differential equations are solved by the Runge-Kutta method [42]. Max *et al.* claim that it is a good approximation for dense vegetation, although it is not suitable for isolated trees.

One of the problems of applying global illumination methods for radiative transfer in regions of vegetation, such as forest canopies and crops, is the scarce amount of analytical or experimental results to evaluate these methods under these circumstances. Usually the results regarding a proposed approach are compared with the results provided by a different approach based on similar assumptions. If they present a good agreement, the accuracy of the proposed approach is considered satisfactory, although they may be both far from the real solution. Nevertheless, since most of these approaches consider the volume elements or leaves as presenting only Lambertian characteristics, the incorporation of physically and biologically-based models including non-Lambertian effects and absorption for these elements is likely to increase the overall accuracy of these approaches.

The efficiency of the radiative transfer methods for plant canopies is directly associated with the efficiency of the global illumination algorithms used in these applications. Improvements on these algorithms would reduce the computational overhead of these applications. For instance, computing the radiosity at all positions necessary to account for all sources of radiation in a vegetation canopy usually involves the iterative solution of a large linear system. Clearly the investigation of faster solutions for these systems (Chapters 7 and 8) would improve the efficiency of the radiative transfer approaches for plant canopies, not only for remote sensing purposes [48], but also for rendering purposes.

3.5 Summary

In this chapter we have summarized the main aspects of light interaction with plants. In order to simulate the mechanisms of photon transport within the foliar tissues, it is necessary to account for the biological and structural characteristics of these tissues and their components. Factors like the presence of pigments and the internal distribution of the tissues have a significant impact on how leaves propagate and absorb light.

Several models have been proposed by remote sensing researchers and plant biologists to describe the reflectance and transmittance characteristics of leaves. These models present different levels of complexity and propose different techniques to describe and simulate light propagation within the foliar tissues. Although these models have not been designed for rendering applications, they provide valuable information that can be incorporated in the design of models aiming at these applications.

A substantial amount of research has also been done to describe and simulate the radiative transfer processes in vegetation. Although most of the research on this topic has been restricted to remote sensing applications, the fundamental aspects involved in these processes are also relevant for the rendering of natural scenes. Not surprisingly global illumination methods used in rendering applications have been incorporated by remote sensing researchers in their studies of radiative transfer in plant canopies.

The local and global levels of light interaction with plants, namely the reflectance models for individual leaves and the radiative transfer processes for canopies, are directly connected. New developments or improvements on either level, or in both, will have a direct influence on the accuracy and efficiency of the rendering methods for natural scenes. In the next chapters we intend to address these issues and present our contributions.

Chapter 4
The Algorithmic BDF Model for Plant Tissue

Although many physically plausible BDF models have been proposed in the computer graphics literature (Section 2.2), there is still a lack of models aimed at biological materials. In this chapter we address this issue through the presentation of an algorithmic BDF model (*ABM*) that accounts for biological factors that affect light propagation and absorption in plant tissues. The model is oriented to leaves, since, as mentioned earlier, they are the most important plant surface interacting with light. However, it can be extended to other plant surfaces like petals and stems since they present similar optical and structural characteristics [80, 134].

Many BDF models used in computer graphics rely on values of reflectance and transmittance either set by the user or obtained from the literature. Although spectral reflectance and transmittance curves are available for leaves [81, 98, 100], they are restricted to a narrow range of illuminating and viewing angles. Thus, it becomes necessary to design a BDF model for plant tissue in which the reflectances and transmittances are calculated by the model itself.

An exact geometrical model of a plant tissue would model individual cells and their interior details explicitly. The geometrical model used by Govaerts et al. [96] (Section 3.3.3) is an example of such an approach. The drawbacks of using such a detailed geometrical model in the simulation of light interaction with foliar tissues are the large number of parameters involved, which makes control difficult, and its significant implementation overhead. Our strategy is to use a higher level of abstraction, which allows our BDF model to be controlled by a small number of biologically meaningful parameters and enables its easy incorporation into global illumination frameworks.

The implementation of the *ABM* is based on an algorithmic process using Monte Carlo methods [107, 152]. Thus, the BDF of a foliar tissue can be calculated and used on the fly during the rendering process, or computed off-line and stored to be reconstructed during rendering (Chapter 5) or used for validation purposes. We compare the results obtained using the *ABM* with available experimental data.

The remainder of this chapter is organized as follows:

- Section 4.1 introduces the algorithmic BDF model.
- Section 4.2 presents the algorithms used to simulate the scattering profile of the foliar tissues.
- Section 4.3 describes the algorithms used to account for the absorption within these tissues.
- Section 4.4 presents an overview of the model's implementation and a summary of its parameters.
- Section 4.5 presents the procedures and parameters used to test the *ABM*, focusing on the computer simulation of measurement devices.
- Section 4.6 discusses the results and performance issues.
- The chapter closes with a summary of the main characteristics of the *ABM*.

4.1 Outline of the Algorithmic BDF Model

In the *ABM* light propagation is described in terms of ray optics. The foliar tissues are assumed to be isotropic due to the lack of experimental data on factors affecting the anisotropy of plant leaves (Chapter 3). Moreover, the refractive-reflective scattering is assumed to be the dominant form of scattering within the foliar tissues, which is consistent with the literature regarding this topic (Chapter 3).

Instead of geometrically modeling many cells individually, the propagation of light within these tissues is simulated as a stochastic process whose states are associated with the air-cell wall interfaces represented in Figure 4.1. Once a ray hits a leaf at interface 1 or interface 4 (state 1 or state 4), it can be reflected back to the environment

or refracted to the interior of the leaf. Then, it can be reflected or refracted multiple times until it is absorbed at state 2 or leaves the leaf at the states 1 or 4.

The light interactions in our model may therefore be seen as a *random walk*[1] process in which the transition probabilities are associated with the Fresnel coefficients computed at each interface (using the Fresnel equations presented in Section 2.1), and the termination probabilities are associated with the free path length computed when a ray travels in the mesophyll layer towards or from interface 2. We examine the free path length concept more closely in Section 4.3.

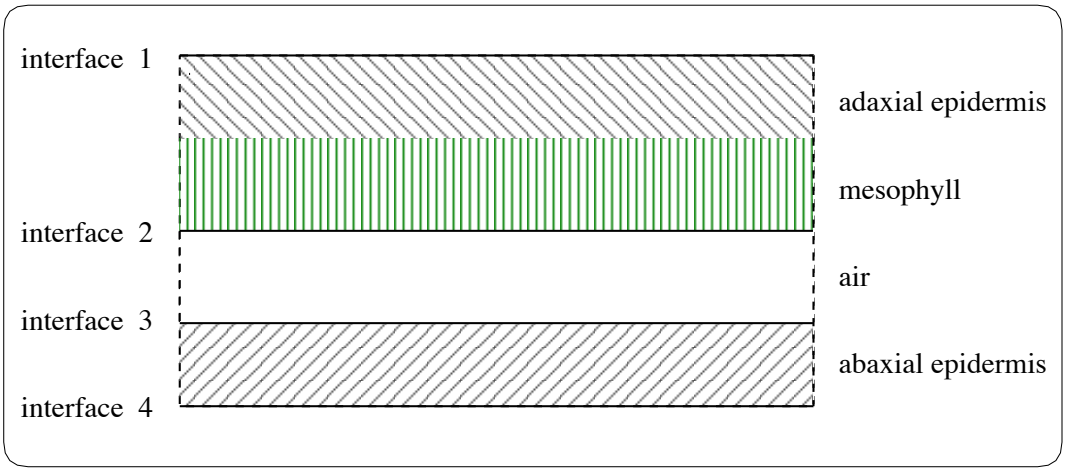

Figure 4.1: Interfaces and tissues considered by the *ABM*.

Although the Fresnel equations are valid only for infinite plane surfaces, in practice, as pointed out by Govaerts [95], these equations may still be used since a cell is large with respect to the wavelength of the incident light, and the portion of the cell wall interacting with a ray can be considered locally flat. Each time a ray hits an interface we compute the corresponding coefficient (R). We consider only the real parts of the refractive indexes, since for the spectral region covered in this work

[1] A *random walk* is a *Markov chain*, whose successive states (connected by the transition and termination probabilities) comprise all possible positions of the particle within the boundary, together with a special "absorbing" state, that the system enters as soon as the particle hits the boundary and remains in for ever after [107].

($400nm$ to $700nm$), and for the materials considered, the complex parts are so small that they can be neglected [95, 96, 173, 241].

Using these simplifying assumptions the equation used to compute the Fresnel coefficients at each interface, assuming unpolarized light [198], reduces to Equation 2.11. After computing the Fresnel coefficient at an interface, we generate a uniform random number $\xi \in [0, 1]$. If the random number ξ is smaller than or equal to the Fresnel coefficient R, we generate a reflected ray $\vec{r}$, otherwise we generate a refracted ray $\vec{t}$. The reflected ray is obtained using the law of reflection (Equation 2.3). The refracted ray is obtained using Snell's law (Equation 2.6).

The *ABM* takes into account the three components of the BDF of plant tissues: surface reflectance, subsurface reflectance and transmittance. They are affected by the surface roughness, the internal scattering and the light absorption in the mesophyll tissue. In the next sections we describe how we address these issues.

4.2 Scattering Profile

Brakke *et al.* [38] have noted that the scattering profile of a plant leaf can be approximated by an exponentiated cosine function. We have used a similar approach to simulate the distribution of the rays reflected or refracted at the foliar tissues. A rough surface, like a leaf epidermis, can contain more than one distinct scale of roughness. The contour of the epidermal walls represents large surface features with respect to the wavelength of the incident light. The microdetails of the epiticular waxes represent the small surface features relative to the wavelength of the incident light.

For the plant epidermis the large-scale roughness will dominate scattering in the specular direction, as pointed out by Grant [97], and the small-scale roughness will control scattering away from the specular direction. The epiticular waxes exhibit a wide range of geometric configurations, and their contribution to the overall reflectance is not as significant as the overall shape of the epidermal cells. For these reasons we decided to concentrate our simulation efforts on the large scale features.

Govaerts *et al.* [96] have shown that the epidermal cells can be approximated by oblate ellipsoids. An oblate ellipsoid has semi-axes a_1, a_2 and a_3, with $a_1 = a_2$ and

$a_1 > a_3$ [212]. For the plant cells we consider a_1 and a_2 as the axes in the plane of the foliar tissues, with values corresponding to average radius of the cell, a_r. Unlike the oblateness definition used by Govaerts *et al.* we define the oblateness of the cell as $\frac{a_r}{a_3}$. The dimensions of the epidermal cells of several species of plants can be found in the literature [33, 95, 153].

To simulate the effects of the shape of the epidermal cells on the reflected rays at the air→epidermal cells interface, we perturb the rays using a warping function (Equation 4.1). This function corresponds to a PDF based on an exponentiated cosine distribution (Appendix A), and the exponent is given by the oblateness of the epidermal cells. The perturbation is performed through angular displacements, α_e and β_e. The angle α_e corresponds to the polar angle with respect to the ideal reflection or ideal transmission direction. The angle β_e corresponds to the azimuthal angle around the ideal reflection or ideal transmission direction. These angles are given by:

$$(\alpha_e, \beta_e) = (\arccos((1 - \xi_1)^{\frac{1}{ob+1}}), 2\pi\xi_2) \tag{4.1}$$

where:

ξ_1 and ξ_2	=	uniformly distributed random numbers $\in [0, 1]$,
ob	=	oblateness of the epidermal cells.

Therefore, leaves with large epidermal cells' oblateness will have a surface reflectance closer to a specular distribution than leaves with small epidermal cells' oblateness. As pointed out by Shirley [204], the energy conservation properties of a PDF based on an exponentiated cosine function become less physical as the exponent gets smaller, and the resulting behavior becomes that of a mirror when the exponent becomes larger. By associating this exponent with a biological meaningful parameter, we set boundaries for the values assumed by this exponent and, consequently, reduce the chances of violating the energy conservation properties of this PDF. Based on the dimensions of the epidermal cells for several species of plants found in the literature, an appropriate range for the oblateness would be [1.4,5].

When light passes to the mesophyll its direction of travel is randomized and it becomes diffuse. We simulate the distribution of the rays in this tissue using another warping function (Equation 4.2). In this case the PDF corresponds to a diffuse

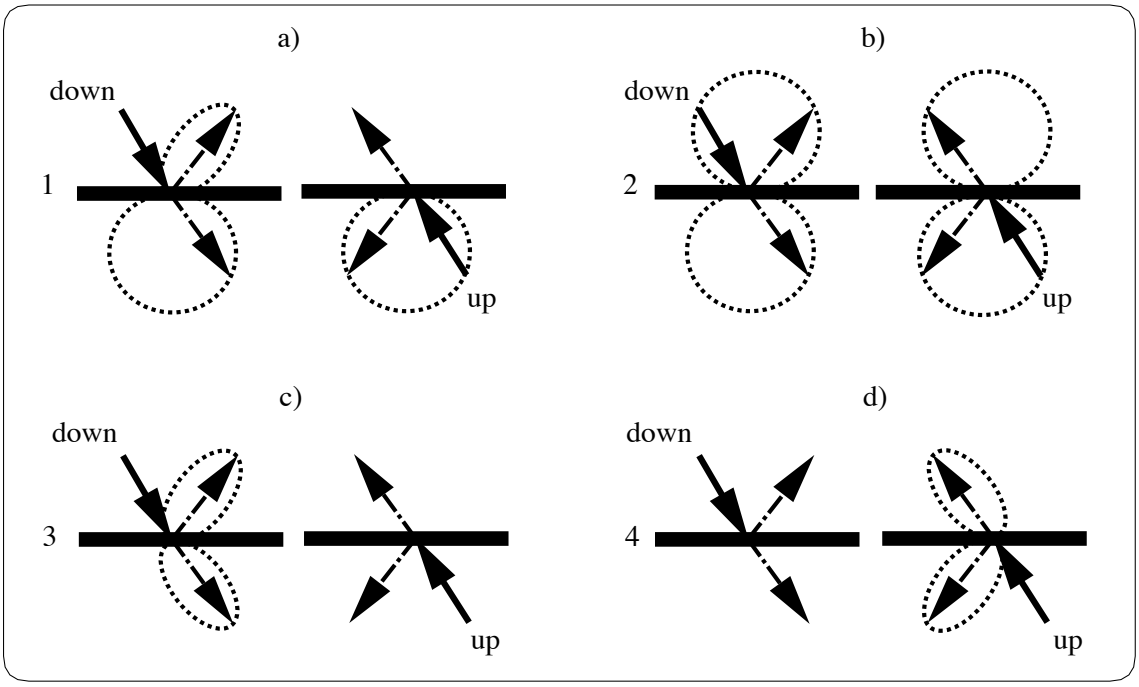

Figure 4.2: Perturbations performed by the *ABM* on the rays distributions at the four interfaces in the upwards and downwards directions of propagation considering the adaxial surface on the top and the abaxial surface on the bottom. a) Interface 1. b) Interface 2. c) Interface 3. d) Interface 4.

or cosine distribution (Appendix A). The perturbation is also performed through angular displacements, α_m and β_m. The angle α_m corresponds to the polar angle with respect to the reflection or transmission direction of the propagated ray. The angle β_m corresponds to the azimuthal angle around the propagation direction. These angles are given by:

$$(\alpha_m, \beta_m) = (\arccos(\sqrt{(1-\xi_3)}, 2\pi\xi_4) \tag{4.2}$$

where:

ξ_3 and ξ_4 = uniformly distributed random numbers $\in [0, 1]$.

Figure 4.2 presents a sketch showing the perturbations performed in each interface in both directions, upwards and downwards. In order to be consistent with available biological information and avoid undue complexity, we adopt a conservative strategy.

In other words, we do not perturb the rays when the impact of the perturbation is not significant to the overall BDF.

4.3 Absorption

A ray may encounter the mesophyll tissue after interacting with interface 1 or interface 2. Once it enters the mesophyll tissue it proceeds through the *mesophyll loop* (Figure 4.3), where it may be propagated diffusively or absorbed. As mentioned in Chapter 3, the diffuse propagation is due to multiple interactions with the many reflective-refractive discontinuities of the air-cell walls interfaces, and the absorption is associated with the presence and concentration of pigments.

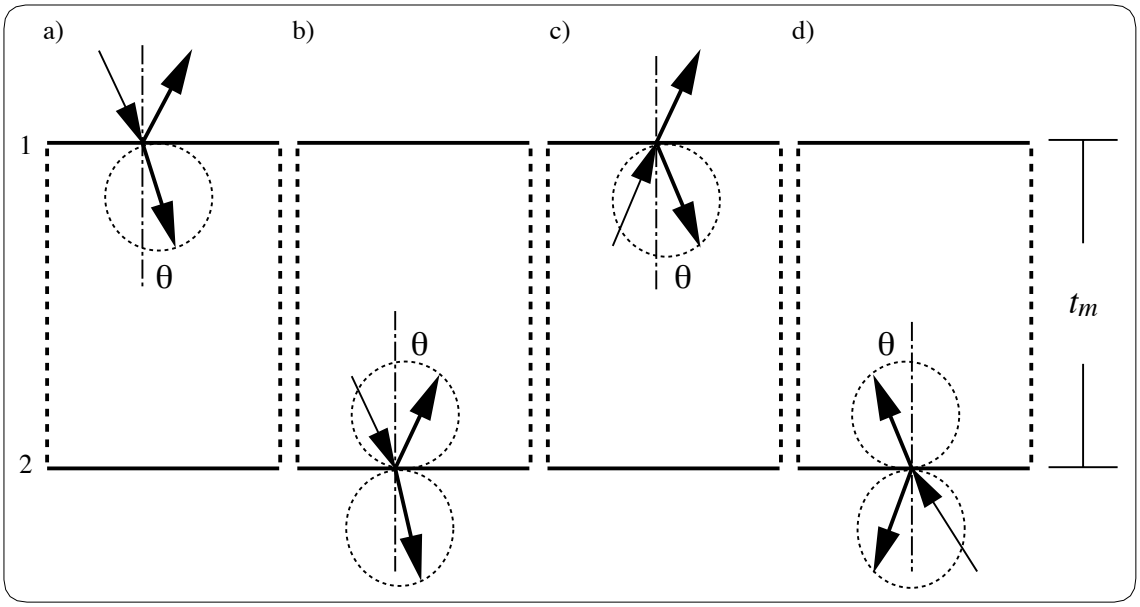

Figure 4.3: *Mesophyll loop.* a) Ray coming from outside interacts with interface 1. b) Refracted ray from interface 1 interacts with interface 2. c) Reflected ray from interface 2 interacts with interface 1. d) Ray coming from interface 3 interacts with interface 2.

When a perturbed refracted ray comes from interface 1 (Figure 4.3a), it is tested for absorption. If it is not absorbed, the ray is tested for reflection or refraction at interface 2 (Figure 4.3b). If the outcome of this test is the refraction of the ray, it

is perturbed and transmitted to interface 3, otherwise the reflected ray is perturbed and tested for absorption. If it is not absorbed, it may be reflected back to the mesophyll at interface 1 (Figure 4.3c), restarting the *mesophyll loop*, or refracted to the environment.

When a ray comes from interface 3 (Figure 4.3d), it is also tested for reflection or refraction at interface 2. If the outcome of the test is a reflected ray, it is perturbed and sent back to interface 3 without accounting for absorption. Otherwise, the refracted ray is perturbed and tested for absorption. If the ray is not absorbed, it proceeds in the *mesophyll loop*.

The absorption of light in dye solutions varies with the thickness of the sample and the concentration of pigments in accordance with a combination of Beer's law and Bouguer's law (Section 2.3). This combination, however, assumes that the incident ray is perpendicular to the interface between the incident medium and the transmission medium. Allen *et al.* [4] have included an additional term to this combination to account for different angles of incidence. The resulting expression states that the transmissivity of a plate along a slant ray is given by:

$$T = \epsilon^{-a(\lambda)\ c\ h\ \sec\theta} \tag{4.3}$$

where:

$a(\lambda)$	=	absorption coefficient,
c	=	concentration of pigment,
h	=	thickness of plate,
θ	=	angle of slant ray with respect to normal direction.

The absorption testing performed by the *ABM* is based on Equation 4.3, and assuming a homogeneous distribution of pigments. It is performed probabilistically every time a ray starts a run in the mesophyll tissue. It consists of the estimation of the ray geometrical path length[2], p, through the following expression:

[2]In this context the ray geometrical path length represents the spatial distance covered by a slant ray transmitted through a medium until it is absorbed or leaves the medium.

$$p = -\frac{1}{a_e(\lambda)} ln(\xi) \cos\theta \tag{4.4}$$

where:

ξ	=	uniformly distributed random number $\in [0, 1]$,
$a_e(\lambda)$	=	effective absorption coefficient of pigments,
θ	=	angle between the ray direction and the normal direction.

The effective absorption coefficient, as described by Jacquemoud and Baret [125], is given by:

$$a_e(\lambda) = \sum_{i=1}^{n} a_i(\lambda)\ c_i \tag{4.5}$$

where:

$a_i(\lambda)$	=	specific absorption coefficient of a given pigment i ($\frac{cm^2}{\mu g}$),
c_i	=	concentration of a given pigment i ($\frac{\mu g}{\mathrm{cm}^2}$).

If p is greater than the thickness of the pigmented medium, t_m (both expressed in cm), then the ray is propagated, otherwise it is absorbed. The thickness value used in this comparison does not correspond to the total thickness of the leaf, but to a fraction of it. Since the palisade tissue presents a higher concentration of pigments than the spongy tissue, and it is largely responsible for the internal scattering, we suggest to use its thickness (around 45% of the total thickness) for most cases. For species that present an undifferentiated mesophyll tissue, *e.g.* characterized by the presence of spongy cells only [81, 230], we suggest using the entire thickness of the mesophyll tissue (around 75% of the thickness).

A comprehensive set of leaf optical experiments, called *LOPEX* [120], provides the pigment concentrations and thickness values for 120 leaf samples representative of more than fifty species. The absorption spectra of chlorophyll and carotenoids pigments can be found in the literature [34, 43, 53, 66, 125, 187, 193, 228, 247, 248]. We examine this issue in more detail in Section 4.5.2. Additional information on optical parameters of foliar tissues is available in a survey by Gausman and Allen [81].

4.4 Implementation Overview and Summary of Parameters

The *ABM* was implemented using the *C++* language and the *ggLibrary* [202], a set of *C++* utilities designed to be used in computer graphics applications. In the *ggLibrary* materials are classified as families, such as metals and dielectrics, and grouped by parameters that affect their behavior, following the ray tracing framework for global illumination proposed by Shirley *et al.* [201]. A set of routines specific for each family computes the BDF for that family.

In the case of the *ABM*, the new family corresponds to the plant tissues, and each of its four states is implemented through a different routine. These routines are called recursively by a controlling routine until the ray is either absorbed or abandons the foliar tissues. The parameters used in the *ABM* are the following:

- η_c - refractive index of the external cutinized wall of the epidermis: used in the Fresnel computations in states 1 and 4;
- *ob* - oblateness of epidermal cells: used in the perturbation of the rays in state 1 (Figure 4.2a), in state 3 (Figure 4.2c), and in state 4 (Figure 4.2d);
- η_m - refractive index of the mesophyll cell wall: used in the Fresnel computations in state 2;
- t_m - thickness of mesophyll tissue: used in the absorption testing performed in state 2 (measured in cm);
- c - concentration of pigments: also used in the absorption testing performed in state 2 (measured in $\frac{\mu g}{\mathrm{cm}^2}$);
- η_a - refractive index of the antidermal wall: used in the Fresnel computations in state 3.

The refractive indexes come from experiments described in the literature [82, 125, 241]. The remaining parameters are also available in the literature, as mentioned earlier. The user may also adjust the optical parameters, *e.g.* t_m, depending on the conditions at hand. For example, for some species, a leaf in the sunlit and a leaf in the shady portions of a tree, may present differences of thickness of around 25% [3].

Notice that we use the same parameters for the adaxial and abaxial epidermis. This approach is accurate for most cases, but for some species it may be more appropriate to use a set of parameters with slightly different values.

We use the absorption spectra of chlorophyll a and b provided by Jacquemoud and Baret [125] in the absorption computations. Although these curves refer to pigments commonly present in the foliar tissues of most species, they may be adjusted according to the lengthening of the optical path of a given plant leaf, also called ratio of intensification [43] or factor of intensification [183]. This factor represents a combination of light that passes through plant tissues without encountering an absorber (sieve effect) and the light that is scattered and has an increased path length (detour effect). As mentioned by Vogelmann [229], these two phenomena have opposite outcomes: the sieve effect lowers absorption (specially at or near wavelengths for which the absorption has a maximum value), whereas the detour effect increases absorption (specially at or near wavelengths for which the absorption has a minimum value). Nonetheless, in dispersive samples, such as plant leaves, the absorption is enhanced by the combination of these two effects [183].

The factor of intensification of a plant leaf depends on the chlorophyll content. The greater the chlorophyll content, the shorter the length of the optical path. For example, McClendon and Fukshansky [157] made estimates of factors of intensification for plant leaves of nine different species with a chlorophyll content between 40 and $70 \frac{\mu g}{cm^2}$, and the resulting value, *i.e.* the mean and the standard deviation, was 2.30 ± 0.25.

4.5 Testing Issues

One possible approach for validating a BDF model is visual inspection by comparing images generated using the model with images generated using previous models. This approach is not suitable in our case, since the graphics literature lacks models aimed at plants. Recall that, although the multiple-layer model proposed by Hanrahan and Krueger [109] can be applied to plants (Section 2.6.3), it does not compute the reflectances and transmittances by itself, which are introduced to the model as

input parameters. Therefore, comparing such a model with the *ABM*, which actually computes the reflectances and transmittances[3], is not appropriate.

Another alternative would be to compare images generated using the *ABM* with photographs of real plants. In this case, since the computer generation of images is highly dependent on rendering parameters, the validation of the *ABM* could be biased by errors generated by other stages of the rendering pipeline, such as geometrical modeling, spectral modeling and directional sampling. As appropriately stated by DeYoung and Fournier [67], the shape of an object, the lighting, and the viewing position can all greatly affect how much BDF properties affect an image, and attempting to measure how much effect these properties will have on generated images would necessitate a separate investigation.

Since our goal at this point of our research is not necessarily to generate the best possible image of a plant, but to contribute to its synthesis respecting the laws of physics, we chose to test our model as a separate unit of the rendering pipeline and compare the results with the best available experimental data. In order to perform a comprehensive validation of a BDF model it would be necessary to consider all possible illuminating and viewing geometries. The very large number of measurements needed as well as the lack of experimental data for all possible geometries precludes the use of this approach.

Judd and Wyszecki [128] pointed out that in many goniophotometric measurements the most interesting and informative data for practical samples are taken in the plane given by the direction of the incident light and the normal of the specimen. Therefore, due to the practical reasons mentioned above and the isotropic characteristic of the *ABM*, we chose to use this plane in its testing.

4.5.1 Simulation of Devices for Measurement of Appearance

We are interested not only in the spectrophotometric curves of reflectance, ρ, and transmittance, τ, but also in the goniophotometric BDF curves, given in terms of BRDF, f_r, and BTDF, f_t. The spectrophotometric measurements are taken to

[3]The reflectances and transmittance computed by the *ABM* can be used as input parameters by other models, including, but not limited to the *H-K* multiple-layer model. We examine this possibility in more detail in Chapter 6.

determine the spectral energy distribution of the propagated light and the goniophotometric measurements are made to find the spatial distribution of that light. As mentioned in Section 2.5, while these two groups of measurements can be performed separately, they work together to establish the *measurement of appearance* of an object (Section 2.5).

Actual measurements of ρ and τ are performed using spectrophotometers[4] equipped with an integrating sphere, and the actual measurements of f_r and f_t are performed using goniophotometers [122, 128]. These devices are important basic tools for fundamental research in colorimetry [151], solar engineering [74], remote sensing [63] and plant biochemistry [34].

An integrating sphere is usually used to measure the hemispherical reflectance factor, which is defined as the ratio of the flux reflected by a sample under specified conditions of irradiation and collection to that reflected by the ideal completely reflecting isotropically diffusing surface under identical conditions of irradiation and viewing [246]. However, the numerical values of reflectance (Section 2.2) and reflectance factor are identical under the conditions of hemispherical collection [86].

For absolute measurements, the sphere wall is the standard, and the integrating sphere theory [87] compensates for the absolute reflectance of the sphere wall by mathematically treating the wall reflectance as unity. Hence, as pointed out by Zerlaut and Anderson [246], the hemispherical measurements made with such integrating spheres are the absolute reflectance (or the absolute transmittance) subject to small errors associated with aperture losses, small values of non-uniformity of sphere wall reflectance, sample mounts and so on.

Besides the BDF computations, the measurement of light distributions of specimens performed by goniophotometers also allows the verification of reciprocity and anisotropy properties of these specimens. Moreover, this device can also be used to measure the glossiness (Section 2.5) of a specimen given by its gloss factor. In this case the measurement is made relative to a perfect mirror surface instead of a perfect diffuser surface [128].

[4]The spectrophotometers can also be used to determine the absorption characteristics of a specimen as a function of wavelength [161].

The reflectance, transmittance and BDF measurements for testing purposes described in this dissertation were performed using computer simulations of a spectrophotometer and goniophotometer, henceforth called virtual measurement devices. The use of such virtual devices gives us control over the spectral data generation and allows us to capture a BDF at different sampling resolutions as pointed out by Lalonde and Fournier [141]. Moreover, it also allows us to avoid measurement errors such as perturbations in the incident and viewing angles due to the fact that specimens tend not to lie flat on the specimen holders of real measurement devices.

In our simulation of a spectrophotometer with integrating sphere the rays are shot from an emitter placed above the specimen following an incident solid angle, $\vec{\omega_i}$. Considering a total number of N rays fired towards the specimen for a given wavelength λ, one can assume that each ray carries the same amount of power, Φ. If the total power to be shot is Φ^i, then, as stated by Shirley [198, 201], the power carried by each ray is given by:

$$\Phi_{ray}(\lambda) = \frac{\Phi^i(\lambda)}{N} \tag{4.6}$$

Recall that reflectance describes the ratio of reflected power to incident power, and transmittance describes the ratio of transmitted power to incident power. Considering this ratio, if m_r rays are reflected towards the upper hemisphere, Ω_r, the spectral reflectance of the specimen with respect to a given wavelength λ of the incident light will be given by:

$$\rho(\lambda, \vec{\omega_i}, \Omega_r) = \frac{m_r}{N} \tag{4.7}$$

Therefore, since we can simply count the number of rays reflected to the upper hemisphere to determine a specimen's absolute spectral reflectance, we do not need to use an integrating sphere in our virtual spectrophotometer in order to collect the reflected rays. The absolute spectral transmittance is calculated similarly considering the lower hemisphere.

In our simulation of a goniophotometer we use a slightly different arrangement. Instead of considering the entire upper and lower hemisphere as collectors, we use

several radiance detectors represented by the patches of a collector sphere placed around a specimen. Using this arrangement, the BRDF regarding a direction associated with a given radiance detector placed in the upper hemisphere is determined in terms of radiant power. More specifically, it is given by the ratio between the radiant power reaching the detector, Φ^r, after interacting with the specimen, and the incident radiant power, Φ^i [17, 95].

The corresponding expression used to compute the BRDF for light incident at wavelength λ, considering the solid angle regarding the direction of incidence, $\vec{\omega_i}$, and the solid angle regarding the direction associated with the radiance detector, $\vec{\omega_r}$, is given by:

$$f_r(\lambda, \vec{\omega_i}, \vec{\omega_r}) = \frac{\Phi^r(\lambda)}{\Phi^i(\lambda)\, \vec{\omega_r}^p} \tag{4.8}$$

where:

$\vec{\omega_r}^p$ = projected solid angle regarding the direction associated with the radiance detector.

In turn, the projected solid angle $\vec{\omega_r}^p$ is given by:

$$\vec{\omega_r}^p = \frac{A_r \cos\psi_r}{L^2} \tag{4.9}$$

where:

A_r = area of the radiance detector,

L = distance from the specimen to the radiance detector,

ψ_r = angle between the direction associated with the radiance detector and the specimen normal.

Recall that the radiant power reaching the radiance detector can be written as:

$$\Phi^r(\lambda) = m_r \Phi_{ray}(\lambda) \tag{4.10}$$

Thus, replacing Equation 4.6 and Equation 4.10 in Equation 4.8, the expression to compute the BRDF reduces to:

$$f_r(\lambda, \vec{\omega_i}, \vec{\omega_r}) = \frac{m_r}{N\, \vec{\omega_r}^p} \tag{4.11}$$

Similarly, the BTDF is calculated considering radiance detectors placed in the lower hemisphere.

The origins of the rays are random points uniformly chosen from a disk used to represent the surface of the emitter. The coordinates of the points are given by the pair (Θ, l), which is computed using the following warping function suggested by Shirley [198]:

$$(\Theta, l) = (2\pi\xi_1, \mathcal{R}\sqrt{\xi_2}) \tag{4.12}$$

where:

ξ_1 and ξ_2 = uniformly distributed random numbers $\in [0, 1]$.

$\mathcal{R}$ = radius of the disk.

The targets of the rays are random points uniformly chosen from a pair of triangles used to represent the specimen. To choose a random point q on a triangle defined by the vertices q_0, q_1 and q_2 we use the following expression:

$$q = q_0 + \varphi(q_1 - q_0) + \gamma(q_2 - q_0) \tag{4.13}$$

where φ and γ are obtained using another warping function suggested by Shirley [198]:

$$(\varphi, \gamma) = (1 - \sqrt{1 - \xi_3}, (1 - \varphi)\xi_4) \tag{4.14}$$

where:

ξ_3 and ξ_4 = uniformly distributed random numbers $\in [0, 1]$.

The precision of real measurement devices is estimated by the ability of the instrument to repeat a measurement of a given specimen as mentioned by Judd and Wyszecki [128]. For instance, a spectrophotometer is considered to be of high precision if the spectral measurement, reflectance or transmittance, is repeated to achieve a tolerance better than $\Delta(\lambda) = \pm 0.001$. In our simulations we select the number of rays N to be shot using this tolerance value to avoid undesirable fluctuations in the measurements.

4.5.2 Testing Parameters

The set of leaf optical experiments cited in Section 4.3, *LOPEX* [120], includes measured spectral curves of reflectance and transmittance. In order to compare our results with these measured curves, we attempt to reproduce the actual measurement conditions used in the *LOPEX* as faithfully as possible. We use the same area for the specimen, $40mm^2$, and the same angle of incidence, $8°$, with respect to the normal of the specimen. Our emitter has a radius of $8mm$, and it is positioned at a distance of $30mm$ from the center of the specimen. These values correspond respectively to the area of the apertures and radius of the integrating sphere of the spectrophotometer used by *LOPEX*.

We selected, without loss of generality, a leaf from the soya species (*Glycine max, Soja hispida*), commonly known as a soybean leaf, because of its standard foliar characteristics and the large variety of experimental data available for comparison [40, 98, 100, 120, 156, 232, 240] besides *LOPEX*. The parameters used in our testing experiments are presented in Table 4.1. The refractive indexes were obtained from the work by Wooley [241], and the oblateness corresponds to the average dimensions of the epidermal cells of plant leaves belonging to the dicotyledons group ($a_r = 12.5$, $a_3 = 2.5$), which are available in the texts by Govaerts *et al.* [96] and Norman [172].

η_c	ob	η_m	c	t_m	η_a
1.6	5.0	1.41	43.62	0.0072	1.42

Table 4.1: Parameters used in the testing of the *ABM*.

The concentration of chlorophyll was obtained directly from *LOPEX*, and the thickness was computed using 45% of the total soybean leaf thickness, also provided by *LOPEX*, as proposed in Section 4.3. The combined absorption spectra of chlorophylls $a + b$ used in the testing of the *ABM* (Figure 4.4) is based on the curve provided by Jacquemoud and Baret [125], but adjusted according to a factor of intensification of 2.13 [157, 183] and values of free path length (Section 4.3) given in cm.

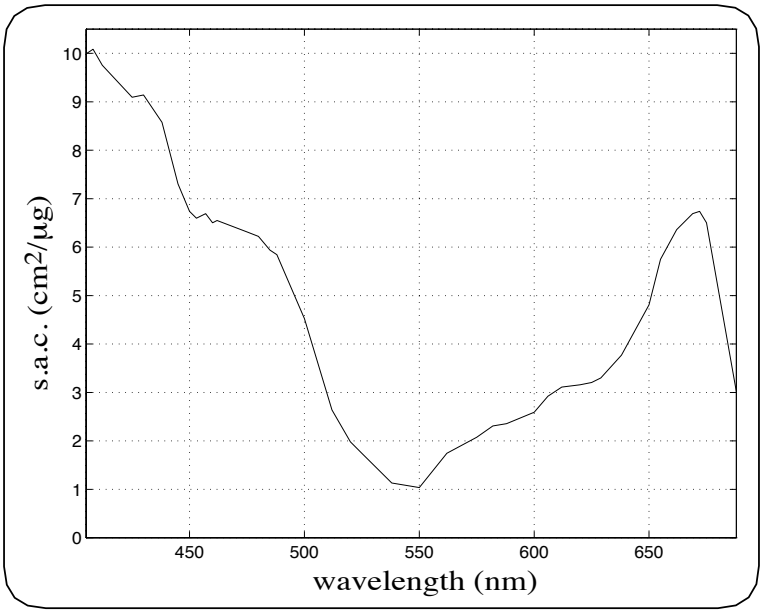

Figure 4.4: Absorption spectra of chlorophylls $a + b$ used in the testing of the *ABM*.

4.6 Results and Discussion

The spectrophotometric curves obtained using the *ABM* are qualitatively in agreement with the actual measured values (Figure 4.5). The quantitative discrepancies may be due in part to the fact that we did not consider the carotenoids (which account for 20% of the total amount of pigments) and that we did not separate different chlorophyll pigments. Moreover, the underestimation of surface microdetails, and shadowing and masking effects [76] may also contribute.

Nevertheless, it shall be noted that the curves presented in Figure 4.5 represent mainly qualitative comparisons between modeled and measured values, since some parameters used in the simulation, such as the refractive indexes and the oblateness, correspond to average values published in the literature. Besides, as stated by Salisbury and Ross [187], the exact positions of the absorption peaks depend on the solvents in which the pigments are dissolved, and one can expect small shifts considering *in vivo* values. For these reasons one must account for possible variations between the average values used by the *ABM* and the exact values regarding the leaf specimen used in the actual measurements performed by *LOPEX*.

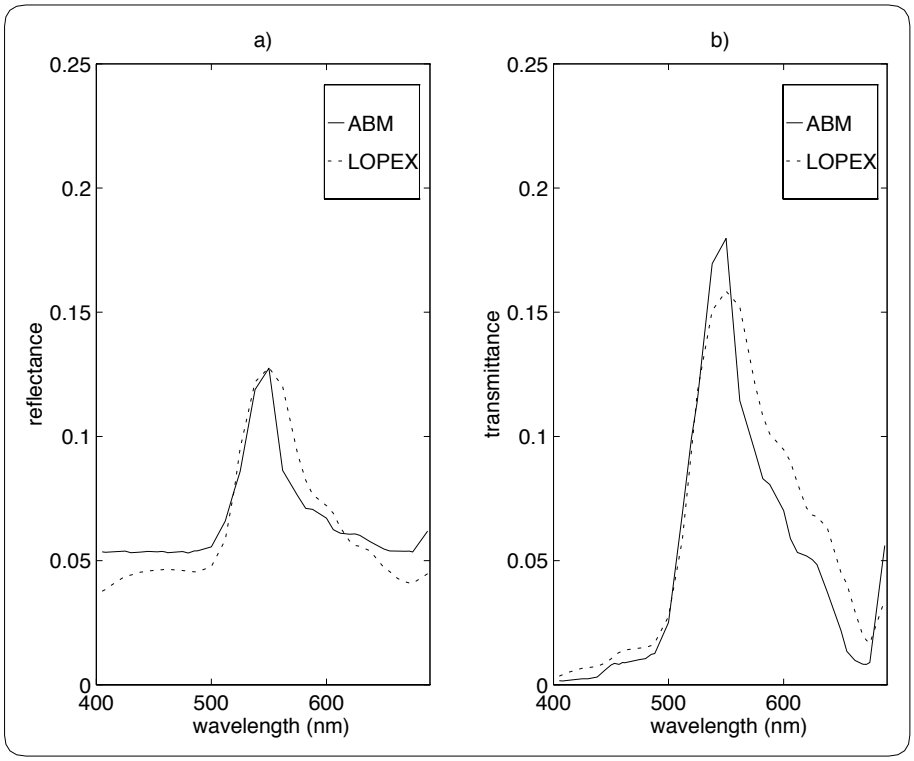

Figure 4.5: Comparison of spectral curves of a soybean leaf computed with the *ABM* with measured spectral curves provided by *LOPEX*, for an angle of incidence of 8° and 10^6 rays. a) Absolute spectral reflectance. b) Absolute spectral transmittance.

As pointed out by Vogelmann [229], the asymmetry in a bifacial leaf anatomy gives different reflectance and transmittance readings for its adaxial (front) and abaxial (back) surfaces. Spectrophotometric curves obtained experimentally by Wooley [240] showed greater reflectance for the back than for the front in the visible region of the spectrum. Figure 4.6a shows that the *ABM* can capture this aspect of foliar tissues optics, which is responsible for the fact that the backs of bifacial leaves appear pale to the eye. The experiments by Wooley also showed greater transmittance when the leaf's back is toward the light, although in the visible region the differences of magnitude of the transmittances are not so high as the differences of magnitude of the reflectances. Figure 4.6b shows that the *ABM* can capture these characteristics of the transmittance curves as well.

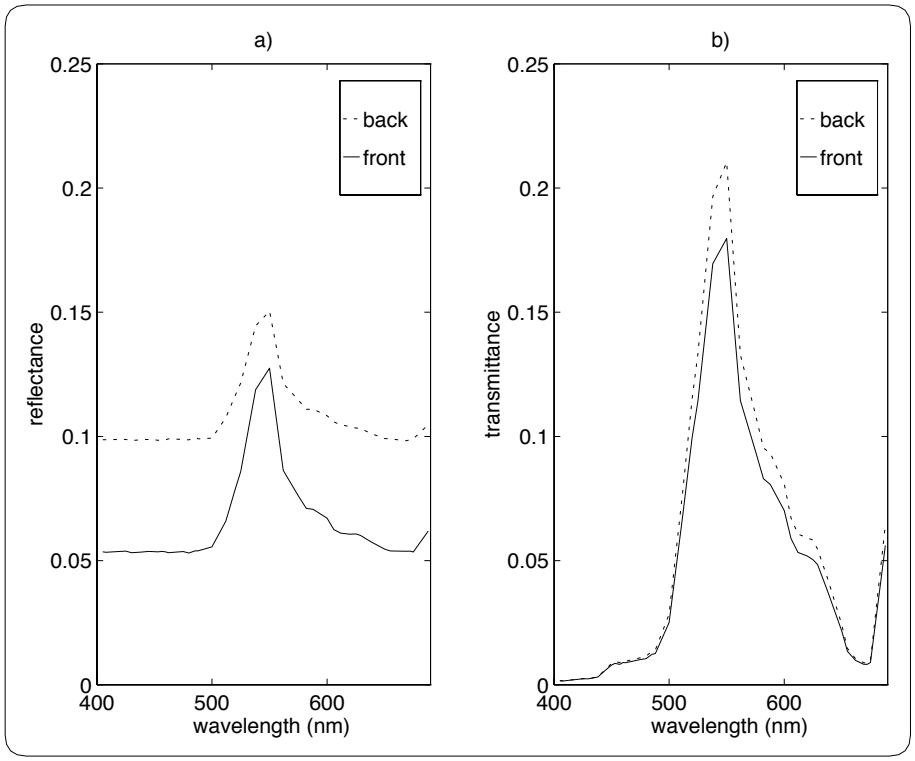

Figure 4.6: Spectrophotometric curves of a soybean leaf obtained using the *ABM* and considering its front surface (adaxial surface) towards the light and its back surface (abaxial surface) towards the light, for an angle of incidence of 8° and 10^6 rays. a) Absolute spectral reflectance. b) Absolute spectral transmittance.

Figure 4.7 (upper part) shows that the simulated BRDF exhibits an angular dependency on the incident angle intermediate to that expected of diffuse and specular reflectors, which corresponds to the characteristics of the real BRDF of leaves [97]. It also shows (lower part) that the simulated BTDF has a near-Lambertian distribution, which is also a characteristic of the real BTDF of leaves [97]. Moreover, the curves in Figure 4.7 agree with the experimental BDF curves of soybean leaves published by Breece and Holmes [40] and Wooley [240].

Even though a performance analysis should in general not be based only on counting the number of interactions, since it does not account for differing amounts and types of work performed on each interaction, we believe that it may be illustrative

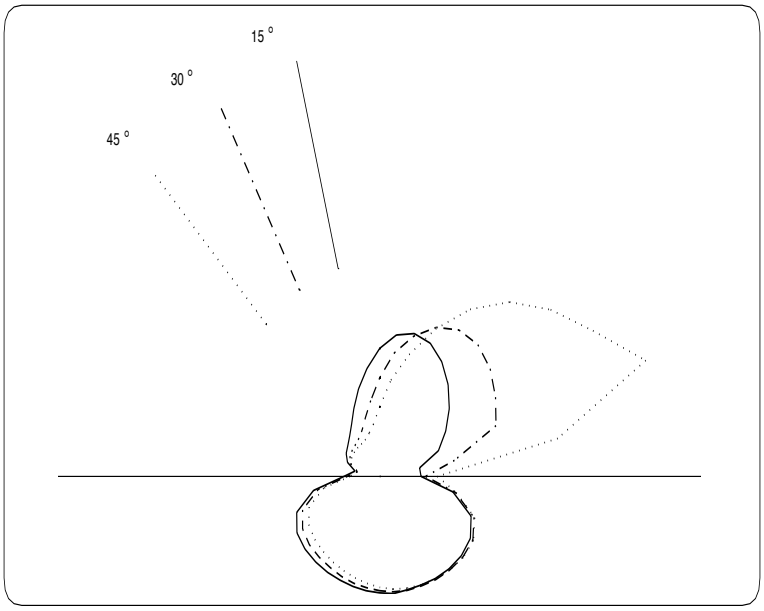

Figure 4.7: BDF of a soybean leaf at a wavelength of $550nm$ (which corresponds approximately to the reflectance and transmittance peaks), for angles of incidence of 15°, 30° and 45°, in the plane given by the incidence direction and the normal of the specimen, 10^7 rays, and the collector sphere divided into 20 patches along its latitude and 40 patches along its longitude.

of the behavior of the *ABM*. Figure 4.8 shows that, despite the stochastic nature of the *ABM* and the number of interfaces involved, the number of interactions (state transitions) is usually low. This aspect suggests that the cost/accuracy ratio of the *ABM* is compatible with physically-based rendering requirements. Furthermore, in applications where the accuracy requirements for individual elements, such as a leaf or a petal, are not very high, *e.g.* the rendering of a tree with hundreds of leaves, the model may be reduced to its most important states, namely states 1 and 2.

The *ABM* is aimed at the visible region of the spectrum, but it may be adapted to be used in the applications regarding other regions, *e.g.* in the near-infrared and infrared applications [185]. For example, in the infrared region (beyond $1300nm$ to $2500nm$) the absorption is controlled by the water content. Then, for this region, one would have to consider the absorption coefficients of water [173], instead of the

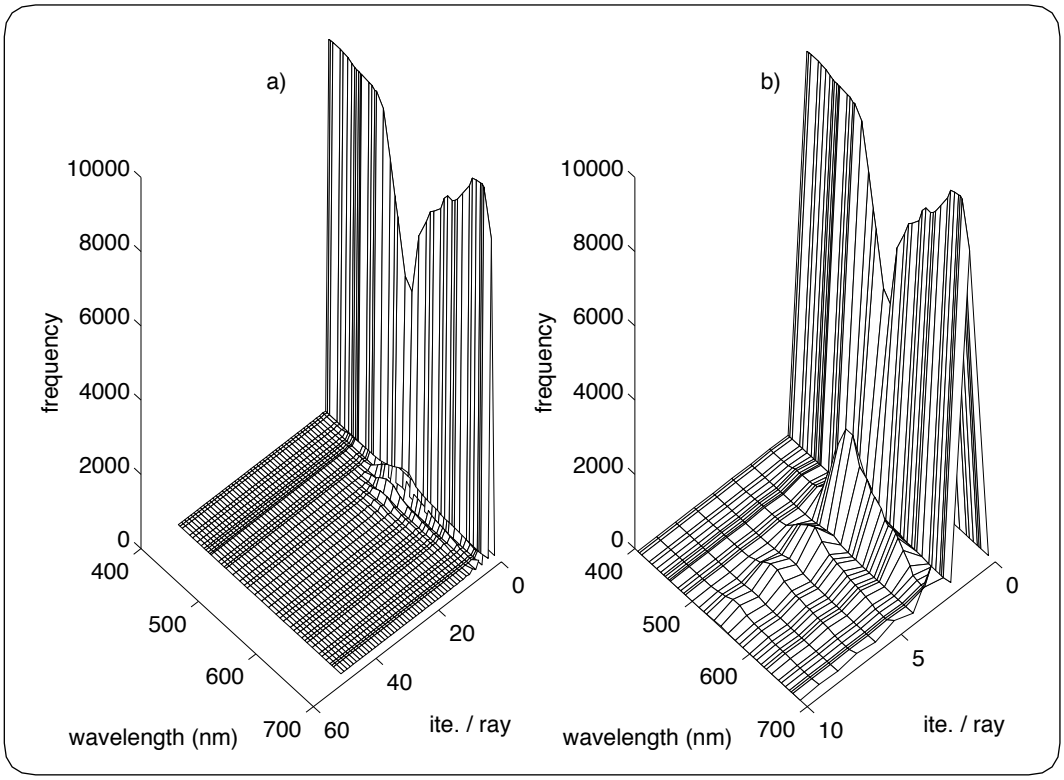

Figure 4.8: Number of interactions per ray and per wavelength for 10^4 rays. a) Graph for zero to fifty interactions per ray, b) Zoom in of the region with high frequency of interactions per ray.

absorption coefficients of pigments, in the absorption testing performed in state 2.

The near-infrared region ($700nm$ to $1300nm$), however, is characterized by a lack of absorption and high values for reflectance and transmittance [223]. For this region weights (initially unit) may be associated with the rays at each interaction according to the Fresnel coefficient. In this case, instead of tracking just one ray after each interaction, we would have to track two. This would represent an additional implementation overhead, but it would reduce considerably the number of rays needed to satisfy the precision criteria. Moreover, it would also allow the use of a fractional method, such as the *Russian roulette*, to stop the rays probabilistically.

The *Russian roulette*, a fractional method named by Kahn [129] after the lethal game, has been used in rendering applications with this purpose [86]. In order to cut

down the number of particles (rays), we select one whose weight is below some threshold due to multiple bounces during the *random walk*. Then we use some probability $\mathcal{P}$ [62] to discard it (*i.e.* we say the particle is absorbed) or to allow it to continue. In the later case its weight is multiplied by $(1-\mathcal{P})^{-1}$. This procedure can be repeated (with the same or different values for $\mathcal{P}$) until the number of particles is reduced to a manageable size [107].

4.7 Summary

In this chapter we have presented a plausible and biologically-based BDF model for plant tissue, the *ABM*. This model is controlled by a small number of biologically meaningful parameters and presents a cost/accuracy ratio that suffices for physically-based rendering applications. We examine this issue in more detail in the next chapters. Moreover, the *ABM* can easily fit into global illumination frameworks due to its simple formulation based on standard Monte Carlo methods.

Differently from other BDF models available in the computer graphics literature, the *ABM* does not rely on inputed values of reflectance and transmittance. Indeed, these values are computed by the *ABM* itself. This aspect and its design specifically tailored to foliar tissues make its comparison with other models used in rendering applications troublesome. The comparison of images generated using the *ABM* with photographs of real plants would not be appropriate either, since it could be biased by errors introduced in other stages of the rendering pipeline. For these reasons, we have tested it as a separate unit of the rendering pipeline through comparisons with real measured data. In order to perform these comparisons we have implemented virtual measurement devices.

Our simulations show that the spectral curves obtained using the *ABM* are in good agreement with measured values. The surface reflectance readings provided by the *ABM* may be improved through the simulation of wax configurations and incorporation of shadowing and masking effects. Moreover, although there is little experimental data on the anisotropy of plant tissues, the incorporation of factors affecting it, such as the presence of venation and hair, could further improved the spectral curves provided by the *ABM*.

It is important to note that we did not perform an *ad hoc* adjustment of the parameters to obtain the spectral curves through the *ABM*. In other words, we have used values as close as possible to the real values associated with the specimen tested. These values were obtained from the remote sensing and botany literature. We have also attempted to reproduce, as faithfully as possible, the real experimental conditions in which the actual measured curves were obtained.

In this chapter we have also pointed out suggestions to improve the efficiency of the *ABM* and to extend its range of applications. In the next chapters we examine alternatives to reduce the computational costs involved in the simulations of light interaction with foliar tissues. We also compare the images obtained using the different proposed approaches, including the *ABM*.

Chapter 5
A Nondeterministic Reconstruction Approach

The in-line application of a BDF model, such as the *ABM*, may add a significant computational overhead to the rendering process. For some applications, *e.g.* a scene with many leaves with the same biophysical characteristics, it is more convenient to measure the BDF, store its values in memory and reconstruct it during rendering saving valuable time. As mentioned in the previous chapter, the BDF of a surface can be measured using a measurement device known as a goniophotometer, or through a computer simulation of such a device, which is called a virtual goniophotometer. In both cases, a large number of measurements is usually required to represent the BDF accurately.

Fortunately, we can often exploit some characteristics of materials and BDF models used in rendering applications to reduce the number of measurements, without decreasing the accuracy of the BDF curves provided by these models significantly. An example is given by plant leaves. As mentioned in Chapters 3 and 4, currently there is not enough information to support the design of a biologically-based anisotropic BDF model for these materials. Therefore, since in practice we are working with an isotropic BDF model, we need to use only one directional input parameter, which reduces the number of measurements considerably.

In this chapter an approach to reconstruct isotropic BDFs is presented. It consists of pre-computing the BDF for a given material using an appropriate BDF model, and storing its corresponding spectral and directional values. Afterwards, these values are used to reconstruct the BDF during rendering. Instead of applying a deterministic[1]

[1]In this context we use the terms deterministic and nondeterministic according to how the different approaches select the PDF used to generate the outgoing distribution of rays associated with a given BDF.

approach involving mathematical tools like wavelets [51, 139, 140, 141], spherical harmonics [44, 235] or interpolation [93], we use a nondeterministic algorithm, based on standard Monte Carlo techniques [107], to reconstruct the BDF. Our goal is to preserve the behavior and the accuracy of the original BDF provided by the BDF model while reducing the costs of its in-line application.

Although the isotropy requirement may seem restrictive, it is appropriate for the rendering of plants due to the circumstances mentioned earlier and has the advantage of reducing storage and computational time considerably. In fact, DeYoung and Fournier [67] have proposed techniques to transform anisotropic tabulated BDFs to match the isotropy property in order to benefit from these computational savings.

We use the virtual goniophotometer described in Section 4.5.1 to perform the BDF measurements, and to test the proposed reconstruction approach through comparisons of spectral curves obtained using a BDF model in-line, the *ABM*, with reconstructed spectral curves. Although the proposed reconstruction approach works in the same way for any wavelength channel, a single channel is assumed in the next section in order to simplify the presentation.

The sections of this chapter are organized as follows:

- Section 5.1 introduces the reconstruction approach.
- Section 5.2 presents the testing parameters and procedures.
- Section 5.3 discusses the results and performance issues.
- Section 5.4 summarizes the contents of this chapter.

5.1 The Reconstruction Approach

5.1.1 Data Generation

The radiance detectors in the virtual goniophotometer used in our investigation are represented by the patches of a collector sphere placed around a specimen. This sphere is divided into n_a patches along its latitudinal axis and n_b patches along its longitudinal axis. The parameter space of the BDF function is represented by the input directional parameter space, S_i, and the output directional parameter space, S_o.

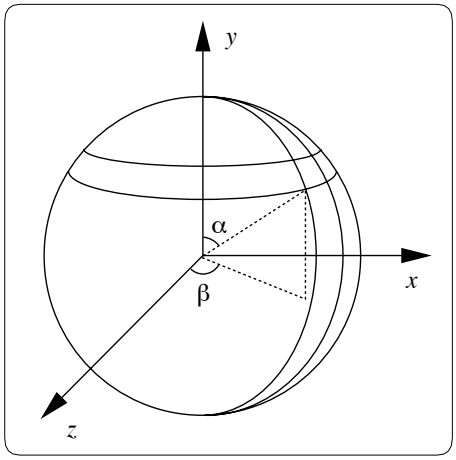

Figure 5.1: Collector sphere and virtual goniophotometer coordinate system.

The first is given by the angle of incidence θ_i ($[0, \pi]$) of an incoming ray with respect to the normal of the specimen. The second is given by the polar angle α ($[0, \pi]$) and the azimuthal angle β ($[0, 2\pi]$), which define the coordinates of the sphere patches with respect to the virtual goniophotometer coordinate system (Figure 5.1).

During the data collection we divide S_i into n_t regular intervals which correspond to selected angles of incidence. For each interval we shoot N rays towards a specimen. The origin of each ray is randomly chosen within the interval. After interacting with the specimen, the rays that are not absorbed hit the collector sphere. The total number of rays that were not absorbed are stored in a vector V of size $[n_t]$. The number of hits on each patch is stored in a temporary array *temp* of size $[n_a \times n_b]$.

Each entry in the array *temp* corresponds to a patch, and it is defined by its row number, which is given by the patch's latitudinal coordinate (α), and column number, which is given by the patch's longitudinal coordinate (β). Afterwards, *temp* is linearized and it becomes a row in a table T of size $[n_t \times n_a n_b]$ (Figure 5.2a), and each patch becomes associated with a number, j, that corresponds to its position in a row.

In order to obtain the number of hits received by a certain patch, later on, during the reconstruction process, we need to access the table entry T_{ij} where i is given by the angle of incidence θ_i and j is given by the patch's number. Before it can be used

in the reconstruction process, however, the table T has to be pre-processed. This pre-processing consists of sorting each row in ascending order, such that each element will contain the sum of the hits of all previous elements plus its own number of hits (Figure 5.2b). Patches that do not receive any hit are automatically eliminated during this procedure, reducing the storage costs. Since the order of the patches is changed during the sorting, the patches' coordinates, α and β, are also stored in a table U of size $[2 \times n_t \times n_a n_b]$ (Figure 5.2c). The time required for the pre-processing operation on T is negligible compared to the data collection or rendering times.

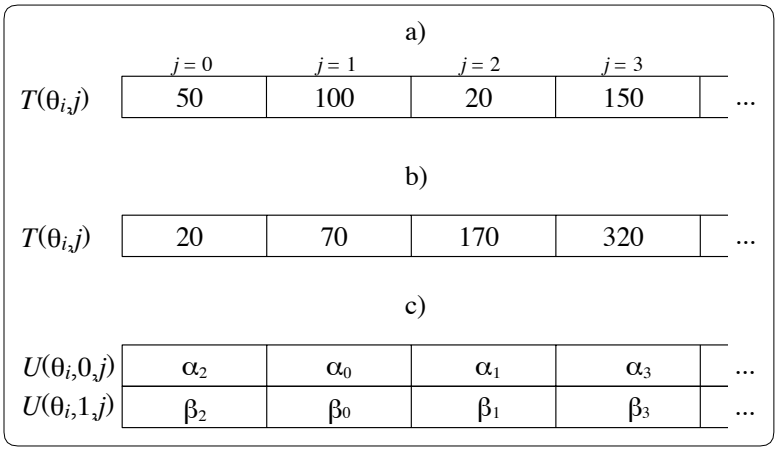

Figure 5.2: Example of pre-processing on T. a) Original row. b) After sorting. c) The corresponding entries in U.

5.1.2 Reconstruction

Initially, when a ray hits a specimen during the rendering process, we determine the angle of incidence:

$$\theta_i = \arccos\left(\frac{\vec{n} \cdot \vec{r}}{|\vec{n}||\vec{r}|}\right) \tag{5.1}$$

where:

$\vec{n}$ = specimen's normal,

$\vec{r}$ = direction of the incident ray.

Afterwards we generate a uniformly distributed random number $\xi \in [0, N]$. If this random number ξ is greater than the number of non-absorbed rays stored in $V(\theta_i)$ we proceed to the next ray, saving valuable time. Otherwise, we proceed to the kernel of the algorithm, where the reconstruction of the BDF takes place.

In the kernel we perform a search in the array T to determine the smallest $T(\theta_i, j) \geq \xi$. In this way the elements associated with the largest numbers of hits will have the highest probabilities of being selected. Once we determine the row element $T(\theta_i, j)$, we use j to obtain the patch's coordinates α and β stored in $U(\theta_i, 0, j)$ and $U(\theta_i, 1, j)$ respectively. Using these coordinates the direction of the outgoing ray is given by the angles α_m and β_m:

$$(\alpha_m, \beta_m) = \left(\alpha + \frac{\pi}{n_a}\zeta_1, \beta + \frac{2\pi}{n_b}\zeta_2\right) \tag{5.2}$$

where:

ζ_1 and ζ_2 = uniformly distributed random numbers $\in [0, 1]$.

These angles, in turn, correspond to angular displacements with respect to the y-axis and the z-axis of the virtual goniophotometer coordinate system (Figure 5.1). Considering that the incident rays are generated during the rendering using a different coordinate system, namely the coordinate system associated with a virtual camera, the generation of the outgoing ray has to be preceded by an appropriate coordinate system transformation [75].

5.1.3 Searching

The searching process mentioned above can be performed using binary search, which has a worst case cost of $\lfloor log_2(n_a n_b) \rfloor + 1$ comparisons. However, initial experiments showed that this strategy was not fast enough for our application. We, therefore, decided to use instead a hashing technique in which the column indexes of T are mapped to another table H of size $[n_t \times n_r]$ (Figure 5.3a).

For an entry of T represented by $T(i, j)$ the mapping is performed using the following approach:

$$H(i, \nu) = j \tag{5.3}$$

where the index ν is given by:

$$\nu = \lfloor T(i,j)\, n_r\, (V(\theta_i))^{-1} \rfloor \tag{5.4}$$

Notice that we may have loss of information related to the mapping from T to H. This loss occurs when several column indexes j of T are mapped to the same entry ν of H. In this case only the last value mapped is kept in H, and the others are lost. In order to minimize the chances for that to happen, the dimension n_r has to be greater than $n_a n_b$. However, the greater the value assigned for n_r, the larger the storage requirements for H. A reasonable trade-off choice for n_r is discussed in Section 5.3.

Before the mapping of the elements of T to H, the entries of H are initialized with zeroes (except for the first entry, which can contain zero as shown in Figure 5.3c). After the mapping some of them may still have zero as their stored value. In order to eliminate the zeroes, we post-process H by replacing the zero entries in each row by the value of the previous nonzero entry, starting from the row's rightmost element (Figures 5.3b and 5.3c). The time required for the post-processing operation on H is also negligible compared to data collection or rendering times.

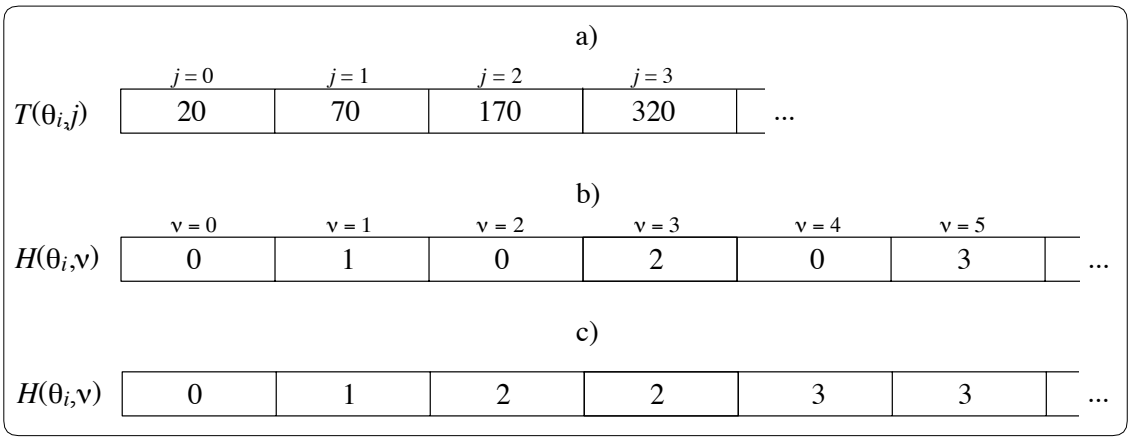

Figure 5.3: a) Example of mapping from T to H. a) T's row. b) The corresponding row entries in H. c) After post-processing.

During the reconstruction, we use table H instead of table T, to access the coordinates of a patch associated with ξ. The patch's position in U is stored in $H(\theta_i, \nu)$, and the index ν is computed using the following expression:

$$\nu = \lfloor \xi \, n_r \, (V(\theta_i))^{-1} \rfloor \tag{5.5}$$

5.2 Testing Parameters and Procedures

The data used in our testing experiments corresponds to a soybean leaf's BDF, which was computed using the *ABM* (Chapter 3). The collector sphere used by the virtual goniophotometer was divided into 20 patches along its latitudinal ($n_a = 20$) and 40 patches along its longitudinal axis ($n_b = 40$). The data collection was performed in intervals of 1° for S_i ($n_t = 180$). The testing measurements were made using 10^8 incident rays per interval, and considering the plane given by the incident light and the specimen's normal.

Lilley *et al.* [148] mention that to provide the correct color in high quality computer graphics, the CIE XYZ values should be converted to the RGB color space of a monitor using the SMPTE[2] monitor chromaticity coordinates. In fact, many monitors used in the current workstations use these coordinates. Thus, we decided to sample the absorption spectra of chlorophyll (Figure 4.4) in the dominant wavelengths corresponding to these coordinates (Table 5.1) in order to generate the curves and the images presented in the next section. Note that the dominant wavelength regarding the green channel corresponds approximately to the wavelength for which the absorption of light by chlorophyll is minimum (Figure 4.4).

	x	y	wavelength
Red	0.630	0.340	$608nm$
Green	0.310	0.595	$551nm$
Blue	0.155	0.070	$455nm$
white (D65)	0.313	0.329	

Table 5.1: Chromaticity coordinates and wavelength values.

[2]Society of Motion Picture and Television Engineers.

The spectral curves of reflectance and transmittance used to represent veins' chromatic attributes in the images presented in the next section were also obtained using the *ABM*. In this case, however, the absorption spectra of chlorophyll (Figure 4.4) was scaled by a factor of 0.1 according to biological data provided by McClendon and Fukshansky [157].

In order to obtain the geometrical descriptions of soybean leaves, used in the images presented in the next section, we grew a soybean plant. After the plant was fully maturated, we removed a few leaves and used a digital scanner to obtain black and white texture maps regarding their contour (contour map) and venation system (venation map). This positional information is mapped to polygons to create the leaves, using a technique similar to the one applied by Bloomenthal [31] and Hanrahan and Krueger [109]. For areas of the contour map outside the leaf contour the polygon is considered transparent and the leaf is not visible. Similarly, the veins are not visible for transparent areas of the venation map.

The root-mean-square (RMS) errors [86], ϵ, of the difference images presented in the next section were computed from normalized pixel values (scaled to $[0, 1]$), using the following expression:

$$\epsilon = \sqrt{\frac{1}{n} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} (d(x_i, y_j))^2} \tag{5.6}$$

where:

n	=	total number of pixels,
n_x	=	number of pixels in the x axis,
n_y	=	number of pixels in the y axis,
$d(x_i, y_j)$	=	difference regarding the pixels with coordinates x_i and y_j.

5.3 Results and Discussion

Figures 5.4, 5.5 and 5.6 show that the BDF curves for a front lit specimen obtained through the reconstruction approach agree closely with the measured curves provided by the BDF model. The curves for different angles of incidence, a larger number of incident rays and a back lit specimen show a similar agreement. The quantitative discrepancies shown in these curves have a small magnitude, and suggest that the se-

lected sampling resolution determined by the choice of n_t, n_a, and n_b was appropriate for this application. Increasing their values may further improve the accuracy of the results, specially for a BDF that presents discontinuities or high frequency effects.

Moreover, for rendering purposes, these quantitative discrepancies can be considered negligible, as we can see in the images presented in Figure 5.7. These images were generated using a modified path tracing rendering algorithm [198]. The corresponding difference image presents a root-mean-square (RMS) error for the three RGB channels of 0.0123, 0.0131 and 0.0132 respectively, which are fairly low considering that the images were generated through a stochastic method.

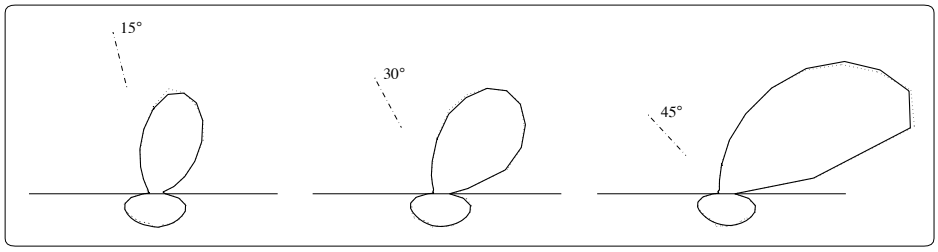

Figure 5.4: Comparison of BDF curves at wavelength of $608nm$ (red region) for a front lit specimen obtained using the *ABM* (solid lines) with curves obtained using the reconstruction approach (dotted lines).

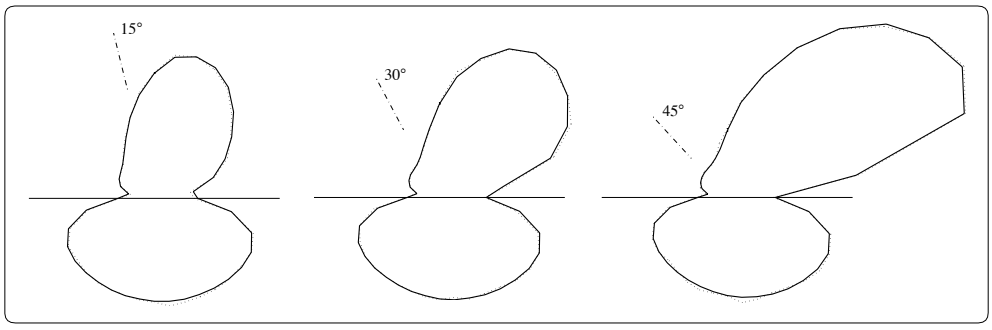

Figure 5.5: Comparison of BDF curves at wavelength of $551nm$ (green region) for a front lit specimen obtained using the *ABM* (solid lines) with curves obtained using the reconstruction approach (dotted lines).

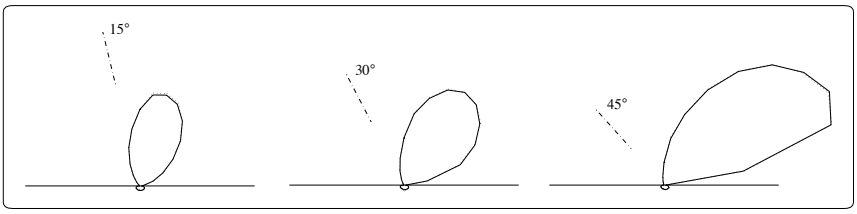

Figure 5.6: Comparison of BDF curves at wavelength of $455nm$ (blue region) for a front lit specimen obtained using the *ABM* (solid lines) with curves obtained using the reconstruction approach (dotted lines).

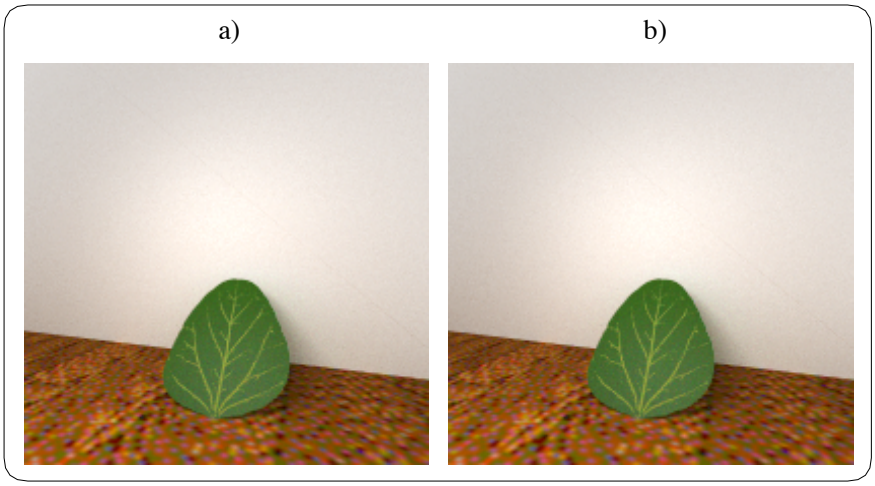

Figure 5.7: Comparison of images obtained using: a) the *ABM* in-line, and b) the nondeterministic reconstruction approach.

The speed up gain of the rendering process using the proposed approach to reconstruct the leaves' BDF over the process using the BDF model in-line was 51%. The overhead of pre-computing the BDFs is minimized by the fact that, for a given material, this operation must be performed only once. Moreover, the resulting tables can be used several times, not only by the proposed nondeterministic approach, but also by deterministic ones [44, 51, 139, 140, 141, 235].

In order to guarantee that the loss of information was not a relevant issue in our experiments, we selected a value for n_r equal to 3560. Then, the data structures (vector V and tables U and H) used by the nondeterministic reconstruction approach

occupied 3.1 Mb of memory space for each wavelength channel. This figure represents an acceptable trade-off between storage requirements and accuracy, specially considering the size of the memories available nowadays and their decreasing cost. However, for applications involving a large number of BDF representations, *e.g.* a scene with several species of plants, the proposed reconstruction approach may become inefficient, since if the data structures cannot be stored in memory the performance will suffer from excessive paging.

As pointed out by Lalonde and Fournier [140], Monte Carlo path tracing rendering techniques [131, 127, 137, 198], including multipass methods that use path tracing as one of the passes [50], can benefit dramatically from accurate generation of reflected and transmitted directions according to the distribution of the BDF. These benefits arise due to the reduction of the variance associated with the process. Further investigation is required to establish the full range of applications and the performance of the proposed approach for different materials and graphics settings. Nonetheless, its stochastic formulation, based on the application of probability techniques, enables its straightforward incorporation into Monte Carlo based rendering systems. Moreover, due to its "ray" nature the generation of reflected and transmitted directions is more intuitive than the one performed by deterministic approaches.

It is important to notice that the main contribution presented in this chapter corresponds to the reconstruction process itself, since the scheme presented in this chapter to store the BDF representation of a given specimen has been somewhat explored by previous works in the computer graphics literature. For example, Cabral *et al.* [44] used a discrete set of buckets that cover the hemisphere above a specimen to represent its BRDF, and Gondek *et al.* [93] used a geodesic sphere around a specimen to represent its BRDF and BTDF.

Gondek *et al.* also used an adaptive approach to subdivide the facets of the geodesic sphere in order to reduce the storage costs related to the BDF representation. Although intuitively this approach seems to be more efficient than the regular grid of patches over a sphere described in this chapter, it requires the implementation of trees and efficient routines for tree traversal. This aspect makes its implementation more complex and may increase the computational overhead substantially.

Recall that during the pre-processing of the storage data structure, described in Section 5.1.1, the patches that do not receive any hit are eliminated, implicitly discarding elements of the grid that are not relevant for the BDF representation of a given specimen. In other words, the storage scheme presented in this chapter, based on a regular grid of patches, also allows an adaptive reduction of the memory costs depending on the BDF representation at hand.

Unfortunately, experimental data regarding performance and storage requirements concerning the application of deterministic approaches is usually not readily available in the literature. This aspect makes the comparisons of the computational costs involving the application of the proposed nondeterministic reconstruction approach and the deterministic ones more difficult. A comprehensive assessment of their advantages and drawbacks would require a separate investigation, which, in turn, would likely involve the implementation of the different reconstruction techniques and their application to different test cases. We discuss this issue in more detail in Chapter 9.

5.4 Summary

We have presented a nondeterministic approach, based on standard Monte Carlo techniques, for accessing arbitrary isotropic reflectance and transmittance functions. The results obtained so far show that the reconstructed BDF curves present a high degree of accuracy and the proposed reconstruction approach may provide a noticeable gain in performance while keeping computational costs within reasonable limits.

As mentioned in Chapter 3, we might increase the accuracy of the spectral curves provided by the *ABM*, *e.g.* by adding features to account for masking and shadowing effects. In this case the use of the reconstruction approach described in this chapter could result in higher speed ups. But would the increase in image quality be sufficiently noticeable to justify the increase on the model's complexity and the risk of losing it during the reconstruction process? The answer to this question is not simple, and involves the evaluation of several aspects related to the rendering process which are beyond the scope of this thesis.

At this point of our research, instead of refining the *ABM* and relying solely on reconstruction approaches to minimize its computational overheads, we decided to

follow a different path through the development of an alternative model. In this alternative model, presented in the next chapter, the processes of light interaction with foliar tissues are simplified to a level in which the quality of the images is not significantly affected, and the gains in performance are sufficiently high to justify the reduction in accuracy.

Chapter 6
A Simplified Model for Light Interaction with Plant Tissue

The accuracy of simulations of light interaction with plants depends on the reflectance and transmittance models for foliar tissues. The use of physically-based models is important for ensuring that rendering methods used in these simulations do not violate the laws of physics [145]. Moreover, these models have to be biologically-based in order to appropriately account for the natural processes involved in these simulations. Simplicity is also an important requirement for a reflectance model, since, as pointed out by Ward [233], a model may otherwise become computationally impractical. Thus, it is necessary to design practical reflectance and transmittance models that allow us to render these materials fast without undermining the image quality.

In this chapter we present a simplified model for light interaction with plant tissue, which aims to provide a balance between two seemingly conflicting goals, namely accuracy and efficiency. This model accounts for the three components of plant tissues' BDF (surface reflectance, subsurface reflectance and transmittance), and uses pre-computed reflectances and transmittances as scale factors in a stochastic simulation of the scattering profile of these tissues. The use of these scale factors replaces the time consuming random walk process used by the *ABM* to simulate the randomization and the absorption of light within the foliar tissues. Moreover, this approach reduces the number of rays needed to achieve a desired accuracy level in the results.

These scale factors are computed off-line using the *ABM*, since the spectral curves provided by this model have already been compared with experimental data of real foliar specimens, showing a good agreement (Section 4.6). Although the proposed simplified model is oriented to leaves, like the *ABM* it can easily be extended to

other plant surfaces like petals and stems. We compare the results obtained using the proposed simplified model with results obtained applying the *ABM* in-line.

The remainder of this chapter is organized as follows:

- Section 6.1 presents the proposed simplified model for light interaction with plant tissue.
- Section 6.2 provides an overview of aspects related to the implementation of the simplified model and a summary of its parameters.
- Section 6.3 describes the testing parameters and procedures.
- Section 6.4 examines the testing results, focusing on comparisons with the *ABM*.
- Finally, Section 6.5 closes this chapter with a summary of the main characteristics of the proposed simplified model.

6.1 The Simplified Model

In the simplified model light propagation is also described in terms of ray optics, and the interaction of light with the foliar tissues is described in terms of their SPF (Section 2.2). As pointed out by Glassner [86], the BDF (or, in this case, the SPF) is a difficult function to work with due to its dependence on several variables. Fortunately, we can make some simplifying assumptions about the foliar tissues that give us a more computationally convenient expression to manipulate.

First, we assume that the physical properties describing the light propagation are identical everywhere within the foliar tissues, *i.e.* they can be considered as homogeneous interacting media [96]. This assumption allows us to leave out the positional argument. Second, since the anisotropy of plant leaves is considered to be associated with their venation system (Section 3.2), and considering that the biological data regarding these systems is scarce to support the design of a biologically-based anisotropic reflectance model for these materials, we also assume that they are isotropic. This assumption allows us to work with only one parameter for the incidence direction ψ_i, which is given by the angle of incidence, θ_i, of an incident ray with respect to the normal of a leaf.

The proposed simplified model takes into account the three components of the SPF of a plant tissue, namely surface reflectance, subsurface reflectance and transmittance. The contribution from each of these components is scaled using the respective values of absolute spectral surface reflectance ($\rho_s(\theta_i, \lambda)$), absolute spectral subsurface reflectance ($\rho_d(\theta_i, \lambda)$) and absolute spectral transmittance ($\tau(\theta_i, \lambda)$), as sketched in Figure 6.1. These spectral values are computed off-line using the ABM[1] and stored in a table.

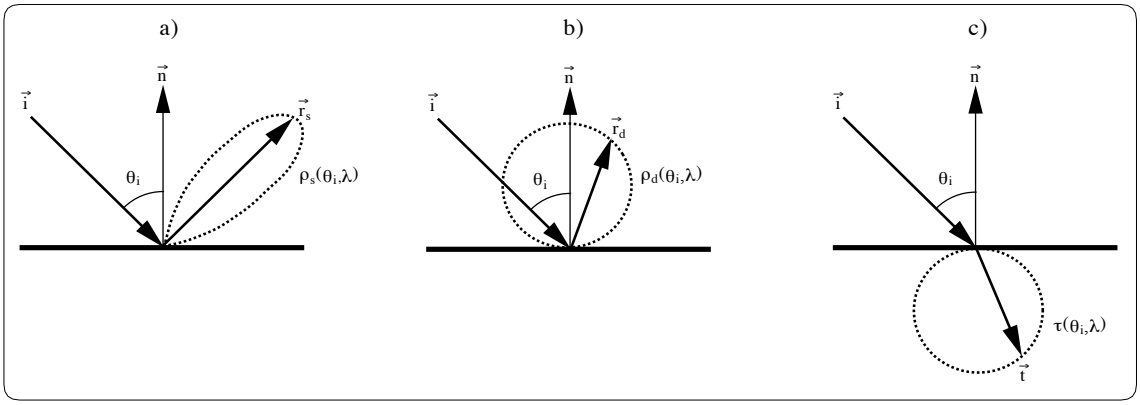

Figure 6.1: Scattering distribution performed by the simplified model to represent the three components of a foliar specimen's SPF: a) surface reflectance b) subsurface reflectance and c) transmittance.

The parameter space of a foliar specimen's SPF is represented by the input directional parameter space and the output directional parameter space. The first, considering the assumption regarding isotropy made earlier, is given by the angle of incidence θ_i ($[0, \pi]$), and the second is given by the azimuthal angle α ($[0, 2\pi]$) and the polar angle β ($[0, \pi]$). The domain chosen for θ_i accounts for differences in the reflectance and transmittance curves of foliar specimens that differ markedly in the structure of their two sides. For specimens that do not present this characteristic (Section 3.1), or for applications that assume identical optical properties for both

[1]The outgoing rays generated by the *ABM* have tags which allow the computation of separated values for the surface reflectance, subsurface reflectance and transmittance (Figures 6.2 and 6.3).

sides [155], the domain of θ_i can be narrowed to $[0, \frac{\pi}{2}]$.

Initially, to account for the surface component of a foliar specimen's SPF for a given incident ray, we obtain the corresponding reflected ray using the law of reflection (Section 2.1). Then, to simulate the effects of the epidermal cells' shape on the reflected rays at the air→epidermal cells interface, we perturb the reflected rays using a warping function (Equation 4.1) based on an exponentiated cosine distribution (Appendix A) employed by the *ABM*.

As mentioned earlier, when light passes to the internal foliar tissues its direction of travel is randomized and it becomes diffuse. This randomization of the incident rays results in a near-Lambertian distribution for the subsurface reflectance of a foliar specimen, and a near-Lambertian distribution for its transmittance. In order to simulate the distribution of rays regarding these two components, we perturb the normal of the foliar specimen instead of the incident ray. The orientation of the normal used in the perturbation depends on the incidence geometry. If the incident ray hits the foliar specimen's front (adaxial epidermis) we use the normal for the subsurface reflectance component and its opposite vector for the transmittance component. Otherwise, we use the normal and its opposite vector the other way around. For these perturbations we use a warping function (Equation 4.2) based on a cosine distribution (Appendix A), also employed by the *ABM*.

6.2 Implementation Overview and Summary of Parameters

The simplified model was also implemented *C++* in order to take advantage of the object oriented features of this programming language. This model, like the *ABM*, can be incorporated as a self-contained class in a graphics library, such as the *ggLibrary* [202], and be used by different global rendering methods.

Since the values for spectral reflectance and transmittance are pre-computed, the set of model parameters is reduced to the oblateness of the epidermal cells, *ob*, used in the perturbations of the reflected rays, and the angle of incidence, θ_i, used to access the table of absolute spectral reflectances and transmittances. The angle of incidence is computed using Equation 5.1.

The images presented in Section 6.4 were generated using a modified version of Kajiya's path tracing [131, 198]. For the computation of the direct light contribution we selected the scale factors $\rho_d(\theta_i, \lambda)$ and $\tau(\theta_i, \lambda)$ according to the position of the light source with respect to the foliar specimen. This selection was made using the angle of the shadow ray with respect to the specimen's normal, θ_s, and applying the following criteria:

– for light coming from above $0^\circ < \theta_s \leq 90^\circ$
$$\begin{cases} \frac{\rho_d(\theta_i,\lambda)}{\pi} & \text{if } 0^\circ < \theta_i \leq 90^\circ \\ \\ \frac{\tau(\theta_i,\lambda)}{\pi} & \text{if } 90^\circ < \theta_i < 180^\circ \end{cases}$$

– for light coming from below $90^\circ < \theta_s < 180^\circ$
$$\begin{cases} \frac{\tau(\theta_i,\lambda)}{\pi} & \text{if } 0^\circ < \theta_i \leq 90^\circ \\ \\ \frac{\rho_d(\theta_i,\lambda)}{\pi} & \text{if } 90^\circ < \theta_i < 180^\circ \end{cases}$$

While the *ABM* uses an explicit mechanism to simulate the absorption of light, the simplified model relies on the absorption probabilities implicitly associated with the scaling factors. In order to perform fair comparisons between these two models (Section 6.4), we use an adaptive tree-depth control [104] based on cumulative ray attenuation (attenuation, for short) in our implementation of the modified path tracing algorithm. In this context the attenuation of a given ray is obtained through the product of reflectances and/or transmittances of the surfaces which are hit in the ray's path [198].

During rendering this attenuation is compared with a cutoff attenuation threshold, ς_c, to control the depth of the tree during the ray tracing. If the attenuation is smaller than ς_c, the ray's path is terminated. For a given scene we select a value for ς_c to limit the maximum depth of both trees (one associated with the process using the simplified model and the other with the process using the *ABM*) to the same percentage of rays to be propagated (0.01%). We examine this issue in more detail in the next section. Since the attenuation has three values corresponding to the three RGB channels, we convert it to luminance in order to compare it with a selected ς_c. This conversion is

made using the following SMPTE formula [148]:

$$Y = 0.21222\,R + 0.7013\,G + 0.0865\,B \tag{6.1}$$

6.3 Testing Issues

For our experiments we selected foliar data regarding a soybean leaf (Section 4.5.2). The spectral measurements needed to compute the table of absolute spectral reflectances and transmittances were made using a virtual spectrophotometer (Section 4.5.1) and 10^6 rays in regular intervals of 1° for the angle of incidence θ_i ($[0, \pi]$). The comparisons regarding the foliar specimen's BDF presented in the next section were performed using a virtual goniophotometer (Section 4.5.1) with a collector sphere divided into 20 patches along its latitude and 40 patches along its longitude, and using 10^8 rays to generate each BDF curve associated with a given wavelength.

For the reasons presented in the previous chapter, we sampled the absorption spectra of chlorophyll (Figure 4.4) in the dominant wavelengths corresponding to SMPTE monitor chromaticity coordinates (Table 5.1). Then, applying the *ABM*, we compute the table of absolute spectral reflectances and transmittances used to generate the curves and images presented in the next section. Figure 6.2 presents the graphs correspondings to the entries of this table.

The table of absolute spectral reflectances and transmittances for the veins, as described in the previous chapter, was also obtained using the *ABM* and the SMPTE monitor chromaticity coordinates, but with the absorption spectra of chlorophyll (Figure 4.4) scaled by a factor of 0.1 according to biological data provided by McClendon and Fukshansky [157]. Figure 6.3 presents the graphs corresponding to the entries of this table.

We used the geometrical descriptions of soybean leaves presented in Section 5.2 to generate the images presented in the next section. For a given graphics setting the measurements regarding both models were performed on the same machine. Moreover, the models were implemented using the same software guidelines to avoid differences that could affect the timing. The root-mean-square (RMS) errors [86] of the difference images, presented in the next section, were computed from normalized pixel values (scaled to $[0, 1]$), and using Equation 6.2.

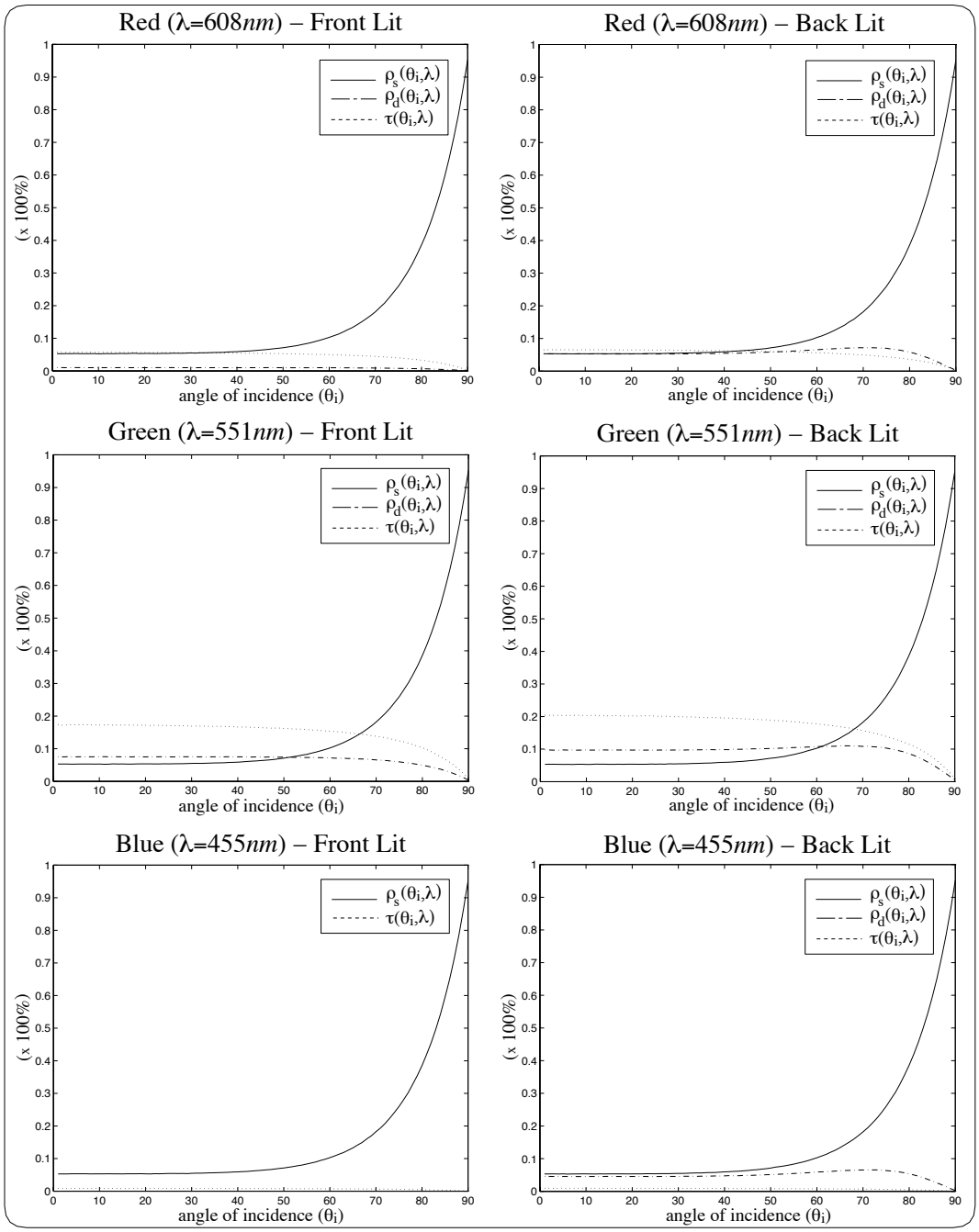

Figure 6.2: Graphs corresponding to the entries of the table of absolute spectral reflectances and transmittances for a soybean leaf.

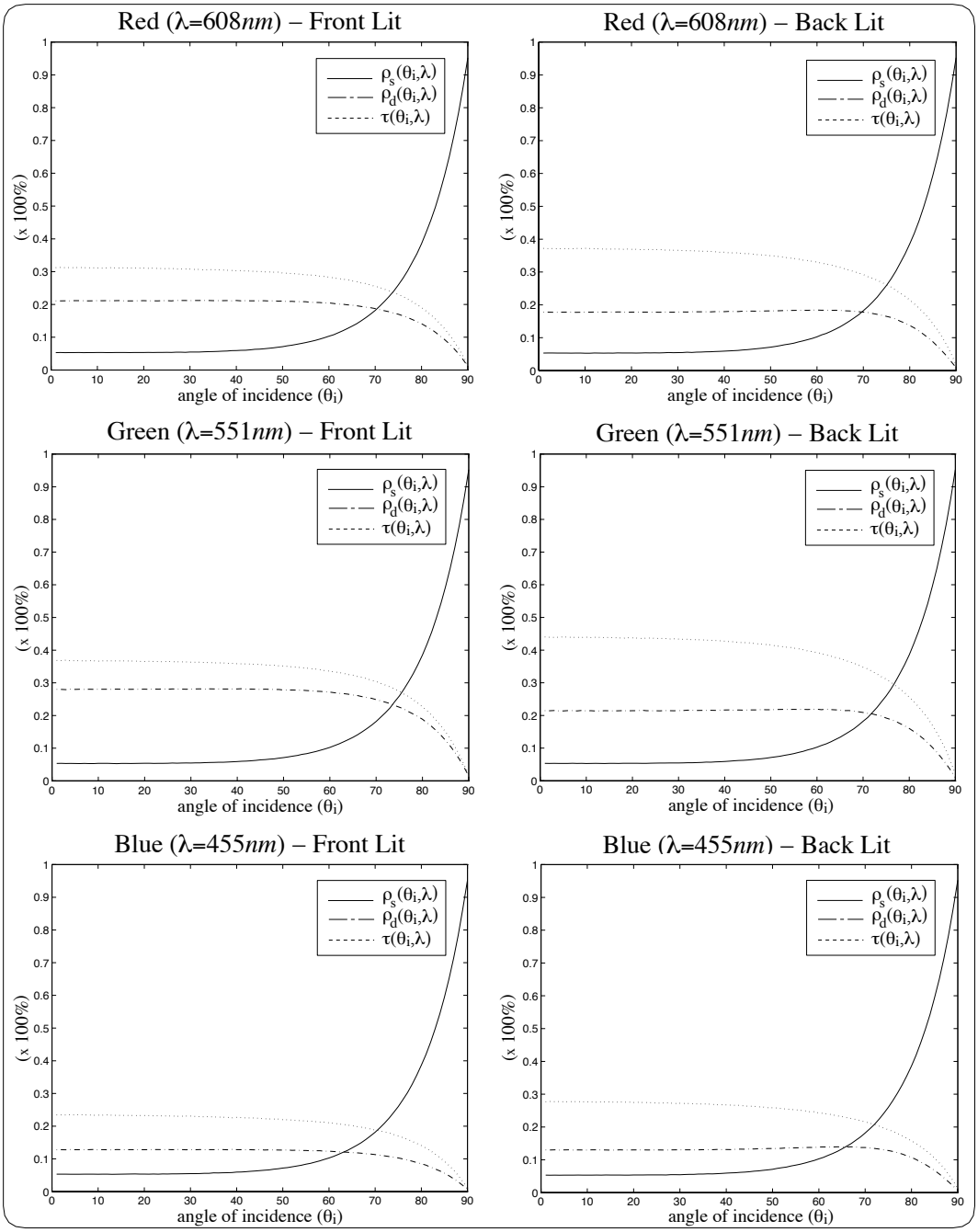

Figure 6.3: Graphs corresponding to the entries of the table of absolute spectral reflectances and transmittances for the veins of a soybean leaf.

The convergence graphs presented in the next section were obtained using the following approach. After choosing a value for the attenuation ς_c we run both processes. During the execution time we use identical "probes" inserted in the respective codes to collect the number of rays shot at each depth of the respective ray tracing trees. We use the data collected by the "probes" to statically fine-tune our initial selection of ς_c. Once the stopping criterion is satisfied by both processes we use the data collected by the "probes" to plot the convergence graphs (Figure 6.6).

These convergence graphs (Figure 6.6) show the percentage of rays that still need to be propagated after each depth of propagation. At depth zero 100% of the rays still need to be propagated. For a certain depth k, where $0 < k \leq n$ (n being the total number of depths considered), this percentage is computed using the following ratio:

$$\frac{\sum_{i=0}^{n}(rpd_i) - \sum_{i=0}^{k}(rpd_i)}{\sum_{i=0}^{n}(rpd_i)} \times 100\% \tag{6.2}$$

where rpd_i represents the number of rays propagated at depth i (Figure 6.7).

6.4 Results and Discussion

As mentioned in Section 4.5, in order to perform a comprehensive evaluation of spectral curves provided by a BDF model it would be necessary to consider all possible testing geometries. However, due to the large number of measurements needed, we limit our examination of the simplified model to selected representative cases. Figure 6.4 shows the BDF curves generated using the *ABM* and the simplified model, for angles of incidence of 30° and 45°, and considering the plane given by the direction of incidence and the specimen's normal[2].

Notice that the curves provided by the simplified model capture the main characteristics of the foliar specimen's BDF, namely an angular dependency on the incident angle for the BRDF intermediate to that expected of diffuse and specular reflectors and a near-Lambertian distribution for the BTDF. They also present a good qualitative agreement with the curves provided by the *ABM*, which agree with the experimental curves published by Breece and Holmes [40] and Wooley [240]. The small

[2]The choice of this plane is consistent with the isotropic characteristic of the simplified model.

discrepancies are mainly related to the simplified nature of the proposed model, and do not significantly affect the image quality as we can see in Figure 6.5.

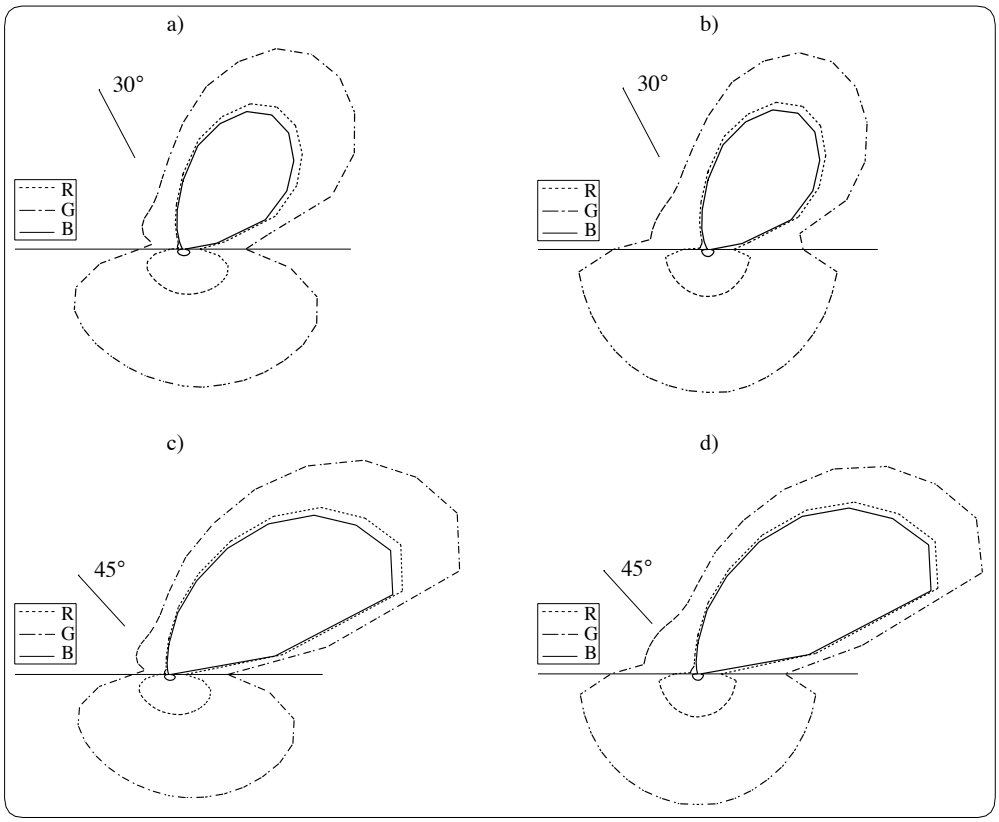

Figure 6.4: BDF curves for a front lit leaf obtained at the wavelengths associated with the RGB channels (Table 6.1). a) and c) Using the *ABM*. b) and d) Using the simplified model.

Figures 6.5a and 6.5b show the first set of images with front lit leaves. The curves presented in Figure 6.6a show that for this graphics setting the rendering process using the simplified model converges faster, in terms of the ray tracing tree-depth, than the rendering process using the *ABM*. Figures 6.5c and 6.5d show the second set of images with back lit leaves. Similarly, Figure 6.6b shows that for this graphics setting we can also observe a faster convergence for the rendering process using the simplified model.

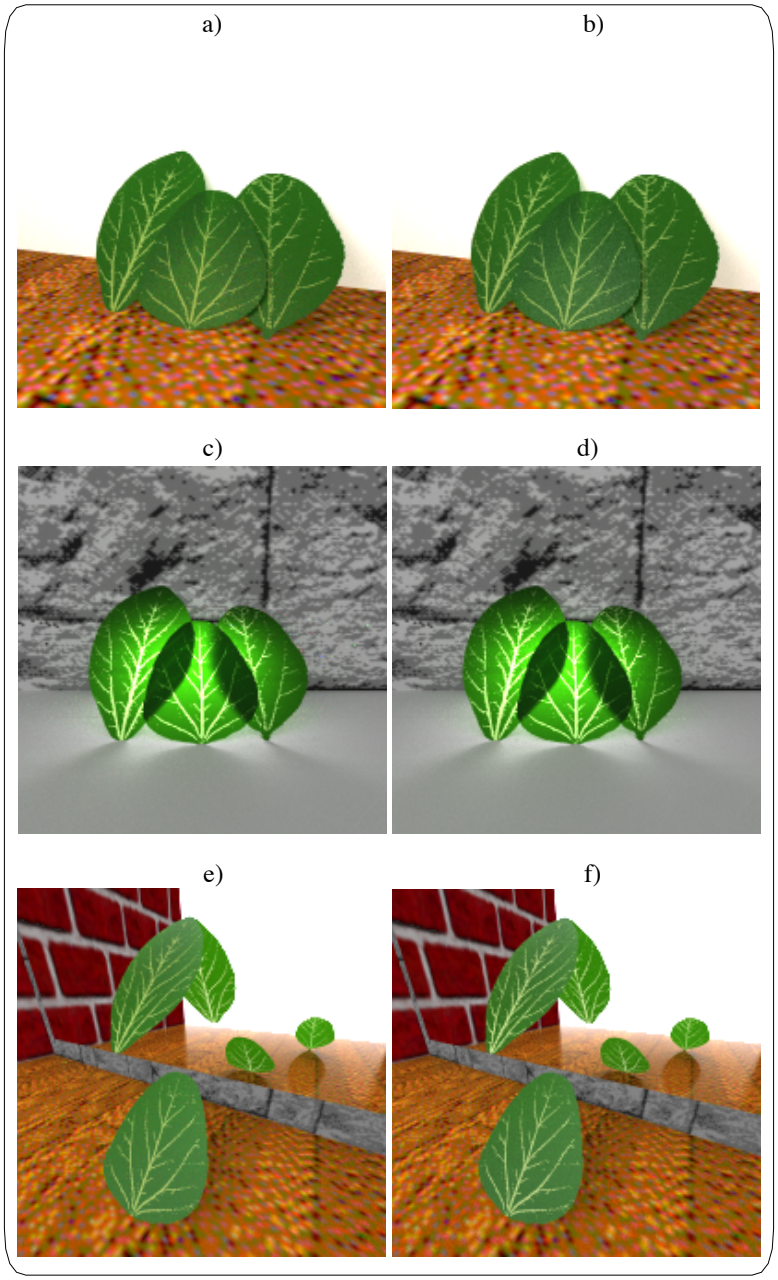

Figure 6.5: Top row: front lit leaves (1^{st} set) using the *ABM* (a) and the simplified model (b). Middle row: back lit leaves (2^{nd} set) using the *ABM* (c) and the simplified model (d). Bottom row: Images with ambient light only (3^{rd} set) and using the *ABM* (e) and the simplified model (f). For all three scenes we used $\varsigma_c = 0.01$.

Figures 6.5e and 6.5f show the third set of images where only ambient light is used, *i.e.* direct lighting calculations are involved in the rendering processes. Figure 6.6c shows that, as in the previous settings, the rendering process using the simplified model converges faster. However, for this graphics setting the curves for the simplified model and the *ABM* are closer.

The spikes in the simplified model curves presented in Figures 6.7a and 6.7b are caused by the fact that at each ray/interface interaction the *ABM* propagates at most one ray, while the simplified model always propagates three rays. The presence of non-diffuse (glossy) surfaces other than the foliar tissues in the third graphics setting has a similar effect on the process using the *ABM* (Figures 6.7c), *i.e.* more than one ray is propagated after an interaction with a glossy surface. This aspect, in turn, makes the curves regarding the simplified model and the *ABM* (Figure 6.6c) closer.

Although the convergence graphs for the three sets of images (Figure 6.6) are illustrative of the behavior of both models, they do not account for different types and amounts of work performed at each depth of the ray tracing trees. In order to extend our performance evaluation, we have measured the speed up gains of the simplified model over the *ABM* for the three graphics settings. These speed up gains depend on a number of factors: the illuminating and viewing angles, the ratio of the number of pixels associated with the specimen(s) to the total number of pixels (in our case, called foliar ratio), the scene geometry, and the loss of quality threshold (in our case, given by the RMS errors).

As mentioned earlier, the large number of measurements required for an in depth comparison precludes the consideration of all the factors and their combinations at this point of our research. However, the figures presented in Table 6.1 indicate that the use of the simplified model can provide noticeable performance gains without a significant loss of image quality. This aspect, in turn, suggests that the simplified model is more suitable than the *ABM* for applications involving several foliar primitives. The figures presented in Table 6.1 are also consistent with the curves presented in Figure 6.7. These curves show that, to have a relatively small RMS error in the difference images, the simplified model involves a considerably smaller number of propagated rays than the ABM.

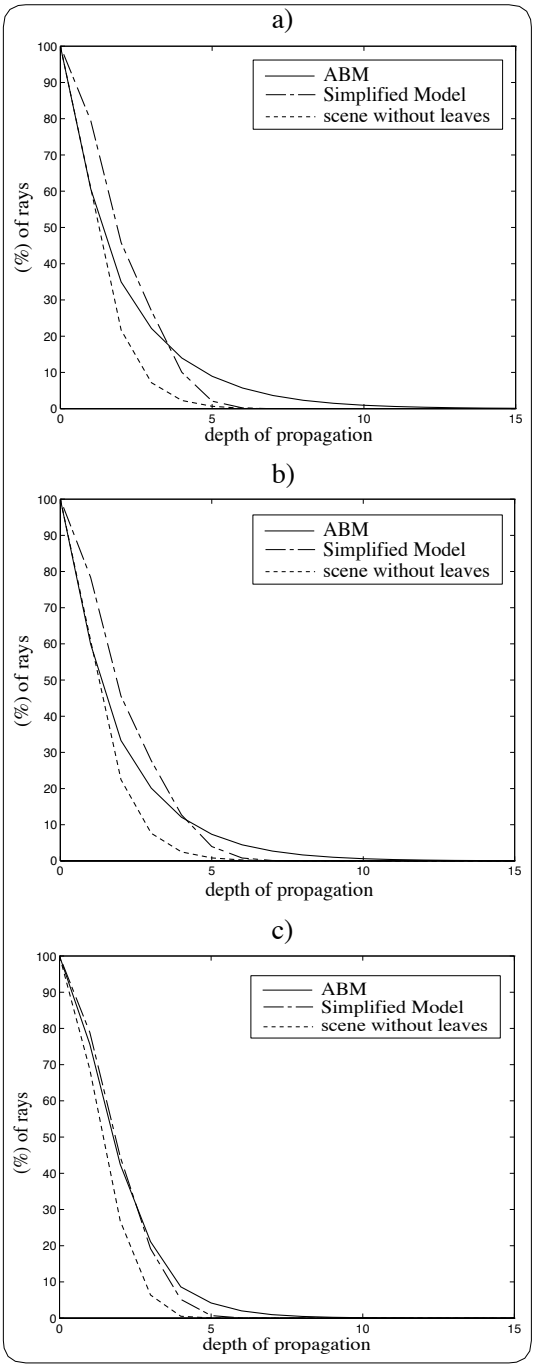

Figure 6.6: Convergence graphs for: a) 1^{st} set, b) 2^{nd} set and c) 3^{rd} set, showing the percentage of rays that still need to be propagated after each depth of propagation.

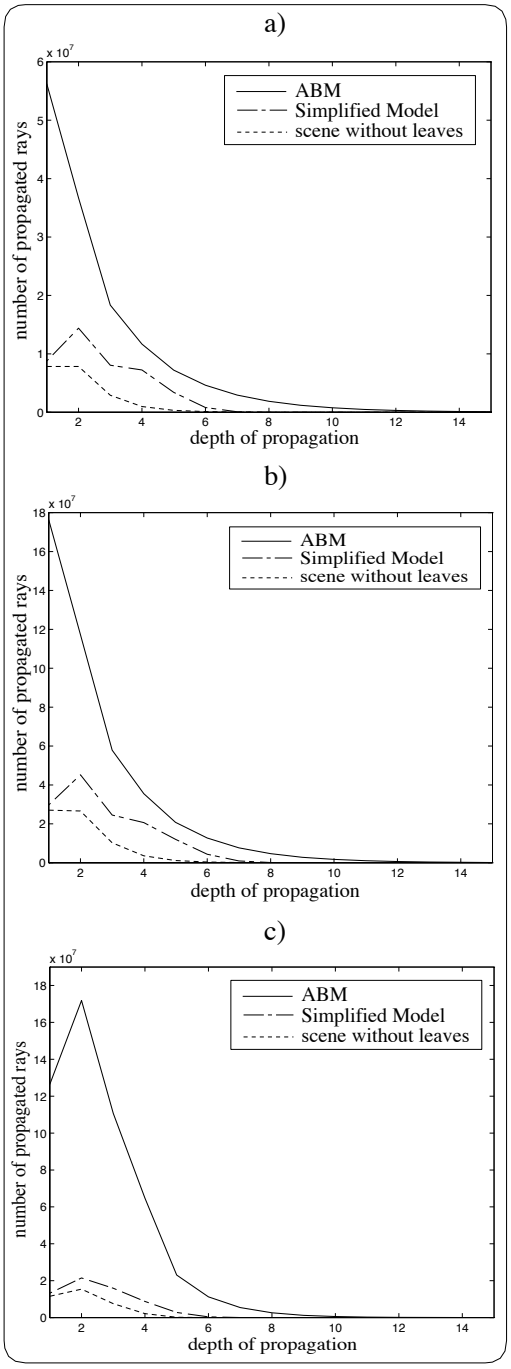

Figure 6.7: Number of rays propagated at each depth of propagation for: a) 1^{st} set, b) 2^{nd} set and c) 3^{rd} set.

Image set	foliar ratio	RMS error			speed up
		Red	Green	Blue	
1^{st}	24.97%	0.009	0.009	0.012	5.14
2^{nd}	19.42%	0.011	0.014	0.010	4.47
3^{rd}	37.61%	0.020	0.016	0.012	9.77

Table 6.1: Comparison of accuracy *vs.* performance gain.

In order to reduce noise in the images due to Monte Carlo path tracing integration, we used a large number of sample points per pixel, which increased the absolute time measurements. For instance, the image presented in Figure 6.5f was generated using 400 sample points per pixel and it took 95 minutes (elapsed CPU time) on a SGI R10000. The incorporation of the model into more efficient global illumination frameworks may considerably reduce the overall rendering time.

The overhead of pre-computing the table of reflectances and transmittances is reduced by the fact that, for a given foliar specimen, this operation must be performed only once. Then, the resulting table can be used several times not only in the rendering of individual leaves, but also in global illumination calculations involving vegetation canopies [155]. Furthermore, the table look-ups are performed through direct indexing, and its storage requirements are within reasonable limits. For instance, the table used in our testing experiments requires only 14.5 Kb of storage space. Even if finer sampling resolutions are used in the measurement of reflectances and transmittances, we believe that the use of such a table provides a worthwhile trade-off between accuracy and computational costs, specially considering the sizes of the memories available nowadays and their decreasing costs.

The quality of images of organic materials is affected by several factors beyond the scope of this dissertation. Certainly, the use of more realistic geometric models would improve the realism of these images. However, it is important to notice that despite the sophisticated geometrical models of foliar tissues available in the graphics literature [32, 69, 106, 108, 180], to the best of our knowledge there is still a lack of geometrical models that accurately account for their venation system as well.

Bump mapping [30] can be used to reduce this problem, by perturbing the normal

of the surface to simulate a wrinkle or a vein. However, as pointed out by Glassner [86], it precludes the simulation of phenomena such as the self-blocking and self-shadowing of incident light by the veins, since it does not alter the geometry of the surface. Displacement mapping [59], a technique that actually alters the geometry of the surface, allows the simulation of these phenomena, but, as pointed out by Watt and Watt [234], it presents the drawback of being difficult to incorporate in a standard polygonal mesh renderer.

The use of accurate luminaire data, including its geometrical description, positional information, and spatial and spectral intensity information [104], can also improve the quality of these images significantly. Moreover, since the rendering of organic materials like plants have a strong wavelength dependency, the use of more efficient methods for guiding the wavelength selection, such as the strategy proposed by Meyer [160], as well as an investigation to determine the group of wavelengths that fits best the perceptual requirements of these applications could also improve the overall quality of the images of these materials.

6.5 Summary

In this chapter we have presented a physically and biologically-based model for light interaction with plant tissue that can be incorporated into rendering systems without a significant computational overhead. Although, as mentioned in Chapter 4, the evaluation of a computer model is less predictable than measuring physical phenomena, our experiments suggest that despite of its simplified nature, the proposed model captures the general behavior of foliar BDFs and does not introduce significant errors along the rendering pipeline. Moreover, the low storage requirements and the significant speed ups provided by the simplified model make it more suitable than the *ABM* for applications involving several foliar primitives. Although the scale factors used by the simplified model were computed using the *ABM*, their computation can be performed through more accurate BDF models as they become available.

Like the *ABM*, the simplified model can be extended to incorporate anisotropic characteristics of the foliar tissues. The current limited knowledge about these characteristics, however, has driven us to treat these materials as isotropic. In fact we

have exploited the assumption of isotropy of foliar specimens' BDF to compress the reflectance and transmittance data. Although the isotropy assumption may seem restrictive, it is relevant to notice that techniques have been proposed in the computer graphics literature [67] to transform a tabulated anisotropic BDF to match the isotropy property in order to reduce storage and computational costs in rendering applications. Incidentally, transformation techniques based on single value decomposition (SVD) type approaches [92] can also be used to make a BDF separable [67]. As appropriately mentioned by Fournier [77] and DeYoung and Fournier [67], transforming a BDF to be separable allows radiosity computations (Chapters 7 and 8) to be performed more efficiently for surfaces whose BDFs present complex directional characteristics [167].

We have also outlined in this chapter some aspects beyond the scope of this dissertation that can contribute to the realism of images of plants. Among these aspects are the spectral sampling and the geometrical representation of veins. Another important aspect related to the quality of these images is the global illumination framework through which they are generated. This aspect brings back the trade-off between accuracy and efficiency. By proposing plausible BDF models for plants we have contributed to the improvement of the overall accuracy of global illumination applications involving these materials. In the next chapters we intend to present our contributions regarding the efficiency of global illumination applications associated with radiosity with faster iterative solutions for radiosity systems.

Chapter 7
Eigen-Analysis for Radiosity Systems

Determining the radiance of each surface of a closed environment using the radiosity method demands the solution of a linear system of equations. This system represents the balance of energy in the environment. Because of the large size of the linear systems associated with radiosity applications and the relatively low accuracies required in the solutions (only 1% to 10%, since humans do not perceive finer variations of light [170]), iterative methods are commonly used.

In this chapter we first describe the radiosity system of linear equations and present some selected iterative methods usually applied to solve these systems. Then, we use this mathematical background to examine the spectral properties of these systems which affect the performance of the iterative methods directly. More specifically, we look at the spectrum or set of eigenvalues of the radiosity coefficient matrix and its relationship to the convergence of the iterative methods.

Furthermore, we show analytically that all the eigenvalues of the radiosity coefficient matrix are *real* and *positive*. Although one could in theory apply any iterative method to solve the radiosity system of linear equations, this fact enables the application of fast methods, such as the Chebyshev method (described in Chapter 8), with confidence of their convergence.

We focus the discussion presented in this chapter on standard, or classical [86], radiosity applications involving diffuse BRDFs. The insights drawn from our analysis, however, are also valid for applications involving materials that present BDFs with Lambertian and non-Lambertian characteristics and handled by multipass methods that use the radiosity method in one of the passes [50, 168, 231].

The sections of this chapter are organized as follows:

- Section 7.1 presents the radiosity system of linear equations.
- Section 7.2 describes iteratives methods usually applied to solve these systems.
- Section 7.3 outlines the definition and basic properties of the eigenvalues and their relationship to the convergence of the iterative methods.
- Section 7.4 presents the proof that all eigenvalues of the radiosity coefficient matrix are real and positive.
- Sections 7.5 presents estimates for the eigenvalues of the radiosity coefficient matrix.
- Section 7.6 summarizes the contents of this chapter.

7.1 The Radiosity System of Linear Equations

7.1.1 Mathematical Formulation

Assuming a closed environment whose surfaces are divided into n patches, the total spectral radiant power leaving a patch depends on the spectral radiant power emitted by the patch plus the spectral radiant power that is reflected. The spectral radiant power depends, in turn, on the total spectral radiant power leaving the other patches in the environment (recall Equation 2.21). If we consider only the Lambertian reflective behavior of surfaces in the environment, we can drop the directional and positional dependences on the spectral measures of light. In this case the process of spectral radiant power transfer in a closed environment can be represented by:

$$\Phi_j(\lambda) = \Phi_j^E(\lambda) + \rho_j(\lambda) \sum_{i=1}^{n} F_{ij} \Phi_i(\lambda) \qquad \text{for each } j = 1, 2...n \qquad (7.1)$$

where:

$\Phi_j(\lambda)$ = total spectral radiant power leaving patch j,
$\Phi_j^E(\lambda)$ = spectral radiant power emitted by patch j,
$\rho_j(\lambda)$ = reflectance of patch j,
F_{ij} = form factor between patch i and patch j,
$\Phi_i(\lambda)$ = total spectral radiant power leaving patch i.

Since we are considering only the Lambertian reflective behavior of the surfaces in the environment, the process of radiant power transfer accounts only for diffuse BRDFs given by $\frac{\rho_j(\lambda)}{\pi}$. In this case the reflectance $\rho_j(\lambda)$ (Section 2.2) represents the fraction of incident radiant power which is reflected back to the environment at wavelength λ by a patch j, and π is incorporated in the definition of the form factor F_{ij}.

The form factor F_{ij}, also called configuration factor, indicates how a patch i "sees" a patch j. In other words, it specifies the fraction of radiant power that leaves patch i and arrives at patch j. Form factors depend on the shape and relative orientation of the patches as well as on the presence of obstacles between them, which is indicated by the visibility term $V(x_i, x_j)$ (Section 2.4). The form factor F_{ij} between a patch i and a patch j (Figure 7.1) can be determined using the following expression:

$$F_{ij} = \frac{1}{A_i} \int_{A_i} \int_{A_j} \frac{cos\theta_i \, cos\theta_j}{\pi ||\vec{l_{ij}}||^2} \, dA_i \, dA_j \, V(x_i, x_j) \tag{7.2}$$

where:

A_i	=	area of patch i,
$\vec{l_{ij}}$	=	vector connecting a point x_i on patch i to a point x_j on patch j,
θ_i	=	angle between the vector $\vec{l_{ij}}$ and the normal of patch i,
θ_j	=	angle between the vector $\vec{l_{ij}}$ and the normal of patch j,
dA_i	=	differential area surrounding on patch i,
dA_j	=	differential area surrounding on patch j,
$V(x_i, x_j)$	=	visibility term.

Although there is in general no closed form solution for the equation above, there are useful analytical formulae for simple geometrical configurations [191, 205, 215, 237]. These formulae are, however, usually not general enough to be used in global illumination. For example, they usually do not take into account occlusion between elements. For this reason, several numerical algorithms have been developed to compute form factors. Loosely speaking, the numerical algorithms to compute form factors can be divided in two groups: deterministic, based on quadrature methods, and nondeterministic, based on Monte Carlo methods. We examine some issues regarding

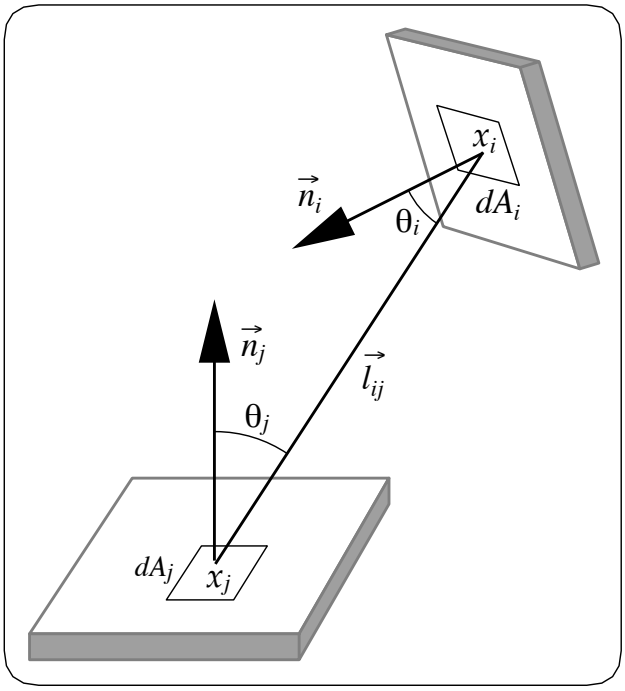

Figure 7.1: Geometry of the form factor between two patches.

the accuracy of these algorithms in Appendix B.

Equation 7.1 holds for each patch in the environment. Then, in order to find the total spectral radiant power of each patch, we need to solve the following linear system:

$$H\Phi = \Phi^E \tag{7.3}$$

where the matrix H and the vectors Φ and Φ^E are given by:

$$\begin{bmatrix} 1-\rho_1(\lambda)F_{11} & -\rho_1(\lambda)F_{12} & \dots & -\rho_1(\lambda)F_{1n} \\ -\rho_2(\lambda)F_{21} & 1-\rho_2(\lambda)F_{22} & \dots & -\rho_2(\lambda)F_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -\rho_n(\lambda)F_{n1} & -\rho_n(\lambda)F_{n2} & \dots & 1-\rho_n(\lambda)F_{nn} \end{bmatrix} \begin{bmatrix} \Phi_1(\lambda) \\ \Phi_2(\lambda) \\ \vdots \\ \Phi_n(\lambda) \end{bmatrix} = \begin{bmatrix} \Phi_1^E(\lambda) \\ \Phi_2^E(\lambda) \\ \vdots \\ \Phi_n^E(\lambda) \end{bmatrix}$$

Another possibility is to rewrite Equation 7.1 in terms of spectral radiant exitance [5], which represents the total spectral radiant power leaving an element per unit of area, and spectral irradiance [5], which represents the spectral radiant power emitted

by an element per unit of area. In this case, from [198], we have:

$$\Phi_j(\lambda) = \pi M_j(\lambda) A_j \tag{7.4}$$

where:

$M_j(\lambda)$ = spectral radiant exitance of patch j,

A_j = area of patch j.

Also from [198], we have:

$$\Phi_j^E(\lambda) = \pi E_j(\lambda) A_j \tag{7.5}$$

where:

$E_j(\lambda)$ = spectral irradiance emitted from patch j.

Substituting Equation 7.4 and Equation 7.5 into Equation 7.1 gives:

$$\pi M_j(\lambda) A_j = \pi E_j(\lambda) A_j + \rho_j(\lambda) \sum_{i=1}^{n} F_{ij} M_i(\lambda) A_i \qquad \text{for each } j = 1, 2...n \tag{7.6}$$

Before going further in our derivation, we review the following identities derived from the mathematical formulation of form factors.

- Reciprocity relationship:

$$A_i F_{ij} = A_j F_{ji} \tag{7.7}$$

- Summation relationship:

$$\sum_{j=1}^{n} F_{ij} \leq 1 \qquad \text{for each } i = 1, 2...n \tag{7.8}$$

  If we assume a closed environment, the above sum is by definition equal to 1.0. However, in practice, even for closed environments the computed value can be greater than 1.0 depending on the accuracy of the method used to compute the form factors (see Appendix B for experiments regarding this topic).

- A planar or a convex patch can not see itself, which means that the form factor

of a planar or a convex surface regarding itself is:

$$F_{ii} = 0$$

However, there may be situations in which it may be appropriate to consider the form factor F_{ii} of a plane or convex patch different from zero. For example, suppose patches i and j are placed in front of each other and patch j is a perfect reflector[1] (mirror). In this case patch i can "see" itself, and F_{ii} is different from zero.

- Finally, a concave patch can see itself, hence:

$$F_{ii} \neq 0$$

Applying the form factor reciprocity relationship (Equation 7.7) in Equation 7.6 and dividing it by A_j we get:

$$\pi M_j(\lambda) = \pi E_j(\lambda) + \rho_j(\lambda) \sum_{i=1}^{n} \pi F_{ji} M_i(\lambda) \qquad \text{for each } j = 1, 2...n \qquad (7.9)$$

Dividing the above expression by π, we finally get the classical expression in terms of spectral radiant exitance, which holds for each patch in the environment:

$$M_j(\lambda) = E_j(\lambda) + \rho_j(\lambda) \sum_{i=1}^{n} F_{ji} M_i(\lambda) \qquad \text{for each } j = 1, 2...n \qquad (7.10)$$

As mentioned in Chapter 2, radiance is the radiometric quantity more appropriate for rendering applications. Since we are considering only the Lambertian reflective behavior of the surfaces, we can use the simple relationship $M = \pi L$ to transform the radiant exitances to radiances and vice-versa [86].

Radiosity, B, is the term used for radiant exitance in the computer graphics literature. In order to simplify matters, we are going to use this term and suppress the dependence on wavelength in the next expressions. In this case, to determine the

[1] Usually non-Lambertian surfaces are not introduced in a radiosity system. Their effects on the form factors of the surrounding Lambertian surfaces can, however, be accounted for through the use of extended form factors [57, 184].

radiosity of each patch, we need to solve a system of linear equations whose coefficient matrix is represented by the radiosity matrix G, vector of unknowns is represented by the vector of radiosities B and the right-hand side vector is represented by the vector of irradiances. This system, known as the radiosity system, is represented by:

$$GB = E \tag{7.11}$$

or, in an expanded form, given by:

$$\begin{bmatrix} 1-\rho_1 F_{11} & -\rho_1 F_{12} & ... & -\rho_1 F_{1n} \\ -\rho_2 F_{21} & 1-\rho_2 F_{22} & ... & -\rho_2 F_{2n} \\ . & . & ... & . \\ . & . & ... & . \\ . & . & ... & . \\ -\rho_n F_{n1} & -\rho_n F_{n2} & ... & 1-\rho_n F_{nn} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ . \\ . \\ . \\ B_n \end{bmatrix} = \begin{bmatrix} E_1 \\ E_2 \\ . \\ . \\ . \\ E_n \end{bmatrix}$$

where the elements of G are $G_{ij} = \delta_{ij} - \rho_i F_{ij}$ (with δ_{ij} being the Kronecker delta).

7.1.2 Analysis in Matrix Terms

The radiosity coefficient matrix G is commonly represented by:

$$G = I - PF \tag{7.12}$$

where:

I = identity matrix,
P = diagonal matrix whose diagonal entries p_{ii} correspond to ρ_i,
F = form factor matrix.

The matrix G is usually a square matrix ($n \times n$). It may, however, be rectangular if one applies a hierarchical method [57]. Moreover, in general G is well-conditioned [57], which means that the solution B of Equation 7.12 is not very sensitive to small perturbations in G or E [92, 143]. Figure 7.2[2] shows a typical radiosity coefficient matrix corresponding to the scene presented in Figure 7.3, which consists of a sphere centered in a "box". The sphere is divided into 128 patches and each of the six faces

[2]Image generated using *EMILY* [39].

of the surrounding "box" are divided into 144 patches for a total 992 patches.

The density of G, defined as the number of nonzero entries divided by n^2, depends on the environment complexity. For example, an environment divided into a large number of patches (*e.g.* 50000) in such a way that several patches cannot "see" each other may have a 10% dense matrix [54]. On the other hand, in the inside of a sphere all the patches can "see" each other and the corresponding matrix is full even if the sphere is divided into a very large number of curved patches.

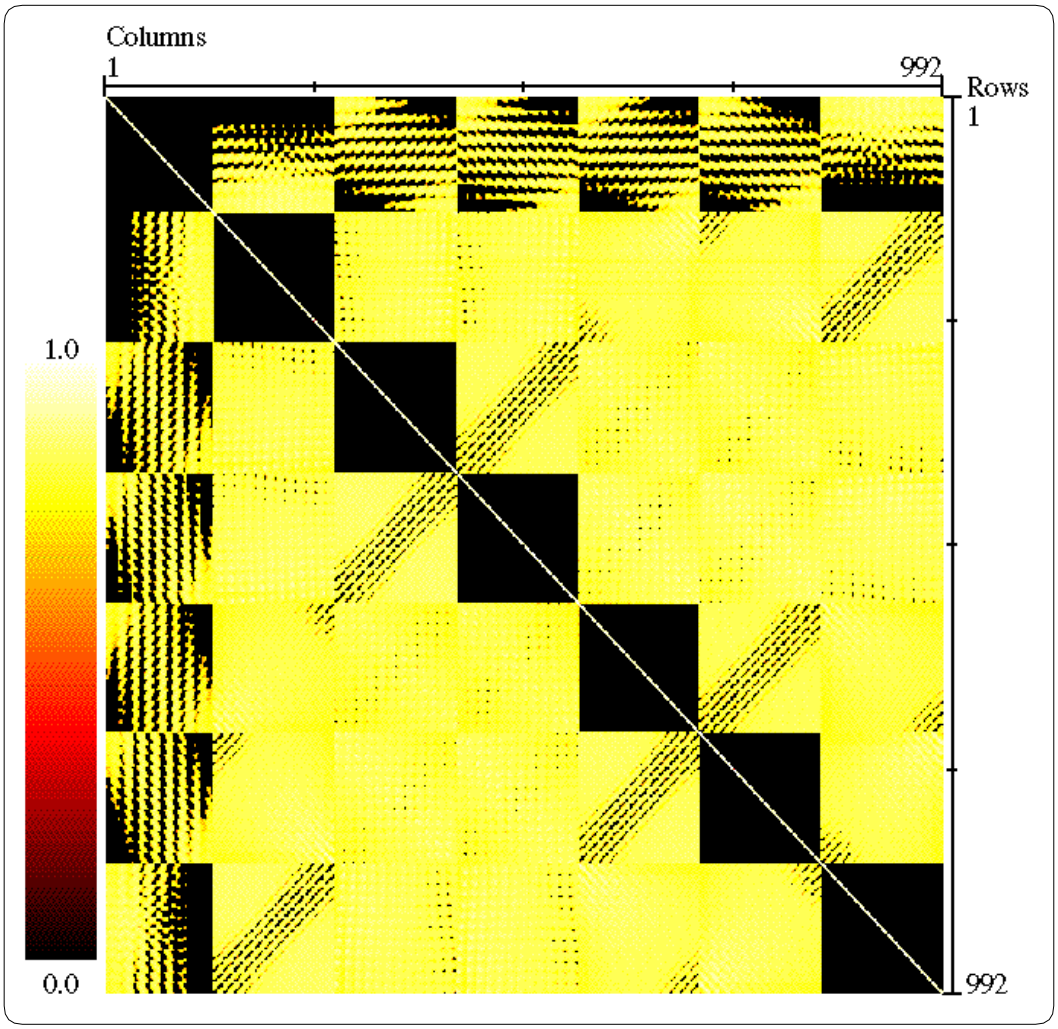

Figure 7.2: A typical radiosity coefficient matrix, *i.e.* presenting diagonal dominance and relative low density.

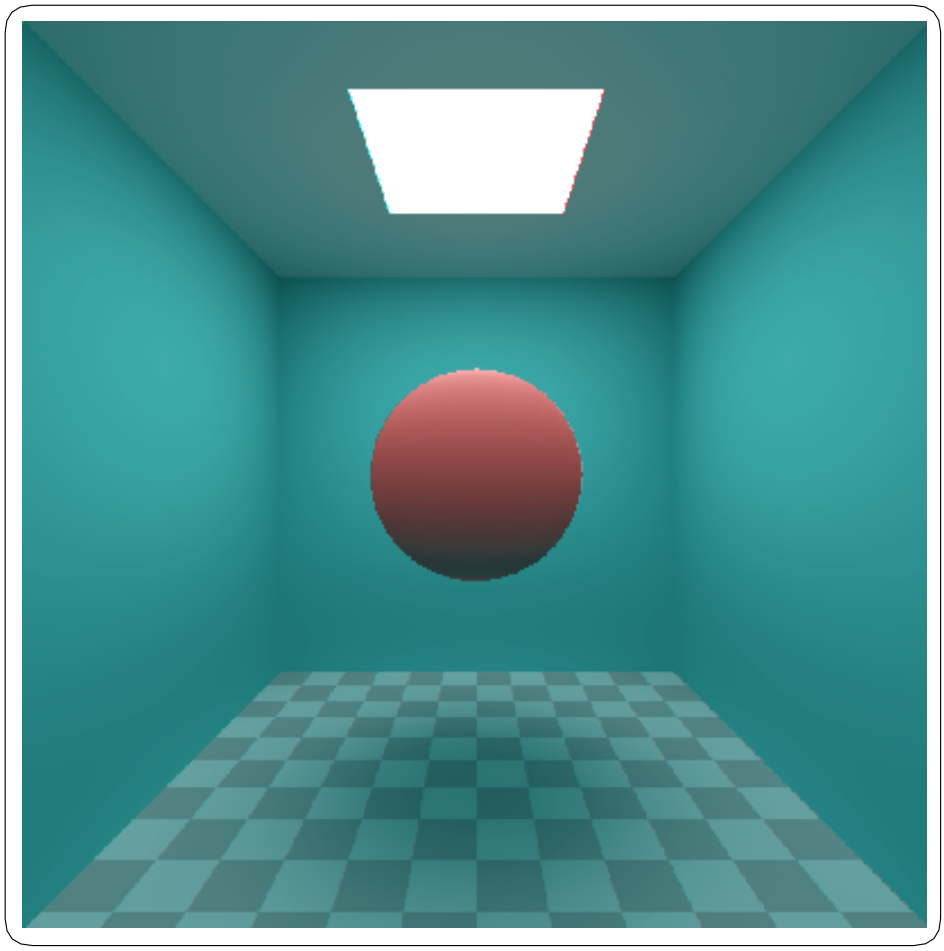

Figure 7.3: An image generated using the radiosity method.

Notice that the density depends more on the placement of patches in the environment than on the number of patches. For example, by changing the level of occlusion in the scene present in Figure 7.3 (*e.g.* by changing the radius of the sphere) we can increase the number of zero entries, represented by the black areas in Figure 7.2, and, in turn, reduce the density of the matrix. A matrix entry is equal to zero when:

- $F_{ij} = 0$ for $i \neq j$ (due to occlusion or patches placed in the same plane);
- $\rho = 0$ (black patches, which does not occur in real environments, although sometimes researchers assign $\rho = 0$ to light sources).

Assuming that $0 \leq \rho < 1$ and that the sum of form factors in any row is equal to one, we can say that the matrix G is strictly diagonally dominant, since the following property (also noticeable in Figure 7.2):

$$|G_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |G_{ij}|$$

holds for each $j = 1, 2, \ldots, n$.

However, if we include, for experimental purposes, a perfect reflector surface[3] ($\rho_i = 1$) in the environment, we cannot ensure dominance of rows. In this case, weak diagonal dominance [245] holds, that is $|G_{ii}| \geq \sum_{j=1, i \neq j}^{n} |G_{ij}|$, unless all $\rho_i = 1$.

7.2 Solutions for Radiosity Systems

Why should we search for faster methods to solve the radiosity system of equations when the most expensive stage of the radiosity pipeline (in some cases up to 90% of the computational time [242]) is the calculation of form factors? The reason is that there are many practical applications in which a radiosity system has to be solved repeatedly, due to changes either to the vector of reflectances or to the vector of irradiances, while the form factors remain unchanged.

Examples of such applications would be sun and shade simulations for environmental design purposes (Figure 7.4) or the simulation of solar radiation over a forest canopy for remote sensing purposes. For a comprehensive sun and shade simulation corresponding to a period of twelve hours, the form factors would have to be computed only once, while the system would have to be solved 720 times to account for solar positions at every minute.

Direct methods for solving linear systems, such as Gaussian Elimination [42] or LU decomposition [42], are not suitable for large radiosity systems of equations because of the relative low density of the coefficient matrix, and because rapid solutions at relatively low accuracy are needed. These aspects plus the special properties of

[3] A perfect reflector surface is a plane surface reflecting all the incident light in a perfect image-forming state, such that we cannot focus on the surface. In other words, there is no visible microstructure, and visually there is no surface at all [128]. Thus, we cannot have $\rho_i = 1$ in any physically realistic system [57]. Nevertheless such a surface can be useful as a gloss standard [122, 128].

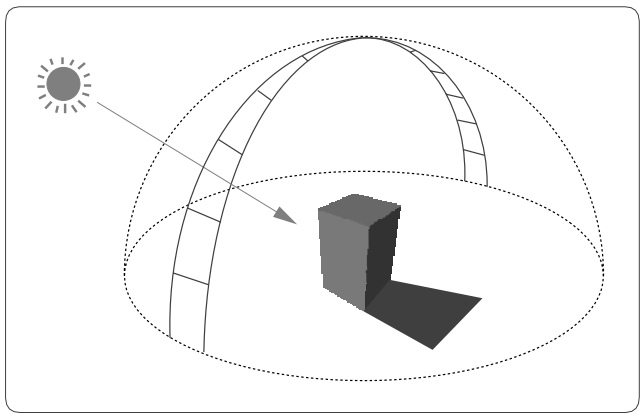

Figure 7.4: Sketch of a sun and shade simulation for a building.

the radiosity coefficient matrix, such as diagonal dominance, make the use of iterative methods more convenient. These methods are generally of the following form[4]:

```
1    initialize B and r
2    while (not converged)
3        select ∇B
4        update B = B + ∇B
5        update r = r − G∇B
```

where ∇B corresponds to the correction vector, and r corresponds to the residual vector given by:

$$r = E - GB \tag{7.13}$$

The iterative methods used to solve the radiosity system of linear equations can be divided into general matrix methods, which update all components of the solution vector on each iteration, and radiosity-specific methods such as progressive refinement and overshooting methods, which update a single component on each iteration. We

[4]For the sake of simplicity and saving space we will omit the obvious **begin** and **end** loop delimiters in the pseudocodes presented in this dissertation. Instead, we will indent the code to indicate the loops. Note also that the equal sign (=) denotes assignment, in the sense of a programming language, throughout the pseudocodes.

use the expression *radiosity-specific methods* to group methods specifically developed to solve the radiosity problem. Although those methods can be considered variations of numerical methods such as Southwell Iteration [91] or SOR (Section 7.2.2), they have been adjusted and fine-tuned to the radiosity case.

7.2.1 The Gauss-Seidel Method

An example of a general matrix method is the Gauss-Seidel method [21, 101], which was introduced to solve the radiosity system of linear equations by Cohen and Greenberg [55]. This method adapted to the radiosity context becomes the following:

```
1    for (all i)
2       B_i = starting guess
3    while (not converged)
4       for i = 1, 2, ..., n
5          B_i = E_i + ρ_i Σ_{j=1,j≠i}^{n} (B_j F_ij)
6    display image using B_i as the radiosity of the patch i.
```

Usually the vector of irradiances is used as the starting guess. This algorithm estimates the radiosity of a patch i by adding its irradiance, and all the radiosity that it reflects (line 5). It uses the most recent estimates of the radiosities of all patches to estimate the radiosity of patch i. The physical interpretation consists of *gathering* the radiosity from other patches to estimate the incoming radiosity. It converges quickly due to the diagonal dominance of the radiosity coefficient matrix. After a few passes through the matrix we get the final image. Gauss-Seidel can have $O(n^2)$ memory requirement if the matrix G is dense and its elements are stored. There are applications (see Chapter 8), however, for which the benefits of having G stored overcome the memory costs.

7.2.2 The Successive Overrelaxation Method

The Successive Overrelaxation method (SOR) [21] consists of applying extrapolation to the Gauss-Seidel method. This extrapolation takes the form of a weighted average between the previous iterate and the computed Gauss-Seidel iterate, for

each component. The physical interpretation is that instead of gathering the correct amount of radiosity, an extrapolation factor w is used to take into account steps that will occur later on in the process.

The structure of the SOR algorithm is similar to that of the Gauss-Seidel algorithm. If $w = 1.0$ then the SOR Method becomes the Gauss-Seidel Method. Its algorithm can be obtained by replacing line 5 of the algorithm presented in Section 7.2.1 by:

$$B_i = (1 - w)B_i + w(E_i + \rho_i \sum_{j=1, j \neq i}^{n} (B_j F_{ij}))$$

The major difficulty of this method is to choose the best value of w which gives the fastest convergence. This factor depends on the problem characteristics. However, the idea of overrelaxation is exploited in the context of progressive refinement methods, and it can be thought of as overshooting (Section 7.2.4).

7.2.3 The Progressive Refinement Method

The progressive refinement approach [54], which was later shown to be equivalent to Southwell's relaxation method [90, 91], allows us to obtain intermediate images and to reduce the storage requirement to $O(n)$ by computing the form factors on the fly:

```
1     for (all i)
2        ∇B_i = 0
3        ΔB_i = E_i
4     while (not converged)
5        select i, such that ΔB_i A_i is largest
6        ∇B_i = ∇B_i + ΔB_i
7        for every other patch j
8           Δrad = ρ_j F_ji ΔB_i
9           ΔB_j = ΔB_j + Δrad
10       ΔB_i = 0
11       display intermediate image using ΔB_k + ∇B_k as the radiosity of patch k.
```

In the previous algorithm A_i represents the area of patch i, ΔB_i represents the amount of "unshot" radiosity and ∇B_i represents the amount of "shot" radiosity ($\nabla B_i = B_i - \Delta B_i$). More formally ΔB_i represents the residual r, and ∇B_i represents the vector of unknowns that we are solving for. During one iteration, the patch with the largest "energy" to shoot, *i.e.* largest $\Delta B_i A_i$, is chosen and its radiosity is shot through the environment. So, in this case, the physical interpretation consists of *shooting* instead of *gathering process*. As a result the other patches may receive some new radiosity Δrad. The intermediate images obtained in the early steps tend to be dim, so an ambient term is added [54]. As the iteration process goes on this ambient term is reduced until the final image is obtained. The ambient term is a weighted sum of the unshot radiosities of all patches in the environment:

$$Ambient = \frac{\mathcal{R}}{\sum_{i=1}^{n} A_i} \sum_{i=1}^{n} (\Delta B_i A_i) \tag{7.14}$$

where A_i corresponds to the area of patch i and $\mathcal{R}$ corresponds to the overall inter-reflection factor $\mathcal{R}$, which is given by:

$$\mathcal{R} = \frac{1}{1 - \rho_{avg}} \tag{7.15}$$

where ρ_{avg} corresponds to an average reflectance for the environment [54], which can be computed as an area-weighted average of the patch reflectances such that:

$$\rho_{avg} = \frac{\sum_{i=1}^{n} (\rho_i A_i)}{\sum_{i=1}^{i=n} A_i} \tag{7.16}$$

7.2.4 The Overshooting Approach

In the original progressive refinement approach a patch may need to shoot radiosity multiple times. Overshooting algorithms based on overrelaxation have been introduced to solve this problem [73, 90, 194, 243]. One such an algorithm was proposed by Feda and Purgathofer [73]. The main idea behind these algorithms consists of shooting the current radiosity of the patch plus an estimate that accounts for radiosity that will return due to the reflection by other patches. This approach can reduce the overall number of iterations required. Feda and Purgathofer use the ambient term not only for display purposes, but also as the overshooting amount. As was mentioned earlier, several methods have been proposed to implement overshooting.

For a comparison of four overshooting algorithms (Feda's, Gortler's, Shao's and Xu's) the reader is referred to [243]. The algorithm proposed by Feda and Purgathofer is given below:

```
1     for (all i)
2        ∇B_i = 0
3        ΔB_i = E_i
4     determine initial Ambient
5     while (not converged)
6        for (each patch i)
7           ΔB'_i = min(ΔB_i + ρ_i Ambient, (Σ_{j=1}^{n} (ΔB_j A_j))/A_i)
8        select i, such that | ΔB'_i A_i | is largest
9        ∇B_i = ∇B_i + ΔB'_i
10       for (every other patch j)
11          Δrad = ΔB'_i ρ_j F_ji
12          ΔB_j = ΔB_j + Δrad
13       ΔB_i = −ρ_i Ambient
14       determine new Ambient
15       display intermediate image using ∇B_k + ΔB_k as the radiosity of patch k.
```

7.3 Eigenvalues and Eigenvectors

7.3.1 Definition and Basic Properties

An *eigenvector* v of a matrix G is a nonzero vector that does not rotate when G is applied to it. In other words, there is some scalar constant Λ, an *eigenvalue*[5] of G, such that $Gv = \Lambda v$. Every square matrix G of order n has n possibly nondistinct complex eigenvalues $\Lambda_1, \Lambda_2, ..., \Lambda_n$. When G is symmetric the eigenvalues are real-valued. The set $\sigma(G)$ of eigenvalues of a matrix is called its spectrum.

[5]In numerical analysis texts an eigenvalue is usually represented by a lowercase lambda, λ. In order to avoid misunderstandings with wavelength, which is also represented by λ in computer graphics, we will use the uppercase lambda, Λ, to represent eigenvalues throughout this dissertation.

In order to find the eigenvalues of a square $n \times n$ matrix like G, we may resort to concepts from the theory of determinants [143], such as the characteristic polynomial of G given by:

$$p(\Lambda) = det(G - \Lambda I) \tag{7.17}$$

The zeros of $p(\Lambda)$ represent the *eigenvalues* or characteristic values of the matrix G. If there is a vector $v \neq 0$ such that $(G - \Lambda I)v = 0$ holds, then v represents an *eigenvector* or characteristic vector of G corresponding to the eigenvalue Λ. This eigenvector is only determined up to a scalar multiple.

7.3.2 Relationship with Iterative Methods

The eigenvalues of certain matrices determine the convergence of iterative methods used to solve the radiosity system $GB = E$. For linear stationary methods of the form:

$$B^{(k+1)} = TB^{(k)} + \tilde{E} \tag{7.18}$$

where k is the number of iterations made, which includes Gauss–Seidel and SOR, the eigenvalues of the iteration matrix T are the relevant ones. The matrix T is derived from the coefficient matrix G. For example, in the Gauss-Seidel iteration, T is given by:

$$T = (D + L)^{-1}U \tag{7.19}$$

where:

D	=	diagonal part of G,
L	=	strictly lower triangular part of G,
U	=	strictly upper triangular part of G.

There needs to be no connection between the eigenvalues of T and those of G. Linear stationary methods converge if $\varrho(T) < 1$, where $\varrho(T)$ is the spectral radius of T[6]. Furthermore, convergence is faster for smaller $\varrho(T)$.

[6]The spectral radius $\varrho(K)$ of a matrix K is defined by $\varrho(K) = max \mid \Lambda \mid$, where Λ is an eigenvalue of K [42]. The spectral radius is usually represented in numerical analysis text by ρ, but, to avoid misunderstanding with the reflectance, also represented by ρ, we are going to represent it by ϱ throughout this dissertation.

For nonstationary methods such as Conjugate Gradient or Chebyshev (Chapter 8), the eigenvalues of the coefficient matrix G are the important ones. The Chebyshev method has convergence determined by the radius of the convex hull of the spectrum of G, which is determined by the extreme eigenvalues. For a matrix with positive real eigenvalues, which is the case of G (see proof presented in the next section), the largest (Λ_{max}) and smallest (Λ_{min}) eigenvalues completely determine convergence, which, as described by Hageman and Young [101], is faster for larger values of the following ratio:

$$\frac{\Lambda_{max} + \Lambda_{min}}{\Lambda_{max} - \Lambda_{min}} \tag{7.20}$$

The convergence of the Conjugate Gradient method is determined by the overall distribution of eigenvalues, and even for a given $\sigma(G)$ it is impossible to predict the exact number of iterations that this method will require. However, Conjugate Gradient generally requires only $k + 1$ iterations when the eigenvalues occur in only $k < n$ clusters, and has faster convergence for larger values of the ratio presented above (Equation 7.20). We examine this issue in more detail in Section 8.2.3.

7.4 Proof that All Eigenvalues of the Radiosity Coefficient Matrix are Real and Positive

Given a $n \times n$ matrix G, the $n \times n$ matrix obtained by interchanging the rows and columns of G is called the transpose of G and is denoted by G^t [143]. In other terms, $G^t_{ji} = G_{ij}$. A square matrix G is said to be symmetric if $G = G^t$. The radiosity coefficient matrix is not symmetric; however, if it is multiplied by a diagonal matrix Q, where the diagonal entry Q_{ii} represents the quotient of the area and the reflectance of patch i, then the resulting matrix:

$$G^s = QG \tag{7.21}$$

is symmetric.

Since QG is symmetric, its eigenvalues are real-valued. Moreover, by applying the Gerschgorin Circle Theorem (Section 7.5), one can verify that they are also positive

(see the eigenvalue estimates of G^s presented in the next section). Hence, QG is a positive definite matrix by Theorem 9.12 of Burden and Faires [42] which states:

> A symmetric matrix A is positive definite if and only if all the eigenvalues of A are positive.

The definition of positive definite means that:

$$x^H QGx > 0 \tag{7.22}$$

for all $x \in \mathcal{C}^n$, where $\mathcal{C}$ is the complex plane and x^H is the conjugate transpose of the column vector x.

Let x be an eigenvector of G and Λ be an eigenvalue. Then:

$$Gx = \Lambda x, \tag{7.23}$$

where Λ and $x \neq 0$ are possibly complex.

Then:

$$QGx = \Lambda Qx \tag{7.24}$$

and

$$x^H QGx = \Lambda x^H Qx \tag{7.25}$$

The left side of Equation (7.25) is necessarily real and positive from Equation (7.22). Furthermore, the definition of an eigenvector x implies that it is nonzero, hence:

$$x^H Qx = \sum_{i=1}^{n} (\overline{x}_i x_i Q_{ii}) \geq Q_{min} \sum_{i=1}^{n} x_i \overline{x}_i = Q_{min} x^H x = Q_{min} \parallel x \parallel > 0 \tag{7.26}$$

where

$$Q_{min} = \min_i Q_{ii}$$

Since all $Q_{ii} > 0$, Equations (7.25) and (7.26) imply that Λ is real and $\Lambda > 0$. Therefore, all of the eigenvalues of G are real and positive.

Arvo [7, 9] has shown that several fundamental operators that arise in global illumination can be uniformly approximated by matrices. Although it may seem that this proof applies directly to the continuous radiosity operator, more details need to

be considered. The critical point is that the diagonal entry i of the matrix Q is the ratio of the area and reflectance of the i-th patch.

For the continuous case, the area is zero and the above argument cannot be used directly. However, the continuous operator is a compact operator [9], sometimes also called a completely continuous operator in functional analysis. Because of this, we can construct a sequence of finite dimensional operators G_k that converge uniformly to the continuous operator G_∞, each with a spectrum consisting only of positive real eigenvalues. That sequence can be constructed using a sequence of uniformly refined discretizations of the scene, for example. The limit G_∞ will necessarily have a spectrum that is real and nonnegative. Because G_∞ is nonsingular[7] [131], it cannot have zero as an eigenvalue. Its compactness also implies that it has only a point spectrum, which implies that G_∞ has only positive real eigenvalues in its spectrum.

7.5 Eigenvalues Estimates for Radiosity Applications

As discussed earlier (Section 7.3.2), the convergence properties of iterative solvers rely extensively on the set of eigenvalues of the corresponding coefficient matrix. It was also noted (Section 7.3.1) that the eigenvalues of G can be obtained, in theory, by finding the roots of its characteristic polynomial. In practice, as pointed out by Burden and Faires [42], it is difficult to determine the roots of an n^{th}-degree polynomial except for small values of n. Then, approximation techniques, such as the Power method [41, 42], Lanczos algorithm [64, 92] or QR Factorization [41, 42], are commonly used to compute the eigenvalues.

Usually to calculate the eigenvalues of a matrix G through these techniques requires more computation than that required to solve the corresponding linear system. However, we can obtain relatively inexpensive estimates of the eigenvalues using the Gerschgorin Circle Theorem [41]. This theorem says that the eigenvalues of G are contained within the union of n circles S_i, which have center G_{ii} and radius $\sum_{j\neq i}^{n} |G_{ij}|$

[7] An $n \times n$ matrix K is said to be nonsingular if an $n \times n$ matrix K^{-1} (the inverse of K) exists such that $KK^{-1} = K^{-1}K = I$ [42]. Alternatively a matrix K is also said to be nonsingular if its determinant is nonzero [143].

and are defined by:

$$S_i = \{z \in \mathcal{C} \parallel |z - G_{ii}| \leq \sum_{j \neq i}^{n} |G_{ij}|\} \tag{7.27}$$

where $\mathcal{C}$ is the complex plane. The union of any k of these circles that do not intersect the remaining $(n-k)$ must contain precisely k (counting multiplicities) of the eigenvalues.

In the case of the radiosity matrix G, if we assume an environment formed by planar or convex surfaces ($F_{ii} = 0$) then all the main diagonal entries will be equal to 1.0. So the centers of the circles are also equal to 1.0. In addition, if we assume a closed environment ($\sum_{j=1}^{n} F_{ij} = 1$), then the radius is given by ρ_i. In practice, if the numerical method used to compute the form factors does not provide exact results (Appendix B), then the radius is in fact given by $\rho_i(1 \pm \delta)$, where δ corresponds to the numerical error associated with the summation of form factors of patch i.

Neumann [165] and recently Nievergelt [169] have applied the Gerschgorin Circle Theorem to obtain estimates for the eigenvalues of the symmetric matrix G^s. Due its symmetry, G^s has only real eigenvalues. Then, instead of circles in the plane of complex numbers, we only have to deal with real intervals. The proof that all eigenvalues of the nonsymmetric radiosity matrix G are real and positive, presented in the previous section, allows us to apply the Gerschgorin Circle Theorem in the same manner to G. In other words, we need to consider only real intervals, instead of circles in the plane of complex numbers, to estimate the largest and the smallest eigenvalues of G. One can use these estimates to apply the Chebyshev method to the solution of radiosity system. We discuss this issue in more detail in the next chapter.

7.6 Summary

The average radiosity (or radiant exitance) of n patches of a closed environment can be approximated by solving a linear system with n unknowns. When n is small (*i.e.* less than few thousands of patches), general matrix methods like Gauss-Seidel can be used where the explicit $n \times n$ matrix can be pre-computed and stored [55]. When n is large, progressive techniques are used where the matrix rows or elements are recomputed as needed [54]. When n is very large (*i.e.* hundreds of thousands of

patches), stochastic techniques can avoid computing or storing the n^2 elements of the matrix [166].

In applications where n is small enough to store the entire matrix in main memory, general matrix techniques will be faster than progressive techniques[8]. For example, for sun and shade simulations [99] the lighting can be examined on simple geometric approximations of the environment being designed, and n can be very small. When the color scheme and lighting are being designed, the computationally expensive part (form factors) of the matrix in the linear system can be reused as the material properties are changed. For these applications the fastest possible general matrix solution is desirable.

In this chapter we have outlined important characteristics of the linear systems arising in radiosity applications and described iterative methods usually applied to solve theses systems. Moreover, we have proved that all eigenvalues of the radiosity coefficient matrix area real positive. This fact is essential for the application of faster iterative solvers to the radiosity systems such as the Chebyshev method [217].

The concepts presented in this chapter will be built upon in the following chapter to extend our investigation on the spectral characteristics of the radiosity systems and their effects on the convergence of the iterative methods. In addition, we are going to present two nonstationary methods, namely the Conjugate Gradient [118, 195] (previously examined in the graphics literature by Heckbert [116, 117]) and the Chebyshev [217] methods. We expect that these non-stationary methods may outperform the stationary and radiosity-specific methods previously described, which is borne out by the comparison results presented in the next chapter.

[8]Progressive techniques will initially converge faster because they can begin iterations before computing the entire matrix. If general matrix techniques are modified to gradually construct the matrix during initial iterations, then this advantage of progressive method goes away.

Chapter 8
Searching for Fast Solutions for Radiosity Systems

As we mentioned in the previous chapter, radiosity-specific methods (Sections 7.2.3 and 7.2.4) used to solve the radiosity system allow us to have a rough preview of the overall illumination of a scene after few steps of iteration. However, as mentioned by Cohen *et al.* [54], since few patches contribute to the overall illumination in the early stages of the iterative process (unless the iteration is started very close to the solution), detailed illumination effects, such as color bleeding, are lost. Radiosity-specific methods have to iterate until the desired convergence is reached to avoid that loss. This means that, even though they do not need to store all the form factors, radiosity-specific methods still have to compute all of them, possibly many times.

Suppose we want to generate several similar images to produce a computer animated sequence. Since there is some significant coherence from frame to frame, we do not need to preview all the frames. The first images can be used as a preview of the whole sequence. For these applications we would like to generate the images in the shortest possible time, i.e. to solve the radiosity linear systems as fast as possible. In this chapter we are going to show that general matrix methods (Section 7.2) can be faster than methods that converge in a smaller number of steps of iteration, namely the radiosity-specific methods[1].

Usually the form factors are computed on the fly when one uses a radiosity-specific method. The same approach can be applied if one uses a general matrix method. However, by storing the form factors we have the additional advantage of avoiding the cost of recomputing them repeatedly. In fact, several applications

[1]For this discussion we assume a conventional RISC architecture, where coherent memory access is vital.

present characteristics that make the off-line computation and storage of the form factors more convenient than their computation on demand:

- Simulations involving lighting and/or reflectance changes do not require new form factor computations in order to solve the radiosity system. Under these circumstances the only restriction to store the form factor matrix may be its size if a given application involves a large number of primitives and/or finer meshes. For these applications one can use techniques, such as hierarchical radiosity [110, 111, 206] and clustering [208, 209, 216], to reduce the the form factors storage costs. Although the order of the radiosity matrix associated with these applications is usually high, the nonzero entries are mostly located in diagonal blocks that can be stored without a significant memory cost and solved in parallel.

- Scene subdivision techniques, such as those used by the *FIAT* [78] and *Lucifer* approaches [76, 147], can be used to separate a large equation system into a collection of small subsystems. For example, recently Shinya *et al.* [196] proposed a modified hierarchical algorithm using virtual walls, which regards a large radiosity system as a collection of small subsystems. The subsystems interact with others only through input and output functions. Since the data necessary for each subsystem computation is completely localized, the database can be stored on disk. Shinya *et al.* claim that their algorithm can work in time complexity $O(n)$ for scenes divided into n subsystems, while classical hierarchical radiosity algorithms run with $O(n^2)$ complexity.

- The view-independent radiosity solution regarding the first pass of a multipass-method usually requires the tessellation of the environment in a coarse mesh [198], since view-dependent passes are going to be applied later to handle the high frequency signals. For these coarse meshes the dimensions of the corresponding radiosity matrices are fairly small, and they can be stored without a significant memory cost.

- For applications in which only small geometrical changes are made, acceleration techniques [22, 49, 83] can be used to speed the process of recomputing the

form factors. For example, if the geometry of a scene changes from one frame to another in the computer animation application mentioned earlier, only a small fraction of the form factors need to be recomputed due to the animation coherence. In this case most of the form factors can be precomputed in an off-line operation and stored. The remaining ones can be recomputed, without a significant overhead, using the techniques cited above.

For environments with high average reflectance and high level of occlusion, which occur very often in architectural and remote sensing applications [48], the iteration process takes a longer time to converge due to the characteristics of the radiosity matrix. For example, when the overall reflectance of an environment approaches 1.0, the spectral radius of the iteration matrix (for a stationary method like Gauss-Seidel) also approaches 1.0, and the linear solvers converge slowly. In this chapter we test the performance of a nonstationary general matrix method, the Conjugate Gradient method [118], and introduce another nonstationary general matrix method, namely the Chebyshev method [217], which can converge faster than any other iterative method mentioned in the graphics literature to date in those situations.

The remaining of this chapter is organized as follows:

- Section 8.1 discusses mathematical tools used to analyze most iterative solvers for linear systems.
- Section 8.2 presents the Conjugate Gradient method.
- Section 8.3 outlines the main properties of the Chebyshev polynomials.
- Section 8.4 introduces the Chebyshev method.
- Section 8.5 describes the parameters and procedures used in our investigation of faster solutions for radiosity systems.
- Section 8.6 examines the convergence of the iterative methods in small amounts of time, as opposed to a small number of iterations, and the effects of reflectance and occlusion changes on the performance of these methods.
- Finally, Section 8.7 summarizes the contents this chapter.

8.1 Residual Polynomials

Using the same notation as before, we consider a linear system in the form given by Equation 7.11. Moreover, an approximate solution $B^{(k)}$ (where k is the number of iterations made) for this system is assumed known. A new approximation $B^{(k+1)}$ is found by choosing a search direction $d^{(k)}$ and a stepsize α_k giving the length of the move in that direction, obtaining:

$$B^{(k+1)} = B^{(k)} + \alpha_k d^{(k)}, \quad k = 0, 1, \ldots \tag{8.1}$$

Recall that the residual vector $r^{(k)}$ is given by $r^{(k)} = E - GB^{(k)}$, and the error vector can be defined as:

$$e^{(k)} = B^{(k)} - B^* \tag{8.2}$$

where:

B^* = exact solution of the system.

In addition, assuming that G is a well-conditioned matrix (Section 7.1.2), it can be shown (Theorem 7.26 of Burden and Faires [42]) that the direction $d^{(k)}$ where the error vector $e^{(k)}$ decreases most rapidly is associated with the residual vector $r^{(k)}$. The residual vector is, therefore, commonly used for $d^{(k)}$. Taking this aspect into account, we can define $d^{(k)}$ as:

$$d^{(k)} = E - GB^{(k)} \tag{8.3}$$

An iterative method then consists of choosing an initial approximate vector $B^{(0)}$ and computing $B^{(k)}$, $k = 1, 2, ...$ until a satisfactory solution is reached. Different iterative methods are found depending on how α_k is computed in each step. One class of methods is constructed by choosing the stepsizes α_k to minimize a measure of the error in $B^{(k)}$. As described by Stiefel [217], recursively substituting Equation 8.1 into the definition of $d^{(k)}$ (Equation 8.3), gives:

$$d^{(k)} = P_k(G)d^{(0)} \tag{8.4}$$

where $P_k(G) = (I - \alpha_{k-1}G)(1 - \alpha_{k-2}G)\cdots(1 - \alpha_0 G)$.

The matrix polynomial $P_k(G)$ is called the residual matrix polynomial. If the matrix G is replaced by a scalar Λ and the identity matrix I by the real value 1, then we obtain:

$$P_k(\Lambda) = (1 - \alpha_{k-1}\Lambda)(1 - \alpha_{k-2}\Lambda)\cdots(1 - \alpha_0\Lambda) \tag{8.5}$$

which is the *residual polynomial* [101]. Note that $P_k(0) = 1$.

If G is diagonalizable then $G = SDS^{-1}$ where S is a nonsingular matrix and D is the diagonal matrix with the eigenvalues $\Lambda_1, \Lambda_2, ..., \Lambda_n$ of G on its diagonal. In this case $P_k(G) = SP_k(D)S^{-1}$, and to obtain the fastest reduction in the residual norm a residual polynomial method needs to present an optimality property, *i.e.* to select polynomials whose values $P_k(\Lambda_i)$ quickly go to zero as the degree of the polynomial increases. These values give the reduction in the i^{th} eigen-component of the residual after the k^{th} step of the iteration. In addition to this optimality property, a good residual polynomial should be computed using short recursions so that only a few of the previous residual vectors need to be stored. We use these properties to analyze the performance of the residual polynomial methods described in the next sections.

8.2 The Conjugate Gradient Method

8.2.1 Fundamentals

Linear stationary methods, such as Gauss-Seidel (Section 7.2.1) and SOR (Section 7.2.2), implicitly update solutions of a linear system of the form presented in Equation 7.11 by:

$$B^{(k+1)} = TB^{(k)} + \tilde{E} \tag{8.6}$$

where:

T = iteration matrix.

Nonstationary methods, such as the Conjugate Gradient method (CG), have an implicit iteration matrix T which changes with each iteration. When combined with a preconditioner[2], CG, originally developed by Hestenes and Stiefel [118], is probably

[2]Iterative methods usually involve a second matrix, a preconditioner, that transforms the coefficient matrix into one with a more favorable spectral distribution [21, 195].

the most powerful and effective iterative method for solving systems of the form:

$$G^s B = E \tag{8.7}$$

where:

G^s = known square symmetric positive definite (SPD) matrix,
B = vector of unknowns,
E = known vector of constants.

CG is a residual polynomial based method, which means that:

$$r^{(k)} = P_k(G^s) r^{(0)} \tag{8.8}$$

where:

$r^{(k)}$ = k-th residual vector,
P_k = k-th degree polynomial satisfying $P_k(0) = 1$.

Furthermore, CG is optimal in the sense that P_k minimizes the error vector (Equation 8.2) in the *energy norm* [195] given by:

$$||e^{(k)}||^2_{G^s} = {e^{(k)}}^T G^s e^{(k)} \tag{8.9}$$

The minimization is over a space of dimension k, so after k steps, at most, the solution is found. CG requires a certain number of iterations (which varies according to the distribution of the eigenvalues of G^s) before the accumulated information allows a rapid decrease in the residual norm. This is highly effective for systems requiring relatively accurate solutions. Moreover, the CG algorithm can be implemented using three–term recursions and hence using only three vectors of additional storage.

Before presenting the algorithm, it is useful to understand where some of its formulas come from. The reader interested in a more detailed derivation of these formulas, however, is referred to more comprehensive texts regarding the CG method [92, 118, 195].

Initially consider the definition of a quadratic form:

$$f(B) = \frac{1}{2} B^T G^s B - E^T B \tag{8.10}$$

where G^s is a matrix, and E and B are vectors.

The gradient of a quadratic form is defined to be:

$$f'(B) = \begin{bmatrix} \frac{\partial}{\partial B_1} f(B) \\ \frac{\partial}{\partial B_2} f(B) \\ \vdots \\ \frac{\partial}{\partial B_n} f(B) \end{bmatrix} \tag{8.11}$$

Applying Equation 8.10 to Equation 8.11, gives:

$$f'(B) = \frac{1}{2} G^{sT} B + \frac{1}{2} G^s B - E \tag{8.12}$$

Since G^s is symmetric, Equation 8.12 reduces to

$$f'(B) = G^s B - E \tag{8.13}$$

One can minimize $f(B)$ by setting $f'(B)$ equal to zero. Setting the gradient to zero, in turn, gives Equation 8.7. Therefore CG can be viewed as a minimization method in which solving a system described by Equation 8.7 is equivalent to minimizing Equation 8.10

In each step of the process CG generates a search direction $d^{(k)}$, then a step size α_k so that the updated solution vector given by:

$$B^{(k+1)} = B^{(k)} + \alpha_k d^{(k)} \tag{8.14}$$

will have minimal residual, over all α, represented by:

$$r^{(k)} = r^{(k-1)} - \alpha_k G^s d^{(k)} \tag{8.15}$$

or

$$r^{(k)} = E - G^s (B^{(k-1)} + \alpha_k d^{(k)}) \tag{8.16}$$

In order to prevent the "zig–zag" effect[3] that causes minimization methods to

[3]The function $f(B)$ (Equation 8.10) forms a trough, and the search directions bounce back and forth between the sides of the trough while making little progress along its length [195].

have slow convergence, search directions are chosen to be G-conjugate [92, 189]:

$$d^{(i)^T} G d^{(j)} = 0 \qquad for\ i \neq j \tag{8.17}$$

It can be shown [195] that it suffices to make $d^{(k+1)}$ G–conjugate to $d^{(k)}$, in which case $d^{(k+1)}$ is G–conjugate to all previous search directions. It can also be shown [92] that choosing α to minimize $f(B^{(k-1)} + \alpha d^{(k)})$ will also make $B^{(k)}$ minimize f over all $B \in \text{span}[d^{(1)}, d^{(2)}, \ldots, d^{(k)}]$.

Then, according to the derivation presented by Golub and Van Loan [92] and by Shewchuk [189], the choice of α_k that minimizes $f(B)$ is given by:

$$\alpha_k = \frac{r^{(k)^T} r^{(k-1)}}{d^{(k)^T} G^s d^{(k)}} \tag{8.18}$$

The search directions, in turn, are updated using the residuals such that:

$$d^{(k+1)} = r^{(k+1)} + \beta_{k+1} d^{(k)} \tag{8.19}$$

where β represents a Gram-Schmidt conjugation constant used to ensure that r^k and r^{k-1} are orthogonal [21, 195], and it is given by:

$$\beta_k = \frac{r^{(k+1)^T} r^{(k+1)}}{r^{(k)^T} r^{(k)}} \tag{8.20}$$

In summary, CG, as pointed out by Shewchuk [189], is simply a method of conjugate directions where the search directions are constructed by conjugation of the residuals. Equations 8.14, 8.16, 8.19, 8.18 and 8.20 form the basis of the CG algorithm, as described in the original paper by Hestenes and Stiefel [118]. It is relevant to note that, by computing residuals recursively, only one matrix-vector multiplication is required per iteration [92].

8.2.2 Application to the Radiosity Problem

As mentioned in the previous chapter, although the radiosity coefficient matrix G is not symmetric, it can be made symmetric (Section 7.4), *i.e.* $G^s = QG$, where Q is a diagonal preconditioner matrix [117, 243]. This preconditioning operation performed on G gives a transformed system [21] of the form $G^s B = QE$. Then, the algorithm

used to implement CG in a radiosity context becomes:

```
1     compute G^s
2     scale E
3     for (all i)
4           B_i^(0) = starting guess
5     compute ΔB^(0) = E − G^s B^(0)
6     temp1 = 0
7     for (each i)
8           D_i = ΔB_i
9           temp1 = temp1 + ΔB_i ΔB_i
10    while (not converged)
11          for (each i)
12                W_i = 0
13                for (each j)
14                      W_i = W_i + G^s_ij D_j
15          temp2 = 0
16          for (each i)
17                temp2 = temp2 + W_i D_i
18          α = temp1/temp2
19          for (each i)
20                B_i = B_i + α D_i
21                ΔB_i = ΔB_i − α W_i
22          temp2 = 0
23          for (each i)
24                temp2 = temp2 + ΔB_i ΔB_i
25          β = temp2/temp1
26          temp1 = temp2
27          for (each i)
28                D_i = ΔB_i + β D_i
```

Recall that the diagonal entries of Q correspond to the quotient of the area, A, and the reflectance, ρ, of the patches. Then, it is not necessary to explicitly form Q to apply CG to the radiosity problem. It is only necessary to be able to compute the effect of applying it to a vector. For instance, in order to obtain G^S we can simply multiply each element G_{ij} of a row of G by the quotient $\frac{A_i}{\rho_i}$. Similarly, to scale the vector E in the right-hand side we can simply multiply each element E_i of E by the quotient $\frac{A_i}{\rho_i}$.

In the above algorithm ΔB represents the residual and D represents the vector of search directions. Usually we use the vector of irradiances, E, as our starting guess for the vector of unknowns (radiosities, B). However, we have performed experiments that showed that the use of a starting guess which takes into account a modified ambient term may result in a faster convergence. This modified ambient term, $Ambient_E$, is computed replacing ΔB_i by E_i in Equation 7.14. Its computation does not add a significant cost to the algorithm, specially considering that in a given environment there are usually few patches that emit light, *i.e.* few patches i with $E_i \neq 0$. To apply CG with this starting guess we replace line 4 of the above algorithm by:

$$B_i^{(0)} = E_i + \rho_i Ambient_E$$

8.2.3 Convergence Considerations

The speed of convergence of CG depends on the overall distribution of eigenvalues of the matrix G^s. Upper bounds for the number of iterations are often based on the spectral condition number[4] of G^s [21, 92, 189]. Since G^s is symmetric and positive definite, its spectral condition number, as mentioned by Hageman and Young [101], is given by:

$$\kappa(G^s) = \frac{\Lambda_{max}(G^s)}{\Lambda_{min}(G^s)} \tag{8.21}$$

where estimates for the extremal eigenvalues (Λ_{max} and Λ_{min}) can be obtained using the Gerschgorin Circle Theorem (Section 7.5).

According to Shewchuk [189] the number of iterations k needed to reduce the

[4]The spectral condition number of a matrix G is defined by $\kappa(G) = \| G \|_2 \| G^{-1} \|_2$, where the spectral norm $\| \cdot \|_2$ of G is defined as $\| G \|_2 = (\varrho(GG^T))^{\frac{1}{2}}$.

residual by a factor of ϵ in the *energy norm* can be estimated by:

$$k \leq \lceil \frac{1}{2} * \sqrt{\kappa(G^s)} * \ln(\frac{2}{\epsilon}) \rceil \tag{8.22}$$

The relation above, based on Theorem 10.2.5 of Golub and Van Loan [92], infer that the number of iterations is proportional to $\sqrt{\kappa(G^s)}$. However, if the extremal eigenvalues are well separated, CG can provide a superlinear convergence rate[5] since it implicitly builds up information about the spectrum of the matrix G^s as its iterations proceed. As mentioned by Barret *et al.* [21], CG tends to eliminate the components of the error in the direction of eigenvectors associated with extremal eigenvalues first. After these have been eliminated, the method progressively moves towards the interior of the region containing the spectrum, *i.e.* the convergence rate becomes dependent on a reduced system with a smaller condition number [65].

It is not *a priori* clear that CG is optimal for radiosity systems, which require low accuracy solutions with only a few sweeps through the matrix. The fact that the radiosity matrices tend to have the eigenvalues tightly clustered around a few values (section 8.6.2), however, suggests that upper bound for the number of iterations needed for radiosity problems may be lower than the one given by the relation above. We examine this issue in more detail in Section 8.6.2.

8.3 Chebyshev Polynomials

Before introducing the Chebyshev method, we need to present the definition and main properties of the Chebyshev polynomials. These polynomials represent the basis of the Chebyshev method, and the knowledge about their behavior is fundamental for the analysis of the convergence of this iterative method.

The Chebyshev polynomials Y_m [42] for the interval $[-1, 1]$ are defined by:

$$Y_m(x) = \cos[m \cos^{-1}(x)] \qquad for\ each\ m \geq 0 \tag{8.23}$$

Since they are orthogonal with respect to the weight function $w(x) = (1 - x^2)^{-1/2}$, they can be computed using the three term recursion:

$$Y_0(x) = 1,\ Y_1(x) = x,\ Y_{m+1}(x) = 2xY_m - Y_{m-1};\ \ m \geq 1. \tag{8.24}$$

[5]Convergence at a rate that increases per iteration [21, 58].

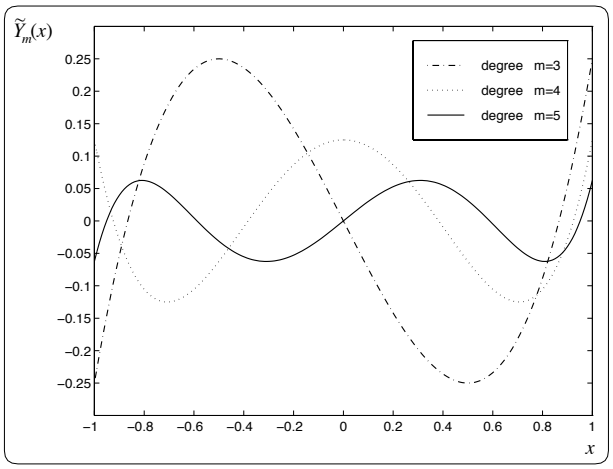

Figure 8.1: Monic Chebyshev polynomials.

Most importantly, Chebyshev polynomials have a *minimax* property: of all m^{th}–degree polynomials with leading coefficient 1, $2^{1-m}Y_m$ has the smallest maximum norm in $[-1, 1]$. The value of its maximum norm is 2^{1-m}. Figure 8.1 shows the graphs of monic Chebyshev polynomials[6] ($\tilde{Y}$) of degree 3, 4 and 5.

Given an interval $[\mu, \nu]$ containing the eigenvalues of a matrix G, the Chebyshev method uses residual polynomials based on monic Chebyshev polynomials shifted from the interval [-1,1] to the interval $[\mu, \nu]$, and scaled so that $\tilde{Y}_m(0) = 1$. This gives:

$$\tilde{Y}_m(\Lambda) = \frac{Y_m\left(\frac{\nu+\mu-2\Lambda}{\nu-\mu}\right)}{Y_m\left(\frac{\nu+\mu}{\nu-\mu}\right)} \tag{8.25}$$

The three term recursion can similarly be translated to the new variables, giving the Chebyshev algorithm presented in the next section.

8.4 The Chebyshev Method

8.4.1 Fundamentals

The Chebyshev method is also a nonstationary method based on residual polynomials. It is directly applicable to nonsymmetric matrices like the radiosity coefficient matrix. However, it requires estimates of the smallest and largest eigenvalues of

[6]Polynomials with leading coefficient 1 [42].

the corresponding coefficient matrix [21, 217]. The iterative process used by the Chebyshev method to solve a linear system of the form presented in Equation 7.11 is characterized by:

$$B^{(k+1)} = B^{(k)} + \Delta D^{(k)} \qquad \Delta D^{(k)} = (\frac{1}{q_k})(r^{(k)} + p_k \Delta D^{(k-1)}) \tag{8.26}$$

where $r^{(k)}$ is the residual, and the scalars q_k and p_k are the coefficients of the residual polynomials. Equation 8.26 is known as the Stiefel iteration [186, 217].

Recall that to obtain fastest reduction in the residual norm a residual polynomial method needs to select polynomials whose ordinates quickly go to zero on the spectrum of the coefficient matrix G as the degree of the polynomial increases. As we proved in Section 7.4, for radiosity problems the eigenvalues are all real and positive. Thus, given a knowledge of an interval $[\mu, \nu]$ containing the spectrum of G, we select polynomials $\tilde{Y}_k$ that have their maximum absolute value on $[\mu, \nu]$ minimal over all monic polynomials of degree k.

The classical algorithm corresponding to the Chebyshev method is given by:

1 $B^{(0)} arbitrary$
2 $r^{(0)} = E - GB^{(0)}$
3 $B^{(1)} = B^{(0)} + \gamma^{-1} r^{(0)}$
4 $r^{(1)} = E - GB^{(1)}$
5 $\omega^{(0)} = 2/\gamma$
6 $k = 1$
7 **while** (not converged)
8 $\quad \Delta D^{(k)} = \omega^{(k)} r^{(k)} + (\gamma \omega^{(k)} - 1) \Delta D^{(k-1)}$
9 $\quad B^{(k+1)} = B^{(k)} + \Delta D^{(k)}$
10 $\quad r^{(k+1)} = E - GB^{(k+1)}$
11 $\quad k = k + 1$

In the above algorithm:

$$\gamma = \frac{\beta}{\alpha} \tag{8.27}$$

$$\alpha = \frac{2}{\Lambda_{\max} - \Lambda_{\min}} \tag{8.28}$$

$$\beta = \frac{\Lambda_{\max} + \Lambda_{\min}}{\Lambda_{\max} - \Lambda_{\min}} \tag{8.29}$$

$$\omega^{(k)} = (\gamma - \frac{1}{4\alpha^2}\omega^{(k-1)})^{-1} \tag{8.30}$$

As mentioned in the previous section, in addition to the "minimax" property, Chebyshev polynomials can also be computed using a three term recursion which implies that the iteration can be implemented using only three additional vectors of storage. Note that the updates in lines 8-10 are vector updates. Although, in practice, $\Lambda_{\max}$ and $\Lambda_{\min}$ may be replaced by ν and μ, where $0 < \mu \leq \Lambda_{\min} < \Lambda_{\max} \leq \nu$, the rate of convergence of Chebyshev methods is maximal when $\mu = \Lambda_{\min}$ and $\nu = \Lambda_{\max}$, and it can diverge if, for example, Λ_{max} is underestimated by ν. Therefore, to implement the Chebyshev Method successfully, we need estimates of the extremal eigenvalues of the matrix G. However, since the Chebyshev polynomials grow very rapidly outside of the interval [$\Lambda_{\max}$,$\Lambda_{\min}$] (Figure 8.13), if we underestimate the maximal eigenvalues, we will notice very quickly the sudden increase of the error norm. This would allow us to immediately reset the estimates and proceed with the new estimates.

8.4.2 Application to the Radiosity Problem

Since the radiosity coefficient matrix G is a square nonsingular real matrix, the adaptation of the Stiefel iteration (Equation 8.26) to the radiosity problem becomes straightforward, resulting in the iterative process characterized by:

$$B^{(k+1)} = B^{(k)} + \Delta D^{(k)} \qquad \Delta D^{(k)} = q_k^{-1}(\Delta B^{(k)} + p_k \Delta D^{(k-1)}) \tag{8.31}$$

where $\Delta D^{(k)}$ is a correction vector, $\Delta B^{(k)} = E - G * B^{(k)}$ is the residual, and q_k and p_k are coefficients of the residual polynomials.

In order to adapt the Chebyshev algorithm to the radiosity context, $\Lambda_{\max}$ and $\Lambda_{\min}$ have been replaced by ν and μ, where $0 < \mu \leq \Lambda_{\min} < \Lambda_{\max} \leq \nu$. As described in Section 7.5, using the Gerschgorin Circle Theorem the extremal eigenvalues may be

approximated by $1.0 \pm \rho_{max}$, where ρ_{max} is the highest reflectance in the environment, which may correspond to the reflectance of a single patch or of a group of patches.

Using the parameters ν and μ above, the Chebyshev algorithm adapted to the radiosity context becomes:

1 **for** (each i)
2 $B_i^{(0)} =$ starting guess
3 **compute** $\Delta B^{(0)} = E - G\,B^{(0)}$
4 $\alpha = 2/(\nu - \mu)$
4 $\beta = (\nu + \mu)/(\nu - \mu)$
6 $\gamma = \beta/\alpha$
7 **for** (each i)
8 $\Delta D_i = \gamma \Delta B_i$
9 $B_i^{(1)} = B_i^{(1)} + \Delta D_i$
10 **compute** $\Delta B^{(1)} = E - G\,B^{(1)}$
11 $\omega = 4/(\nu + \mu)$.
12 $k = 1$
13 **while** (not converged)
14 $\omega = (\gamma - \frac{1}{4\alpha^2}\omega)^{-1}$
15 **for** (each i)
16 $\Delta D_i = \omega * \Delta B_i^{(k)} + (\gamma\omega - 1)\Delta D_i$
17 $B_i^{(k+1)} = B_i^{(k)} + \Delta D_i$
18 **compute** $\Delta B^{(k+1)} = E - G\,B^{(k)}$
19 $k = k + 1$

The increase of the highest reflectance of a given environment does not necessarily change its overall or average reflectance considerably, specially if only the reflectance of few small patches is changed. In this case, it is very likely that the increase of the highest reflectance does not prompt a noticeable change in the convergence either, as shown by preliminary experiments. On the other hand, if we increase the average reflectance, expressed in terms of the ρ_{avg} (Equation 7.16), then slow convergence

results. Since the increase of the highest reflectance may not change the average reflectance considerably, as we described above, we performed experiments using ρ_{avg} instead of ρ_{max} to estimate the eigenvalues. The better results obtained for all the cases tested (Section 8.5.2), determined the following choice of Chebyshev parameters:

$$\nu = 1.0 + \rho_{\text{avg}} \tag{8.32}$$

$$\mu = 1.0 - \rho_{\text{avg}} \tag{8.33}$$

In all the cases tested, the use of a starting guess which takes into account the modified ambient term, $Ambient_E$ (section 8.2.2), gives better results than a starting guess which uses only the vector of emittances. To use this starting guess we replace line 2 of the above algorithm by:

$$B_i^{(0)} = E_i + \rho_i Ambient_E$$

8.4.3 Convergence Considerations

Heuristically, we can compare the CG and Chebyshev methods in terms of their underlying residual polynomials [21]. For CG the polynomials tend to rapidly converge to zero at the extremal eigenvalues, then progressively move towards the interior of the region containing the spectrum, zeroing out eigenvalues as iterations proceed. Chebyshev polynomials tend to uniformly drive all the eigenvalues to zero at the same rate. For solutions requiring high accuracy, the CG algorithm is usually the faster of the two since it implicitly creates the "optimal" polynomial on each step. For the low accuracy solutions needed in radiosity, however, it is more important to quickly reduce the residual corresponding to all the eigenvalues. So we expect the Chebyshev iteration to be faster, which is borne out by the testing results presented in the next section.

8.5 Testing Parameters and Procedures

We compared the following algorithms, using explicitly stored form factors:

- Gauss-Seidel (GS) described in Section 7.2.1,
- Standard Progressive Refinement (PR) described in Section 7.2.3,

- Overshooting (FEDA) described in Section 7.2.4,
- Conjugate Gradient (CG) described in Section 8.2,
- Chebyshev (CHEBY) described in Section 8.4.

The starting guesses were chosen in order to obtain the best possible rate of convergence for each tested algorithm. Consequently, the initial error norm is not the same for curves on a single graph. The starting guess used for Gauss-Seidel and Conjugate Gradient algorithms was the vector of emittances and for the radiosity-specific methods we used a vector of zeros. For the Chebyshev algorithm we used the starting guess given by $B_i^{(0)} = E_i + \rho_i Ambient_E$ (Section 8.4.2).

8.5.1 Performance Measurement

The general matrix methods check the convergence after a complete sweep of the coefficient matrix, *i.e.* one iteration. The radiosity-specific methods perform this check after one relaxation step, *i.e.* one step of iteration. To make our measurement comparable we count steps of iteration. In this context an iteration of a general matrix method corresponds to n steps of iteration, where n is the order of the coefficient matrix.

To measure the time we start the clock at the beginning of a cycle of k steps of iteration, and stop it after k steps of iteration. For the general matrix methods tested we use $k = n$. We check for k at each step in order to make the timing overhead the same for all methods. In addition, the error norm used as stopping criterion is computed outside of the timing cycles. The time measurements are given by elapsed CPU time on a SGI Challenge (20-R4400). All the algorithms were implemented using the same software guidelines to avoid differences that could affect the timing.

8.5.2 Test Cases

The test scene used in our experiments consists of a sphere centered in a "Cornell box" (Figure 8.2). The sphere was divided into 128 patches and the faces of the surrounding cube were divided into 144 patches forming a total of 992 patches. The light source corresponds to 16 patches on the center of the box "ceiling" (face 6). The values used for d and h were 3.0 and 6.0 respectively.

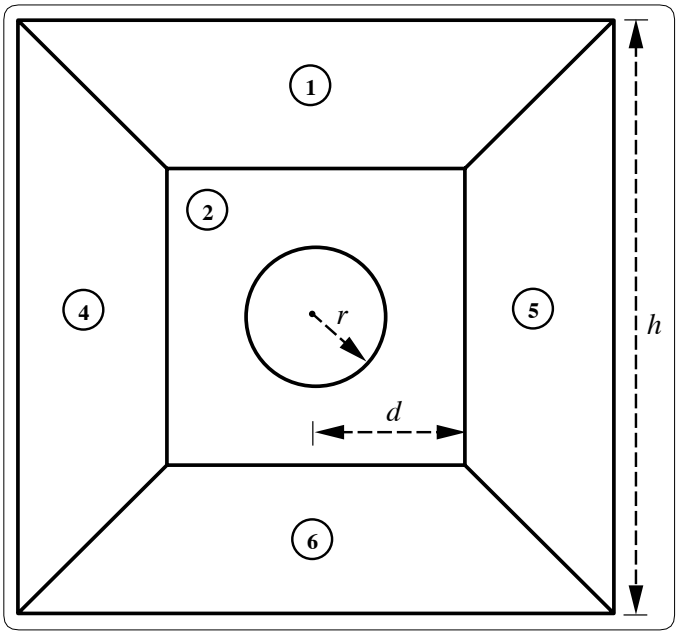

Figure 8.2: Sketch of the test scene (face 3 is frontal).

By assigning different values for the reflectances of the surfaces of the test scene, cube faces and sphere, we can vary ρ_{avg}. By changing the sphere radius, using $r = 1.0$ and $r = 2.0$, we can change the level of occlusion in the scene, which results in different densities (δ) for the coefficient matrix, 70% and 53% respectively. The representative test cases used in our experiments are shown in Table 8.1.

case	ρ_{avg}	$\delta(\%)$
A	0.24	53
B	0.46	53
C	0.77	53
D	0.88	53
E	0.78	70
F	0.89	70

Table 8.1: Test cases.

8.5.3 Convergence Checking

We use as our stopping criterion the largest unshot energy, *i.e.* the L_∞ norm of the vector with components $r_i A_i$, in which r_i represents the residual and A_i represents the area of patch i, given by:

$$\xi_\infty = \max_{1 \leq i \leq n} |r_i A_i| \tag{8.34}$$

If ξ_∞ is smaller than a given tolerance we stop the iterations. The value assigned to the tolerance depends on how visually close to true solution one wants the final image be. In general it is not necessary to use very low tolerance as in most numerical applications. We used a tolerance, *tol*, equal to 10^{-3}, but present the full convergence histories so that the methods can be compared for larger tolerances.

For the sake of completeness, besides the L_∞ norm of the unshot energy used as the stopping criterion in the experiments described in the next section, we also computed three other error norms: the L_2 norm of the residual, the *RMS* error adapted to the radiosity context, and the percentage of "reflected" radiosity (*error*).

The L_2 norm of the residual r is given by:

$$\varphi_2 = (\sum_{i=1}^{n} r_1^2)^{\frac{1}{2}} \tag{8.35}$$

Finally, the RMS error, as defined by Feda and Purgathofer [73], is also based on the L_2 norm , and consists in calculating the square root of the area weighted mean of the square of individual errors:

$$RMS_{error} = \sqrt{\frac{\sum_{i=1}^{n} (B_i^* - B_i)^2 A_i}{\sum_{i=1}^{n} A_i}} \tag{8.36}$$

where:

B_i = computed radiosity of patch i.
B_i^* = exact (converged) radiosity of patch i.
A_i = area of patch i.

The *error* as defined by Goertler *et al.* [90], is given by:

$$error = \frac{\sqrt{\sum_{i=1}^{n}(B_i^* - B_i)^2}}{\sqrt{\sum_{i=1}^{n}(B_i^* - E_i)^2}} \tag{8.37}$$

where:

B_i = computed radiosity of patch i.
B_i^* = exact (converged) radiosity of patch i.
E_i = irradiance of patch i.

The norm of the residual does not account for the areas of the patches, and the difficulty with the last two error norms is that the converged true solution used to compute them is not available in practice. For testing purposes we can approximate the exact solution using the Gauss-Seidel method iterated until the residual vector has norm less than 10^{-10} or smaller. The conclusion drawn from using these error norms were basically the same as the conclusions obtained using the largest unshot energy as the stopping criterion. For this reason, we did not include the results regarding the use of these three error norms in this dissertation. However, the reader interested in these experiments is referred to [16].

8.6 Testing Results and Discussion

Testing was performed to compare the performance of the five algorithms previously described. In particular, we examine the effect of the matrix spectrum (which depends on the reflectances and level of occlusion in the scene). Numerical testing is necessary because most theory about the convergence rates of these methods deals with *asymptotic* convergence. For the low accuracy solutions needed in graphics radiosity problems, however, oftentimes an adequate solution is available before the asymptotic convergence region is approached.

The Gauss-Seidel and CG methods are *parameter-free*, that is, there are no algorithmic parameters which must be set by the user. Chebyshev requires estimates of the smallest and largest eigenvalues, but for this application those can be set automatically as was done here, by using $1 \pm \rho_{avg}$. The overshooting method can be fine-tuned by different choices of the overshooting parameter, in the same way that

Gauss-Seidel is generalizable to the SOR method (Section 7.2.2). However, like the SOR method, the optimal choice of parameters is unknown except for a few special cases. The version tested here automatically selects the overshooting parameter.

8.6.1 Steps of Iteration *vs* Time

Usually progressive refinement or overshooting converge in fewer steps of iteration than the other three methods. However, counting steps of iteration does not account for the different amounts and types of work performed on each step. So a method that converges in fewer steps of iteration may in fact require more overall time. The experiments, whose results are presented in Figures 8.3, 8.4, 8.5 and 8.6, show that this distinction does occur in radiosity applications.

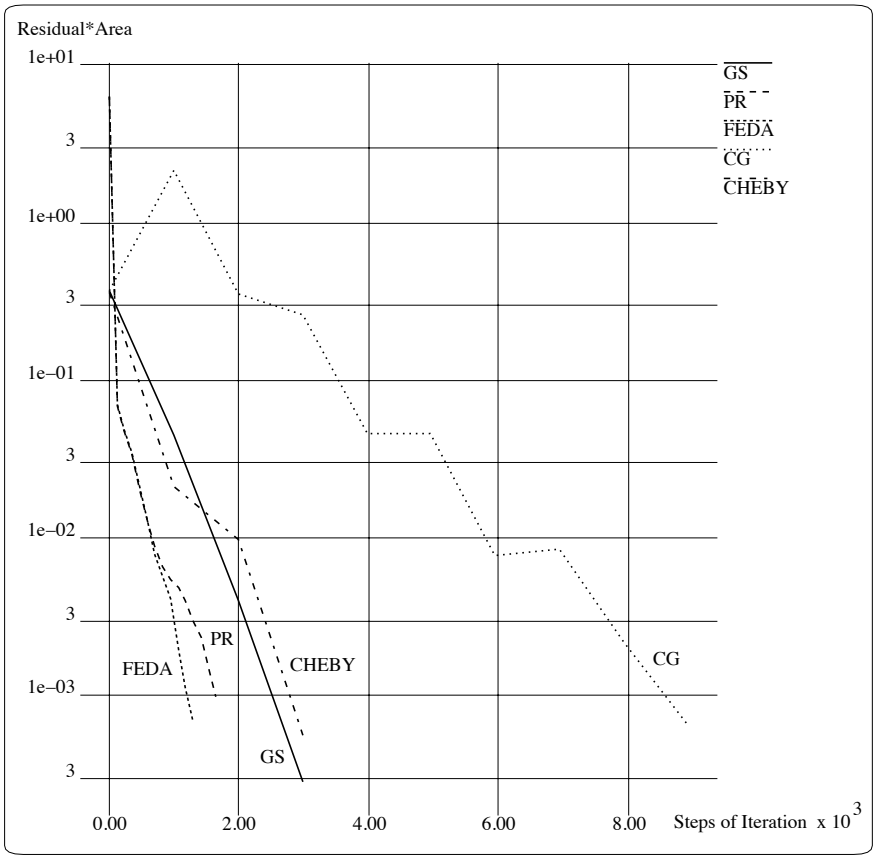

Figure 8.3: Case A ($\rho_{avg} = 0.24$, $\delta = 53\%$) with performance measured in steps of iteration.

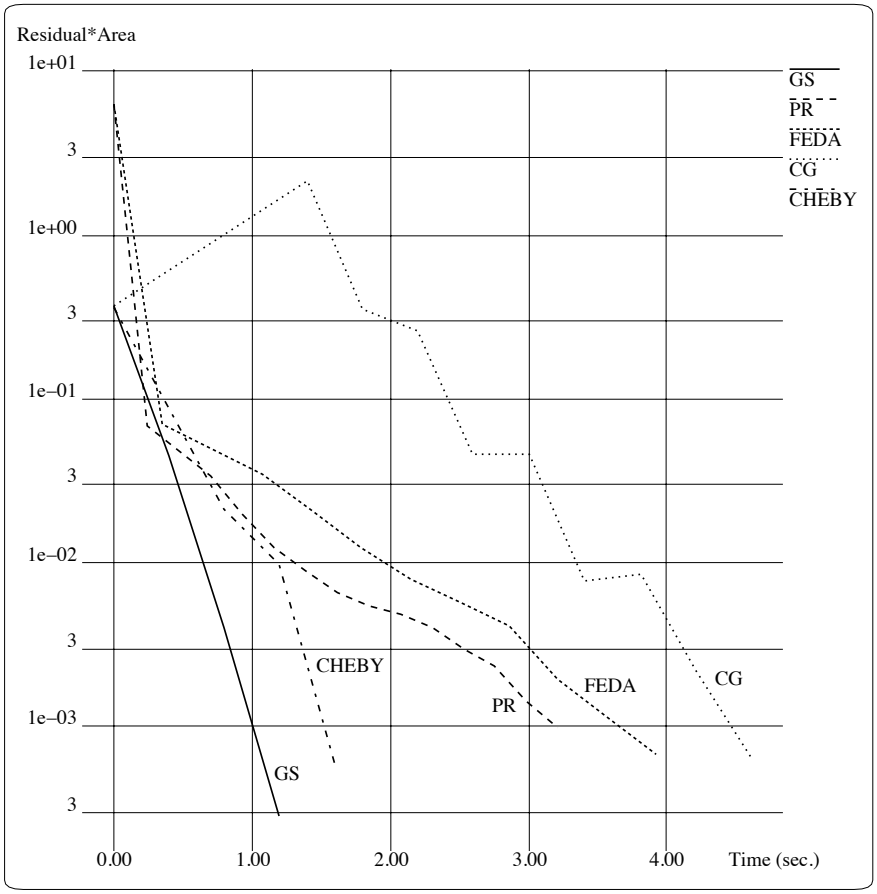

Figure 8.4: Case A ($\rho_{avg} = 0.24$, $\delta = 53\%$) with performance measured in CPU time.

Figure 8.3 shows that, for test case A ($\rho_{avg} = 0.24$, $\delta = 53\%$), progressive refinement (PR) and overshooting (FEDA) methods converge in fewer steps of iteration for $tol = 10^{-3}$. However, Figure 8.4, showing the same convergence history as in Figure 8.3, but plotted against elapsed CPU time, shows that Gauss-Seidel (GS) and Chebyshev (CHEBY) methods converge in less CPU time. When the average reflectance of the environment is increased, the difference becomes even more noticeable, as shown in Figures 8.5 and 8.6 regarding test case B ($\rho_{avg} = 0.46$, $\delta = 53\%$).

Counter-intuitively, Gauss-Seidel and Chebyshev methods require more operations than the radiosity-specific methods, but require less CPU time. The main reason is the different amounts of pipelining and data locality the algorithms allow. Note that in the progressive refinement method, we search for the patch with largest amount of

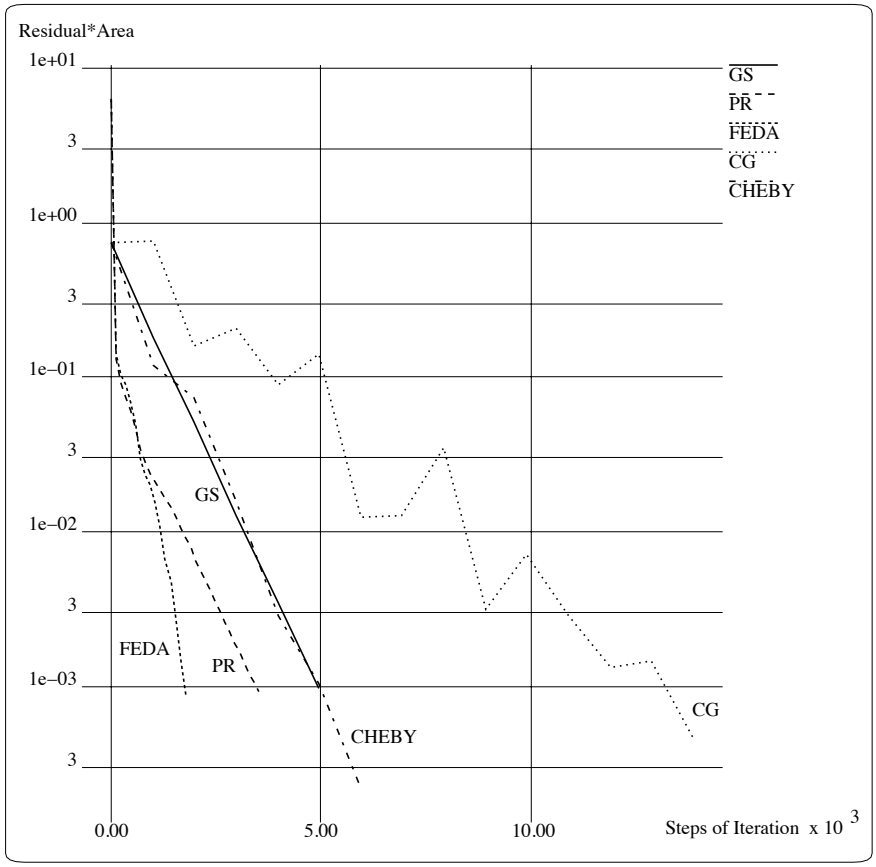

Figure 8.5: Case B ($\rho_{avg} = 0.46$, $\delta = 53\%$) with performance measured in steps of iteration.

unshot radiosity, which involves traversing a potentially large amount of data without performing any operations on it that decrease the residual. The general matrix methods, by contrast, simply process each row of the matrix in order. Although this may mean processing rows whose corresponding patch has no unshot radiosity remaining, in practice performance is enhanced. By avoiding the search phase, the computations can be better pipelined by compilers, and all data which is brought into the processor is actually used in improving the solution rather than searching for the next row to handle.

Furthermore, the innermost loop of the progressive refinement and overshooting methods consists of a *saxpy* operation, *i.e.* a vector update of the form $\vec{y} = \vec{y} + \alpha\vec{x}$,

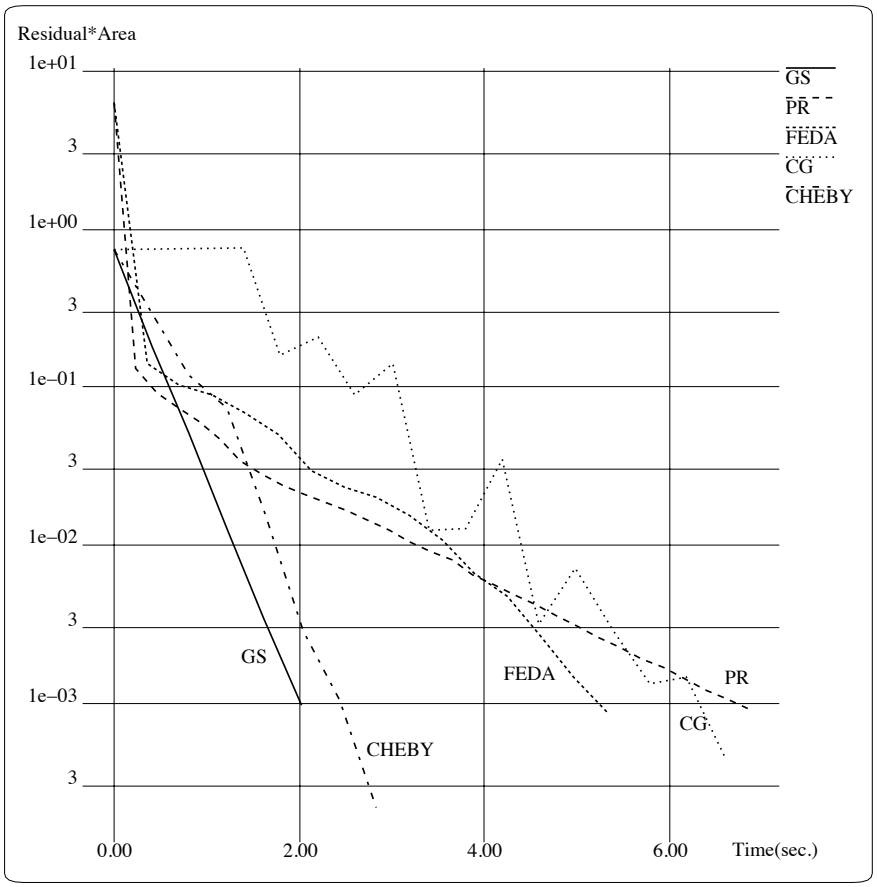

Figure 8.6: Case B ($\rho_{avg} = 0.46$, $\delta = 53\%$) with performance measured in CPU time.

where $\vec{x}$ and $\vec{y}$ are vectors and α is a scalar. This operation entails $4n$ memory references (n each for reading ρ_j, $\triangle B_j$, F_{ij}, and an additional n for writing $\triangle B_j$) and $3n$ floating point operations (flops). By contrast the Gauss-Seidel, Chebyshev and CG methods have an inner product as the innermost loop. For the last two methods this entails $2n$ memory references and $2n$ floating point operations, because quantities not indexed by the innermost loop are kept in registers and so do not require a memory reference. In particular, the carry-around scalar that the inner product is summed into, and the reflectance ρ_i, are kept in registers. Hence, the ratio of memory references to flops is 4/3 for the radiosity-specific methods, while the ratio is 1 for the general matrix methods.

This means that the general matrix methods better utilize data locality, getting

more flops out of data in the cache or registers before having to read or write new cache lines. Note that the better data reuse of the general matrix methods is not *a priori* evident from examining the algorithms. It is possible that in progressive refinement and overshooting methods, only a few patches are selected to shoot out radiosity over and over again. In that case, the data associated with those patches would likely remain in cache, potentially giving better data locality properties. Our experiments show that this is not the case in practice, however.

Usually over 90% the patches are selected the same number of times ± 1. Furthermore, the patches selected most often are only selected a few more times than the average. For example, Figure 8.7 shows the number of times each patch is selected to shoot out radiosity in order to solve the system regarding test case E ($\rho_{avg} = 0.78$, $\delta = 70\%$).

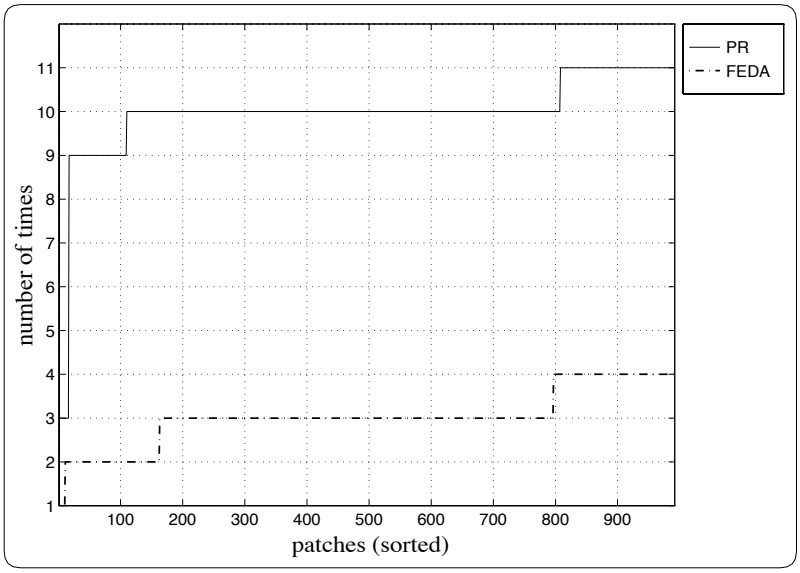

Figure 8.7: Number of times that patches are selected by progressive refinement (PR) and overshooting (FEDA) methods to shoot out radiosity in order to solve the system regarding test case E ($\rho_{avg} = 0.78$, $\delta = 70\%$). Patches are sorted on ascending order of selection times.

8.6.2 Effects of Reflectance and Occlusion on the Convergence

Figures 8.8 and 8.9 show the performance of the tested methods as the average reflectance is increased and matrix density (associated with the level of occlusion in the environment) is kept fixed at $\delta = 53\%$. Figures 8.10 and 8.11 do the same for $\delta = 70\%$. High average reflectances cause a larger number of interreflections, causing the eigenvalues to become more spread out, which in turn slows the convergence (Section 7.3.2). Figure 8.12 shows the distribution of eigenvalues of the radiosity matrix as the overall reflectance increases and the matrix density is kept fixed at $\delta = 53\%$.

The increase of reflectance is especially deleterious for PR. Because PR selects

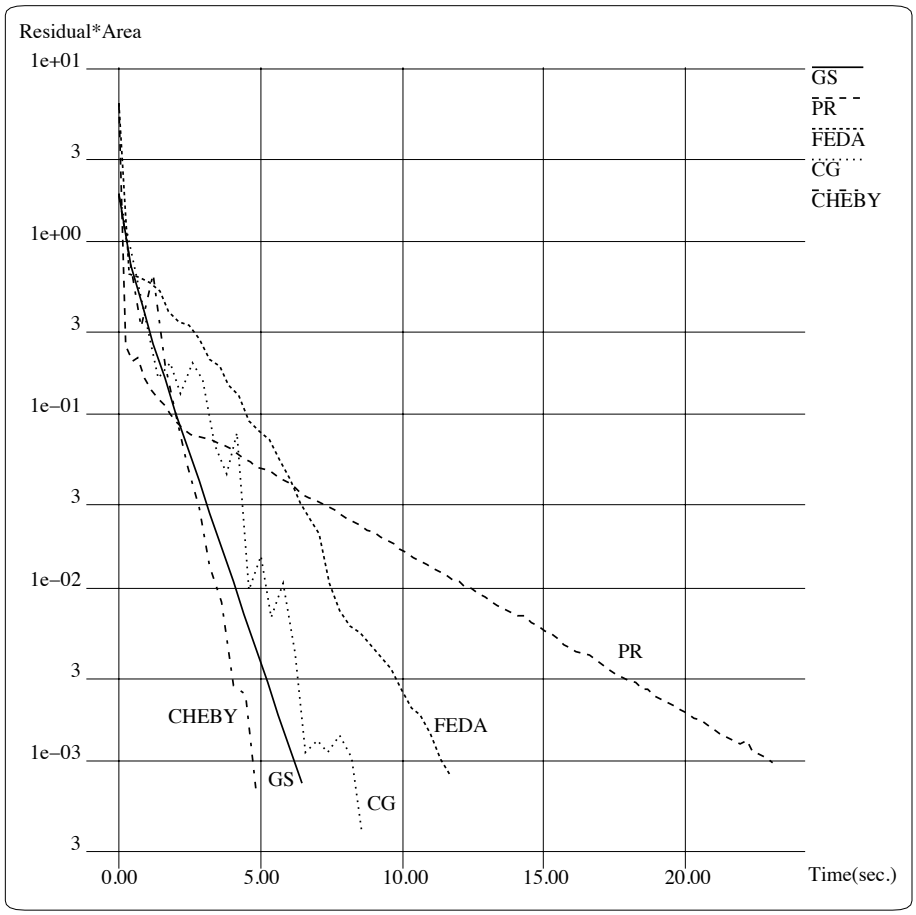

Figure 8.8: Case C ($\rho_{avg} = 0.77$, $\delta = 53\%$) with performance measured in CPU time.

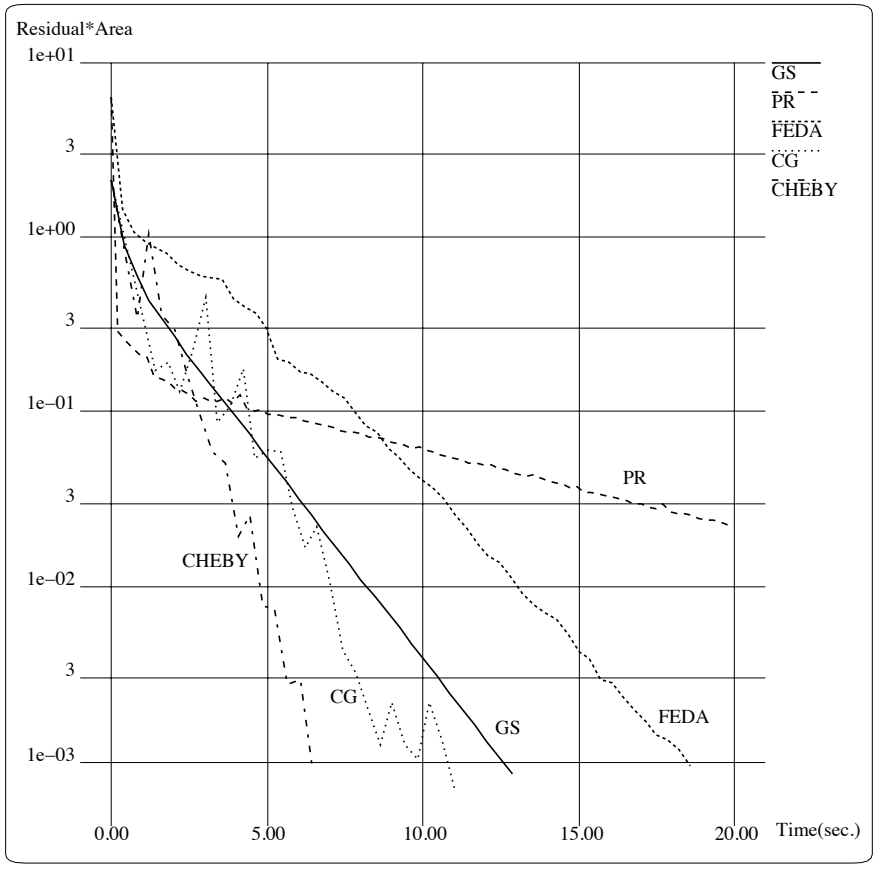

Figure 8.9: Case D ($\rho_{avg} = 0.88$, $\delta = 53\%$) with performance measured in CPU time.

which patch to process on each step, it is a nonstationary method that actually changes its innermost loop depending on the specific data of the problem. This means that it is not as amenable to analysis as the Chebyshev or CG methods, which are expected to take more steps of iteration as the eigenvalues get spread out.

As reflectance increases the spectral radius of the Gauss-Seidel iteration matrix approaches 1.0, and its relative performance decreases. The Gauss-Seidel method has a linear convergence rate that directly depends on the spectral radius of the iteration matrix. Although, relative to Gauss-Seidel, the Chebyshev method is not as sensitive to the increase in the average reflectance, it also has worsening performance. Figure 8.13 shows that, for scenes with higher values for the average reflectance, the Chebyshev method needs to use polynomials of higher degree in order to converge.

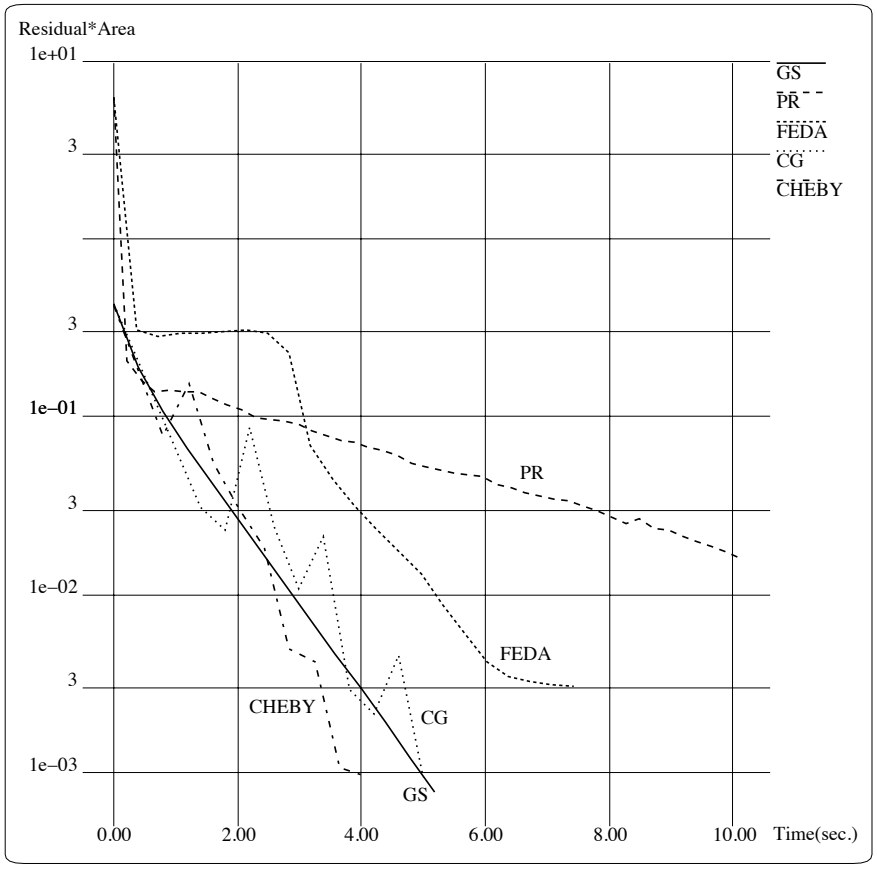

Figure 8.10: Case E ($\rho_{avg} = 0.78$, $\delta = 70\%$) with performance measured in CPU time.

The vertical bars are used in Figure 8.13 to mark the limits of the spectrum (minimum and maximum eigenvalues) of the radiosity matrix. As we can see, the ordinates of the polynomials go to zero in a region not entirely coincident with the spectrum of matrix. Recall that, to obtain fastest reduction in the residual norm, the Chebyshev method needs to select polynomials whose ordinates quickly go to zero on the spectrum of the coefficient matrix as the degree of the polynomial increases. These aspects indicate that the use of tighter estimates for the eigenvalues could reduce the gap between the region in which the polynomials go to zero and the spectrum of the matrix. This, in turn, would result in a faster convergence for the Chebyshev method.

By increasing the level of occlusion in a scene, we reduce the density of the radiosity coefficient matrix. Figure 8.14 shows the eigenvalues distribution as the den-

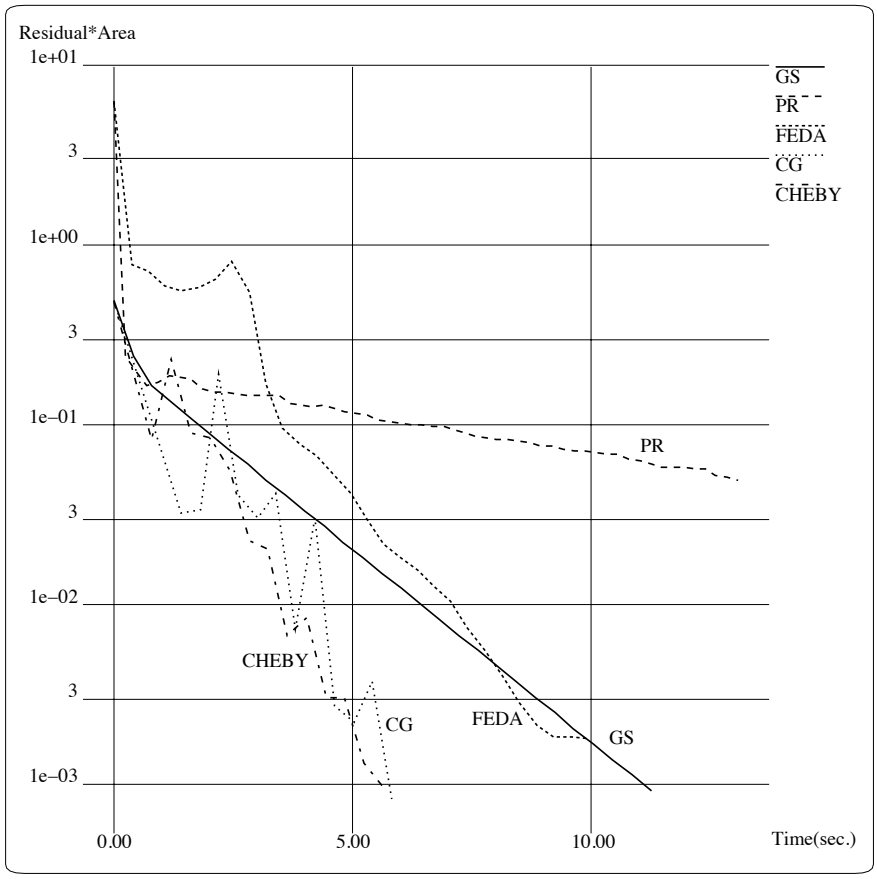

Figure 8.11: Case F ($\rho_{avg} = 0.89$, $\delta = 70\%$) with performance measured in CPU time.

sity decreases and the average reflectance is kept fixed at $\rho_{ave} = 0.77$. As we can see, the eigenvalues become more spread out. This effect, as the theory suggests (Section 7.3.2), reduces the convergence for all methods. Table 8.2 summarizes the performance of the tested methods for the six test cases. Comparing the results regarding test cases E and C and the results regarding test cases F and D, we can verify, in practice, the slower convergence due to the density decrease. Because CG is adaptive and takes some number of iterations to implicitly accumulate the eigenvalue information it needs, the higher average reflectance and lower density (higher level of occlusion) problems give it more a chance to accumulate that information. However, as Figure 8.11 shows, only for the highest average reflectance and density levels tested ($\rho_{avg} = 0.89$, $\delta = 70\%$) did it become competitive with the Chebyshev method.

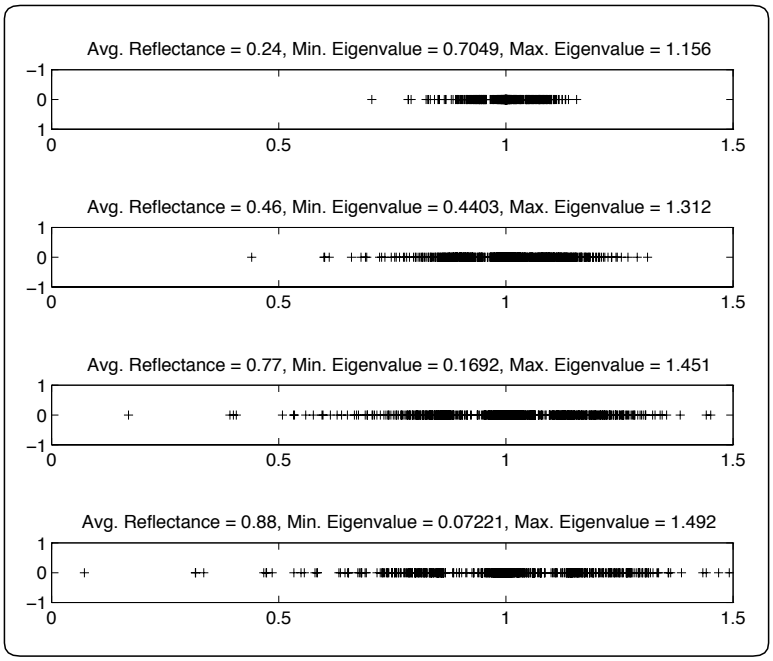

Figure 8.12: Eigenvalue distribution as the overall reflectance increases.

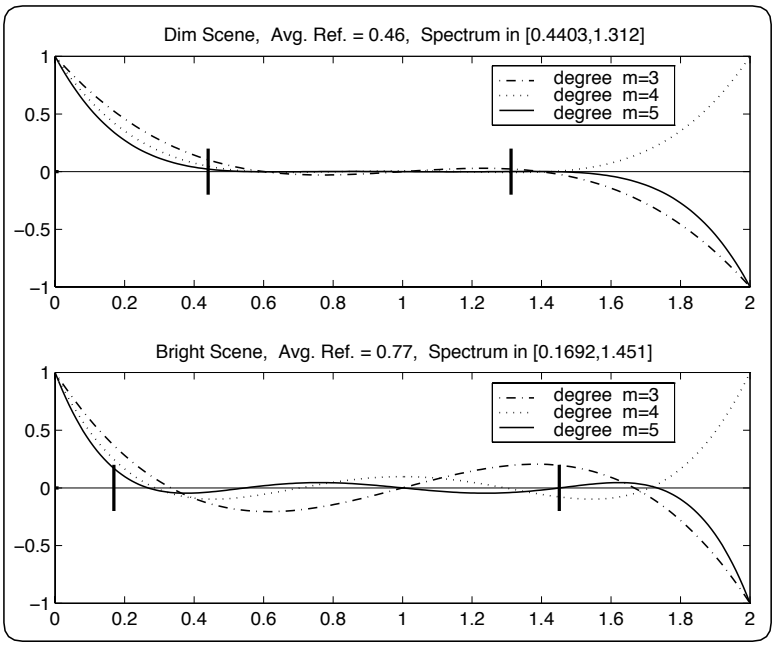

Figure 8.13: Chebyshev polynomials for scenes with different average reflectances.

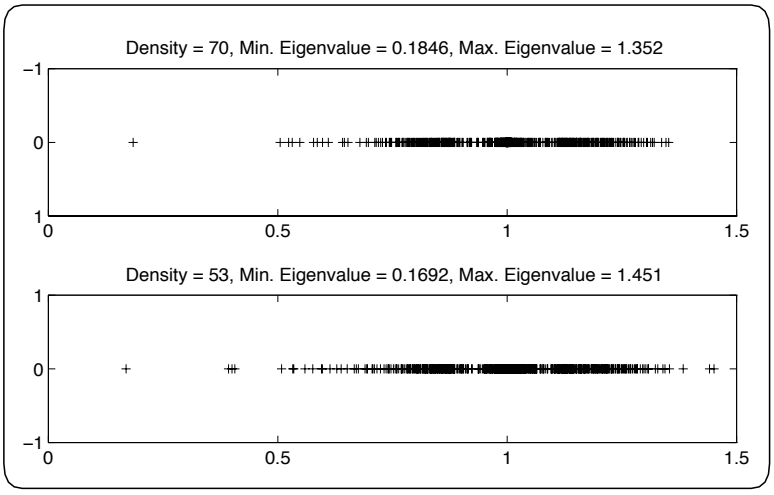

Figure 8.14: Eigenvalue distribution as the density of the radiosity matrix decreases due to the increase on the level of occlusion in the scene.

The time of the fastest algorithm for each test case is presented in boldface in Table 8.2. In general, the Gauss-Seidel and Chebyshev methods are the fastest overall. Note that the Feda method implemented failed on test problems E and F, which have higher density and average reflectance [7].

Methods	Test Cases					
	A	B	C	D	E	F
CG	4.68	6.60	8.63	11.02	5.05	5.80
CHEBY	1.61	2.81	**4.81**	**6.45**	**4.05**	**5.64**
FEDA	3.92	5.32	11.73	18.47	7.45^{+}	9.92^{*}
GS	**1.20**	**2.01**	6.43	12.83	5.23	11.26
PR	3.21	6.85	23.03	50.56	22.53	50.43

Table 8.2: Algorithms performance (total time in seconds). The symbols $*$ and $+$ indicate failure to converge after 2478 and 3304 steps of iteration respectively.

[7] Feda and Purgathofer [73] mentioned that in highly reflecting scenes the ambient term gets too high, which could lead to a negative term in the next iteration. This, in turn, could cause divergence. As suggested by Feda and Purgathofer, we have limited the distributable energy to the total unshot energy to prevent negative ambient terms. Nevertheless, divergence has occured.

8.7 Summary

Our experiments using explicitly stored form factors have shown that although radiosity-specific methods make rapid initial improvement, faster than any other method for limited tolerances (10^{-1}), they are slower than the general matrix methods for higher tolerances. Radiosity-specific methods require searching for a patch to shoot on each step, which can require traversing a large data structure. The disadvantage of general matrix methods of storing the form factors is compensated by the regularity of the computations which allows good pipelining and data locality.

The performance advantages of the general matrix methods are not as attractive when the form factors are computed on the fly, *e.g.* when n is large. In that case, the innermost loop consists of computing the form factors, which generally requires more flops than the matrix solving algorithms themselves. Avoiding even a few extra form factor computations by searching through rows for the most unshot radiosity may then give the edge to radiosity-specific methods such as the overshooting method propose by Feda and Purgathofer (Section 7.2.4).

For environments with high average reflectance, which may occur in several applications, including those involving regions of vegetation [48], the rate of convergence is slower for all of the iterative methods used. Our numerical experiments have also shown that the CG method and particularly the Chebyshev method, with the estimates of the maximal eigenvalues described previously, represent the fastest approaches to handle those cases.

The experiments also show that selecting the "best" method is delicate, and no single method is superior in all cases. The relative performance depends on architectural performance features such as pipelining and data locality as well as problem characteristics. Developing practical solution strategies will likely require implementing a variety of linear solvers, with the one actually used chosen at runtime dependent on problem parameters such as reflectance and occlusion, and determining the best parallel implementation for shared memory multiprocessor workstations. It will also be necessary to bring more numerical linear algebra tools to bear on the problem.

During the investigation presented in this chapter we have addressed important practical issues related to the application of iterative methods to radiosity problems,

such as convergence conditions and stopping criteria. We have also pointed out that iterative methods can have a better performance with different choices of initial approximation for the solution vector. Intuitively we may say that the closer the initial approximation, the faster the iterative method will converge[8]. The problem is how to come up with a closer initial approximation without knowing the true solution before hand, *i.e.* solving the system. In Appendix C we present experiments, in which eigenvectors are used as solution vectors in graphics settings, that that may lead to a solution for this problem in the radiosity context.

[8]There are methods, such as the Conjugate Gradient Squared (CGS) [21], that tend to diverge if the initial approximation is close to the solution.

Chapter 9
Conclusions

9.1 Summary of Contributions

Recent developments in physically-based rendering have provided very realistic images of a variety of environments composed mostly of inorganic materials. This dissertation aimed to extend physically-based rendering to some organic materials. Specifically, biologically and physically-based BDF models for plants were introduced. Moreover, as we discussed in the introduction in Chapter 1, there is an important interplay between local and global illumination, which affects the versatility, accuracy and efficiency of the rendering methods directly. Considering this aspect, in this dissertation, instead of focusing only on local illumination issues related to BDF models (Chapters 4 to 6), we also investigated solutions for global illumination problems (Chapters 7 and 8). More specifically, we sought faster alternatives to solve radiosity systems. The relevance of this investigation resides in the fact that the radiosity method corresponds to the basic layer of several global illumination frameworks, including those designed for radiative transport in vegetation [48].

In Chapter 2 we laid the physical foundation and outlined the physically-based rendering concepts used in this research. We also reviewed the standard reflectance models used in computer graphics. The theoretical contributions of this dissertation to the field of computer graphics started in Chapter 3, where we presented an extended review of the main factors affecting the interaction of light with foliar tissues. Understanding these factors is the first and fundamental step towards biologically-based simulation of light transport by plants through rendering methods. Furthermore, we presented a concise and structured survey of BDF models used in botany and remote sensing, and discussed the use of global illumination methods for radiative transfer

in regions of vegetation. Although the BDF models and global illumination methods used in remote sensing are aimed at different applications, such as inversion procedures used to determine the biophysical characteristics of plants, they provide valuable insights for the design of models and methods tailored to rendering applications.

The original contributions of this dissertation started in Chapter 4 with the presentation of a plausible algorithmic BDF model (*ABM*) for plants. The *ABM* is controlled by a small number of biological meaningful parameters and its formulation is based on standard Monte Carlo methods. The main factors affecting the interaction of light with foliar tissues, such as the absorption of light by chlorophyll, were incorporated into the new model. The *ABM* was designed primarily for physically-based rendering applications, being the first BDF model specifically oriented to plants presented in the computer graphics literature.

In contrast to many standard BDF models used in computer graphics, the *ABM* does not rely on external values for reflectance and transmittance. In other words, the reflectances and transmittances regarding different viewing and illuminating geometries are computed by the model itself. This aspect indicates that besides its on-line applications, the *ABM* can be used off-line in order to generate spectral curves of reflectance and transmittance required as input parameters to other models.

In its evaluation the *ABM* was tested as a separate unit of the rendering pipeline to avoid bias from errors introduced in others units, *e.g.* the geometrical modeling and the spectral sampling. We have compared the spectral curves of reflectance and transmittance provided by the *ABM* with measured curves obtained thorough a set of leaf optical experiments (*LOPEX*) [120]. Is important to notice that, to obtain the spectral curves through the *ABM*, we did not performed an *ad hoc* adjustment of the parameters. Instead, we used exact measured values for some parameters, such as pigment concentration and leaf thickness, and the average of measured values for others, such as the dimensions of the epidermals cells used to calculate their oblateness. This data was obtained from the remote sensing and botany literature as well as from *LOPEX*. The spectrophotometric curves obtained through the *ABM* showed a good agreement with the curves provided by *LOPEX*, and the goniophotometric curves also presented a good agreement with the curves from previous experiments

reported in the remote sensing literature [40, 240]. To the best of our knowledge, no other BDF model oriented to the rendering of biological materials have had its readings compared against real measured data.

To apply this testing approach we had to search for spectrophotometric and goniophotometric experimental data for plants and to implement virtual measurement devices. These two aspects represent peripheral contributions associated with the development of the *ABM*. The large amount of data collected during this research, provided mainly by *LOPEX* and other sources cited throughout this dissertation, was introduced to the graphics community and can be used as input parameters in the application of existing models or in the testing of new ones. Although virtual spectrophotometers and goniophotometers have already been used before in rendering applications [44, 91, 139, 235], we have presented a structured derivation of the formulation used to these devices, which is usually not readily available in rendering papers. Using this formulation, these virtual measurement devices can be easily implemented and used for data collection and testing of BDF models.

The on-line application of an algorithmic BDF model like the *ABM* may become computationally expensive if one needs to render several foliar primitives. Alternatively, the BDF of a given specimen can be precomputed through the off-line application of an an appropriate BDF model, stored in memory, and reconstructed during rendering, saving valuable time. Image synthesis algorithms based on Monte Carlo path tracing [50, 127, 131, 137, 198] have been often used in physically-based rendering applications. In this context, the reconstruction of a BDF involves the generation of reflected and transmitted directions for the rays reaching the specimen.

Instead of applying existing deterministic approaches involving mathematical tools like wavelets [51, 139, 140, 141], spherical harmonics [44, 235] and interpolation [93], we decided to investigate a nondeterministic approach to reconstruct isotropic BDFs, general enough to be not only applied to the rendering of foliar tissues, but also to access arbitrary isotropic BDFs. This approach consists of an algorithm based on probability methods, which makes its incorporation into Monte Carlo based rendering frameworks straightforward. The testing of this new algorithm indicated that it can provide reconstructed BDF curves which are very close to the original BDF curves,

and it can also prompt a noticeable gain in performance. Further testing is required, however, to determine its full range of applications.

In Chapter 6 we presented a simplified model of light interaction with foliar tissues also aiming at applications involving several foliar primitives. This model, like the *ABM*, accounts for the three components of plant tissues' BDF, namely surface reflectance, subsurface reflectance and transmittance. The simulation of the processes of light propagation and absorption within the foliar tissues, however, are simplified to obtain gains in performance without undermining the quality of the images. It uses precomputed reflectances and transmittances as scale factors to represent stochastically the scattering profile of the foliar tissues.

The use of these factors replaces the random walk process used by the *ABM* to simulate the randomization and the absorption of light within the foliar tissues. The tests performed on the simplified model indicated that the reduction in image quality is sufficiently small and the gains in performance sufficiently high to justify the reduction in the accuracy of the simulations. As suggested earlier, the scale factors are computed off-line using the ABM. The resulting table of spectral absolute reflectances and transmittances can also be used by other BDF models relying on the input of these parameters.

As mentioned earlier, the simulation of radiative transport in plant canopies depends not only on the BDF models used to simulate the direct interaction of radiation with foliar primitives, but also on the global illumination approaches used to describe quantitatively the indirect interactions of radiation among these primitives. The plausible models proposed in this work were designed to accurately and physically fulfill the first requirement. The radiosity method [91], addressed in Chapter 7 and 8, or its stochastic variation [200] are commonly used to fulfill the second requirement [35, 47, 83, 89, 158].

The computationally expensive part of the radiosity methods is the computation of form factors. There are several applications, however, in which the form factor matrix is computed only once and the resulting system of linear equations must be solved many times. Among these applications we can include environmental and architectural design [99] as well as the study of radiation effects on regions of vegetation

such as forest canopies or crops [47, 48]. Usually in these applications only the parameters associated with the emittance, reflectance and transmittance characteristics of the primitives are changed. After each change is performed, the system of linear equations representing the balance of energy in the environment is solved again. The performance of the iterative methods used to solve this system, therefore, becomes an important issue.

With this in mind we provided the mathematical background for the analysis of the spectral characteristics of the radiosity matrices in Chapter 7. The relationship between the eigenvalues of the coefficient and iteration matrices with the convergence of the iterative methods used to solve the radiosity system of linear systems has been emphasized. The importance of this investigation was not restricted to the immediate purpose of finding faster solutions to the radiosity systems. It also aimed at a broader understanding of the fundamental operators that arise in global illumination, such as the radiosity matrix, which may lead to more efficient approaches to solve the rendering equation [131]. We have also presented a concise and original proof that all the eigenvalues of the radiosity matrix are real and positive. This proof allowed us to use the Gerschgorin Circle Theorem [41] to obtain tighter estimates for the eigenvalues of the radiosity matrix using real intervals instead of circles in the plane of complex numbers. These estimates, in turn, allow us to apply selected nonstationary methods to solve the radiosity system with confidence of their convergence.

In Chapter 8 we introduced the Chebyshev method [217] to the graphics literature, aiming at fast solutions for radiosity systems. This nonstationary method, with the eigenvalues estimates proposed in this dissertation, can converge faster for environments with high average reflectance and high level of occlusion than the iterative methods previously used. Besides this practical contribution, we have also examined and compared the performance of radiosity-specific methods and general matrix methods. Differently from previous works in the computer graphics literature involving iterative methods, we have focused this analysis on practical computational aspects such as pipelining and data locality.

Among the observations drawn from this analysis, we highlight the fact that the measure usually applied to evaluate the performance of the iterative solvers in the

radiosity context, namely counting steps of iteration, may lead to wrong conclusions. Since counting steps of iteration does not account for differing amounts and types of work performed on each step, methods that converge in fewer steps of iteration may in fact require more time. In other words, general matrix methods, such as Gauss-Seidel, Conjugate Gradient or Chebyshev, can outperform radiosity-specific methods like progressive refinement and overshooting.

We also performed an original investigation of the effects of increases in the average reflectance and level of occlusion in a given environment on the performance of the iterative solvers. Our numerical experiments have shown that an increase on either of these factors causes a larger number of interreflections, causing the eigenvalues to become more spread out, which, in turn, slows the convergence of the iterative methods. These effects are less deleterious to the Chebyshev method, which may have its convergence improved even further with the use of tighter estimates for the eigenvalues of the radiosity matrix.

The BDF models, the reconstruction algorithm and the virtual measurement devices have been implemented using an object oriented programming approach. The modular code used to implement them corresponds to classes written in $C{+}{+}$. This approach allows the encapsulation of the local illumination code so that it can be incorporated into different global renderers. Indeed, in order to test the BDF models within global rendering frameworks, we have implemented a distributed ray tracer [60], a standard path tracing [131], a modified path tracing [198] and a wavelength dependent path tracing. These renderers have also been implemented using $C{+}{+}$. The radiosity code, on the other hand, has been implemented using C. Nonetheless, it can be easily ported to $C{+}{+}$ to take advantage of the objected oriented features of this programming language.

Although the main purpose of this research was to provide means for the enhancement of the physically-based rendering of natural environments, its contributions are not limited to computer graphics applications. They can also be applied in remote sensing, illuminating engineering, environmental design and architecture. Moreover, since the the radiative transfer processes can also be applied to other forms of energy, such as heat, the investigation regarding the radiosity method can also find

applications in thermal and mechanical engineering.

9.2 Further Research

In this dissertation we have examined some issues in more depth than previous publications, and we also attempted to take a broad view of the problems related to the design of BDF models and the application of radiative transfer methods. Many questions remain to be answered, however. Fortunately so, since, as appropriately stated by Hammel [106], "the avenues of open research that a piece of work creates are just as important as the accomplishments it produced". In this section we examine a number of ways that the models and algorithms introduced in this dissertation can be extended to add new features. We also look at more general open problems, their implications and recent developments.

The *ABM*, presented in Chapter 4, is an isotropic model. In order to make it capable of capturing the anisotropic behavior of the foliar tissues, it would be necessary to account for the characteristics of the venation systems. Unfortunately, as of today, very few publications have presented experiments and data regarding this topic. The modular and algorithmic nature of this model, however, allow the incorporation of these factors as more information becomes available in the literature.

Besides the veins, other factors may be taken into account to improve qualitatively the surface reflectance readings of the *ABM*. Among these factors are the presence of hairs and the simulation of shadowing and masking effects associated with different wax configurations. Moreover, the effects of deposition of water, dust and chemical substances may also be incorporated to the model by adding new states to its stochastic simulation of light propagation. Accounting for wave optics phenomena, such as interference, diffraction and polarization[1], however, would require more substantial

[1]In the context of plant leaves polarization would be the most relevant wave optics phenomenon to be taken into account [98, 150]. Since the perpendicular component of reflectance drops to zero at a particular angle of incidence, θ_f, known as Brewster's angle, the effects of polarization on the overall reflectance become extremely pronounced for certain viewing and illuminating geometries. This angle is given by Brewster's law [86]: $tan\theta_f = \frac{\eta_t}{\eta_i}$. Moreover, from a remote sensing point of view, since polarization is altered by surface effects, the measurement of this phenomenon allows the decoupling between surface and subsurface components of reflectance.

changes, since the model is based on geometrical optics. It would also be interesting to explore applications of the *ABM*, as well as the simplified model, presented in Chapter 6, to senescent leaves and other plant surfaces such as petals and stems. These applications would also require specific biological data, which, as mentioned in Chapter 3, either does not exist or is still not readily available.

In this dissertation we have focussed on aspects associated with the visible region of the light spectrum. In remote sensing applications involving reflectance signatures, however, the near-infrared and the infrared regions are also important. For applications in the near-infrared region the *ABM* could be adapted by replacing the absorption testing by a *Russian roulette* procedure. For the infrared region the *ABM* could be adapted by considering the absorption coefficients of water instead of the absorption coefficients of pigments.

The efficiency of the *ABM* could also be improved through a detailed analysis of the ray density required to obtain spectral curves with a certain accuracy. This analysis would allow the use of the minimal possible number of rays in data collection and measurement procedures. In addition, its performance could be further maximized through the use of specialized parallel software [46, 221] or hardware [177]. The simplified model, whose design based on a fixed number of outgoing rays is very suitable to parallelism, would also benefit from the application of parallel processing techniques.

The reconstruction algorithm, presented in Chapter 5, was originally developed to access isotropic reflectance and transmittance functions. Another interesting direction of future research would be to extend this algorithm to the reconstruction of anisotropic reflectance and transmittance functions. In terms of its use of computational resources, a comparison with deterministic approaches would also be appropriate to determine the complete extension of the advantages derived from it, besides its accuracy, easy implementation and intuitive nature. Moreover, since the gain in performance in rendering applications provided by this algorithm and by the simplified model depends on a number of factors (the illuminating and viewing angles, the number of pixels occupied by the specimen(s), the scene geometry and the loss of quality threshold), a detailed evaluation of the impact of these factors on their

performance would provide a more precise picture of their advantages and range of applications.

The analysis of the spectral properties of radiosity systems, presented in Chapters 7 and 8, can also be further extended. It may be possible to use the Perron-Frobenius theorem (Appendix C) to provide an upper bound on the smallest eigenvalue of the radiosity matrix, thereby analytically proving that the eigenvalues spread as the average reflectance of the environment increases. Furthermore, if one can determine what the eigenvectors of the radiosity matrix represent in terms of the physical application (an open problem in rendering), it may be possible to obtain low rank approximations of them. These eigenvectors, in turn, could be used to obtain low rank approximations of the radiosity matrices and faster global illuminations solutions. We present preliminary experiments regarding this topic and examine its implications in Appendices C and D. Also, while in the investigation presented in Chapters 7 and 8 we have focused on the diffuse operators of the rendering equation, further developments in this area shall extend this investigation to the non-diffuse operators[2] of as well.

Finally, the contributions of this dissertation and its proposed extensions can be combined in the development of a multipass method specially tailored to biological materials. The first pass of this method, used to determine the indirect illumination between the primitives, could be performed by a standard radiosity algorithm. The advantages of using such an algorithm, instead of a radiosity via ray-tracing approach [200], are the possibility of reussing the numerical data if only the physical characteristics of the materials are changed and the straightforward elimination of the direct illumination contribution.

In closing, we realize that there is still a lot of research that has to be done to achieve a true physically-based photorealism of natural environments using reasonable computational resources. We have listed many problems and issues that are worth looking into. Several aspects that affect the realism of computer generated images are beyond the scope of this dissertation, such as geometrical modeling, spectral sampling,

[2]In the case of non-diffuse operators one can foresee a greater difficulty, since these operators, differently from the diffuse operators, are not compact [9]. The research in this area is just starting. Bringing more sophisticated functional analysis tools to bear in this investigation may allow us to consider these operators.

use of accurate luminaire data, participating media and so on. One could question the benefits of improving specific stages of the rendering pipeline when these benefits can be diluted by the current limitations in other stages. In this case, the research presented in this dissertation might be seen as a growing branch in a growing tree. The other branches of the tree would represent other areas of research related to the different stages of the rendering pipeline. They are equally important and we hope that, eventually, they will form a well balanced canopy, whose final size, shape and reflectance signature remain to be seen.

References

[1] Allen, W., Gayle, T., and Richardson, A. Plant canopy irradiance specified by the Duntley equations. *Journal of the Optical Society of America 60*, 3 (1970), 372–376.

[2] Allen, W., and Richardson, A. Interaction of light with a plant leaf canopy. *Journal of the Optical Society of America 58*, 18 (1968), 1023–1028.

[3] Allen, W., and Richardson, A. Willstater-stoll theory of leaf reflectance evaluated by ray tracing. *Applied Optics 12*, 10 (1973), 2448–2453.

[4] Allen, W., Richardson, A., and Thomas, J. Plant canopy irradiance specified by the duntley equations. *Journal of the Optical Society of America 59*, 10 (1969), 1376–1379.

[5] ANSI. Nomenclature and definitions for illuminating engineering. In *ANSI/IES RP-6-1986*. Illuminating Engineering Society of North America, New York, 1986.

[6] Appel, A. Some techniques for shading machine renderings of solids. In *Spring Joint Computer Conference* (1968), pp. 37–45.

[7] Arvo, J. *Analytic Methods for Simulated Light Transport*. PhD thesis, Yale University, 1995.

[8] Arvo, J. Applications of irradiance tensors to the simulation of non-lambertian phenomena. In *SIGGRAPH Proceedings, Annual Conference Series* (August 1995), pp. 335–342.

[9] Arvo, J. The role of functional analysis in global illumination. In *Rendering Techniques'95 (Proceedings of the Sixth Eurographics Rendering Workshop)*

(Dublin, June 1995), P. M. Hanrahan and W. Purgathofer, Eds., Springer-Verlag, pp. 115–126.

[10] Ashdown, I. Eigenvector radiosity. unpublished manuscript, 1994.

[11] Ashdown, I. *Radiosity: A programmers's perspective*. John Wiley & Sons, New York, NY, 1994.

[12] Asrar, G., and Mynemi, R. Applications of radiative transfer models for remote sensing of vegetation conditions and states. In *Photon - Vegetation Interactions - Applications in Optical Remote Sensing and Ecology* (Berlin, 1991), R. Mynemi and J. Ross, Eds., Springer-Verlag, pp. 538–558. ch. 17.

[13] Baranoski, G. The parametric differential method: An alternative to the calculation of form factors. *Computer Graphics Forum (EUROGRAPHICS Proceedings) 11*, 3 (September 1992), 193–204.

[14] Baranoski, G., Bramley, R., and Rokne, J. Eigen-analysis for radiosity systems. In *Proceedings of the Sixth International Conference on Computational Graphics and Visualization Techniques - Compugraphics'97* (Algarve, Portugal, December 1997), H. P. Santo, Ed., Graphic Science Promotions and Publications, pp. 193–201.

[15] Baranoski, G., Bramley, R., and Shirley, P. Fast radiosity solutions for environments with high average reflectance. In *Rendering Techniques'95 (Proceedings of the Sixth Eurographics Rendering Workshop)* (Dublin, June 1995), P. M. Hanrahan and W. Purgathofer, Eds., Springer-Verlag, pp. 345–356.

[16] Baranoski, G., Bramley, R., and Shirley, P. Iterative methods for fast radiosity solutions. Tech. Rep. TR 429, Computer Science Department, Indiana University, April 1995.

[17] Baranoski, G., and Rokne, J. An algorithmic reflectance and transmittance model for plant tissue. *Computer Graphics Forum (EUROGRAPHICS Proceedings) 16*, 3 (September 1997), 141–150.

[18] Baranoski, G., and Rokne, J. A nondeterministic reconstruction approach for isotropic reflectances and transmittances. In *Proceedings of the Sixth Pacific Conference on Computer Graphics and Applications - Pacific Graphics'98* (Singapore, October 1998), IEEE Computer Society, pp. 134–139.

[19] Baranoski, G., and Rokne, J. A simplified model for light interaction with plant tissue. In *Proceedings of the Eighth International Conference on Computer Graphics and Visualization - GraphiCon'98* (Moscow, September 1998), pp. 154–161.

[20] Baret, F., Andrieu, B., and Guyot, G. A simple model for leaf optical properties in visible and near infrared: application to the analysis of spectral shifts determinism. In *Applications of Chlorophyll Fluorescence* (1988), Kluwer Academic, pp. 345–351.

[21] Barrett, R., Berry, M., Chan, T., Demmel, J., Donato, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C., and Vorst, H. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. SIAM, Philadelphia, 1994.

[22] Baum, D., Wallace, J., Cohen, M., and Greenberg, D. The back-buffer algorithm: an extension of the radiosity method to dynamic environments. *The Visual Computer 2*, 5 (September 1986), 298–306.

[23] Beckmann, P., and Spizzichino, A. *The Scattering of Electromagnetic Waves from Rough Surfaces*. MacMillan, New York, 1963.

[24] Benson, L. *Plant Classification*. D.C. Heath and Company, Boston, 1957.

[25] Berman, A., and Plemmons, R. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, New York, 1987.

[26] Biolley, J., and Jay, M. Anthocyanins in modern roses: Chemical and colorimetric features in relation to colour range. *Journal of Experimental Botany 44*, 268 (1993), 1725–1734.

[27] Bjorn, L. Interception of light by plant leaves. In *Crop Photosynthesis: Spatial and Temporal Determinants* (Amsterdam, 1992), N. R. Baker and H. Thomas, Eds., Elsevier, pp. 253–276.

[28] Blackwell, D. Minimax and irreducible matrices. *Journal of Mathematical Analysis and Applications 3* (1961), 37–39.

[29] Blinn, J. Models of light reflection for computer synthesized pictures. *Computer Graphics (SIGGRAPH Proceedings) 11*, 2 (1977), 192–198.

[30] Blinn, J. Simulation of wrinkled surfaces. *Computer Graphics (SIGGRAPH Proceedings) 12*, 3 (1977), 286–292.

[31] Bloomenthal, J. Modeling the mighty maple. *Computer Graphics (SIGGRAPH Proceedings) 19*, 3 (July 1985), 305–311.

[32] Bloomenthal, J. *Skeletal Design of Natural Forms*. PhD thesis, Department of Computer Science, University of Calgary, January 1995.

[33] Bone, R., Lee, D., and Norman, J. Epidermal cells functioning as lenses in leaves of tropical rain-forest shade plants. *Applied Optics 24*, 10 (May 1985), 1408–1412.

[34] Bonner, J., and Varner, J. *Plant Biochemistry*. Academic Press, New York, 1965.

[35] Borel, C., Gerstl, S., and Powers, B. The radiosity method in optical remote sensing of structured 3-d surfaces. *Remote Sensing of Environment 19* (1991), 13–44.

[36] Bowes, B. *A Colour Atlas of Plant Structure*. Manson Publishing, 1997.

[37] Bradley, J. *XV - Interactive Image Display for the X Window System*. Bryn Mawr, PA, USA, December 1994. Version 3.10a.

[38] Brakke, T., Smith, J., and Harnden, J. Bidirectional scattering of light from tree leaves. *Remote Sensing of Environment 29* (1989), 175–183.

[39] Bramley, R., and Loos, T. Emily: A visualization tool for large sparse matrices. Tech. Rep. TR 412b, Computer Science Department, Indiana University, July 1994.

[40] Breece, H., and Holmes, R. Bidirectional scattering characteristics of healthy green soybean and corn leaves in vivo. *Applied Optics 10*, 1 (January 1971), 119–127.

[41] Burden, R., and Douglas, J. *Numerical Analysis*, forth ed. PWS-KENT Publishing Company, Boston, 1988.

[42] Burden, R., and Faires, J. *Numerical Analysis*, fifth ed. PWS-KENT Publishing Company, Boston, 1993.

[43] Butler, W. Absorption spectroscopy in vivo: Theory and applications. *Annual Review of Plant Physiology 15* (1964), 451–470.

[44] Cabral, B., Max, N., and Springmeyer, R. Bidirectional reflection from surface bump maps. *Computer Graphics (SIGGRAPH Proceedings) 21*, 4 (July 1987), 273–281.

[45] Chandrasekar, S. *Radiative Transfer*. Dover Publications, New York, 1960.

[46] Chandy, K., and Kesselman, C. CC++: A declarative concurrent object-oriented programming notation. In *Research Directions in Concurrent Object-Oriented Programming* (Cambridge, Massachusetts, 1993), G. Agha, P. Wegner, and A. Yonezawa, Eds., The MIT Press, pp. 282–313. Chapter 11.

[47] Chelle, M. *Développement d'un modèle de radiosité mixte pour simuler la distribution du rayonnement dans les couverts végétaux*. PhD thesis, L'Université de Rennes I, France, March 1997.

[48] Chelle, M., Andrieu, B., and Bouatouch, K. Nested radiosity for plant canopies. *The Visual Computer 14*, 3 (1998), 109–125.

[49] Chen, S. Incremental radiosity: An extension of progressive radiosity to an interactive image synthesis system. *Computer Graphics (SIGGRAPH Proceedings) 24*, 4 (August 1990), 135–144.

[50] Chen, S., Rushmeier, H., Miller, G., and Turner, D. A progressive multi-pass method for global illumination. *Computer Graphics (SIGGRAPH Proceedings) 25*, 4 (July 1991), 164–174.

[51] Christensen, P., Stollnitz, E., and Salesin, D. Global illumination of glossy environments using wavelets and importance. *ACM Transactions on Graphics 15*, 1 (July 1995), 37–71.

[52] Clarke, F., and Parry, D. Helmholtz reciprocity: Its validity and application to reflectometry. *Lighting Research and Technology 17*, 1 (1985), 1–11.

[53] Clayton, R. *Molecular Physics in Photosynthesis*. Blaisdell Publishing Company, Cambridge, England, 1965.

[54] Cohen, M., Chen, S., Wallace, J., and Greenberg, D. A progressive refinement approach to fast radiosity image generation. *Computer Graphics (SIGGRAPH Proceedings) 22*, 4 (July 1988), 75–84.

[55] Cohen, M., and Greenberg, D. The hemi-cube: A radiosity solution for complex environments. *Computer Graphics (SIGGRAPH Proceedings) 19*, 3 (July 1985), 31–40.

[56] Cohen, M., Greenberg, D., Immel, D., and Brock, P. An efficient radiosity approach for realistic image synthesis. *IEEE Computer Graphics and Applications 6*, 3 (March 1986), 26–35.

[57] Cohen, M., and Wallace, J. *Radiosity and Realistic Image Synthesis*. Academic Press Professional, Cambridge, 1993.

[58] Concus, P., Golub, G., and O'Leary, D. A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations. In

Sparse Matrix Computations (New York, 1976), J. Bunch and D. Rose, Eds., Academic Press, pp. 309–322.

[59] Cook, R. Shade trees. *Computer Graphics (SIGGRAPH Proceedings) 18*, 4 (July 1984), 223–231.

[60] Cook, R., Porter, T., and Carpenter, L. Distributed ray tracing. *Computer Graphics (SIGGRAPH Proceedings) 18*, 4 (July 1984), 165–174.

[61] Cook, R., and Torrance, K. A reflectance model for computer graphics. *ACM Transactions on Graphics 1*, 1 (January 1982), 7–24.

[62] Coveyou, R., Cain, V., and Yost, K. Adjoint and importance in Monte Carlo application. *Nuclear Science and Engineering 27* (1967), 219–234.

[63] Crowther, B. Computer modeling of integrating spheres. *Applied Optics 35*, 30 (1996), 5880–5886.

[64] Cullum, J., and Willoughby, R. *Lanczos algorithms for large symmetric eigenvalue computations*. Birkhauser, 1985. vols I and II.

[65] der Sluis, A. V., and der Vorst, H. V. The rate of convergence of conjugate gradients. *Numerical Mathematics 48* (1986), 543–560.

[66] Devlin, R., and Barker, A. *Photosynthesis*. Van Nostrand Reinhold Company, New York, 1971.

[67] DeYoung, J., and Fournier, A. Properties of tabulated bidirectional reflectance distribution functions. In *Proceedings of Graphics Interface* (1997), pp. 47–55.

[68] DiLaura, D., and Franck, P. On setting up and solving large radiative transfer systems. *Journal of the Illuminating Engineering Society 22*, 2 (October 1996), 3–7.

[69] Dimian, D. Physically-based model of folded surfaces. Master's thesis, Department of Computer Science, The University of Calgary, November 1997.

[70] Duntley, S. The optical properties of diffusing materials. *Journal of the Optical Society of America 32*, 2 (February 1942), 61–70.

[71] E284-91C, A. S. Standard terminology of appearance. In *Physics-Based Vision Principles and Practice: Radiometry* (Boston, 1992), L. Wolff, S. Shafer, and G. Healey, Eds., Jones and Bartlett Publishers, pp. 146–161.

[72] Esau, K. *Plant Anatomy*. John Wiley & Sons, Inc., New York, 1953.

[73] Feda, M., and Purgathofer, W. Accelerating radiosity by overshooting. In *Proceedings of the Third Eurographics Rendering Workshop* (Barcelona, June 1992), A. Chalmers, Ed., pp. 21–32.

[74] Fendley, J. An analysis of the measuring procedure for the integrating sphere spectrophotometer. *Solar Energy 35*, 3 (1985), 281–282.

[75] Foley, J., van Dam, A., Feiner, S., and Hughes, J. *Computer Graphics: Principles and Practice*, second ed. Addison-Wesley Publishing Company, Reading, Massachusetts, 1990.

[76] Fournier, A. From local to global illumination and back. In *Rendering Techniques'95 (Proceedings of the Sixth Eurographics Rendering Workshop)* (Dublin, June 1995), P. M. Hanrahan and W. Purgathofer, Eds., Springer-Verlag, pp. 127–136.

[77] Fournier, A. Separating reflection functions for linear radiosity. In *Rendering Techniques'95 (Proceedings of the Sixth Eurographics Rendering Workshop)* (Dublin, June 1995), pp. 298–305.

[78] Fournier, A., Fiume, E., Ouellette, M., and Chee, C. K. *FIAT LUX: Light driven global illumination*, 1989. Technical Memo DGP89-1, Dynamic Graphics Project.

[79] Fourty, T., Baret, F., Jacquemoud, S., Schmuck, G., and Verdebout, J. Leaf optical properties with explicit description of its biochemical

composition: Direct and inverse problems. *Remote Sensing of Environment 56* (1996), 104–117.

[80] Gates, D., Keegan, H., Schelter, J., and Weidner, V. Spectral properties of plants. *Applied Optics 4*, 1 (January 1965), 11–20.

[81] Gausman, H., and Allen, W. Optical parameters of leaves of 30 plant species. *Plant Physiology 52* (1973), 57–62.

[82] Gausman, H., Allen, W., and Escobar, D. Refractive index of plant cell walls. *Applied Optics 13*, 1 (1973), 109–11.

[83] George, D., Sillion, F., and Greenberg, D. Radiosity redistribution for dynamic environments. *IEEE Computer Graphics and Applications 10*, 4 (July 1990), 26–34.

[84] Gerstl, S., and Borel, C. Principles of the radiosity method for canopy reflectance modeling. In *International Geoscience and Remote Sensing Symposium - IGARSS'90* (Washington, May 1990), pp. 1735–1737.

[85] Gilks, W., Richardson, S., and Spiegelhalter, D. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London, 1996.

[86] Glassner, A. *Principles of Digital Image Synthesis*. Morgan Kaufmann, San Francisco, 1995.

[87] Goebel, D. Generalized integrating-sphere theory. *Applied Optics 6*, 1 (January 1967), 125–128.

[88] Goel, N. Models of vegetation canopy reflectance and their use in the estimation of biophysical parameters from reflectance data. *Remote Sensing Review 4*, 1 (1988), 221.

[89] Goel, N., Rozehnal, I., and Thompson, R. A computer graphics based model for scattering from objects of arbitrary shapes in the optical region. *Remote Sensing of Environment 36* (1991), 73–104.

[90] Goertler, S., Cohen, M., and Slusallek, P. Radiosity and relaxation methods. Tech. rep., Princeton University, 1993.

[91] Goertler, S., Cohen, M., and Slusallek, P. Radiosity and relaxation methods. *IEEE Computer Graphics and Applications 14*, 6 (November 1994), 48–58.

[92] Golub, G., and Loan, C. V. *Matrix Computations*, second ed. John Hopkins University Press, Baltimore, 1989.

[93] Gondek, J., Meyer, G., and Newman, J. Wavelength dependent reflectance functions. In *Proceedings of SIGGRAPH* (Florida, July 1994), A. Glassner, Ed., Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH, ACM Press, pp. 213–220.

[94] Goral, C., Torrance, K., Greenberg, D., and Battaile, B. Modelling the interaction of light between diffuse surfaces. *Computer Graphics (SIGGRAPH Proceedings) 18*, 3 (July 1984), 212–222.

[95] Govaerts, Y. *A model of light scattering in three-dimensional plant canopies: a Monte Carlo ray tracing approach.* PhD thesis, Département of Physique, Faculté des Sciences, Université Catholique de Louvain-la-Neuve, 1995.

[96] Govaerts, Y., amd M. Verstraete, S. J., and Ustin, S. Three-dimensional radiation transfer modeling in a dycotyledon leaf. *Applied Optics 35*, 33 (November 1996), 6585–6598.

[97] Grant, L. Diffuse and specular characteristics of leaf reflectance. *Remote Sensing of Environment 22* (1987), 309–322.

[98] Grant, L., Daughtry, C., and Vanderbilt, V. Polarized and specular reflectance variance with leaf surface features. *Physiologia Plantarum 88*, 1 (1993), 1–9.

[99] Greenberg, D. Computers and architecture. *Scientific American 264*, 2 (February 1991), 104–109.

[100] Gupta, R., and Wooley, J. T. Spectral properties of soybean leaves. *Agronomy Journal 63* (1971), 123–126.

[101] Hageman, L., and Young, D. *Applied Iterative Methods*. Academic Press, New York, 1981.

[102] Hall, R. A characterization of illumination models and shading techniques. *Visual Computer 2*, 5 (November 1986), 268–277.

[103] Hall, R. *Illumination and Color in Computer Generated Imagery*. Springer-Verlag, New York, 1989.

[104] Hall, R., and Greenberg, D. A testbed for realistic image synthesis. *IEEE Computer Graphics and Applications 3*, 8 (November 1983), 10–20.

[105] Hamilton, D., and Morgan, W. Radiant-interchange configuration factors. Tech. Rep. TN-2836, NASA, 1952.

[106] Hammel, M. *Differential L-systems and their application to the simulation and visualisation of plant development*. PhD thesis, Department of Computer Science, University of Calgary, June 1996.

[107] Hammerley, J., and Handscomb, D. *Monte Carlo Methods*. Wiley, New York, 1964.

[108] Hanan, J. *Parametric L-Systems*. PhD thesis, University of Regina, June 1992.

[109] Hanrahan, P., and Krueger, W. Reflection from layered surfaces due to subsurface scattering. *Computer Graphics (SIGGRAPH Proceedings)* (August 1993), 165–174.

[110] Hanrahan, P., and Salzman, D. A rapid hierarchical algorithm for unoccluded environments. In *Eurographics Workshop on Photosimulation, Realism and Physics in Computer Graphics* (Amsterdam, June 1990), K. Bouatouch and C. Bouville, Eds., Elsevier, pp. 151–171.

[111] Hanrahan, P., Salzman, D., and Aupperle, L. A rapid hierarchical radiosity algorithm. *Computer Graphics (SIGGRAPH Proceedings) 25*, 4 (July 1991), 197–206.

[112] He, X., Torrance, J., Sillion, F., and Greenberg, D. A comprehensive physical model for light reflection. *Computer Graphics (SIGGRAPH Proceedings) 25*, 4 (July 1991), 175–186.

[113] Hecht, E., and Zajac, A. *Optics*. Addison-Wesley, Reading, Massachusetts, 1974.

[114] Heckbert, P. Writing a ray tracer. In *An Introduction to Ray Tracing* (San Diego, CA, 1989), A. Glassner, Ed., Academic Press.

[115] Heckbert, P. Adaptive radiosity textures for bidirectional ray tracing. *SIGGRAPH Proceedings, Annual Conference Series 24*, 4 (August 1990), 145–154.

[116] Heckbert, P. *Simulating Global Illumination Using Adaptative Meshing*. PhD thesis, University of California, Berkeley, January 1991.

[117] Heckbert, P., and Winget, J. Finite element methods for global illumination. Tech. rep., University of California, Berkley, January 1991.

[118] Hestenes, M., and Stiefel, E. Methods of conjugate gradients for solving linear systems. *Journal of Research National Bureau of Standards 49* (1952), 409–436.

[119] Hopcroft, J., and Ullman, J. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley, Reading, Massachusetts, 1974.

[120] Hosgood, B., Jacquemoud, S., Andreoli, G., Verdebout, J., Pedrini, G., and Schmuck, G. Leaf optical properties experiment 93. Tech. Rep. Report EUR 16095 EN, Joint Research Center, European Comission, Institute for Remote Sensing Applications, 1995.

[121] Hottel, H. Radiation heat transmission between surfaces separated by non-absorbing media. *Trans. ASME 53* (1931), 265–273.

[122] Hunter, R., and Harold, R. *The Measurement of Appearance*, second ed. John Wiley & Sons, New York, 1987.

[123] Immel, D., Cohen, M., and Greenberg, D. A radiosity method for non-diffuse environments. *Computer Graphics (SIGGRAPH Proceedings) 20*, 4 (1986), 133–142.

[124] Ishimura, A. *Wave Propagation and Scattering in Random Media*. Academic Press, New York, 1978.

[125] Jacquemoud, S., and Baret, F. Prospect: A model of leaf optical properties spectra. *Remote Sensing of Environment 34*, 2 (1990), 75–92.

[126] Jacquemoud, S., Ustin, S., Verdebout, J., Schmuck, G., Andreoli, G., and Hosgood, B. Estimating leaf biochemistry using prospect leaf optical properties model. *Remote Sensing of Environment 56* (1996), 194–202.

[127] Jensen, H. Importance driven path tracing using photon map. In *Rendering Techniques'95 (Proceedings of the Sixth Eurographics Rendering Workshop)* (Dublin, June 1995), P. M. Hanrahan and W. Purgathofer, Eds., Springer-Verlag, pp. 325–335.

[128] Judd, D., and Wyszecki, G. *Color in Business, Science and Industry*, third ed. John Wiley & Sons, New York, 1975.

[129] Kahn, H. Use of different Monte Carlo sampling techniques. In *Symposium on Monte Carlo Methods* (New York, 1956), H. Meyer, Ed., John Wiley & Sons, pp. 146–190.

[130] Kajiya, J. Anisotropic reflection models. *Computer Graphics (SIGGRAPH Proceedings) 19*, 3 (July 1985), 15–22.

[131] Kajiya, J. The rendering equation. *Computer Graphics (SIGGRAPH Proceedings) 20*, 4 (August 1986), 143–150.

[132] Kalos, M., and Whitlock, P. *Monte Carlo Methods*, vol. Volume I: Basics. John Wiley & Sons, New York, 1986.

[133] Kay, D., and Greenberg, D. Transparency for computer synthesized images. *Computer Graphics (SIGGRAPH Proceedings) 13*, 2 (1979), 158–164.

[134] Kay, Q., Daoud, H., and Stirton, C. Pigment distribution, light reflection and cell structure in petals. *Botanic Journal of the Linnean Society 83* (1981), 57–84.

[135] Kumar, R., and Silva, L. Light ray tracing through the leaf cross section. *Applied Optics 12*, 12 (1973), 2950–2954.

[136] Lafortune, E. *Mathematical Models and Monte Carlo Algorithms for Physically Based Rendering*. PhD thesis, Department of Computer Science, Faculty of Engineering, Katholieke Universiteit Leuven, 1996.

[137] Lafortune, E., and Willems, Y. Bi-directional path tracing. In *Proceedings of the Third International Conference on Computational Graphics and Visualization Techniques - Compugraphics'93* (Alvor, Portugal, December 1993), H. P. Santo, Ed., Graphic Science Promotions and Publications, pp. 145–153.

[138] Lafortune, E., and Willems, Y. D. Using the modified phong reflectance model for physically based rendering. Tech. rep., Department of Computer Science, K.U. Leuven, November 1994.

[139] Lalonde, P., and Fournier, A. Filtered local shading in the wavelet domain. In *Rendering Techniques'97 (Proceedings of the Eight Eurographics Rendering Workshop)* (Etienne, June 1997), J. Dorsey and P. Slusallek, Eds., Springer-Verlag, pp. 163–174.

[140] Lalonde, P., and Fournier, A. Generating reflected directions from brdf data. *Computer Graphics Forum (EUROGRAPHICS Proceedings) 16*, 3 (September 1997), 293–300.

[141] Lalonde, P., and Fournier, A. A wavelet representation of reflectance functions. *IEEE Transaction on Visualization and Computer Graphics 3*, 4 (October 1997), 329–336.

[142] Lancaster, P. *The Theory of Matrices*. Academic Press, San Diego, 1969.

[143] Lancaster, P., and Tismenetsky, M. *The Theory of Matrices*, second ed. Academic Press, San Diego, 1985.

[144] Lewis, G. The conservation of photons. *Nature 2981*, 118 (December 1926), 874–875.

[145] Lewis, R. Making shaders more physically plausible. In *Proceedings of the Fourth Eurographics Rendering Workshop* (Paris, June 1993), M. Cohen and C. Puech, Eds., pp. 47–62.

[146] Lewis, R. Making shaders more physically plausible. *Computer Graphics Forum 13* (June 1994), 109–120.

[147] Lewis, R. *Light-Driven Global Illumination with Wavelet Representation of Light Transport*. PhD thesis, The University of British Columbia, 1998.

[148] Lilley, C., Lin, F., Hewitt, W., and Howard, T. *Colour in Computer Graphics*. ITTI Computer graphics and Visualisation, Manchester Computing Centre, The University of Manchester, Manchester, England, December 1993.

[149] Longhurst, R. *Geometrical and Physical Optics*, third ed. Longman Group Limited, London, 1973.

[150] Ma, Q., Nishimura, A., Phu, P., and Kuga, Y. Transmission, reflection and depolarization of an optical wave for a single leaf. *IEEE Transactions on Geoscience and Remote Sensing 28*, 5 (September 1990), 865–872.

[151] MacAdam, D. *Color Measurements Theme and Variations*. Springer Verlag, Berlin, 1981.

[152] Marchuk, G., Mikhailov, G., Nazaraliev, M., Darbinjan, M., Kargin, B., and Elepov, B. *The Monte Carlo Methods in Atmospheric Optics*. Springer Verlag, Berlin, 1980.

[153] Martin, G., Myers, D., and Vogelmann, T. Characterization of plant epidermal lens effects by a surface replica technique. *Journal of Experimental Botany 42*, 238 (May 1991), 581–587.

[154] MathWorks. *MATLAB User's Guide*. Natick, MA, USA, November 1994. Version 4.2c.

[155] Max, N., Mobley, C., Keating, B., and Wu, E. Plane-parallel radiance transport for global illumination in vegetation. In *Rendering Techniques'97 (Proceedings of the Eight Eurographics Rendering Workshop)* (Etienne, June 1997), J. Dorsey and P. Slusallek, Eds., Springer-Verlag, pp. 239–250.

[156] McClendon, J. The micro-optics of leaves. i. patterns of reflection from the epidermis. *American Journal of Botany 71*, 10 (1984), 1391–1397.

[157] McClendon, J., and Fukshansky, L. On the interaction of absorption spectra of leaves - ii. the non-absorbed ray of the sieve effect and the mean optical pathlength in the remainder of the leaf. *Photochemistry and Photobiology 51*, 2 (1990), 211–216.

[158] Mech, R. *Modeling and Simulation of the Interaction of Plants with the Environment using L-systems and their Extensions*. PhD thesis, Department of Computer Science, University of Calgary, November 1997.

[159] Menzel, D. *Selected Papers on the Transfer of Radiation*. Dover Publications, New York, 1966.

[160] Meyer, G. Wavelength selection for synthetic image generation. *Computer Vision, Graphics, and Image Processing 41* (1988), 57–79.

[161] Meyer-Arendt, J. *Introduction to Modern and Classical Optics*. Prentice-Hall, New Jersey, 1984.

[162] Mobley, C. *Light and Water: Radiative Transfer in Natural Waters*. Academic Press, San Diego, 1994.

[163] Mynemi, R., Ross, J., and Asrar, G. A review on the theory of photon transport in leaf canopies. *Agricultural and Forest Meteorology 45* (1989), 1–153.

[164] Nakamae, E., Kaneda, K., Okamoto, T., and Nishita, T. A lighting model aiming at drive simulations. *Computer Graphics (SIGGRAPH Proceedings) 24*, 3 (August 1990), 395–404.

[165] Neumann, L. New efficient algorithms with positive definite radiosity matrix. In *Photorealistic Rendering Techniques 1994 (Proceedings of the Fifth Eurographics Rendering Workshop* (Darmstadt, June 1994), G. Sakas, P. Shirley, , and S. Muller, Eds., Springer-Verlag, pp. 219–237.

[166] Neumann, L., Feda, M., Kopp, M., and Purgathofer, W. A new stochastic radiosity method for highly complex scenes. In *Photorealistic Rendering Techniques 1994 (Proceedings of the Fifth Eurographics Rendering Workshop* (Darmstadt, June 1994), G. Sakas, P. Shirley, , and S. Muller, Eds., Springer-Verlag, pp. 219–237.

[167] Neumann, L., and Neumann, A. Photosimulation: Interreflection with arbitrary reflectance models and illumination. *Computer Graphics Forum (EUROGRAPHICS Proceedings) 8* (1989), 21–34.

[168] Neumann, L., and Neumann, A. Radiosity and hybrid methods. *ACM Transactions on Graphics 14*, 3 (July 1995), 233–265.

[169] Nievergelt, Y. Making any radiosity matrix symmetric positive definite. *Journal of the Illuminating Engineering Society 26*, 1 (1997), 165–172.

[170] Nievergelt, Y. Radiosity in illuminating engineering. *The UMAP Journal 18*, 2 (1997), 167–179.

[171] Nishita, T., Sirai, T., Tadamura, K., and Nakamae, E. Display of the earth taking into account atmospheric scattering. In *SIGGRAPH Proceedings, Annual Conference Series* (August 1993), pp. 175–182.

[172] Norman, A. *Soybean Physiology, Agronomy, and Utilization.* Academic Press, New York, N.Y., 1978.

[173] Palmer, K., and Williams, D. Optical properties of water in the near infrared. *Journal of the Optical Society of America 64*, 8 (August 1974), 1107–1110.

[174] Pattanaik, S., and Mudur, S. Computation of global illumination by Monte Carlo simulation of the particle model of the light. In *Third Eurographics Workshop on Rendering* (Amsterdam, 1992), A. Chalmers, D. Paddon, and F. Sillion, Eds., Elsevier, pp. 71–83.

[175] Pattanaik, S., and Mudur, S. The potential equation and importance in illumination computations. *Computer Graphics Forum (EUROGRAPHICS Proceedings) 12*, 2 (1993), 131–136.

[176] Phong, B. Illumination for computer generated pictures. *Communications of ACM 18*, 6 (June 1975), 311–317.

[177] Potter, D., and Cline, M. Massive parallel computational simulations in light scattering. *Proceedings of the IEEE 79*, 4 (April 1991), 567–572.

[178] Poulin, P., and Fournier, A. A model for anisotropic reflection. *Computer Graphics (SIGGRAPH Proceedings) 24*, 4 (August 1990), 273–282.

[179] Preisendorfer, R. *Radiative Transfer on Discrete Spaces.* Pergamon, New York, N.Y., 1965.

[180] Prusinkiewicz, P., and Lindenmayer, A. *The Algorithmic Beauty of Plants.* Springer Verlag, New York, 1990.

[181] Ross, J., and Marshak, A. Calculation of canopy bidirectional reflectance using the Monte Carlo method. *Remote Sensing of Environment 24* (1988), 213–225.

[182] Ross, J., and Marshak, A. Monte Carlo methods. In *Photon - Vegetation Interactions - Applications in Optical Remote Sensing and Ecology* (Berlin, 1991), R. Mynemi and J. Ross, Eds., Springer-Verlag, pp. 441–467. ch. 14.

[183] Ruhle, W., and Wild, A. The intensification of absorbance changes in leaves by light-dispersion. differences between high-light and low-light leaves. *Planta 146* (1979), 551–557.

[184] Rushmeier, H., and Torrance, K. Extending the radiosity method to include specularly reflecting and translucent materials. *ACM Transactions on Graphics 9*, 1 (January 1990), 1–27.

[185] Rushmeier, H. E., and Tynor, S. D. Incorporating the BRDF into an infrared scene generation system. In *Conference on Characterization, Propagation and Simulation of Infrared Scenes, SPIE Proceedings* (Florida, April 1990), vol. 1311.

[186] Saad, Y., Sameh, A., and P.Saylor. Solving elliptic difference equations on a linear array of processors. *SIAM Journal of Scientific and Statistical Computing 6*, 4 (1985), 1049–1063.

[187] Salisbury, F., and Ross, C. *Plant Physiology*, third ed. Wadsworth Publishing Company, Belmont, California, 1985.

[188] Schlick, C. An inexpensive brdf model for physically-based rendering. *Proc. of European Association for Computer Graphicas Conference and Exhibition - EUROGRAPHICS 13*, 3 (1994), 233–246.

[189] Schlick, C. A survey of shading and reflectance models. *Computer Graphics Forum (EUROGRAPHICS Proceedings) 13*, 2 (1994), 121–131.

[190] Schramm, M., Gondek, J., and Meyer, G. Light scattering simulations using complex subsurface models. In *Proceedings of Graphics Interface* (1997), pp. 56–67.

[191] Schroder, P., and Hanrahan, P. On the form factor between two polygons. In *SIGGRAPH Proceedings, Annual Conference Series* (1993), pp. 163–164.

[192] Schuster, A. Radiation through foggy atmosphere. *Astrophysical Journal 21*, 1 (January 1905), 1–22.

[193] Seliger, H., and McElroy, W. *Light: Physical and Biological Action*. Academic Press, New York, 1966.

[194] Shao, M., and Badler, N. Analysis and acceleration of the progressive refinement method. In *Proceedings of the Fourth Eurographics Rendering Workshop* (Paris, June 1993), pp. 247–258.

[195] Shewchuck, J. An introduction to the conjugate gradient method without the agonizing pain. Tech. rep., School of Computer Science, Carnegie Mellon University, 1994.

[196] Shinya, M., Mori, T., and Osumi, N. A system subdivision approach for large radiosity computation. *The Visual Computer 14* (1998), 18–30.

[197] Shirley, P. Physically based lighting calculations for computer graphics: A modern perspective. In *Eurographics Workshop on Photosimulation, Realism and Physics in Computer Graphics* (Amsterdam, June 1990), K. Bouatouch and C. Bouville, Eds., Elsevier, pp. 67–81.

[198] Shirley, P. *Physically based lighting for computer graphics*. PhD thesis, Dept. of Computer Science, University of Illinois, November 1990.

[199] Shirley, P. A ray tracing method for illumination calculation in diffuse-specular scenes. In *Graphics Interface* (Toronto, 1990), Canadian Information Processing Society, pp. 205–212.

[200] Shirley, P. Radiosity via ray tracing. In *Graphics Gems II* (1991), Ed. Academic Press, pp. 306–310.

[201] Shirley, P. Time complexity of Monte Carlo radiosity. In *Proceedings of the Annual Conference of the European Association for Computer Graphics - EUROGRAPHICS* (Amsterdam, 1991), F. Post and W. Barth, Eds., North-Holland, pp. 459–465.

[202] Shirley, P., Baranoski, G., Cho, B., Defrancesco, A., Gupta, V., Jones, B., Kamath, R., Kim, J., Kim, K., Loos, T., Ma, H., Meyer, C., Rusinov, D., Sampson, S., Vogl, G., Winnicka, B., and Zimmerman, K. *The GG Library Reference Manual.* Department of Computer Science, Indiana University, Bloomington, Indiana, USA, 1993.

[203] Shirley, P., and Sung, K. A ray tracing framework for global illumination systems. In *Graphics Interface* (Toronto, 1991), Canadian Information Processing Society, pp. 117–128.

[204] Shirley, P., and Wang, C. Distribution ray tracing: Theory and practice. In *Proceedings of the Third Eurographics Workshop on Rendering* (Bristol, England, 1992), pp. 200–209.

[205] Siegel, R., and Howell, J. *Thermal Radiation Heat Transfer.* MacGraw-Hill Kogakusha, Tokyo, 1972.

[206] Sillion, F. A unified hierarchical algorithm for global illumination with scattering volumes and object clusters. *IEEE Transactions on Visualization and Computer Graphics 1*, 3 (1995), 240–254.

[207] Sillion, F., Arvo, J., Westin, S., and Greenberg, D. A global illumination solution for general reflectance distributions. *Computer Graphics (SIGGRAPH Proceedings) 25*, 4 (July 1991), 187–196.

[208] Sillion, F., and Drettakis, G. Feature-based control of visibility error: A multi-resolution clustering algorithm for global illumination. *Computer Graphics (SIGGRAPH Proceedings)* (August 1995), 145–152.

[209] Sillion, F., Drettakis, G., and Soler, C. A clustering algorithm for radiance calculation in general environments. In *Rendering Techniques'95 (Proceedings of the Sixth Eurographics Rendering Workshop)* (Dublin, June 1995), P. M. Hanrahan and W. Purgathofer, Eds., Springer-Verlag, pp. 196–205.

[210] Sillion, F., and Puech, C. A general two-pass method integrating specular and diffuse reflection. *Computer Graphics (SIGGRAPH Proceedings) 23*, 3 (July 1989), 335–344.

[211] Sillion, F., and Puech, C. *Radiosity and Global Illumination*. Morgan Kaufmann Publishers, San Francisco, 1994.

[212] Smail, L. *Analytic Geometry and Calculus*. Appleton-Century-Crofts, New York, 1953.

[213] Smith, B., Boyle, J., Dongarra, J., Garbow, B., Ikebe, Y., Klema, V., and Moler, C. *Matrix Eigensystem Routines - EISPACK Guide*, second ed., vol. 6 of *Lecture Notes in Computer Science*. Springer-Verlag, 1976.

[214] Smits, B., and Meyer, G. Newton's colors: Simulating interference phenomena in realistic image synthesis. In *Eurographics Workshop on Photosimulation, Realism and Physics in Computer Graphics* (Amsterdam, June 1990), K. Bouatouch and C. Bouville, Eds., Elsevier, pp. 185–194.

[215] Sparrow, E., and Cess, R. *Radiation Heat Transfer*. Brooks/Cloe, Belmont, 1966.

[216] Stamminger, M., Slusallek, P., and Seidel, H. Bounded radiosity - illumination on general surfaces and clusters. *Computer Graphics Forum (EUROGRAPHICS Proceedings) 16*, 3 (September 1997), 309–317.

[217] Stiefel, E. Kernel polynomials in linear algebra and their numerical application. In *Further Contributions to the Solutions of Simultaneous Linear Equations and the Determination of Eigenvalues*. National Bureau of Standards, Applied Mathematical Series - 49, 1958.

[218] Suits, G. The calculation of the directional reflectance of a vegetative canopy. *Remote Sensing of Environment 2* (1972), 117–125.

[219] Takagi, A., Takaoka, H., Oshima, T., and Ogata, Y. Accurate rendering technique based on colorimetric conception. *Computer Graphics (SIGGRAPH Proceedings) 24*, 4 (August 1990), 263–272.

[220] Tanaka, S. Estimation of sunlit leaf area in tobacco plant community by the Monte Carlo method. estimation on direct sunlight. In *Photosynthesis and Utilization of Solar Energy. Level III Exp 1968, J IBP/PP - Photosynthesis Level III Group* (Tokyo, 1969), pp. 76–79.

[221] Todler, K., and Hammer, C. PARC++: A parallel CC++. *Software Practice and Experience 25*, 6 (June 1995), 623–636.

[222] Torrance, K., and Sparrow, E. Theory for off-specular reflection from roughned surfaces. *Journal of the Optical Society of America 57*, 9 (1967), 1105–1114.

[223] Tucker, C., and Garrat, M. Leaf optical system modeled as a stochastic process. *Applied Optics 16*, 3 (1977), 635–642.

[224] Turner, G. *Introduction to Paint Chemistry*. Chapman and Hall, New York, 1967.

[225] Veah, E. *Robust Monte Carlo Methods for Light Transport Simulation*. PhD thesis, Stanford University, December 1997.

[226] Verhoef, W. Light scattering by leaf layers with application to reflectance canopy modeling: the sail model. *Remote Sensing of Environment 16* (1984), 125–141.

[227] Verhoef, W. Earth observation modeling based on layer scattering matrices. *Remote Sensing of Environment 17* (1985), 164–178.

[228] Vernon, L., and Seely, G. *The Chlorophylls*. Academic Press, New York, 1966.

[229] VOGELMANN, T. Plant tissue optics. *Annual Review of Plant Physiology and Plant Molecular Biology 44* (1993), 231–251.

[230] VOGELMANN, T., AND MARTIN, G. The functional significance of palisade tissue: penetration of directional versus diffuse light. *Plant, Cell and Environment 16* (1993), 65–72.

[231] WALLACE, J., COHEN, M., AND GRENBERG, D. A two-pass solution to the rendering equation: A synthesis of ray tracing and radiosity methods. *Computer Graphics (SIGGRAPH Proceedings) 23*, 3 (July 1987), 311–320.

[232] WALTER-SHEA, E., NORMAN, J., AND BLAD, B. A two-pass solution to the rendering equation: A synthesis of ray tracing and radiosity methods. *Remote Sensing of Environment 29* (1989), 161–174.

[233] WARD, G. Measuring and modeling anisotropic reflection. *Computer Graphics (SIGGRAPH Proceedings)* (July 1992), 265–272.

[234] WATT, A., AND WATT, M. *Advanced Animation and Rendering Techniques.* Addison-Wesley, New York, 1992.

[235] WESTIN, S., ARVO, J., AND TORRANCE, K. Predicting reflectance from complex surfaces. *Computer Graphics (SIGGRAPH Proceedings) 26*, 2 (July 1992), 255–264.

[236] WHITTED, T. An improved illumination model for shaded display. *Communications of ACM 23*, 6 (1980), 343–349.

[237] WIEBELT, J. *Engineering Radiation Heat Transfer.* Holt Rimehart and Winston, New York, 1966.

[238] WILLSTATTER, R., AND STOLL, A. *Investigations on Chlorophyll.* Science Press, London, 1928.

[239] WOLFF, L. Diffuse-reflectance model for smooth dielectric surfaces. *Journal of the Optical Society of America 11*, 11 (1994), 2956–2968.

[240] Woolley, J. Reflectance and transmittance of light by leaves. *Plant Physiology 47* (1971), 656–662.

[241] Woolley, J. Refractive index of soybean leaf cell walls. *Plant Physiology 55* (1975), 172–174.

[242] Xu, H., Peng, Q., and Liang, Y. Accelerated radiosity method for complex environments. *Computers and Graphics 14*, 1 (1990), 65–71.

[243] Xu, W., and Fussell, D. S. Constructing solvers for radiosity equation systems. In *Photorealistic Rendering Techniques 1994 (Proceedings of the Fifth Eurographics Rendering Workshop* (Darmstadt, June 1994), G. Sakas, P. Shirley, , and S. Muller, Eds., Springer-Verlag, pp. 207–217.

[244] Yamada, N., and Fujimura, S. A mathematical model of reflectance and transmittance of plant leaves as a function of chlorophyll pigment content. In *International Geoscience and Remote Sensing Symposium* (September 1988), pp. 833–834.

[245] Young, D. *Iterative Solution of Large Linear Systems*. Academic Press Inc., New York, 1971.

[246] Zerlaut, G., and Anderson, T. Multiple-integrating sphere spectrophotometer for measuring absolute spectral reflectance and transmittance. *Applied Optics 20*, 21 (November 1981), 3797–3804.

[247] Zscheile, F., and Comar, C. Influence of preparative procedure on the purity of chlorophyll components as shown by absorption spectra. *Botanical Gazette 102* (March 1941), 463–481.

[248] Zscheile, F., Jr., J. W., Beadle, B., and Roach, J. The preparation and absorption spectra of five pure carotenoids pigments. *Plant Physiology 17*, 3 (July 1942), 331–346.

Appendix A
Monte Carlo Techniques for Directional Sampling

In this appendix we outline some Monte Carlo definitions and techniques and present a concise derivation of warping functions used in physically-based rendering. The material presented in this appendix have already been examined by computer graphics researchers [86, 136, 138, 146, 198]. This appendix differs by being oriented to scattering simulations and by providing derivation details frequently omitted in computer graphics papers. For a more comprehensive treatment of Monte Carlo methods the reader is referred to classic Monte Carlo texts [107, 132]. The sections of this appendix are organized as follows:

- Section A.1 briefly describes Monte Carlo techniques for numerical integration commonly used to solve the rendering equation.
- Section A.2 presents probability density functions used in rendering applications.
- Section A.3 derives warping functions used to generate scattered directions.
- Finally, Section A.4 summarizes the main concepts presented in this appendix.

A.1 Importance Sampling and Warping Transformations

The integral term of the rendering equation (Chapter 2) can be estimated using Monte Carlo techniques. Among these techniques one can highlight the *importance sampling* [107]. The idea behind this technique is simple. If the integrand is a product of two functions, and we know one of them, we can use this information to guide our sampling strategy. Usually areas that contribute more (large values) or vary more quickly will have more importance and will be sampled more densely. In this case we

need to compensate for the non-uniform sampling to avoid the introduction of bias in the final result.

Suppose that we need to determine Θ given by the following integral expression involving a real-valued function g:

$$\Theta = \int g(x)\, dx = \int \frac{g(x)}{p(x)}\, p(x)\, dx \tag{A.1}$$

where $p(x)$ represents the importance (real-valued) function, also called the *probability density function* (PDF) [107], which satisfies the following conditions [132, 136]:

- $p(x) \geq 0$ for each $x \in [0, 1]$ for which $f(x) \neq 0$,
- $\int_0^1 p(x)\, dx = 1$,
- $\frac{g(x)}{p(x)} < \infty$ except perhaps on a (countable) set of points.

Thus, to solve Equation A.1, we have two alternatives:

- draw samples from g, attach weights given by p and sum them, or
- draw samples ξ_i with the density given by p, evaluate $g(\xi_i)$ and sum them.

In this appendix we focus on the second alternative, and, in order to apply it, one often uses a technique called warping [86, 198]. This technique consists of generating uniform distributed samples in a canonical space, and, afterwards, deform that space to match the desired density given by p.

The key aspect of any importance sampling application is the selection of the PDF. For example, for the problem represented by Equation A.1, an optimal PDF $p(x)$ would be given by $p(x) = Cg(x)$, with the constant C represented by $\frac{1}{\Theta}$! Clearly this is not an option, since, if already knew Θ, we would not need to use Monte Carlo techniques to estimate it. The practical solution is to choose a function $\tilde{p}(x)$ "close" to $p(x)$. We will examine this aspect in more detail in the next sections.

A.2 Probability Density Functions

For the sake of simplicity, since the transmission is usually handled very similarly to reflection, in the following presentation we will focus on the BRDF (f_r), implicity including the BTDF by analogy.

In order to solve the rendering equation using importance sampling, one usually resorts to stochastic ray-tracing, in which new scattering directions have to be sampled recursively at each intersection point such that:

$$\begin{aligned} L_r(x,\psi) = \; & L_e(x,\psi) + \\ & f_r(x,\psi,\psi_i)L_e(x',\psi') + \\ & f_r(x,\psi,\psi_i)f_r(x',\psi',\psi_i')L_e(x'',\psi'') \; + ... \end{aligned} \tag{A.2}$$

where:

L_r	=	reflected radiance,
L_e	=	emitted radiance,
x	=	point on the surface,
f_r	=	BRDF,
ψ_i	=	direction from which light comes in,
ψ	=	direction in which light is reflected.

In a path tracing implementation of Equation A.2, for example, the x^n is chosen by sending a ray from x^{n-1} in the direction $-\psi^{n-1}$, and ψ^{n-1} is chosen according to a PDF based on the BRDF of the surfaces of the environment. Ideally, to choose reflected ray directions for radiosity [200] or stochastic ray-tracing [131] calculations, one should be able to sample according to the following PDF:

$$PDF(\psi) = \frac{f_r(x,\psi_i,\psi)\, cos\theta_i}{\int_\Omega f_r(x,\psi_i,\psi)\, cos\theta_i d\omega_i} \tag{A.3}$$

where:

Ω	=	hemisphere,
$d\omega_i$	=	differential solid angle around ψ_i,
$cos\theta_i$	=	cosine of the angle between ψ_i and the surface normal at x.

In practice one often has to use an approximating PDF. Before getting to the specifics of the PDFs presented in this appendix, it is relevant to examine more closely a BRDF which constitutes the basis for the derivation of a number of PDFs frequently used in rendering applications involving the simulation of non-Lambertian phenomena. The formulation of this BRDF is based on a modified version of the Phong model (Section 2.6.1).

As mentioned in Chapter 2, the Phong model does not have a physical basis and cannot be used as a BRDF in physically-based rendering applications. Immel *et al.* [123] has suggested a Phong-like BRDF, however, which was later constrainted by Lewis [146] to become physically plausible and used in rendering applications. This BRDF is given by:

$$f_r(x, \psi_i, \psi) = \frac{\rho_d}{S_d} + \frac{\rho_s}{S_s} cos^n \alpha_b \tag{A.4}$$

where:

α_b	=	angle between the halfway vector between the source and viewing directions and the normal,
ρ_d	=	the diffuse reflectance,
ρ_s	=	the specular reflectance,
n	=	the specular exponent,
S_d and S_s	=	constant factors included to conserve energy.

Recall that in order to conserve energy (Section 2.2) the following relation has to be satisfied:

$$\rho(x, \psi) = \int_\Omega f_r(x, \psi_i, \psi)\, cos\alpha\, d\omega \leq 1 \quad \forall x, \psi_i \tag{A.5}$$

with the differential solid angle given by:

$$d\omega = sin\alpha\, d\alpha\, d\beta \tag{A.6}$$

where:

α	=	polar angle ($\alpha \in [0, \frac{\pi}{2}]$ for the upper hemisphere),
β	=	azimuthal angle ($\beta \in [0, 2\pi]$).

Using Equation A.5, the constant factor S_d is given by:

$$\begin{aligned} S_d = \int_\Omega cos\alpha \, d\omega &= \int_{\beta=0}^{2\pi} \int_{\alpha=0}^{\frac{\pi}{2}} cos\alpha \; sin\alpha \, d\alpha \, d\beta \\ &= -\int_{\beta=0}^{2\pi} \int_{\alpha=0}^{\frac{\pi}{2}} cos\alpha \, dcos\alpha \, d\beta \\ &= -2\pi[\tfrac{cos^2\alpha}{2}]_0^{\frac{\pi}{2}} = \pi \end{aligned} \tag{A.7}$$

and the constant factor S_s is given by:

$$\begin{aligned} S_s = \int_\Omega cos^{n+1}\alpha \, d\omega &= \int_{\beta=0}^{2\pi} \int_{\alpha=0}^{\frac{\pi}{2}} cos^{n+1}\alpha \; sin\alpha \, d\alpha \, d\beta \\ &= -\int_{\beta=0}^{2\pi} [\tfrac{-cos^{n+1}\alpha}{n+2}]_0^{\frac{\pi}{2}} \, d\beta \\ &= \tfrac{1}{n+2} \int_{\beta=0}^{2\pi} d\beta \\ &= \tfrac{2\pi}{n+2} \end{aligned} \tag{A.8}$$

Replacing S_d and S_s in Equation A.4 [138] gives:

$$f_r(x, \psi_i, \psi) = \frac{\rho_d}{\pi} + \rho_s \, \frac{n+2}{2\pi} \, cos^n\alpha_b \tag{A.9}$$

Usually specular directions are sampled according to the following PDF based on the specular component of Equation A.9 [198]:

$$PDF(\alpha, \beta) = \frac{n+1}{2\pi} \, cos^n\alpha \tag{A.10}$$

Diffuse directions, on the other hand, are commonly sampled using the following PDF [138, 198]:

$$PDF(\alpha, \beta) = \frac{1}{\pi} \, cos\alpha \tag{A.11}$$

Another PDF based on the specular component of Equation A.9 was proposed by

Shirley and Wang [204], and it is given by:

$$PDF(\alpha_p, \beta_p) = \frac{n+2}{8\pi} cos^n\left(\frac{\alpha_p}{2}\right) \tag{A.12}$$

where:

α_p = angle between the perfect reflection direction and the outgoing direction,

β_p = azimuthal angle around the reflection direction.

A.3 Warping Functions

As pointed out by Lafortune and Willems [138], the space of directions in 3D space is two-dimensional. Consequently, the PDFs presented in the previous section can be sampled by selecting two uniform stochastic variables, ξ_1 and ξ_2, over the interval [0,1], and transforming them using the warping technique mentioned in Section A.1. In this section we present the derivation of the warping functions used in the research described in this dissertation (Chapter 4 and 6), which correspond to the PDFs described by Equations A.10 and A.11.

Before presenting the derivations, we shall briefly review some relevant concepts. If a random variable ξ ranges over some region $\mho$, then the probability that ξ will take on a value in some subregion $\mho_i \subset \mho$ is given by:

$$\mathcal{P}(\xi \in \mho_i) = \int_{\xi' \in \mho_i} PDF(\xi')\, d\mu(\xi') \quad (PDF : \mho \rightarrow \Re^1) \tag{A.13}$$

where $\mathcal{P}(event)$, also called cumulative distribution function [107], is the probability that the *event* is true [198]. In computer graphics applications $\mho$ is typically an area ($d\mu = dA = dx\, dy$) or a set of directions ($d\mu = d\omega = sin\alpha\, d\alpha\, d\beta$).

Initially, consider the PDF given by Equation A.10. The corresponding cumulative distribution function is given by:

$$\mathcal{P}(\alpha, \beta) = \int_0^{\beta} \int_0^{\alpha} \frac{n+1}{2\pi} cos^n\alpha'\, sin\alpha'\, d\alpha'\, d\beta' \tag{A.14}$$

Since the PDF presented in the integrand of Equation A.14 is separable [198], derivation techniques can be applied on each dimension to find the warping function

used to generated the corresponding scattered directions [198]. Solving Equation A.14 in the dimension associated with α results in:

$$\int_0^{\alpha} \frac{n+1}{2\pi} cos^n\alpha' \, sin\alpha' \, d\alpha' = \left[-\frac{n+1}{n+1} cos^{n+1}\alpha'\right]_0^{\alpha} = -cos^{n+1}\alpha \, + \, 1 \tag{A.15}$$

then we have:

$$\xi_1 = -cos^{n+1}\alpha \, + \, 1 \tag{A.16}$$

or:

$$\alpha = arccos(1-\xi_1)^{\frac{1}{n+1}} \tag{A.17}$$

Solving for the dimension associated with β gives:

$$\int_0^{\beta} \frac{1}{2\pi} \, d\beta' = \frac{\beta}{2\pi} \tag{A.18}$$

from which we get:

$$\xi_2 = \frac{\beta}{2\pi} \tag{A.19}$$

or:

$$\beta = 2\pi\xi_2 \tag{A.20}$$

Therefore, the corresponding warping function is represented by:

$$(\alpha, \beta) = (arccos(1-\xi_1)^{\frac{1}{n+1}}, 2\pi\xi_2) \tag{A.21}$$

Finally, consider the PDF given by Equation A.11. Then the corresponding cumulative density function is given by:

$$\mathcal{P}(\alpha, \beta) = \int_0^{\beta} \int_0^{\alpha} \frac{cos\alpha'}{\pi} \, sin\alpha' \, d\alpha' \, d\beta' \tag{A.22}$$

Like in the previous case, the PDF presented in the integrand of Equation A.22 is separable, and derivation techniques can be applied on each dimension to find the warping function used to generated the corresponding scattered directions. Thus, solving Equation A.22 in the dimension associated with α results in:

$$\int_0^{\alpha} 2 \, cos\alpha' \, sin\alpha' \, d\alpha' = 2\left[-\frac{cos^2\alpha'}{2}\right]_0^{\alpha} = 2\left(-\frac{cos^2\alpha}{2} + \frac{1}{2}\right) = -cos^2\alpha \, + \, 1 \tag{A.23}$$

from which we get:

$$\xi_1 = -cos^2\alpha + 1 \tag{A.24}$$

or:

$$\alpha = arccos(\sqrt{1-\xi_1}) \tag{A.25}$$

Solving for the dimension associated with β gives:

$$\int_0^\beta \frac{1}{2\pi}\, d\beta' = \frac{\beta}{2\pi} \tag{A.26}$$

then we have:

$$\xi_2 = \frac{\beta}{2\pi} \tag{A.27}$$

or:

$$\beta = 2\pi\xi_2 \tag{A.28}$$

Therefore, the corresponding warping function is represented by:

$$(\alpha, \beta) = (arccos(\sqrt{(1-\xi_1)}), 2\pi\xi_2) \tag{A.29}$$

A.4 Summary

In this appendix we have described warping functions used to generate scattered directions for stochastic ray-tracing implementations. They are based in standard Monte Carlo techniques and have been used very often in physically-based rendering applications. The derivation of these functions follows the principles of probability theory and can be applied, as needed, to derive other warping functions based on different BRDFs.

Appendix B
Accuracy of Methods for Form Factor Computation

There are several numerical methods for the computation of form factor between finite surfaces or patches. In order to select one these methods, speed and accuracy aspects are taken into account. For off-line computations of form factors, for example aiming at sun and shade simulations [99] and radiative transfer simulations in vegetation [48], accuracy becomes the key element since the errors introduced by these methods affect directly the accuracy of the radiosity solutions (Chapter 7).

Generally speaking, these methods can be divided into two groups: deterministic and nondeterministic. In this appendix we compare the accuracy of two methods which are representatives of these two groups. With these comparisons we intend to highlight some important aspects that affect the accuracy of the numerical methods for form factor computation. We also intend to discuss the differences in the magnitude and in the distribution of the errors associated with the form factors computed through these two different approaches. The remaining of this appendix is organized as follows:

- Section B.1 outlines the implications of the errors introduced during form factor computation.
- Section B.2 briefly describes the two methods used in the comparisons.
- Section B.3 presents the criteria used to compare the accuracy of these methods.
- Section B.4 presents the results of these comparisons.
- Section B.5 summarizes the main conclusions drawn from these comparisons and suggests directions in which they can be extended.

B.1 Implications of the Form Factor Numerical Errors

From the discussion presented in Chapters 7 and 8, it is clear that the errors introduced in the form factor computation affect the radiosity vector B, since, in order to obtain this vector, we must solve, either explicitly or implicitly, the radiosity system given by $GB = E$ (Section 7.1.1). Recall that the radiosity coefficient matrix G can be represented by $I - PF$ (Section 7.1.2), where PF is the scaled form factor matrix with all non-negative entries, and E represents the vector of emittances.

Besides the effect mentioned above, these errors may have other subtle effects, with possibly larger implications for the accuracy of radiosity solutions. Recall that the selection of iterative methods to solve the radiosity system of linear equations was based in two general assumptions, namely the diagonal dominance and the non-singularity of G. Although, in theory, these properties hold for radiosity systems, in practice, it may not be the case.

If the errors introduced in the computation of the form factors of a given surface of the environment are such that their summation relationship (Section 7.1.1) does not hold anymore, then we cannot ensure the diagonal dominance of G (Section 7.1.2). Moreover, in this case, we cannot guarantee that $\varrho(PF)$, the spectral radius of PF, be less than 1. In this case, according to Theorem 4.4 of Young (p. 35) [245], which, adapting to the radiosity notation, states that:

> The matrix $I - PF$ is nonsingular and the series $I + PF + (PF)^2 + \cdots$ converges if and only if $\varrho(PF) < 1$.

we cannot ensure the nonsingularity of G either.

The failure to fulfill the diagonal dominance property makes the iterative process more time consuming. This happens because, in this case, we need to apply techniques, such as pivoting [92], to enforce this property. The failure to fulfill the nonsingularity property has, however, more serious consequences, since, in this case, we cannot guarantee the convergence of the iterative methods at all.

B.2 Statement of the Methods

The Parametric Differential Method (PDM) [13] is an example of a deterministic method for form factor computation between two surfaces or patches. This method

evaluates the integrand of Equation 7.2 using a numerical technique called Gaussian quadrature [42]. Sample points are placed in both patches following a Gaussian distribution (Figure B.1). The value of the visibility term $V(x_i, x_j)$ (Section 7.1.1) is determined by testing the intersection of rays, which connect pairs of sample points, with other objects in the environment. For example, when one uses a 5-points Gaussian quadrature 25 sample points are placed on each patch, and 625 rays are used to test for obstacles between the two patches.

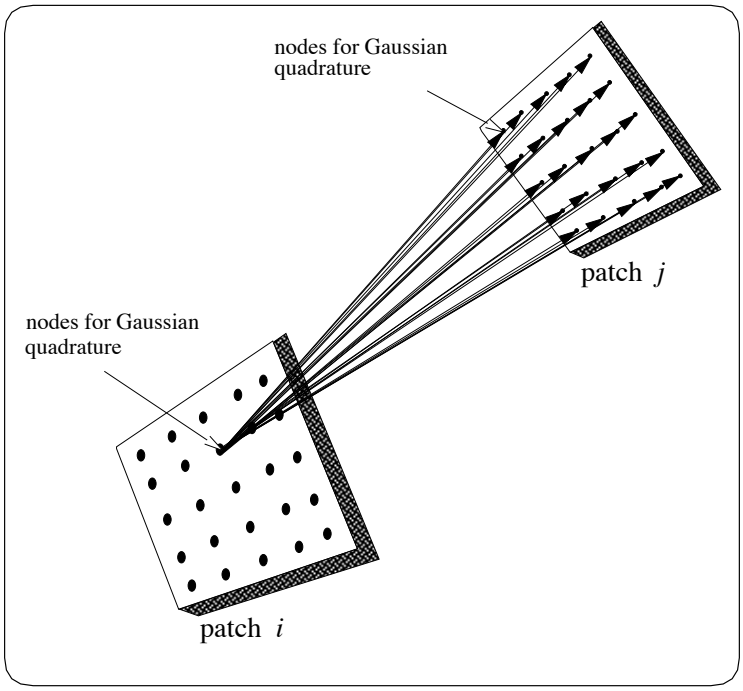

Figure B.1: Geometry regarding a deterministic method for form factor computation using a Gaussian distribution of sample points.

One of the nondeterministic methods for form factor computation uses the ray casting technique [11] and, although it can be extended to specular surfaces, it is usually applied to diffuse surfaces [200]. In order to compute the form factor between diffuse patches a set of sample points is randomly distributed in a source patch and rays are shot from these sample points in random directions having a cosine distribution (Figure B.2). The number of times each patch in the environment is hit by a

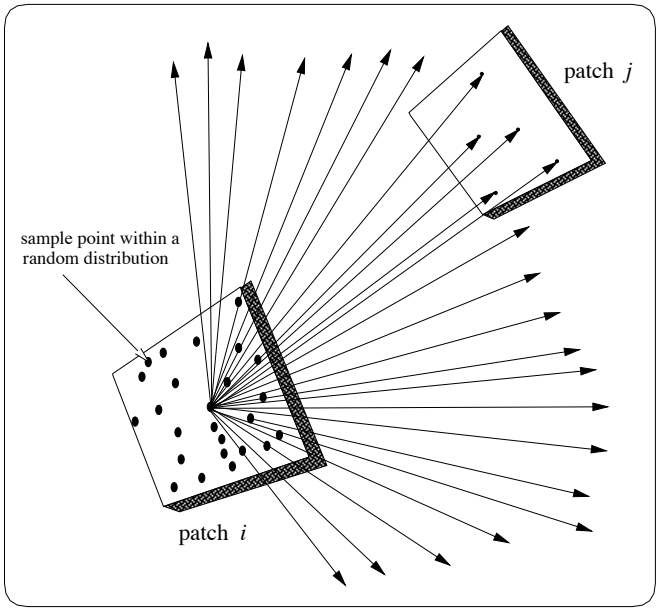

Figure B.2: Geometry regarding a nondeterministic method for form factor computation using a random distribution of sample points.

ray is recorded. In this case the form factor between a source patch i and a patch j is given by the number of rays that hit patch j divided by the total number of rays shot from patch i. Henceforth we will refer to this method as SR, *i.e.* shooting method with a random distribution of sample points.

B.3 Comparison Criteria

In order to evaluate the accuracy of numerical methods for form factor computation it is convenient to use geometries for which the analytical form factors are available to be used as reference. In the tests performed in this appendix we considered an environment formed by a cube, or box (Figure B.3), which allow us to fulfill this evaluation guideline. Recall that the summation of the form factors of one surface or patch regarding the other surfaces or patches of a closed environment must sum to 1 (Section 7.1.1). We use this summation relationship as well as the analytical form factors for parallel and perpendicular surfaces to evaluate the accuracy of the results provided by both methods regarding the test environment.

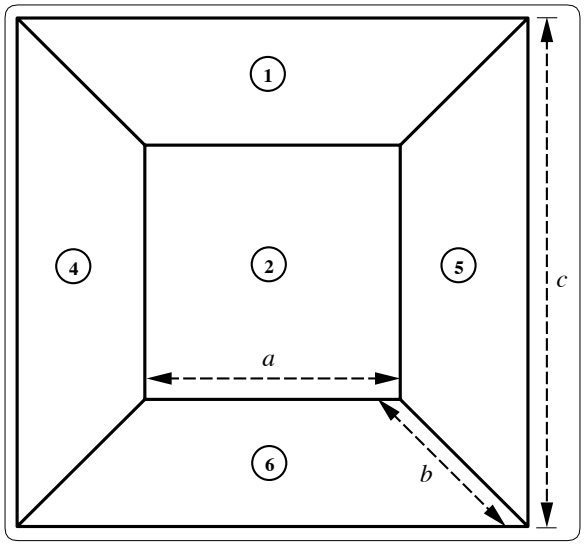

Figure B.3: Sketch of the test environment (face 3 is frontal).

The analytical form factor regarding parallel, directly opposed patches or surfaces, such as the faces 1 and 6 (Figure B.3), is given by the following expression presented by Hottel [121] and Hamilton and Morgan [105]:

$$F_{1-6} = \frac{2}{\pi XY}\left\{ln\left[\frac{(1+X^2)(1+Y^2)}{1+X^2+Y^2}\right]^{\frac{1}{2}} + X\sqrt{1+Y^2}\,tan^{-1}\left(\frac{X}{\sqrt{1+Y^2}}\right)\right. \\ \left. +Y\sqrt{1+X^2}\,tan^{-1}\left(\frac{Y}{\sqrt{1+X^2}}\right) - X\,tan^{-1}(X) - Y\,tan^{-1}(Y)\right\} \tag{B.1}$$

where X and Y are given by:

$$X = \frac{a}{c} \quad and \quad Y = \frac{b}{c} \tag{B.2}$$

The analytical form factor regarding two patches or surfaces of same length, having a common edge, and at angle of 90° to each other, such as the faces 1 and 2 (Figure B.3), can be obtained through a similar expression provided by Hottel [121] and Hamilton and Morgan [105]. Alternatively, assuming that the form factor between the analytical parallel faces have already been computed and using the summation relationship of the form factors, the analytical form factor regarding the perpendicular

surfaces, such as faces 1 and 2, can be obtained through the following expression:

$$F_{1-2} = \frac{1 - F_{1-6}}{4} \tag{B.3}$$

In the tests performed in this appendix we considered the dimensions of the cube presented in Figure B.3 given by $a = b = c = 6.0$. Replacing these values in the Equation B.1, the analytical form factor regarding the parallel faces of the cube is equal to 0.199824. Replacing this value in the Equation B.3, the analytical form factor regarding the perpendicular faces of the cube is equal to 0.200043. For the sake of simplicity, all the figures regarding analytical and numerical form factors presented in this appendix have been truncated after the sixth digit. Moreover, in order to determine the relative error regarding the summation relationship regarding each individual face and the form factors for the different pairs of faces we used the following expression:

$$relative\,error\,(\%) = \frac{|analytical\,value - numerical\,value|}{|analytical\,value|} \tag{B.4}$$

B.4 Results

Intuitively we may consider that the accuracy of both methods can be increased by subdividing the faces into a larger number of patches. Table B.1 shows the effects of finer surface subdivisions on the accuracy of the form factors computed using the PDM. In this experiment we used 25 sample points per patch and 25 rays per sample point. As we can see, there is a direct relationship between the increase in the number of patches and the increase on the accuracy of the form factors computed using the PDM. Such a direct relationship, however, cannot be always observed for nondeterministic methods, as demonstrated by the figures presented in Table B.2. This table shows the effects of finer surface subdivisions on the accuracy of the form factors computed using the SR. In this experiment we used 1000 sample points per patch and one ray per sample point. The increase on the number of rays per source patch can also increase the accuracy of the form factors. Table B.3 illustrates this aspect for an application of the SR (using 1000 sample points per patch and one ray

per sample point) in which the test environment was divided into 54 patches.

	number of patches		
	216	486	1014
Σ of face l	1.013652	1.009103	1.006303
Σ of face 2	1.013652	1.009103	1.006303
Σ of face 3	1.013652	1.009103	1.006303
Σ of face 4	1.013652	1.009103	1.006303
Σ of face 5	1.013652	1.009103	1.006303
Σ of face 6	1.013652	1.009103	1.006303
Relative error (%)	8.19	5.46	3.78

Table B.1: Effects of finer surface subdivisions on the accuracy of the PDM.

	number of patches		
	216	486	1014
Σ of face l	1.0	1.0	1.0
Σ of face 2	0.997333	0.997815	1.00034
Σ of face 3	1.001140	1.0009	1.00208
Σ of face 4	0.998056	1.00023	0.999124
Σ of face 5	1.003670	1.00049	1.00108
Σ of face 6	1.004640	0.998506	0.995243
Relative error (%)	2.43	0.53	0.91

Table B.2: Effects of finer surface subdivisions on the accuracy of the SR.

In order to compare the accuracy of both methods, PDM and SR, we considered the test environment divided into 1014 patches (169 patches per face) and 625 rays per source patch. The results present in Table B.4 seem to indicate that the non-deterministic method is more accurate than the deterministic method for the test environment used in our experiments. However, since the figures presented in this table correspond to summations, it may be possible that a cancellation of error terms

occurs when we perform the sum of form factors regarding each individual face of the cube. This cancellation may, in turn, hide a higher accuracy provided by the deterministic method for certain geometries. Indeed, Table B.5 shows that the deterministic method provides more accurate results than the nondeterministic method for form factors regarding pairs of parallel faces.

	number of rays	
	10000	1000000
Σ of face 1	1.0	0.999994
Σ of face 2	0.998789	0.999691
Σ of face 3	0.997356	0.999901
Σ of face 4	1.00064	0.999982
Σ of face 5	0.999844	0.999935
Σ of face 6	1.00197	1.00009
Relative error (%)	0.66	0.0587

Table B.3: Effects of higher ray densities on the accuracy of the SR.

	PDM	SR
Σ of face 1	1.006303	1.0
Σ of face 2	1.006303	1.00016
Σ of face 3	1.006303	0.97264
Σ of face 4	1.006303	0.999451
Σ of face 5	1.006303	1.00034
Σ of face 6	1.006303	1.00235
Σ of relative error (%)	3.78	0.61

Table B.4: Comparison between the PDM and the SR considering the environment divided into 1014 patches and using 625 rays per source patch.

The accuracy of both methods is affected by the singularity [57] in the integrand of Equation 7.2, which may occur when the two differential areas are too close. When the

singularity does occur, the task of determining which method provides more accurate results becomes more difficult. Table B.6 presents the relative errors regarding the form factors between the pairs of perpendicular faces. As we can see, the average of the relative errors regarding the form factors obtained using the nondeterministic method is smaller than those regarding the deterministic method. Although this aspect may suggest that the random distribution of sample points used by the SR may be less prone to singularity problems, it does not guarantee a smaller error for all the cases due to the stochastic nature of the SR.

$Face_i - Face_j$	PDM	Rel. error (%)	SR	Rel. error (%)
1 - 6	0.199841	0.02	0.199044	0.38
2 - 3	0.199841	0.02	0.199044	0.38
4 - 5	0.199841	0.02	0.199195	0.30

Table B.5: Form factors regarding the parallel faces divided into 169 patches each.

$Face_i - Face_j$	PDM	Rel. error (%)	SR	Rel. error (%)
1 - 2	0.201615	0.78	0.200502	0.23
1 - 3	0.201615	0.78	0.200634	0.30
1 - 4	0.201615	0.78	0.199044	0.50
1 - 5	0.201615	0.78	0.200625	0.29
2 - 4	0.201615	0.78	0.198892	0.58
2 - 5	0.201615	0.78	0.201051	0.50
2 - 6	0.201615	0.78	0.200691	0.32
3 - 4	0.201615	0.78	0.200161	0.06
3 - 5	0.201615	0.78	0.198305	0.86
3 - 6	0.201615	0.78	0.199138	0.45
4 - 6	0.201615	0.78	0.202159	1.06
5 - 6	0.201615	0.78	0.201164	0.56

Table B.6: Form factors regarding the perpendicular faces divided into 169 patches each.

B.5 Summary

In this appendix we have performed comparisons between two numerical methods for form factor computation based on different approaches. The results of these tests show that, even for simple environments, such as the one used in our experiments, a single method cannot provide the most accurate results for all possible geometries. Usually, for geometries not prone to singularity problems, the deterministic methods may provide more accurate results. Moreover, while the error associated with the deterministic methods is uniform, the error associated with the nondeterministic methods is nonuniform. This aspect suggests that statistical analysis tools shall be bring to bear on future comparisons. These comparisons shall take into account more complex environments, including occlusions and curved objects.

Finally, despite the simplicity of the environment used in our experiments, it was shown that selecting the "best" method is delicate, since no single method is superior in all cases. The relative accuracy depends on the geometrical characteristics of the environment. Therefore, developing of accurate practical solutions for off-line computation of form factors will likely require implementing different methods, with the application of a given method determined by the geometry at hand. For example, considering the environment used in our experiments, one could use the deterministic method to compute the form factor regarding the parallel faces and the nondeterministic method for the perpendicular faces.

Appendix C
Experiments with Eigenvectors

Arvo [7, 9] has shown that several fundamental operators that arise in global illumination can be uniformly approximated by matrices. Considering this aspect, if one can determine what the eigenvectors of a global illumination matrix, like the radiosity matrix G, represent in terms of the physical application, it may be possible to obtain close estimates for these eigenvectors. These eigenvectors could, in turn, be used to obtain low rank approximation of this matrix using single value decomposition (SVD) type approaches [92].

In some fields the physical meaning of eigenvalues and eigenvectors is well known. For instance, the eigenvectors of the stiffness matrix associated with a discretized structure of uniform density represent the natural modes of vibration of the structure being studied [195], and the corresponding eigenvalues define the natural frequencies of vibration [143]. In global illumination, however, the physical meaning of the eigenvalues and eigenvectors is still an open problem.

Although we have shown numerically that the distribution of the eigenvalues of radiosity matrices is affected by increases in the overall reflectance and level of occlusion of the environment (Section 8.6), the physical meaning of the corresponding eigenvectors has not been addressed yet. In this appendix we show some interesting features of using eigenvectors of radiosity matrices as solution vectors in graphics settings in order to contribute to the clarification of this question.

The remaining sections of this appendix are organized as follows:

- Section C.1 describes the experiments set up.
- Section C.2 provides the mathematical background for the experiments.

- Section C.3 presents the experiments regarding the eigenvectors corresponding to the smallest eigenvalue of the radiosity matrix G.
- Section C.4 presents the experiments regarding the eigenvectors corresponding to the largest eigenvalue of the radiosity matrix G.
- Section C.5 presents the experiments regarding the eigenvectors corresponding to the largest eigenvalue of the symmetrized radiosity matrix G^s.
- Section C.6 presents the experiments regarding the eigenvectors corresponding to the smallest eigenvalue of the symmetrized radiosity matrix G^s.
- Finally, Section C.7 summarizes the mains aspects and discusses the practical implications of the investigation presented in this appendix.

C.1 Experiments Set-Up

The test scene used in our experiments is described in Section 8.5.2. In this appendix we use two different sets of parameters to allow a wider range of observations. In set 1 we use one light source corresponding to 16 patches, with an irradiance equal to 10.0, and we set the reflectance of the sphere to 0.65. In set 2 we use two light sources, each corresponding to 16 patches, with an irradiance equal to 5.0, and we set the reflectance of the sphere to 0.55. The reflectances of the light sources are set to 0.1, and the values assigned for the reflectances of the other surfaces are presented in Table C.1. The form factors are computed using PDM [13]. The eigenvalues and eigenvectors are computed using *MATLAB* [154] through the QR method [42] whose algorithms are provided by the *EISPACK* routines [213].

The images presented in this paper are rendered using flat shading and greyscale to allow a better detection of the features associated with the eigenvector components. Figure C.1 shows the images corresponding to the solution (radiosity) vectors of the radiosity systems of linear equations regarding the two sets of parameters, which are solved using the Chebyshev method (Section 8.4). These images are used in our experiments as reference images to be compared with images obtained using eigenvectors as solution vectors.

Surface	reflectance
face 1	0.90
face 2	0.75
face 3	0.75
face 4	0.85
face 5	0.85
face 6	0.65

Table C.1: Sets of reflectances used in the experiments.

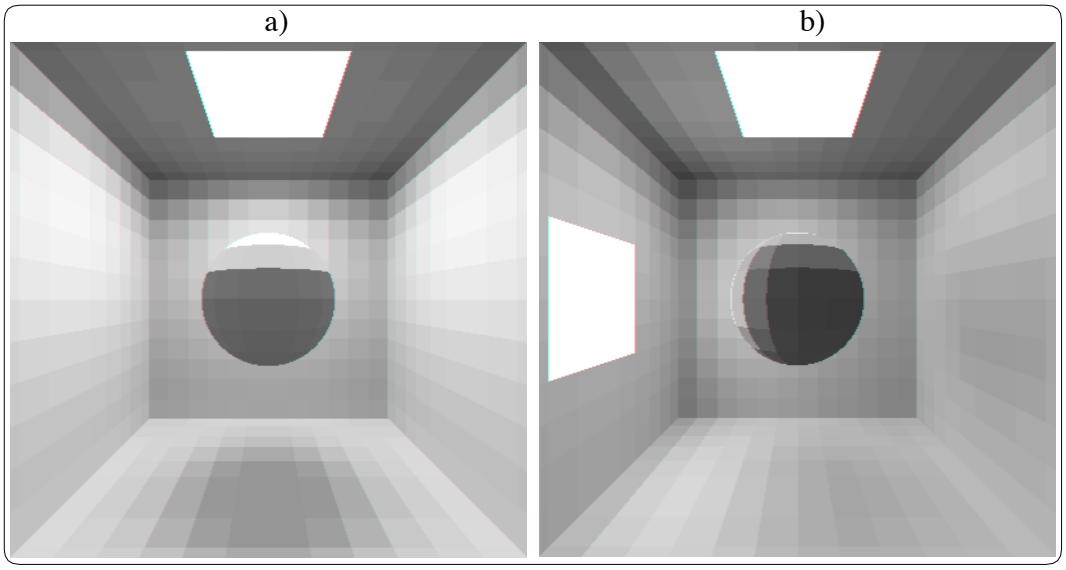

Figure C.1: Images corresponding to the solution (radiosity) vectors of the linear systems associated with: a) set 1 and b) set 2.

C.2 Mathematical Background

Any matrix K that can be expressed in the form:

$$K = sI - C, \quad s > 0, \quad H \geq 0 \tag{C.1}$$

for which $s \geq \varrho(H)$, the spectral radius of matrix H, is called an M-matrix [25].

Recall that the matrix G can be represented by $G = I - PF$ (Equation 7.12) where PF is the scaled form factor matrix with all non-negative entries. When a

scene has no concave patches, the diagonal entries of the matrix G are ones and the off-diagonal entries are the negatives of products of the form factors and the corresponding reflectances. When the reflectances are less than 1.0 in value, the summation of those products in any row is necessarily less than one. This implies that the spectral radius of PF is less than one. Therefore G is an M-matrix.

Since G is a nonsingular M-matrix and $\varrho(PF) < 1$, the matrix version of the Neumann lemma [25] for convergent series gives[1]:

$$G^{-1} = (I - PF)^{-1} = \sum_{k=0}^{\infty} (PF)^k = I + PF + (PF)^2 + (PF)^3 + \cdots \qquad \text{(C.2)}$$

This implies that the inverse G^{-1} of G is a positive matrix, having all positive components. In other words, G^{-1} is a nonnegative matrix.

A directed graph for a square matrix $K \in \mathcal{C}_{n\times n}$ is obtained by connecting distinct points N_i and N_j of the complex plane with a direct line $\vec{N_iN_j}$ for each nonzero element of K. For example, considering a matrix $K \in \mathcal{C}_{3\times 3}$ given by:

$$K = \begin{bmatrix} 1.0 & 0.2 & 0.0 \\ 0.2 & 1.0 & 0.3 \\ 0.0 & 0.3 & 1.0 \end{bmatrix}$$

its directed graph is represented in Figure C.2.

A directed graph is strongly connected if and only if, for each pair of nodes N_i, N_j with $i \neq j$, there are nodes $N_{t_1}, \cdots N_{t_r}$ of the graph forming a direct path

$$\vec{N_iN_{t_1}}, \vec{N_{t_1}N_{t_2}}, ..., \vec{N_{t_r}N_j}$$

connecting N_i to N_j.

The square matrix $K \in \mathcal{C}_{n\times n}$ is irreducible [25, 28, 143] if and only if its directed graph is strongly connected [142]. Since the visibility of patch i from patch j implies the reverse, the directed graphs for G and G^{-1} are strongly connected. Therefore G and G^{-1} are also irreducible.

The Perron-Frobenius theory [143], which concerns square irreducible nonnegative matrices, implies that the largest eigenvalue $\zeta_{\max}$ of G^{-1} is real and positive, and the

[1] Assuming that $\sum_{j=1}^{n} F_{ij} = 1$ holds for all i (Appendix B).

corresponding eigenvector ν of G^{-1} has all components strictly of the same sign [28]. Multiply the relationship:

$$G^{-1}\nu = \zeta_{\max}\nu \tag{C.3}$$

through by G to get $\nu = \zeta_{\max}G\nu$, or

$$G\nu = \Lambda_{\min}\nu \tag{C.4}$$

where $\Lambda_{\min} = 1/\zeta_{\max}$. Hence, the eigenvector of G corresponding to its smallest eigenvalue, Λ_{min}, can be scaled in such way that all its components are positive, and so can be physically interpreted as a radiosity vector. In the next section we show some images obtained this way, and try to give a physical interpretation to them.

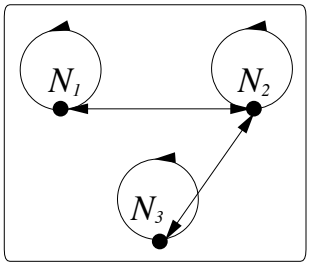

Figure C.2: The directed graph of a square matrix $K \in \mathcal{C}_{3\times3}$.

C.3 Eigenvectors Corresponding to the Smallest Eigenvalue of the Radiosity Matrices

Figure C.3 shows the eigenvectors ν_S corresponding to the smallest eigenvalues of the radiosity matrices G for two different choices of configuration and/or parameters. As expected from the Perron-Frobenius theory, their components have all the same sign. We can notice seven distinguished groups of points which are associated, from left to right, with the sphere and the six faces. The eigenvector components corresponding to the patches with lowest reflectances, which in our experiments correspond to the emitter patches, are represented by the points with lowest absolute values.

After taking the absolute values and normalizing these eigenvectors (by dividing all their components by the absolute value of their largest component), we use them

as solution vectors to display the images of the scenes (Figure C.4, top row). The features presented in these images seem to be associated with the distribution of the reflectances in the scenes (Figure C.4, bottom row). In Figure C.4a we can notice that the top of the sphere, which is closer to an area with low reflectance, is darker than its bottom. In Figure C.4b, areas of the scene that are directly exposed to the luminaires are also darker, possibly due to the low reflectances assigned to the emitter patches. These features seem to indicate that the absolute values of components of the eigenvectors ν_S are directly proportional to the reflectance values of the corresponding patches. Furthermore, they are also associated with the direct interaction of reflectances. In other words, patches directly exposed to areas with low reflectance, shown as darker areas in the images, correspond to components of the eigenvectors ν_S with low absolute values (assuming normalized eigenvectors).

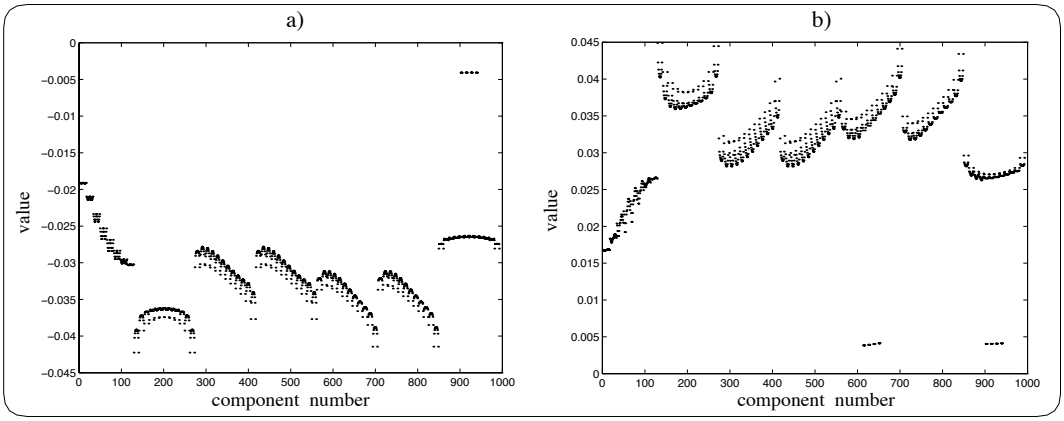

Figure C.3: Eigenvectors ν_S of the matrices G associated with: a) set 1 and b) set 2.

C.4 Eigenvectors Corresponding to the Largest Eigenvalue of the Radiosity Matrices

For the sake of completeness, we also look at the eigenvector ν_L corresponding to the largest eigenvalue of the radiosity matrices G (Figure C.5). After taking the absolute values and normalizing these eigenvectors, we use them to display the images

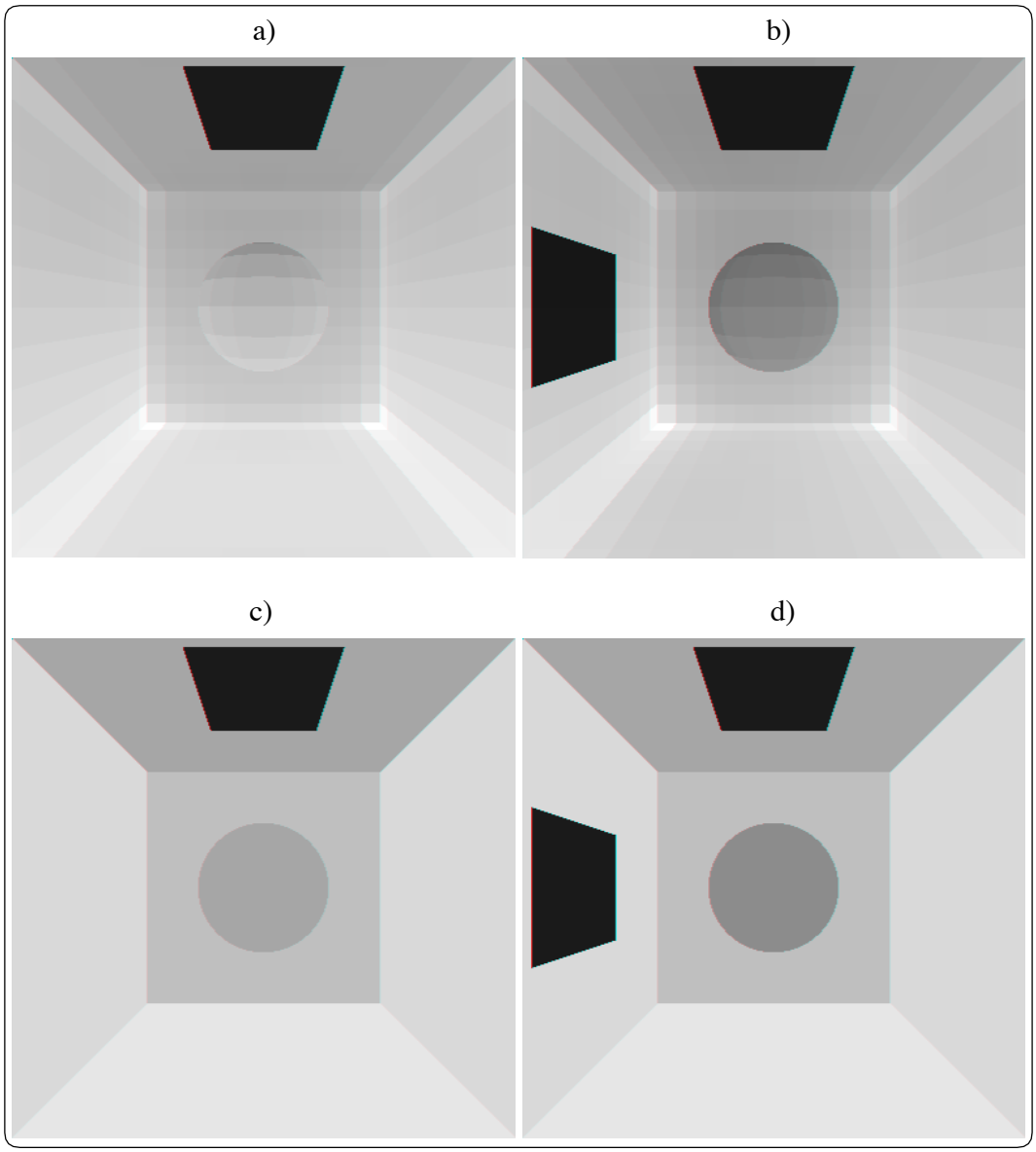

Figure C.4: Images obtained using as solution vectors the eigenvectors ν_S of the matrices G associated with: a) set 1 and b) set 2; and images obtained using as solution vectors the vectors of reflectances associated with: c) set 1 and d) set 2.

regarding the two sets of parameters (Figure C.6). However, these images do not reveal features that can lead to a physical interpretation as in the previous case.

A possible explanation for this is that the eigenvector corresponding to the largest

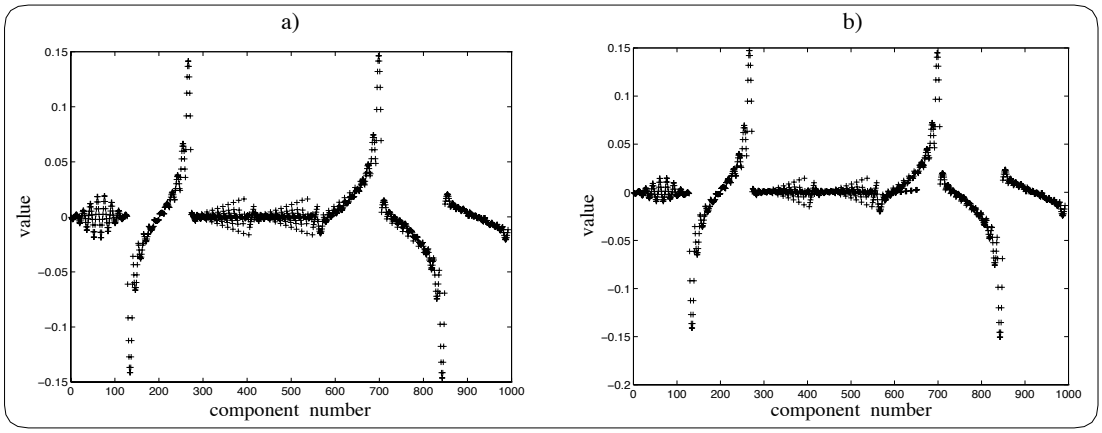

Figure C.5: Eigenvectors ν_L of the matrices G associated with: a) set 1 and b) set 2.

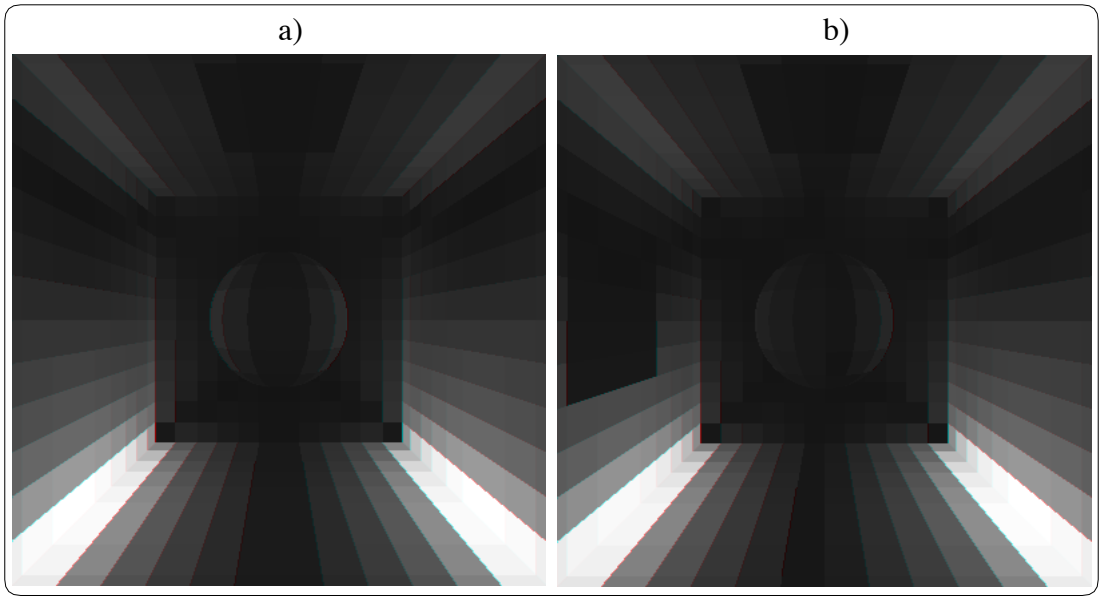

Figure C.6: Images obtained using as solution vectors the eigenvectors ν_L of the matrices G associated with: a) set 1 and b) set 2.

eigenvalue of G is also the eigenvector corresponding to the smallest eigenvalue of G^{-1}. If the radiosity vector B is expanded in terms of the eigenvectors ν of G^{-1}, then B may be approximated reasonably well by a multiple of ν, namely ν_S, due to the sign condition (Section C.2) and ν_S, but not by ν_l also due to the sign condition.

C.5 Eigenvectors Corresponding to the Largest Eigenvalue of the Symmetric Radiosity Matrices

A SVD type approach [92] may be used to provide a low rank approximation for the symmetric matrices G^s (Section 7.4). A low rank approximation of G^s is given by:

$$G^s_p = \Lambda_n \nu_n \nu_n^{'} + \Lambda_{n-1} \nu_{n-1} \nu_{n-1}^{'} + ... + \Lambda_{n-p} \nu_{n-p} \nu_{n-p}^{'} \qquad \text{for} \quad p \leq n-2 \qquad \text{(C.5)}$$

where $\nu_1', \nu_2', ..., \nu_n'$ correspond to the transposes of the eigenvectors of G^s. For a symmetric matrix the SVD is the same as the eigenvalue-eigenvector decomposition. The principal components are then the eigenvectors, ν_L, corresponding to the largest eigenvalues. Because of this, we decided to extend our investigation to the eigenvectors ν_L of the matrices G^s (Figure C.7).

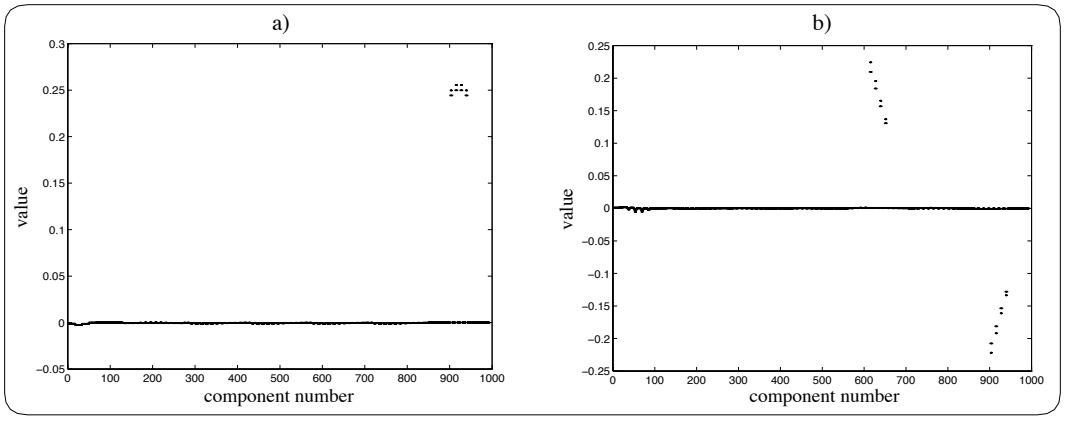

Figure C.7: Eigenvectors ν_L of the symmetric matrices G^s associated with: a) set 1 and b) set 2.

After taking the absolute values and normalizing the eigenvectors ν_L, we use them as solution vectors to display the images regarding the two sets of parameters. As one would expect looking at the plots of eigenvectors (Figure C.7a and C.7b), the images (Figure C.8a and C.8b) are almost completely dark, with the exception of the emitter patches. These images were displayed using a Gamma correction function [75] provided by *XV* [37] in which the Gamma value, γ, is set to 1.0. When we increase

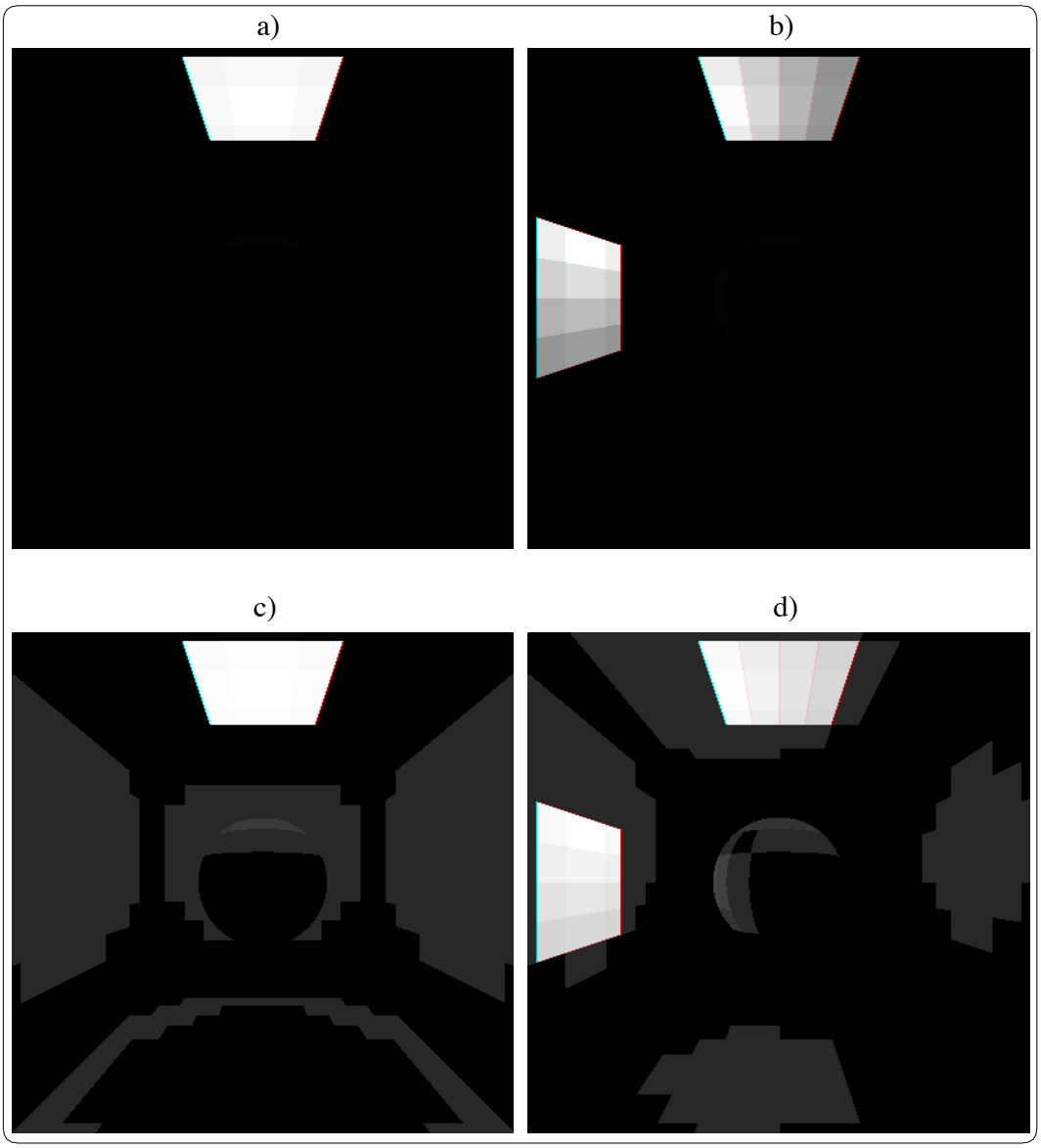

Figure C.8: Images obtained using as solution vectors the eigenvectors ν_L of the matrices G^s associated with: a) set 1 (with $\gamma = 1.0$), b) set 2 (with $\gamma = 1.0$), c) set 1 (with $\gamma = 2.2$) and d) set 2 (with $\gamma = 2.2$).

this value to $\gamma = 2.2$, some interesting features appeared (Figure C.8c and C.8d). These features seem to be related with the paths of direct light propagation. They also show that there is more useful information associated with the components of

the eigenvectors ν_L than the almost straight lines in the plots of Figure C.7 indicate. Furthermore, on face 2 of the image presented in Figure C.8d there is no sign of any feature associated with the paths of direct light propagation, as one would expect since that surface is also exposed to the luminaires.

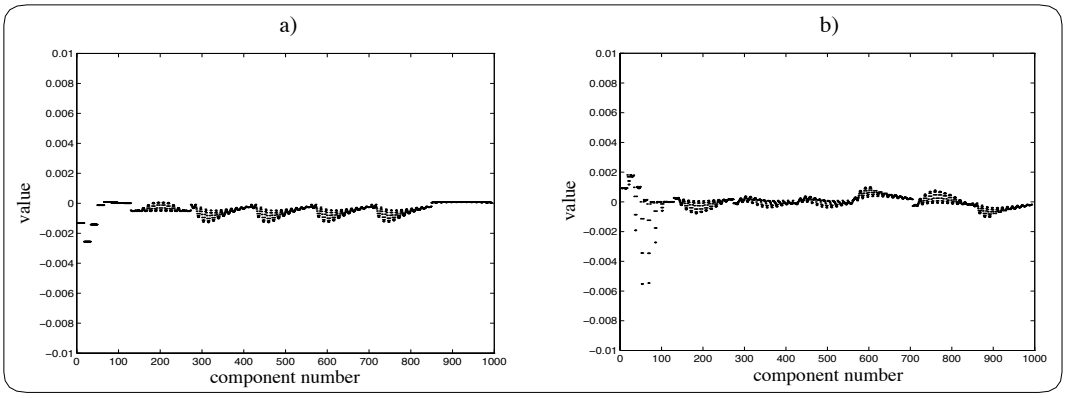

Figure C.9: Zoom in of the components of the eigenvectors ν_L regarding the non emitter patches and associated with a) set 1 and b) set 2.

Figure C.9 presents a zoom in of the components of the eigenvectors ν_L associated with the nonemitter patches, and reveals the patterns associated with the features presented in Figures C.8c and C.8d. To analyze the physical meaning of these patterns more closely, we set the components of the eigenvectors ν_L associated with the emitter patches to 1.0, and took the absolute values and normalized the remaining ones. The resulting images, presented in Figures C.10a and C.10b, are very close to the solution images (Figures C.1a and C.1b). Increasing the value of γ from 1.0 to 2.2, which has the effect of increasing the brightness of the scenes, we can notice that the similarities with the solution images become even more evident (Figures C.10c and C.10d).

Where does the association with the paths of direct light propagation comes from? Looking at the graphs presented in Figures C.7 and C.9, we can notice that the absolute values of the components of the eigenvectors ν_L of G^s are inversely proportional to the reflectance values of the corresponding patches. Moreover, the components with the highest absolute values, henceforth called dominants, are associated with

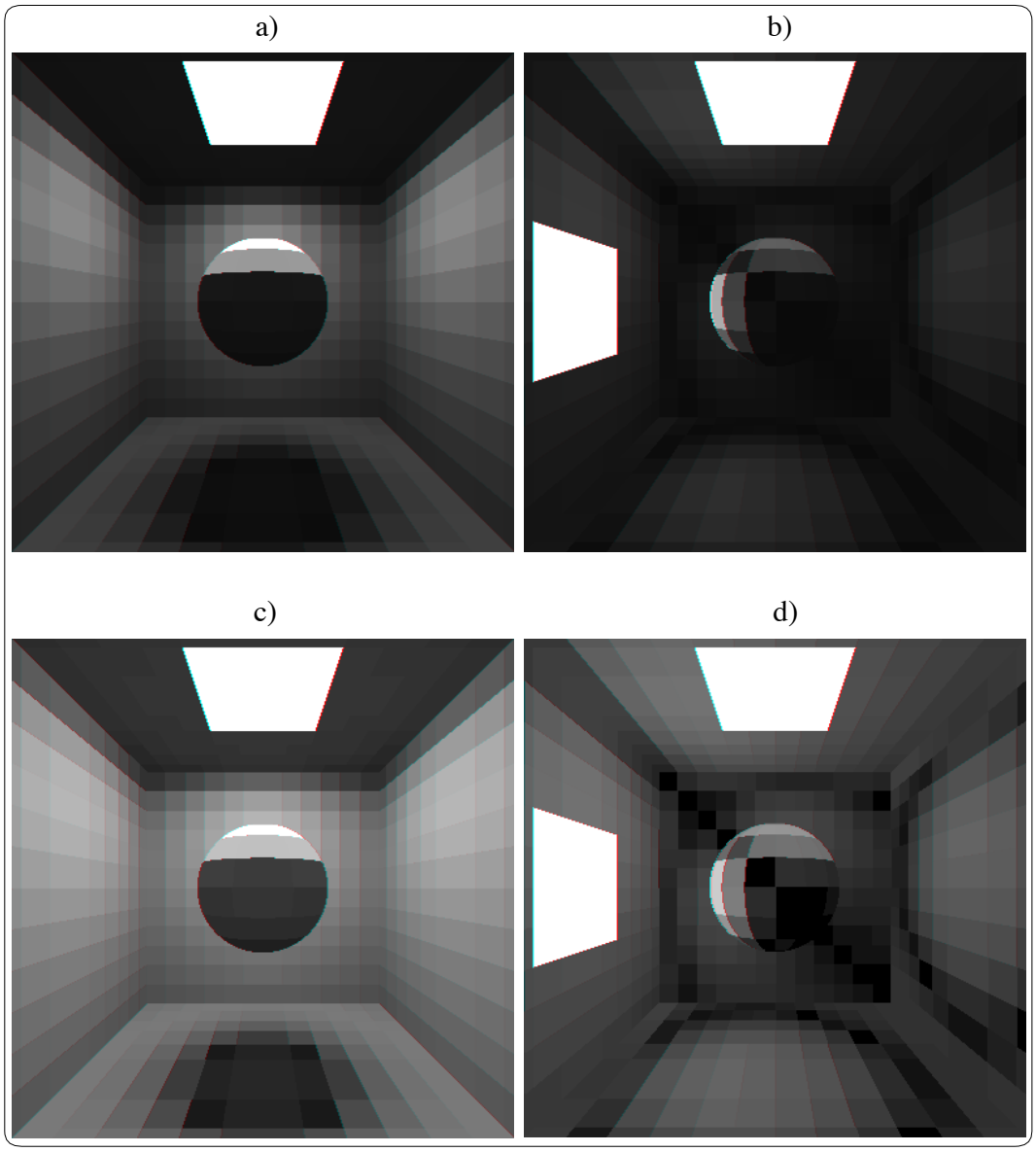

Figure C.10: Images obtained using as solution vectors the adjusted versions of the eigenvectors ν_L regarding the matrices G^s associated with: a) set 1 (with $\gamma = 1.0$), b) set 2 (with $\gamma = 1.0$), c) set 1 (with $\gamma = 2.2$) and d) set 2 (with $\gamma = 2.2$).

the patches with the lowest reflectance in the environment, which correspond in our experiments to the emitter patches. The next components with high absolute values are those whose corresponding patches are directly exposed to the patches associated

with the dominant components (Figures C.8c and C.8d). If we assign different reflectance values to the emitter patches such that they no longer correspond to the dominant components of the eigenvectors ν_L of G^s (Figure C.11), the association with the paths of direct light propagation can not be established, as we can see in the images presented in Figure C.12, where we set the reflectance of the emitter patches to 0.9. In this case, the dominant components will correspond to the sphere patches, since they now present the lowest reflectances in both scenes. Furthermore, the scene in which we assign a lower reflectance value for the sphere (Figure C.12b) presents a higher brightness than the other one (Figure C.12a).

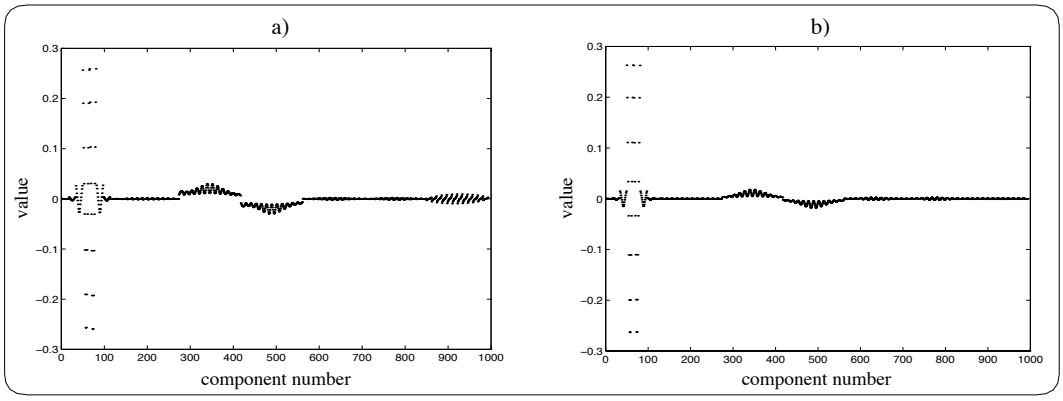

Figure C.11: Eigenvectors ν_L of the symmetric matrices G^s (with the reflectance of the emitter patches set to 0.9) associated with: a) set 1 and b) set 2.

To further investigate the relationship between the components of the eigenvectors ν_L and the reflectances of the patches, we form vectors V, in which the entry V_i corresponds to the diagonal entry Q_{ii} of Q (matrix use to make the radiosity matrix symmetric (Section 7.4), which, in turn, are represented by the quotient of area, A_i, and the reflectance, ρ_i, of a patch i. After normalizing these vectors we use them to display the images regarding the two sets of parameters. Comparing these images (Figure C.13) with the previous ones (Figure C.12), we can notice a similar color gradation on the spheres, especially the dark spots on the top and on the bottom.

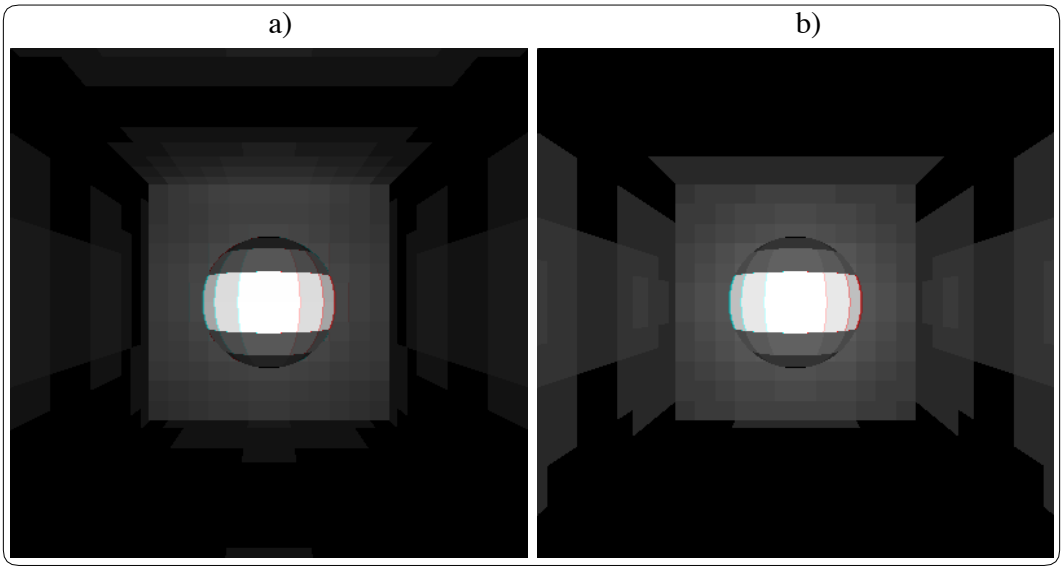

Figure C.12: Images obtained using as solution vectors the eigenvectors ν_L of the matrices G^s (with the reflectance of the emitter patches set to 0.9) associated with: a) set 1 (with $\gamma = 2.2$) and b) set 2 (with $\gamma = 2.2$).

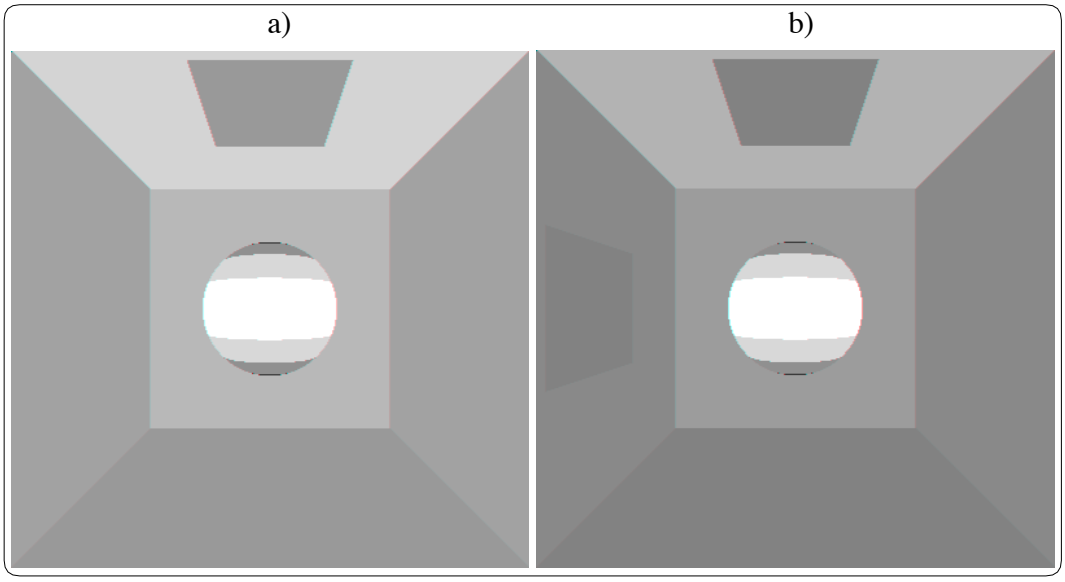

Figure C.13: Images obtained using as solution vectors the vectors V associated with: a) set 1 and b) set 2.

This aspect suggests that the components of the eigenvectors ν_S of the matrices G^s are not only associated with the direct interaction of reflectances, but they are also associated with the areas of the patches. This additional dependency may come from the symmetrization process ($G^s = QG$).

C.6 Eigenvectors Corresponding to the Smallest Eigenvalue of the Symmetric Radiosity Matrices

For the sake of completeness, we also looked at the eigenvectors ν_S of symmetric radiosity matrices corresponding to their smallest eigenvalues (Figure C.14). Similarly, as for the eigenvectors ν_S of the matrices G, all components of the eigenvectors ν_S of matrices G^s have strictly the same sign.

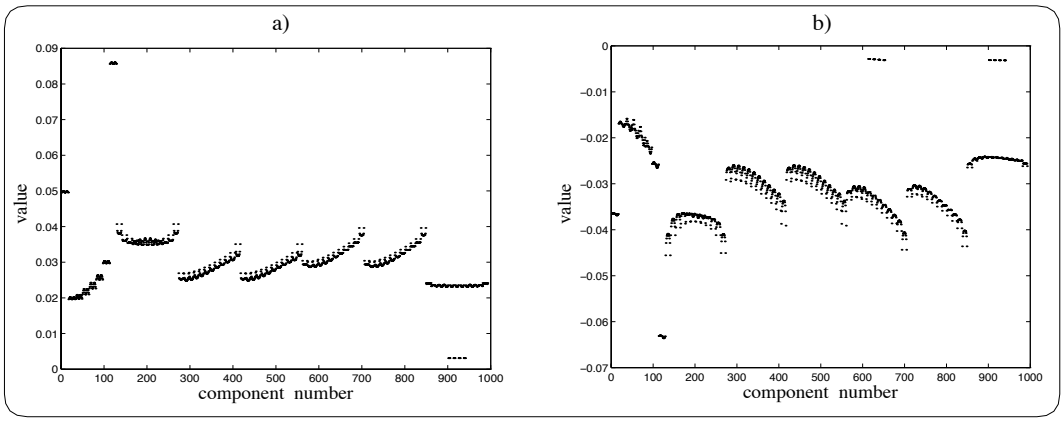

Figure C.14: Eigenvectors ν_S of the symmetric matrices G^s associated with: a) set 1 and b) set 2.

Figure C.15 (top row) presents the images obtained using these eigenvectors as solution vectors, after taking the absolute values and normalizing them. We can notice that the absolute values of the components of the eigenvectors ν_S of G^s are directly proportional to the reflectances of the corresponding patches. Moreover, like in the images regarding the eigenvectors ν_S of G, areas exposed to direct interaction with low reflectance patches are darker. However, we can also notice bright spots on the top and on the bottom of the spheres that are not presented in the previous

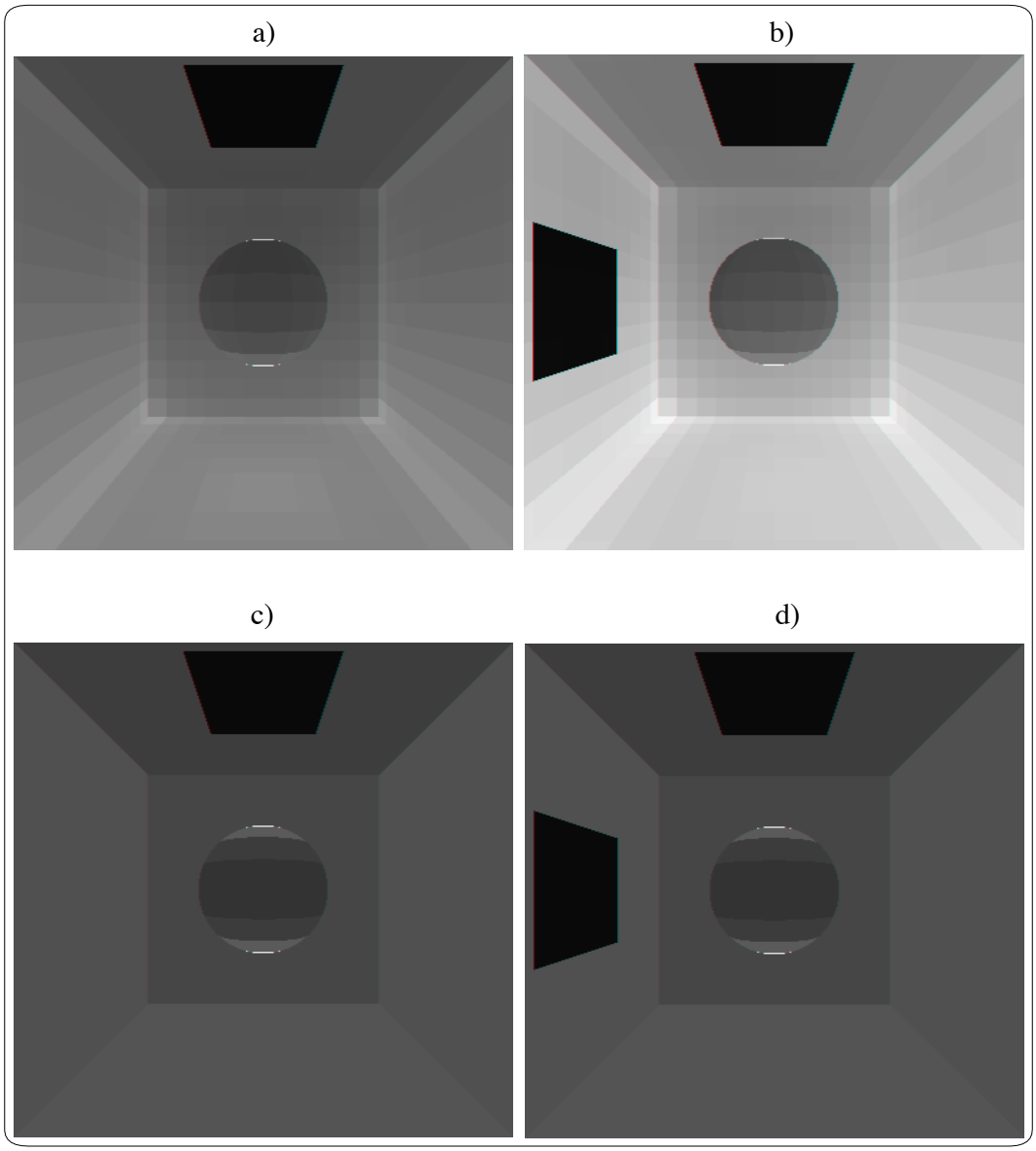

Figure C.15: Images obtained using as solution vectors the eigenvectors ν_S of the matrices G^s associated with: a) set 1 and b) set 2; and images obtained using as solution vectors the vectors P associated with: c) set 1 and d) set 2.

cases. The bright spots are represented in the graphs of Figure C.14 by points notably plotted away from the group of points associated with the spheres (numbers 1-128), and have a direct effect in the eigenvectors normalization. The points corresponding

to the bright spots associated with set 1 present a higher variance than the points corresponding to the bright spots associated with set 2. As a result, the image in Figure C.15a is considerably darker than the image in Figure C.15b, despite set 2 having a smaller average reflectance.

To investigate the origin of these bright spots, we compute vectors U, in which the entry U_i corresponds to the inverse of the diagonal entry Q_{ii} of Q. After normalizing these vectors we use them to display the images regarding the two sets of parameters. In these images (Figure 15c and 15d) we can notice the same bright spots. This aspect suggests that the components of the eigenvectors ν_S of the matrices G^s are not only associated with the direct interaction of reflectivities, but they are also associated with the areas of the patches. This additional dependency may also come from the symmetrization process.

C.7 Discussion and Summary

In this appendix we have shown experiments using eigenvectors of the radiosity coefficient matrix G and its symmetric version G^s as solution vectors. The components of the eigenvectors associated with the smallest eigenvalue of G have values directly proportional to the reflectances of the corresponding patches in the scene. Furthermore, eigenvector components regarding patches exposed to low reflectance regions have lower absolute values than the remaining components. The components associated with the largest eigenvalue of G^s, however, are inversely proportional to the reflectances of the corresponding patches. Moreover, for a scene in which the lowest reflectance values are assigned to the emitter patches, a common procedure in radiosity applications, this eigenvector provides a very close approximation to the solution vector. This aspect suggests three possible applications for this eigenvector:

- To provide an environment lighting preview for illumination engineering purposes such as selection of luminaires types and locations;

- To be used as an initial approximation by iterative methods oriented to the solution of radiosity systems. We examine this possibility in more detail in the Appendix D;

- To provide low rank approximations of G^s.

The first two applications mentioned above incorporate the additional overhead of computing the eigenvalues and eigenvectors. Fortunately, the standard methods used to determine the eigenvalues and eigenvectors of a matrix, such as the Lanczos algorithm [64, 92], are themselves iterative. Then, as appropriately noted by DiLaura and Franck [68] and Ashdown [10], since these methods have the ideal characteristic of determining the eigenvalues in order of their size, they can be formulated to determine the largest eigenvalue first and stop the process.

The third application mentioned above relies on the knowledge of the physical meaning of eigenvalues and eigenvectors in the radiosity context. With this information we could obtain estimates of the eigenvalues and eigenvectors without having to explicitly form G^s. These estimates could, in turn, be used by SVD type approaches to obtain a low approximation of G^s, saving both time and memory requirements for radiosity applications. Unfortunately, the full physical interpretation of radiosity eigenvalues and eigenvectors remains an open problem. The investigation presented in this appendix, however, provides an experimental substrate for further developments in this area, which, as stated by Ashdown [10, 11], remains largely unexplored in computer graphics despite its significant potential.

Appendix D
Finding an Optimal Scaling Parameter

As presented in Appendix C, the eigenvector corresponding to the largest eigenvalue of the symmetric radiosity matrix G^s (Section C.5) may provide a good initial approximation for the solution of the radiosity system $GB = E$ when scaled by a certain value. This observation motivated us to search for an optimal scaling parameter Γ for this eigenvector. In order to achieve this purpose, we present the following derivation.

Since we are dealing with an eigenvector of G^s, which is given by $G^s = QG$ (Section 7.21), we shall consider the system of the form:

$$G^s B = E^s \tag{D.1}$$

where E^s represents the scaled vector of irradiances given by $E^s = QE$. We shall also suppose that an eigenvector ν associated with an eigenvalue Λ is chosen as a solution to the system represented by Equation D.1. Then, we have $G^s\nu = \Lambda\nu$ and, in general, $\Lambda\nu \neq E^s$.

Assume $\Gamma\nu$ is chosen as a solution instead of ν such that Γ is a parameter to be computed. In this case, what would be the optimal choice for Γ so that $\Gamma\Lambda\nu$ is as close to E^s as possible?

For simplicity, we consider the squared Euclidean norm:

$$\varphi(\Gamma) = \| \Gamma\Lambda\nu - E^s \|^2 = \sum_{i=1}^{n_p}(\Gamma\Lambda\nu_i - E_i^s)^2 = \sum_{i=1}^{n_p}(\Gamma^2\Lambda^2{\nu_i}^2) - \sum_{i=1}^{n_p}(2\Gamma\Lambda\nu_i E_i^s + E_i^{s2}) \tag{D.2}$$

where n_p corresponds to the total number of patches.

Minimizing

$$\frac{d}{d\Gamma}\varphi(\Gamma) = \sum_{i=1}^{n_p}(2\Gamma\Lambda^2{\nu_i}^2) - \sum_{i=1}^{n_p}(2\Lambda\nu_i E_i^s) = 0 \tag{D.3}$$

gives:

$$\Gamma = \frac{\sum_{i=1}^{n_p}(\nu_i E_i^s)}{\Lambda \sum_{i=1}(n_p(\nu_i{}^2)} \tag{D.4}$$

Recall that Q corresponds to a diagonal matrix whose diagonal entries Q_{ii} represent the quotient between the area, A_i, and the reflectance, ρ_i, of a patch i. Then, considering that we can replace ν in Equation D.3 by the eigenvector ν_L associated with the largest eigenvalue Λ_L of G^s, we can rewrite the expression for Γ as:

$$\Gamma = \frac{\sum_{i=1}^{n_p}(\nu_{Li}\frac{A_i\ E_i}{\rho_i})}{\Lambda_L \sum_{i=1}^{n_p}(\nu_{Li}{}^2)} \tag{D.5}$$

As we mentioned in the previous Appendix, we can use the Lanczos algorithm [64, 92] to compute only ν_L and Λ_L. Nevertheless, further experiments are required to determine if the gains of performance obtained through the use of $\Gamma\nu_L$ as an initial approximation are sufficiently high to justify the additional cost of computing ν_L and Λ_L.

Index