

# Seeing is Believing: Reality Perception in Modeling, Rendering & Animation



**Alan Chalmers**  
Scott Daly  
**Ann McNamara**  
Karol Myszkowski  
Holly Rushmeier  
Tom Troscianko

Course #21



# Abstract

---

Advances in image synthesis techniques allow us to simulate the distribution of light energy in a scene with great precision. Unfortunately, this does not ensure that the displayed image will have a high fidelity visual appearance. Reasons for this include the limited dynamic range of displays, any residual shortcomings of the rendering process, and the extent to which human vision encodes such departures from perfect physical realism. Conversely, along many parameters, the visual system has strong limitations, and ignoring these leads to an over specification of accuracy beyond what can be seen on a given display system. This gives rise to unnecessary computational expense.

It is increasingly important to provide quantitative data on the fidelity of rendered images. This can be done either by developing computational metrics which aim to predict the degree of fidelity, or to carry out psychophysical investigations into the degree of similarity between the original and rendered images. This course addresses techniques to compare real and synthetic images, identify important visual system characteristics and thus produce benefits to the graphics community such as being able to reduce rendering times significantly.

Case studies involving both static and dynamic images are considered. Their different perception metric requirements are compared and contrasted.

This course address the problems associated with creating and evaluating realism in static and dynamic images. It covers the fundamentals of perception metric design for evaluation by both humans and computational models and discusses the issues of fidelity, relevant psychophysics, visual perception, shape and state-of-the-art metrics. The use of the techniques are illustrated by applying them to case studies investigating image quality between synthetic images and between synthetic images and the real world.

# Lecturers

---



**ALAN CHALMERS** is a Senior Lecturer in the Department of Computer Science at the University of Bristol, UK. He has published over 70 papers in journals and international conferences on realistic graphics. He is currently vice chair of ACM SIGGRAPH. His research is investigating the use of very realistic graphics in the accurate visualisation of archaeological site reconstructions and techniques which may be used to reduced computation times without affecting the perceptual quality of the images.



**SCOTT DALY** has degrees in electrical engineering and bioengineering, with a thesis directed toward retinal neurophysiology. He has worked for RCA, Photo Electronics Corporation, and Eastman Kodak in digital video, laser scanning, image compression, image fidelity models, and data image embedding. He shares an Emmy with Kodak colleagues for a video transceiver used in the 1989 Tianamen Square news coverage. Conference activities include papers, tutorials, and chairing sessions for SPIE, SPSE and RIDT. Currently at Sharp Laboratories of America, he is now applying visual models towards improving digital video and displays. He has 12 patents ranging from tonescales to steganography.



**ANN MCNAMARA** is a lecturer in the Department of Computer Science, Trinity College Dublin where she is a member of the Image Synthesis Group. She received her BSc and PhD degrees from the University of Bristol, in 1996 and 2000 respectively. Her research focuses on the use of psychophysical techniques for the evaluation of virtual environments with respect to the real world scenes they are intended to depict. She has given several (invited) talks on this work. She has also published numerous papers which detail her research.



**KAROL MYSZKOWSKI** is a senior visiting researcher at Max-Planck-Institute for Computer Science, Germany. Since 1993 until recently he served as an Associate Professor in the Department of Computer Software at the University of Aizu, Japan. Before he joined academia he worked for computer graphics industry developing global illumination and rendering software. His current research is investigating the role of human perception to improving the performance of photo-realistic rendering and animation techniques



**HOLLY RUSHMEIER** received the BS, MS and PhD degrees in mechanical engineering from Cornell University in 1977, 1986 and 1988 respectively. She is a research staff member at the IBM T.J. Watson Research Center. Since receiving the PhD, she has held positions at the Georgia Institute of Technology and the National Institute of Standards and Technology. In 1990 she was selected as a US National Science Foundation Presidential Young Investigator. In 1996 she served as the papers chair for the ACM SIGGRAPH conference, in 1998 as the papers co-chair for the IEEE Visualization conference and in 2000 as the co-chair for the Eurographics Rendering Workshop. From 1997 to 1999 she was Editor-in-Chief of ACM Transactions on Graphics. Her research interests include data visualization, rendering algorithms, and acquisition of input data for computer graphics image synthesis.



**TOM TROSCIANKO** has a wide background in vision science, and his current research is investigating the way in which human vision samples the natural visual environment. He has carried out psychophysical investigations on vision in order to develop computational models of how complex environments are represented and decisions are made. He worked at the University of Bristol, as well as IBM, and is currently Professor of Psychology in the School of Cognitive and Computing Sciences at the University of Sussex.

# Contents

---

## **I. Introduction to the Perception of Reality**

- some intuitive examples of applications
- the role of perception
- our focus: synthetic images generated using computer graphic methods

## **II. Psychophysical Image Quality Metrics**

- psychophysics: theory and practice for comparing natural and rendered images
- fidelity assesment
- procedures for comparing real and synthetic images
- case studies

## **III. Important Issues for Automating Image Quality Estimation**

- visual perception
- computer models of visual system

## **IV. Computational Image Quality Metrics**

- state-of-the-art metrics
  - VDP
  - Sarnoff model
  - Animation Quality Metric
- validation of metrics through experiments with subjects
- customising metrics for specific tasks

## **V. Metrics and geometric simplification**

- use of image metrics vs. geometric metrics to guide simplification
- ratings of object quality for different simplification levels
- correlation of image quality metrics with
  - naming times
  - other object quality metrics

## **VI. Applications in Rendering and Animation**

- explicit use: controlling image computation
- implicit use: improving rendering efficiency
- animation and dynamic case studies

# Seeing is Believing: Reality Perception in Modeling, Rendering & Animation



## Introduction

---

1

The aim of realistic image synthesis is the creation of accurate, high quality imagery which faithfully represents a physical environment, the ultimate goal being to create images which are perceptually indistinguishable from an actual scene. Advances in image synthesis techniques allow us to simulate the distribution of light energy in a scene with great precision. Unfortunately, this does not ensure that the displayed image will have a high fidelity visual appearance. Reasons for this include the limited dynamic range of displays, any residual shortcomings of the rendering process, and the extent to which human vision encodes such departures from perfect physical realism. Image quality metrics are paramount to provide quantitative data on the fidelity of rendered images. Typically the quality of an image synthesis method is evaluated using numerical techniques which attempt to quantify fidelity using image to image comparisons (often comparisons are made with a photograph of the scene that the image is intended to depict). Several image quality metrics have been developed whose goals are to predict the *visible* differences between a pair of images. It is well established that simple approaches, such as mean squared error (MSE), do not provide meaningful measures of image fidelity, more sophisticated techniques are necessary. As image quality assessments should correspond to assessments made by humans, a better understanding of features of the **H**uman **V**isual **S**ystem (HVS) should lead to more effective comparisons,

which in turn will steer image synthesis algorithms to produce more realistic, reliable images. Any feature of an image not visible to a human is not worth computing. Results from psychophysical experiments can reveal limitations of the HVS. However, problems arise when trying to incorporate such results into computer graphics algorithms. This is due to the fact that, often, experiments are designed to explore a single dimension of the HVS at a time. The HVS comprises many complex mechanisms, which rather than function independently, often work on conjunction with each other, making it more sensible to examine the HVS as a whole. Rather than attempting to reuse results from previous psychophysical experiments, new experiments are needed which examine the complex response HVS as a *whole* rather than trying to isolate features for individual investigations. This course addresses techniques to compare real and synthetic images, identify important visual system characteristics and help reduce rendering times significantly. The following topics are covered: fidelity of images; human visual perception including important characteristics of the human visual system; computational models of perception including spatial and orientation channels and visual masking; objective metrics including Visual Difference Predictors, the Sarnoff model and Animation Quality Metrics; and psychophysics.

## I.1 Course Syllabus

Introduction to Image Quality (Chalmers 10 mins)

- some intuitive examples of applications:
- the role of perception
- subjective and objective methods of image quality estimation
- our focus: synthetic images generated using computer graphics methods

Psychophysical Image Quality Metrics (McNamara & Troscianko 50 minutes)

- psychophysics
- fidelity of final image
- working with real subjects

- procedures for comparing real and synthetic images
- case studies

Important Issues for Automating Image Quality Estimation (Daly & Troscianko 40 minutes)

- visual perception
- computer models of visual system

Computational Image Quality Metrics (Daly 45 minutes)

- state-of-the-art metrics
- VDP
- Sarnoff model
- Animation Quality Metric
- validation of metrics through experiments with subjects
- customising metrics for specific tasks

Metrics and Geometric Simplification (Rushmeier 20 minutes)

- use of image metrics vs. geometric metrics to guide simplification
- ratings of object quality for different simplification levels
- correlation of image quality metrics with
  - naming times
  - other object quality metrics

Applications in Rendering and Animation (Myszkowski 35 minutes)

- explicit use: controlling image computation
- implicit use: improving rendering efficiency
- animation and dynamic case studies

Summary, discussion and questions (All 20 mins)

# Seeing is Believing: Reality Perception in Modeling, Rendering & Animation



## I. Introduction

**Alan Chalmers**

Scott Daly

Ann McNamara

Karol Myszkowski

Holly Rushmeier

Tom Troscianko

Course #21





## Seeing is Believing: Reality perception in modeling, rendering and animation"

Alan Chalmers  
Scott Daly  
Ann McNamara  
Karol Myszkowski  
Tom Troscianko  
Holly Rushmeier




---

---

---

---

---

---

---

---

## Evaluating Quality in Images

- ✍ Course Syllabus
  - ✍ Introduction to Image Quality
  - ✍ Psychophysical Image Quality Metrics
  - ✍ Important Issue for Automating Image Quality Estimation
  - ✍ Computational Image Quality Metrics
  - ✍ Metrics for Geometric Simplification
  - ✍ Application in Rendering and Animation
  - ✍ Summary, Discussion and Questions




---

---

---

---

---

---

---

---

## Introduction to Image Quality

- ✍ Why do we need high fidelity images?
- ✍ How do I know this image is real?







---

---

---

---

---

---

---

---

### No absolute truths

- ✍ For sites which no longer exist or have yet been built
- ✍ User must have confidence in the image




---

---

---

---

---

---

---

---

### Photo-Realism versus High Fidelity

- ✍ What is reality?
- ✍ Image synthesis techniques allow us to simulate accurately light distribution within a scene ?
- ✍ does NOT imply high fidelity visual appearance




---

---

---

---

---

---

---

---

### Problems

- ✍ Lack of high fidelity due to:
  - ✍ problems with modelling the scene
  - ✍ residual shortcomings of the rendering process
  - ✍ limited dynamic range of displays
  - ✍ extent to which human vision encodes such departures from perfect physical realism




---

---

---

---

---

---

---

---

### Image Quality Metrics

- Trying to provide quantitative data
  - real vs photograph
  - real vs synthetic
  - photograph vs synthetic



MSE:  
3297.343



---

---

---

---

---

---

---

---

### Importance

- Image quality metrics can:
  - provide user confidence in images
  - open substantial application opportunities
  - significantly reduce rendering times



*Realism in Real-Time*



---

---

---

---

---

---

---

---

# Seeing is Believing: Reality Perception in Modeling, Rendering & Animation



## II. Psychophysical Image Quality Metrics

Alan Chalmers  
Scott Daly  
**Ann McNamara**  
Karol Myszkowski  
Holly Rushmeier  
Tom Troscianko

Course #21



**Psychophysical Image Quality Metrics**



Ann McNamara  
Image Synthesis Group  
Trinity College Dublin  
Ireland



---

---

---

---

---

---

---

---

**Subjective Image Quality - Overview**

- Realism
- Psychophysics
- Working with Participants
- Procedures for comparing Real Scenes with Synthetic Image
- Case Studies



---

---

---

---

---

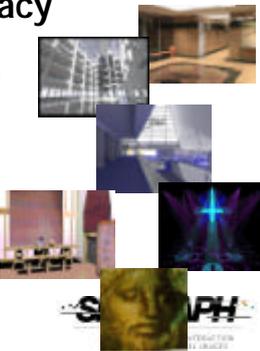
---

---

---

**Need For Accuracy**

- Lighting Engineering
  - Architecture
  - Stage Lighting
- Industry
  - Entertainment
  - Safety Critical
- Archaeology



---

---

---

---

---

---

---

---

### Why Compare Images ?

- ⌘ Compare and validate lighting simulations
- ⌘ Use comparisons to guide rendering more efficiently
  - ⌘ Can we compute less without altering human perception of an image
  - ⌘ While pixel by pixel comparison might be  $> 0$ , human might not see any difference



---

---

---

---

---

---

---

---

### Pixel by Pixel Comparison



---

---

---

---

---

---

---

---

### Visual Psychophysics

- ⌘ Determine the relationship between the physical world and human's subjective experience of that world



---

---

---

---

---

---

---

---

### Visual Psychophysics

- Measure the mind
  - Without bias
  - Systematically
  - Repeat Observations
  - Relationship between mind & matter



---

---

---

---

---

---

---

---

### Experimental Design

- Make inferences without ambiguity
- Rule out alternative causes, leaving only the actual factor that is the real cause



---

---

---

---

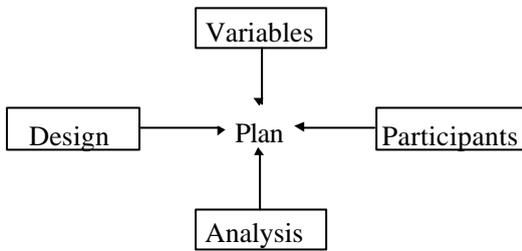
---

---

---

---

### Experimental Design



---

---

---

---

---

---

---

---

### Experimental Design

- ✍ Selecting what to study
- ✍ Selecting who to study
- ✍ Specifying how to study
- ✍ Specifying the sequence of measurements to be recorded
- ✍ What kind of evidence will result



---

---

---

---

---

---

---

---

### Order Effects

- ✍ Order of Presentation
- ✍ Good before Bad
- ✍ Timing



---

---

---

---

---

---

---

---

### Randomisation

- ✍ Not haphazard
- ✍ No event is ever predictable from any of the preceding sequence



---

---

---

---

---

---

---

---

### Counterbalancing

- ✍ Order of Presentation
- ✍ Influence Results
- ✍ Make Experiment "fair"
- ✍ Reduce Bias



---

---

---

---

---

---

---

---

### Control of Extraneous Variables

- ✍ May influence or affect the results of the condition
- ✍ For Example - Outside illumination



---

---

---

---

---

---

---

---

### Experimental Design

- ✍ The Question to be Answered
- ✍ Choice of Task (Measure)
- ✍ Choice & Control of Physical Stimuli
- ✍ Organisation of Participants
- ✍ Sequence of Presentation
- ✍ Instructions to Participants
- ✍ Recording, Presenting & Analysing Results



---

---

---

---

---

---

---

---

### Psychophysics for Judging Image Quality

- Psychophysical methods allow us to ask how close to reality computer images are
- Validate progressive global illumination solution



---

---

---

---

---

---

---

---

### Case Study 1

Comparing a real scene with computer generated images that represent that scene



---

---

---

---

---

---

---

---

### A method of comparing real scenes with graphics

- Task: Estimate the lightness of various regions of a scene
- Lightness perception is known to depend on prior perception of 3-D shape and illumination



---

---

---

---

---

---

---

---

### A method of comparing real scenes with graphics

Lightness is therefore a useful measure of the fidelity of illumination and 3-D reconstruction of a graphics scene



---

---

---

---

---

---

---

### Why Lightness ?



---

---

---

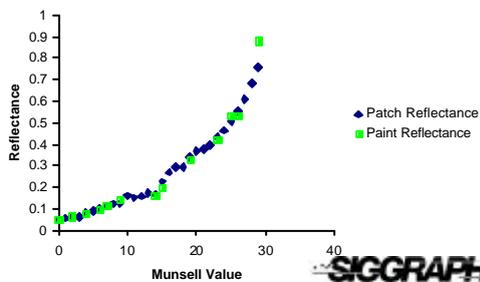
---

---

---

---

### Physical Stimulii



---

---

---

---

---

---

---

### Physical Stimuli: The Real Scene



- ✍ Painted 5-sided Cube
- ✍ Objects painted with different grey paints
- ✍ Complex illumination, with secondary reflections



---

---

---

---

---

---

---

---

### Physical Stimuli: Graphic Reconstructions



---

---

---

---

---

---

---

---

### Participants and Sequencing

- ✍ How Many Participants
- ✍ Randomisation of Participants
- ✍ Randomisation of presentation
- ✍ Time of Day Influence
- ✍ Training on Physical Stimuli



---

---

---

---

---

---

---

---

### Physical Stimuli: Training on Patches



---

---

---

---

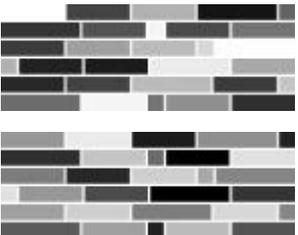
---

---

---

---

### Physical Stimuli: Training on Patches



---

---

---

---

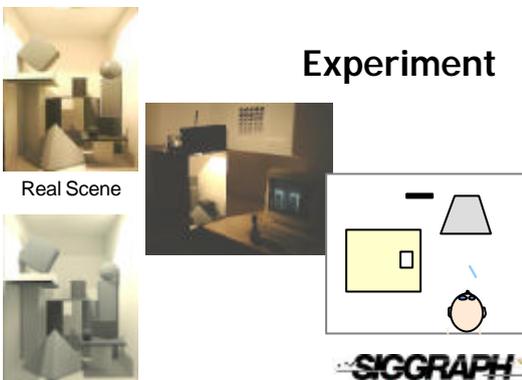
---

---

---

---

### Experiment



Real Scene

Rendered



---

---

---

---

---

---

---

---

### Case Study : Results

✍ Average match in each image is plotted along side average match in the real scene




---

---

---

---

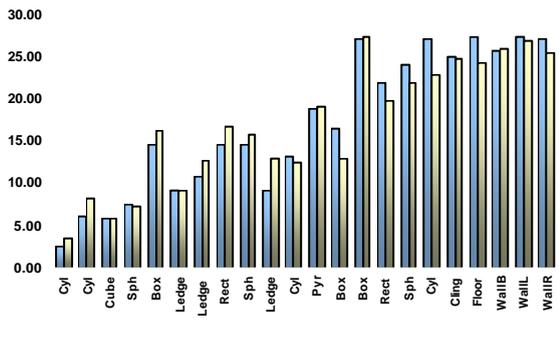
---

---

---

---

### Real & Photograph




---

---

---

---

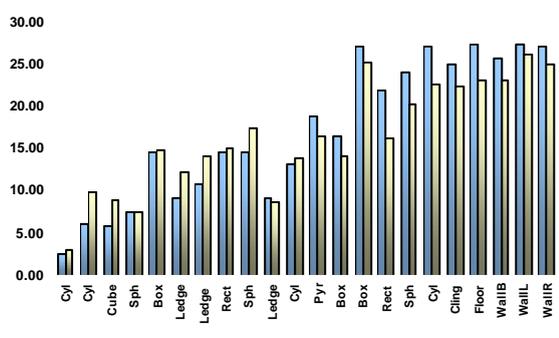
---

---

---

---

### Real & Two Ambient Bounce




---

---

---

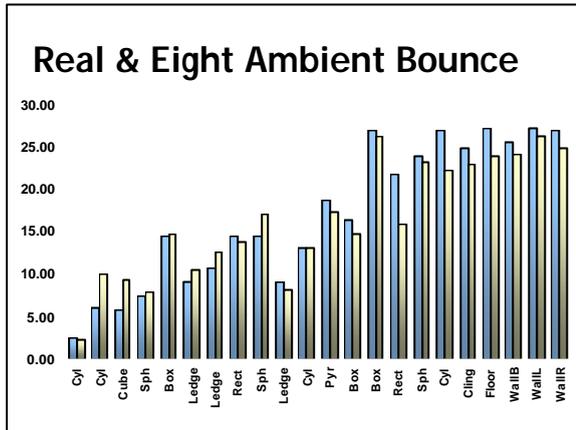
---

---

---

---

---




---

---

---

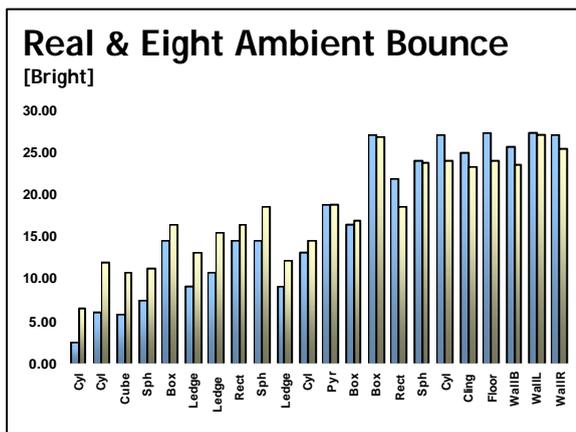
---

---

---

---

---




---

---

---

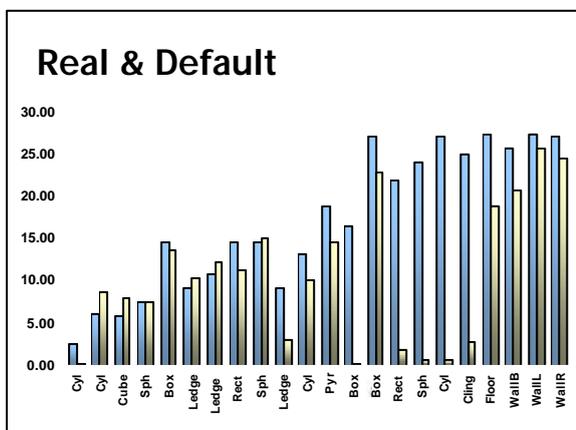
---

---

---

---

---




---

---

---

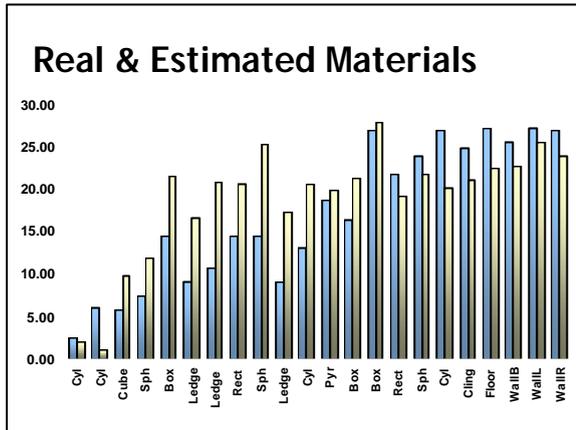
---

---

---

---

---




---

---

---

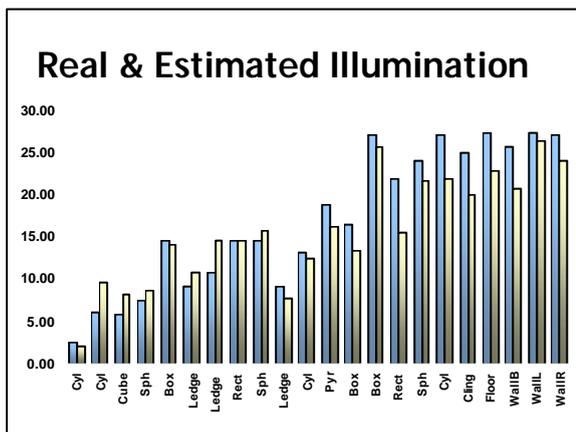
---

---

---

---

---




---

---

---

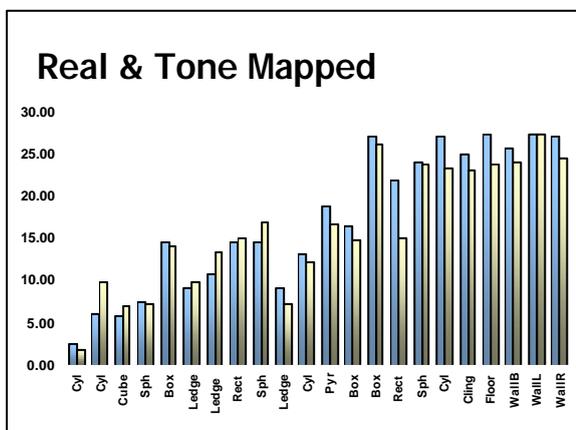
---

---

---

---

---




---

---

---

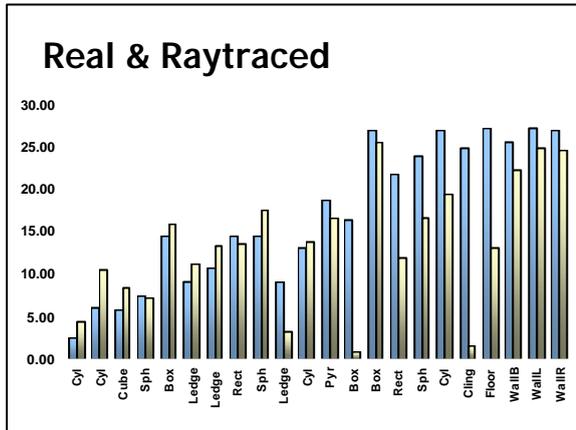
---

---

---

---

---




---

---

---

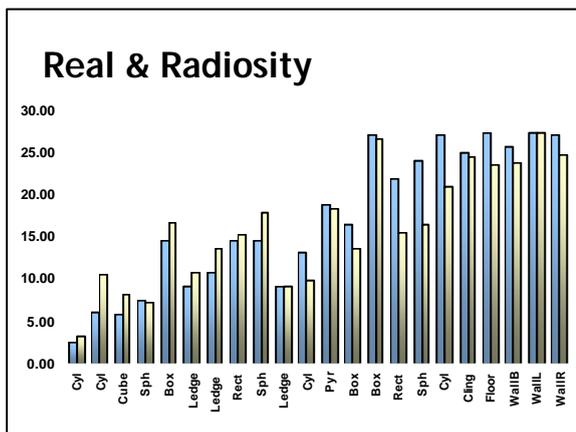
---

---

---

---

---




---

---

---

---

---

---

---

---

### Data Analysis

- ✍ Correlation - indication of how closely related two sets of data are
- ✍ ANOVA - Analysis Of Variance
- ✍ T Tests
- ✍ Statistics reinforce the evidence from the data




---

---

---

---

---

---

---

---

## Case Study 2

Validation of progressive  
global illumination solution



---

---

---

---

---

---

---

## The atrium model



Basic Model:  
751639 Polygons



Model used for global illumination:  
951524 polygons



---

---

---

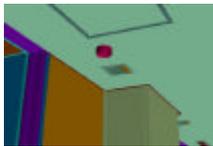
---

---

---

---

## Level of detail



---

---

---

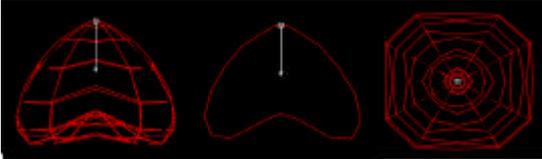
---

---

---

---

### Luminaires



Goniometric diagram of the main light source

---

---

---

---

---

---

---

---

### Selection of materials for BRDF measurements



---

---

---

---

---

---

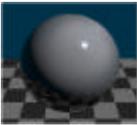
---

---

### Preview of the "blue-tile" material

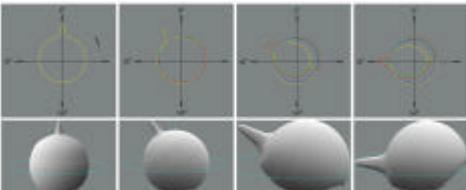


Photograph



Sampled Preview

BRDF characteristics



---

---

---

---

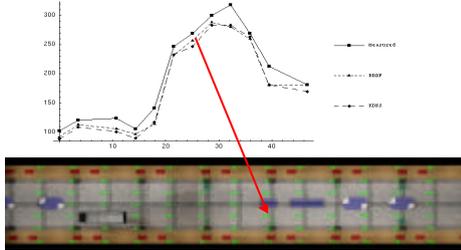
---

---

---

---

### Measurement of illumination



Measurement points sampled following a grid pattern on the main floor of the atrium.

---

---

---

---

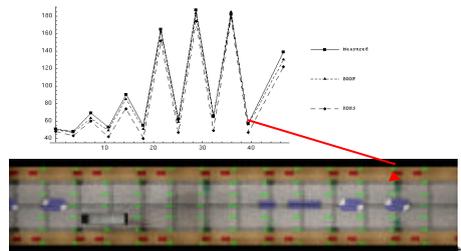
---

---

---

---

### Measurement of illumination



Measurement points sampled following a grid pattern on the main floor of the atrium

---

---

---

---

---

---

---

---

### Photography of the atrium



Daylight film



Tungsten film



Digital camera




---

---

---

---

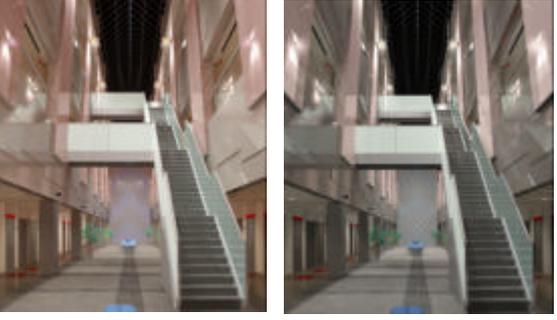
---

---

---

---

### Renderings of the atrium



---

---

---

---

---

---

---

---

### Psychophysical experiment: real world versus computer images and photographs



---

---

---

---

---

---

---

---

### Results

- ≈ Photograph only slightly better in representing real world than graphics
- ≈ 16 subjects preferred images obtained using the "artistic" approach while 9 subjects selected the BRDF based images



---

---

---

---

---

---

---

---

## Results

☞ Detailed statistics concerning quality of shadow reconstruction, highlights and reflections rendering, contrast reproduction, luminaire appearance, texture appearance are under preparation.



---

---

---

---

---

---

---

---

## Summary

☞ Experiments should be designed to produce accurate results

☞ Attention must be paid to a number of subtle experimental issues such as sample size, bias and randomisation



---

---

---

---

---

---

---

---



---

---

---

---

---

---

---

---

# Seeing is Believing: Reality Perception in Modeling, Rendering & Animation



## III. Important Issues for Automating Image Quality Estimation

Alan Chalmers  
Scott Daly  
Ann McNamara  
Karol Myszkowski  
Holly Rushmeier  
**Tom Troscianko**

Course #21



## Important Issues for Automating Image Quality Estimation

Tom Troscianko  
School of Cognitive & Computing Sciences  
University of Sussex  
United Kingdom



---

---

---

---

---

---

---

---

## Collaborators

Ann McNamara  
Alan Chalmers  
David Tolhurst  
Alejandro Párraga



---

---

---

---

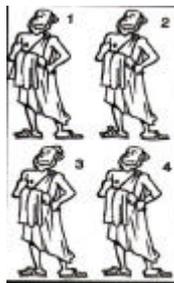
---

---

---

---

## Comparison of images



---

---

---

---

---

---

---

---

### General Question

Two main techniques are available to quantify differences between images:

**Computational**

**Psychophysical**

When might these approaches be useful?



---

---

---

---

---

---

---

---

### Computational approach

- ✍ We began by asking whether human vision is optimised to the statistics of natural scenes
- ✍ We developed a model which predicts the visibility of differences between natural scenes
- ✍ This model was validated in two very different ways



---

---

---

---

---

---

---

---

### How can we define a "natural" scene



Much of the diversity of visual images is described by a very simple statistical relationship:

$$\text{Amplitude}(f) \propto f^{-2}$$

Where *Amplitude* denotes amplitude spectra averaged across all orientations, *f* is spatial frequency and  $\propto$  (also called the *slope parameter*) lies within a fairly narrow range (0.7 - 2.0).



---

---

---

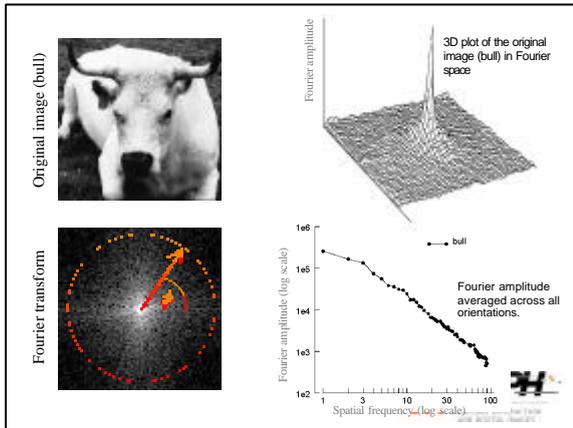
---

---

---

---

---




---

---

---

---

---

---

---

---

### Experimental method?

- ✍ We need a technique which measures visual performance
- ✍ Task consistent with real-life vision
- ✍ Using naturalistic stimuli
- ✍ Shape discrimination!




---

---

---

---

---

---

---

---




---

---

---

---

---

---

---

---

Morph sequence:  
"Car becomes bull"



---

---

---

---

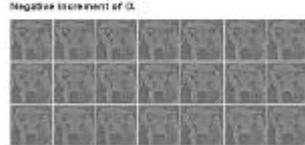
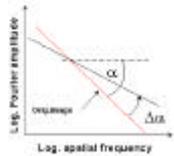
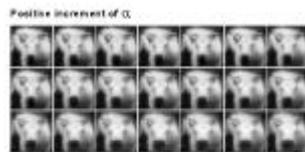
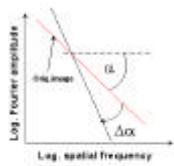
---

---

---

---

Procedure: change slope value of ?



---

---

---

---

---

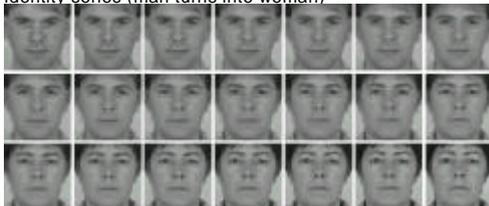
---

---

---

Another set of morphed stimuli

Identity series (man turns into woman)



The IDENTITY morph series, courtesy of Phil Benson



---

---

---

---

---

---

---

---

### A Basic Experiment:

You will see three images:

Image 1

Reference

Image 2

Is Image 1 or Image 2 different from the Reference?



Example of single experimental trial



---

---

---

---

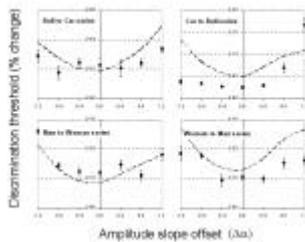
---

---

---

---

### Experimental results and predictions from a model of local contrast discrimination



---

---

---

---

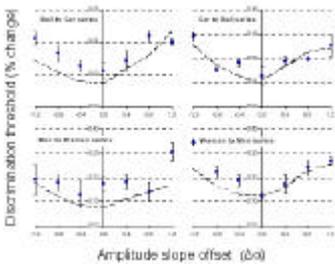
---

---

---

---

### Further results and predictions...



---

---

---

---

---

---

---

---



---

---

---

---

---

---

---

---

### What is that dotted line?

- ✍ A model based on properties of simple cortical neurons
- ✍ These neurons discriminate contrast in different spatial bands
- ✍ No free parameters - model assumes individual's contrast sensitivity function and contrast discrimination function



---

---

---

---

---

---

---

---

### Quality of computer graphics?

- ✍ Our model gives us a metric for comparing images
- ✍ So we can determine differences between a test and reference image
- ✍ This allows us to see how much like a "real" scene a graphics scene may be, and to compare scenes with and without a target



---

---

---

---

---

---

---

---

How is the "difference number" calculated?



---

---

---

---

---

---

---

Synthetic image with vehicle      and without the vehicle

3 cycles

8 cycles

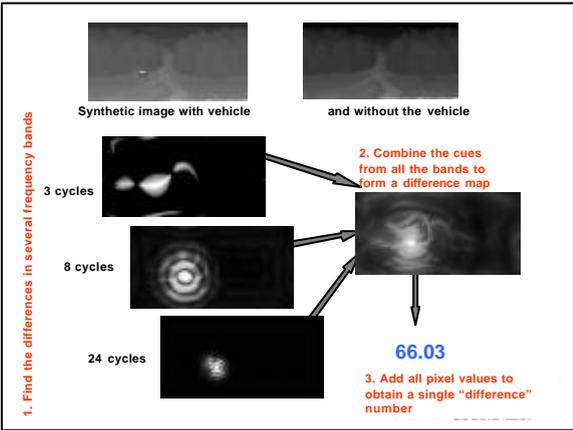
24 cycles

1. Find the differences in several frequency bands

2. Combine the cues from all the bands to form a difference map

3. Add all pixel values to obtain a single "difference" number

66.03



---

---

---

---

---

---

---

What Happens With Real Imagery?

- ✗ The last slide showed synthetic images
- ✗ Here we show some **real** images

Real      Synthetic

- ✗ More complex difference map from real images
- ✗ due to uncontrolled lighting and atmospheric conditions



---

---

---

---

---

---

---

### Validation of model?

- Can the computational model predict image differences which are above threshold?
- How can this be investigated?
- Magnitude estimation???



---

---

---

---

---

---

---

---

### Magnitude estimation



"Let this difference be 50 units"



---

---

---

---

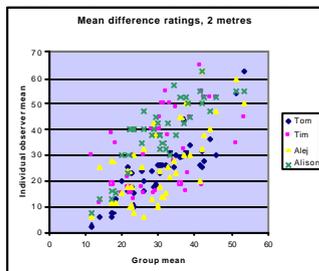
---

---

---

---

### Results for 4 observers



---

---

---

---

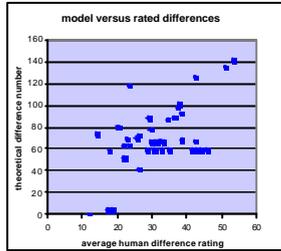
---

---

---

---

### Agreement with model?



---

---

---

---

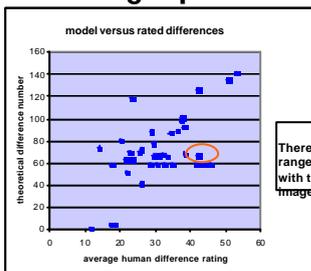
---

---

---

---

### Outliers on graph?



---

---

---

---

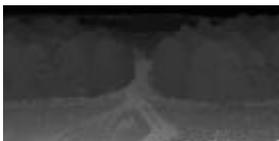
---

---

---

---

### Model can be "fooled" by lightness differences



---

---

---

---

---

---

---

---

### Results of scaling experiment

- ✍ Moderately good agreement between model and ratings
- ✍ Correlation about 0.6
- ✍ Image display software tends to include lightness artefacts
- ✍ Correlation rises to 0.7 if problematic images excluded
- ✍ The computational approach has problems with lightness in pictures



---

---

---

---

---

---

---

---

### Quality of computer graphics?

- ✍ Our model gives us a metric for comparing images
- ✍ So we can determine differences between a test and reference image
- ✍ This allows us to see how much like a "real" scene a graphics scene may be, and to compare scenes with and without a target



---

---

---

---

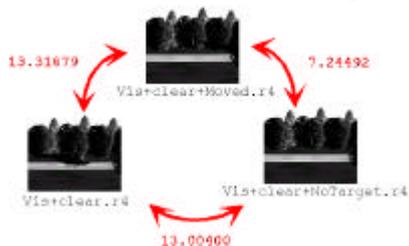
---

---

---

---

### Predicted perceived difference between images



---

---

---

---

---

---

---

---

### Another method of comparing real scenes with graphics

- ✍ Estimation of lightness of parts of a scene
- ✍ Lightness perception is known to depend on prior perception of 3-D shape and illumination
- ✍ Lightness is therefore a useful measure of the fidelity of illumination and 3-D reconstruction of a graphics scene



---

---

---

---

---

---

---

---

### The original scene



- ✍ Painted interior of tea-chest
- ✍ Variety of objects painted with different grey paints
- ✍ Complex illumination, with secondary reflections



---

---

---

---

---

---

---

---

### High quality graphics reconstruction



“Radiance” algorithm  
Allows for secondary reflections



---

---

---

---

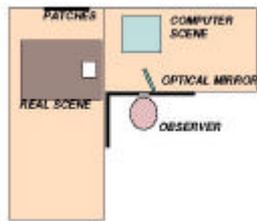
---

---

---

---

### Psychophysical comparison of scenes



---

---

---

---

---

---

---

---

### Why Lightness



---

---

---

---

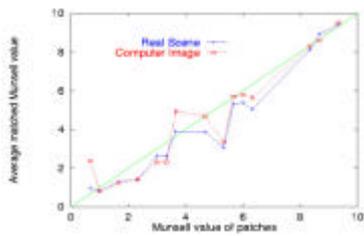
---

---

---

---

### Lightness estimation



---

---

---

---

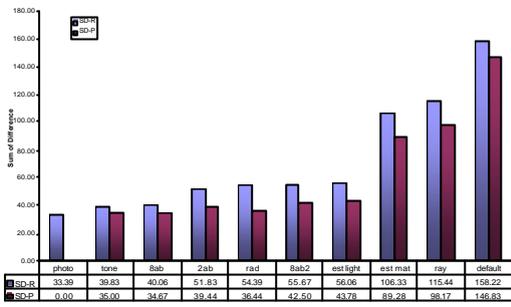
---

---

---

---

Can this method discriminate between graphics of different quality?




---

---

---

---

---

---

---

---

---

---

Can this method distinguish sharp and blurred images?



SIGGRAPH 2001

---

---

---

---

---

---

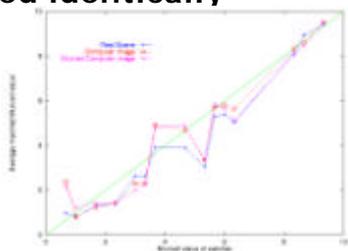
---

---

---

---

Sharp and blurred scenes are judged identically



SIGGRAPH 2001

---

---

---

---

---

---

---

---

---

---

**General conclusions**

- ✍ Computational methods can predict image differences but are susceptible to error where human vision uses “high level” mechanisms
- ✍ Lightness estimation taps into such mechanisms
- ✍ Psychophysics of lightness therefore complements other approaches



---

---

---

---

---

---

---

---

# Seeing is Believing: Reality Perception in Modeling, Rendering & Animation



## IV. Computational Image Quality Metrics

Alan Chalmers

**Scott Daly**

Ann McNamara

Karol Myszkowski

Holly Rushmeier

Tom Troscianko

Course #21



**Current State and Assessments of Applied Visual Models**

Scott Daly  
Center for Displayed Appearance  
Information Systems Technology  
Sharp Laboratories of America




---

---

---

---

---

---

---

---

**Outline**

- ✂ Overview of visual model design and approaches
- ✂ Basic Spatio-temporal properties of detection by the Visual System
- ✂ State-of-the-art visual distortion metrics:
  - Spatial and Chromatic:
    - VDP (Daly)
    - Sarnoff (Lubin and Brill)
    - Efficiency Versions
  - Spatiotemporal (Motion)
    - Animation Quality Metric
- ✂ Validation of metrics:
  - Modelling published psychophysical data
  - Testing with system-based test targets
  - Testing in actual applications




---

---

---

---

---

---

---

---

**Visual Model Design and Approaches**

Visual Modeling uses published work from the following fields of basic research:

- ✂ Anatomical
  - Optics of eye
  - Sampling structure of retina
  - Cellular interconnections of visual pathway
- ✂ Physiological
  - Functional behavior of individual cells
  - Functional behavior of regions in
  - Data from electrophysiology experiments (measurements of electrical responses of neurons)
  - Retina is analog up to ganglion cells
  - For remaining visual pathway, information is conveyed with neural spikes (i.e., digital, like PCM)
- ✂ Psychophysical
  - Experiments using human observer responses
  - Used to test theories based on physiology and anatomy
  - Signal detection theory and signal processing used to model psychophysical results
  - Threshold (can or cannot see signal) vs. Suprathreshold (rank magnitude of signal)
  - Empirical results (without theory) also useful for visual optimization of graphics effects




---

---

---

---

---

---

---

---

### Types of Visual Models

Mathematical, quantitative descriptions of visual response w/ parameters

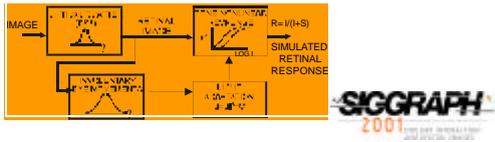
Historical Examples:

- CIELAB standard lightness response (1976):  
 $L^* = 116(Y/Y_0)^{1/3} - 16$ ,  $Y$  is luminance,  $Y_0$  is luminance of white point
- Contrast Sensitivity Function (CSF) = spatial frequency response (Mannos & Sakrison '74)  
 $CSF(u,v) = 2.6 \cdot (0.0192 + 0.144 \cdot r^{1/2}) \cdot \exp(-0.144 \cdot r^{1/2})$  ( $u, v = H$  and  $V$  freq.)  
 $r = (u^2 + v^2)^{1/2}$  (radial frequency)

Image processing models of visual thresholds and appearance (simulations)

Historical Example:

- Visual response in retina (Normann & Baxter '83)




---

---

---

---

---

---

---

---

### Ways to use visual models

Visual Analysis: of complete imaging systems or system components

- provide basic understanding, limitations, and opportunities
- typically extrema parameters of visual system are used

examples:

- maximum spatial frequencies that can be seen (cut-off frequency) to set needed resolution
- maximum temporal frequencies for setting frame update rates
- minimum gray level changes for setting bit-depth
- minimum noise levels

Visual Optimization: used to improve existing designs

- use visual models of key aspects relevant to application like frequency response, luminance response...
- image capture systems: Color Filter Array (CFA) algorithms, field-sequential approaches...
- image processing algorithms: compression, enhancement, watermarking, halftoning,....
- display design: new triad patterns, subtriad addressing, ....

Visual Metrics: used to compare visual effects on actual images (vs. test patterns)

- Image Fidelity: whether any distortions are visible compared to a benchmark system may vary locally throughout image to help engineers improve system
- Image Quality: a graded scale, may not need benchmark so it can be absolute assessment




---

---

---

---

---

---

---

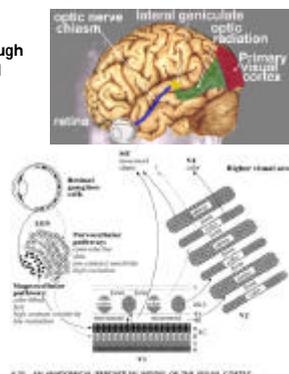
---

### Properties of the Visual System

This talk will proceed through key properties of the Visual System

Properties dissected along these dimensions:

- Luminance Level
- Spatial Frequency
- Local Spatial Content
- Temporal Frequency
- Motion
- Global Color
- Eccentricity




---

---

---

---

---

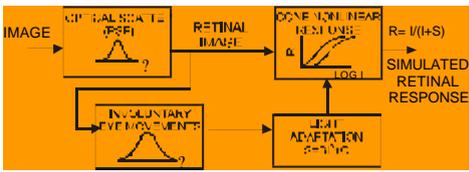
---

---

---

**Properties of Visual System: Luminance Nonlinearity**

- ✂ Luminance proportional to photon flux = "Linear"
- ✂ Pixel and surround effects
  - Photoreceptor and neighboring cells
  - Grey-level nonlinearity (instantaneous)
  - Light Adaptation



**SIGGRAPH 2001**

---

---

---

---

---

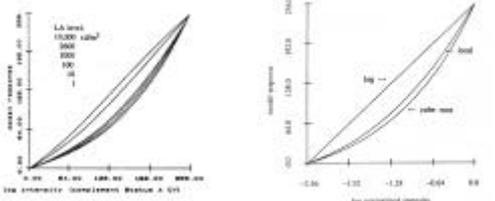
---

---

---

**Properties of Visual System: Luminance Nonlinearity**

- ✂ Local cone model (ignore PSF and eye movements)
  - Visual response in retina close to cube root ( $-L^*$ ) for practical video light levels
  - Cube-root domain is close to gamma-corrected domain ( $L^{1/3} \approx L^{1/2.4}$ )



**SIGGRAPH 2001**

---

---

---

---

---

---

---

---

**Properties of Visual System: Luminance Nonlinearity: Example**

- ✂ Use gamma-corrected domain to process images (or local cone,  $L^*$ , or cube-root)
  - For light levels in typical video range (50-200  $cd/m^2$ )
  - Technique works well for quantization, compression, watermarking



**2001**

---

---

---

---

---

---

---

---





**Properties of Visual System: Local Image Contrast: Masking**

Typical result from masking by noise

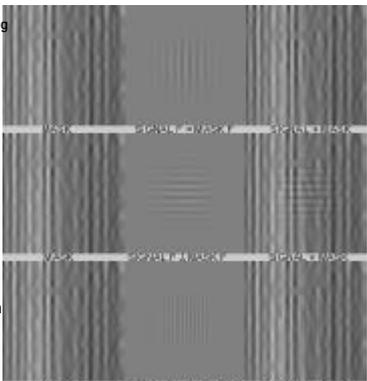
Noise is narrowband

- limited radial freq
- limited orientation

Little masking unless frequencies of mask are close to those of signal

- radially
- orientation

Effect also only occurs locally, with spatial extent depending on frequency




---

---

---

---

---

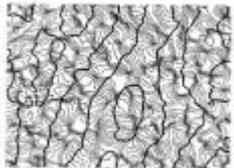
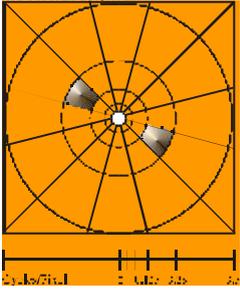
---

---

---

**Properties of Visual System: Masking: Frequency Structure**

Majority of psychophysical masking results consistent with visual system modeled as bank of filter-detectors as shown below to right:

Electrophysiology measures of common orientation response in visual cortex

- gray lines = common orientation
- black lines = orientation boundary

Key Features of frequency "channels": Dyadic radial frequencies, orientation selectivity, baseband, space frequency localization

---

---

---

---

---

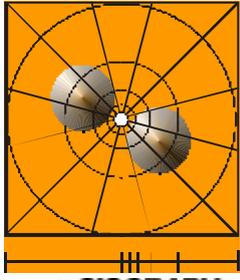
---

---

---

**Properties of Visual System: Masking: Frequency Structure**

Channels overlap, and have the Fourier symmetry of real signals



**SIGGRAPH**  
2001

---

---

---

---

---

---

---

---

### Properties of Visual System: Masking: Transducer Function

- Masking within "channel" = band of wavelet
- Psychophysical results shown to the right:
  - Dashed = Noise masking (phase incoherent)
  - Solid = Sine Masking (phase coherent)
- Results are -same for all frequencies once normalized by frequency's threshold = 1/CSF
- Results modeled as bandpass filter followed by nonlinear transducer function of contrast
- Transducer function,  $f(C)$ , derived from integral of inverse threshold data,  $T(C)$ :
  - Response:  $R \propto f(C)$
  - Threshold:  $T(C) \propto k / R \propto k / f(C)$
  - Transducer function:  $f(C) \propto 1/T(C)$
- Transducer functions derived from sine masking and noise masking data are shown to the right
  - Note that plot is now with linear axes

---

---

---

---

---

---

---

---

---

---

---

---

### Properties of Visual System: Temporal Frequency

- CSF for temporal frequencies also has been measured and modeled
- To right is shown temporal CSF for different light adaptation levels for luminance
  - Top curve is best for mid-bright display applications
- Opponent Color signals temporal CSF also has about 1/2 the bandwidth and sensitivity of the luminance

---

---

---

---

---

---

---

---

---

---

---

---

### Properties of Visual System: Spatiotemporal

- Motion occurs in area V5 of visual cortex
- Most psychophysical data measures spatio-temporal CSF
- Test signal is product of spatial and temporal frequency
- frequency
  - Standing Wave
- Data shows max visible temporal frequency near 50 cy/sec
  - Thus 60 fps usually causes no visible flicker
  - Movie film at 24 fps causes visible flicker, so projectors shutter each frame 2 or 3 times to increase fundamental temporal frequency

---

---

---

---

---

---

---

---

---

---

---

---



**Properties of Visual System: Global Color**

- Global color can be primarily handled through the field of color reproduction... many standards




---

---

---

---

---

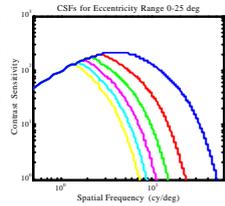
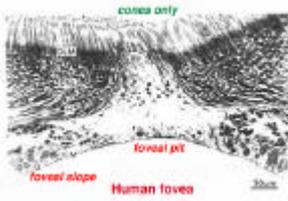
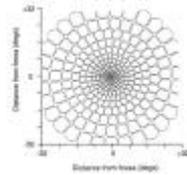
---

---

---

**Properties of Visual System: Eccentricity**

- Eccentricity : Position in visual field**
  - 0 degrees eccentricity refers to where your eyes are pointed, corresponds to fovea in retina
  - 90 degrees eccentricity is near edges of visual field (periphery)
- Spatial Bandwidth of eye reduces in periphery**
- Cones are densely packed in fovea : high spatial sampling -> high bandwidth**
- They become more less dense as eccentricity increases**




---

---

---

---

---

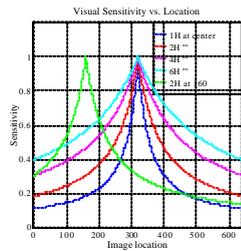
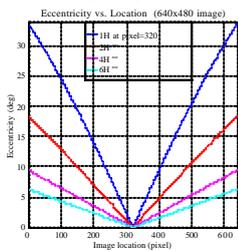
---

---

---

**Properties of Visual System: Eccentricity**

- How eccentricity changes across image as viewing distance changes (left)**
  - Assuming viewer looking at center of image (pixel = 320)
- Eccentricity model predictions of how visual sensitivity varies across image (right)**




---

---

---

---

---

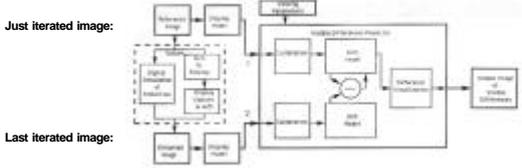
---

---

---

### State of the Art Metrics

Usage of visual difference (a.k.a discrimination) metrics:

Just iterated image: 

Last iterated image:

- Spatial and Chromatic**
  - Visible Differences Predictor (VDP)
  - Sarnoff model
  - Efficiency Versions
    - Bradley, Bolin, UCLA
- Spatiotemporal**
  - Sarnoff
  - Watson's DVQ
  - Animation Quality Metric




---

---

---

---

---

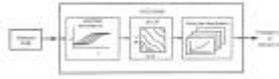
---

---

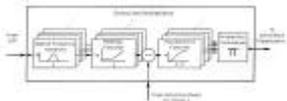
---

### Visible Differences Predictor (VDP)

Models basic visual effects of amplitude variations, spatial variations, and signal dependent variations sensitivity in three separate stages:



- Amplitude variations: modeled as local cone model described previously, entire image is processed through as point-nonlinearity
- Intended to be used with display model (problems with impossible conditions, e.g., if light level = 0)
- Spatial variations: the 2D CSF is modeled as a global filter, filter described previously
- Signal dependent variations (masking): described in block diagram below:





---

---

---

---

---

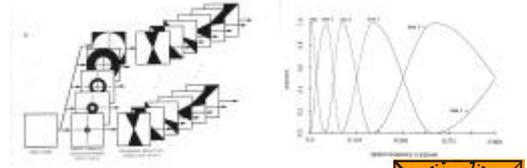
---

---

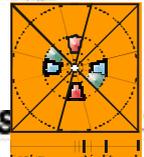
---

### VDP: Channel Design

- Spatial frequency hierarchy: filter bank of spatial frequency "channels" (mechanisms)
- Cascade of isotropic radial filters (DOM filters) with fan filters



- Radial filters shown above right, sum of all filters = 1.0 for reversibility in applications
- Resulting dissection of frequency plane:
- FFT's used to implement filtering, but subsequent steps occur in spatial domain




---

---

---

---

---

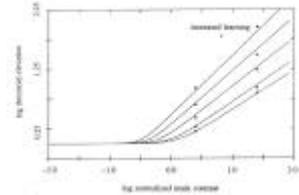
---

---

---

### VDP: Masking and Detection Probability

- ✂ Although masking is caused by transducer function it is actually modeled as separate image processing pathway created in addition to signal pathway
- ✂ Physiologically unrealistic, but makes easier to test and fine-tune parameters
- ✂ Masking modeled with possibility to include learning effects: results in decreasing masking slope
- ✂ Phase uncertainty modeled with LPF of masking images, otherwise detection near zero-crossings is overestimated (omitted in most 2nd party implementations)
- ✂ Signal difference and masking level signals are input to Weibull psychometric function to give a probability detection for each position in each channel
- ✂ Probabilities across channels are summed at each location with probability summation




---

---

---

---

---

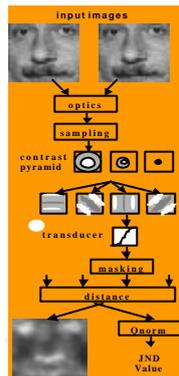
---

---

---

### Sarnoff Model

- ✂ Spatial portion originally developed by Jeff Lubin
- ✂ Color added by Michael Brill
- ✂ Many similarities to VDP
- ✂ Some advancements:
  - Separate Optics Modelling
  - Contrast transducer with facilitation
  - Suprathreshold capability
- ✂ Possible missteps
  - Local contrast calculation in place of amplitude nonlinearity
  - Fewer orientation channels
  - Attempt to do suprathreshold at expense of threshold accuracy
  - Resampling of image needed to convert to cy/deg space (at 120 pixels/degree)



\* Sarnoff Model figures courtesy of Michael Brill

---

---

---

---

---

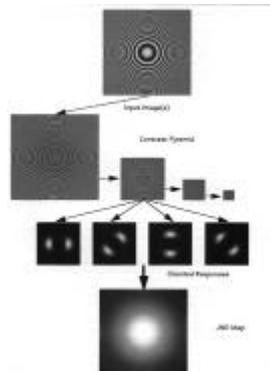
---

---

---

### Sarnoff Model

- Key Components**
- Optical PSF
  - steerable filters via convolution for channels (stay in spatial domain)
  - filters are weighted to get CSF normalization
  - filtered images go through transducer function (point nonlinearity)
  - differences taken to get JNDs (per position, per channel)
  - spatial pooling of JNDs which varies with eccentricity (periphery)
- ✂ JND are accumulated with various methods (average, peak, Minkowski) to get single number difference rating or, Visualization
- JNDs only aggregated across channels
  - no contrast polarity
  - therefore no spatial gauge of appearance




---

---

---

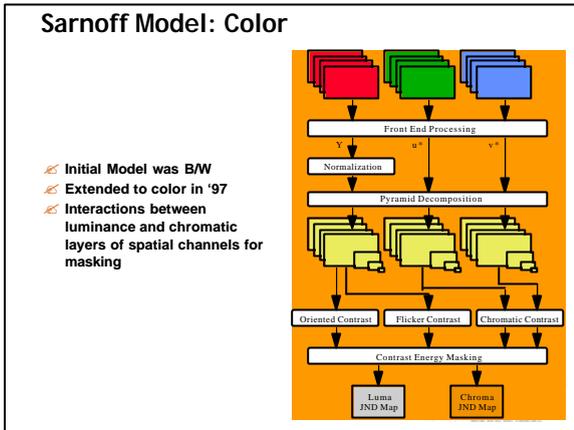
---

---

---

---

---




---

---

---

---

---

---

---

---

### Key differences between VDP and Sarnoff Model

- ✂ See paper by Li, Meyer, Klassen (1998) (albeit with some misunderstandings)
- ✂ **Scope:**
  - VDP is threshold and achromatic
  - Channel filters sum to 1.0: so is reversible for applications such as watermarking, lossless compression, graphics rendering
  - Sarnoff is suprathreshold, and now has color and temporal, meant as metric more than as model to place within applications
- ✂ **Physiological Soundness**
  - VDP channels are defined in digital frequency domain, engineering approach
  - Sarnoff channels defined in cy/deg domain, and image must be resized for mapping
- ✂ **Efficiency in implementation**
  - VDP least efficient in memory
  - Sarnoff least efficient in computations
- ✂ **Accuracy in predicting psychophysical results**
  - VDP most tested on core psychophysical threshold experiments
  - Sarnoff most tested on practical distortions at suprathreshold
- ✂ **Visualization Strategy**
  - VDP uses contrast polarity: allows shape of distortion to be simulated, but can't imply suprathreshold distortions (except where proportional to **SIGGRAPH**)
  - Sarnoff used JND scale: magnitude only, can do suprathreshold but shapes of distortions not simulated

---

---

---

---

---

---

---

---

### Key differences between VDP & Sarnoff Models (details)

- ✂ **Grey-level nonlinearity and contrast**
  - VDP models cone nonlinearity and calculated contrast against global mean
  - Sarnoff models local contrast (recently changed to global)
- ✂ **Optics and filter-nonlinearity sandwich**
  - VDP consolidates similar visual effects into basic stages (nonlinearities, spatial filtering, masking)
  - Sarnoff has front-end optics to affect LPF before any neural amplitude nonlinearity (correct physiologically)
- ✂ **CSF adaptivity**
  - VDP models CSF adaptive to light-level, image size, accommodation, other viewing conditions and applies as global filter prior to channel decomposition
  - Sarnoff achieves CSF by weighting individual channels making it harder to adapt CSF
- ✂ **Spatial phase and phase uncertainty**
  - VDP has all phases combined in each channel (but models phase uncertainty separately by LPF filtering masking signal in each cortex band)
  - Sarnoff has Hilbert pair for each channel (a sine and cosine phase, doubles # channels)
- ✂ **Spatial Pooling**
  - VDP has no spatial pooling other than channel filter support
  - Sarnoff models spatial pooling: further step of spatial averaging of signal after transducer
- ✂ **Cross-channel masking (Heeger, D'Zmura)**
  - Neither models this effect since increases memory load and hard to **SIGGRAPH 2001**

---

---

---

---

---

---

---

---

**Efficiency Versions**

Due to computational burden of VDP and Sarnoff models, numerous versions have been developed that aim to be much more efficient, acknowledging loss of accuracy

**Bradley's wavelet VDP for compression**

- 2D cartesian-separable wavelet
- explored overcomplete wavelet

**Mark Bolin's Hadamard-based model for computer graphics**

- 2D cartesian-separable wavelet using Hadamard basis functions
- Applied to direct and Monte Carlo light source sampling

**UCLA Hadamard model for signal processing**

- Uses oriented 1D Hadamard basis functions

**Watson DCT-based DVQ (Digital Video Quality) metric**

- Considers gray-scale nonlinearity as a form of masking (at DC)
- Models global CSF by mapping to DCT coefficients
- Models masking by grouping coefficients and applying nonlinearity
- Models color by a luminance channel, a single opponent color channel (R/G), and an S-cone channel
- Models spatial pooling, Minkowski summation
- Models temporal CSF with IIR so no frame buffers needed (Infinite Impulse Response)
- Performed as well as Sarnoff for MPEG2 distortions in VQEG study




---

---

---

---

---

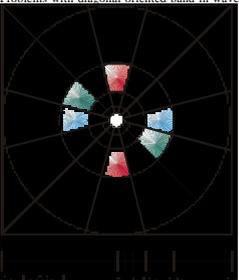
---

---

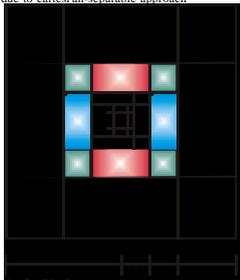
---

**Efficiency Versions: Wavelet problems**

- ✗ Close in structure to the visual filter bank (H and V dyadic), Cartesian-separable wavelet
- ✗ Problems with diagonal oriented band in wavelet due to cartesian-separable approach



Visual System Frequency Decomposition  
(Cortex Transform)



Wavelet Algorithm




---

---

---

---

---

---

---

---

**Moving Images & Spatiotemporal Issues**

- ✗ **Non-Spatiotemporal separability: cannot use product of spatial CSF and temporal CSF**
- ✗ **Temporal channels as well as spatial**
  - Sustained and Transient
  - Some evidence for 3rd temporal channel at low SFs
  - Spatio-temporal covariance: possible spatio and temporal channels configurations

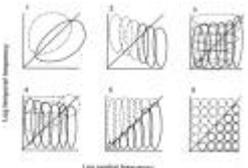


Fig. Spatio-temporal Covariance

- Line with slope = 1, 1 deg/sec
- Solid lines = nondirectional, dashed lines = directional
- Version 4 is consistent with most amount of data = velocity channels + flicker channel

✗ **Spatiovelocity separability**




---

---

---

---

---

---

---

---

### Spatiotemporal: Sarnoff Model

- Spatial model is basically doubled, each one having different temporal prefilter
- No accounting for eye movements
- No temporal masking, such as is often taken advantage of in adaptive video compression at scene cuts
- Modeled as spatiotemporal separable
- A buffer of 4 fields is used to implement temporal filter

---

---

---

---

---

---

---

---

### Spatiotemporal : Sarnoff Model: Results

- Sarnoff model prediction of two slices of the spatiotemporal surface of Van Doorn and Koenderinck:
- Errors are acknowledged and due to implementation efficiencies such as number of channels

---

---

---

---

---

---

---

---

### Motion Imagery

- Animation Quality Metric (AQM): advanced model for moving images (Myszkowski)
- Key visual effect that must be considered is high spatial frequencies as velocity increased
- Begins with Eriksson's gumbo of Watson, VDP, Sarnoff, Heeger techniques
- Computer graphics has luxury of knowing object (and pixel) velocities on display

---

---

---

---

---

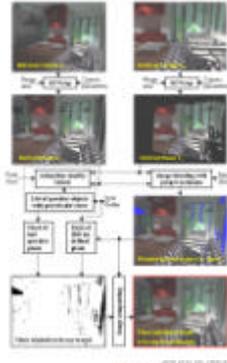
---

---

---

### Motion Imagery: Animation Quality Metric

- ✦ AQM best suited to computer graphics rendering
- ✦ Used to determine when lower quality will suffice per frame and per local region
- ✦ Could theoretically be applied to video compression
  - estimated motion vectors as needed by the motion compensation portions of the algorithm would replace the 3D warp and pixel flow modules
- ✦ Uses higher quality but computation-ally more expensive rendering such as ray tracing for key frames and selected glossy & transparent objects
  - higher velocities prevent their selection
  - Image Based Rendering used when lower quality will not be visually detected
  - threshold aspects of visual models are more important here




---

---

---

---

---

---

---

---

### Validation of Metrics

- ✦ Initial testing of the visual model is best to proceed from final detection stage toward front-end processing (reverse visual pathway)
  - Clever psychophysical experiments attempt to isolate internal processing stages of vision
  - These kinds of stimuli can allow finetuning of model parameters
- ✦ Example of testing given for VDP (achromatic and still images)
  - Fit of model to published psychophysical data
  - Once model is tested against psychophysical data for key stimuli, it can be tested for actual image quality applications
- ✦ Testing of visibility of distortions (more useful for visually lossless applications)
- ✦ Testing of image quality




---

---

---

---

---

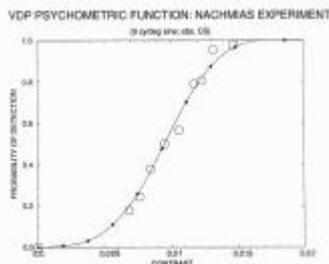
---

---

---

### Validation of Metrics: Psychometric Function

- ✦ Tests psychometric function component
- ✦ Tests detection of a single frequency on a uniform surround
- ✦ No preceding components adversely affect its modelling
- ✦ No cone nonlinearity, no CSF, no masking, etc.




---

---

---

---

---

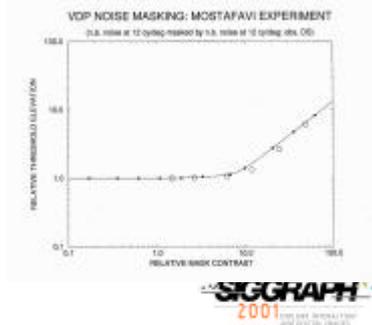
---

---

---

**Validation of Metrics: Masking Function**

- ✦ Test masking function and integration of energy within a channel
- ✦ Test masking function and integration of energy within a channel
- ✦ Tests CSF value and amplitude nonlinearity calibrations for 12 cpd single frequency




---

---

---

---

---

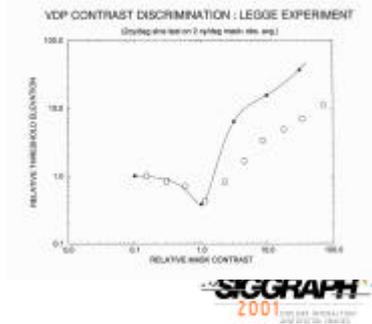
---

---

---

**Validation of Metrics: Contrast Masking**

- ✦ Tests ability model contrast masking where signal and mask are both narrowband and phase coherent signals
- ✦ Special modifications in input set-up required: if mask contrast (i.e. reference image) is less than threshold, a uniform field replaces reference image
- ✦ Magnitude of facilitation effect predicted
- ✦ Masking slope is too high, closer to 1.0 since based on noise masking results




---

---

---

---

---

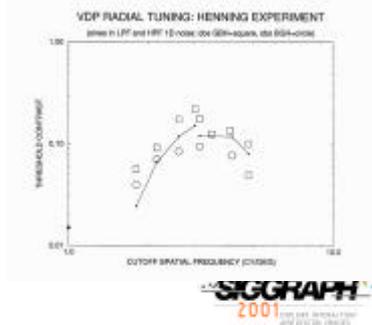
---

---

---

**Validation of Metrics: Radial Frequency Channels**

- ✦ Test radial frequency channels, i.e. their "tuning"
- ✦ Also tests interaction of psychometric function, probability summation, and CSF effects
- ✦ Note individual variations




---

---

---

---

---

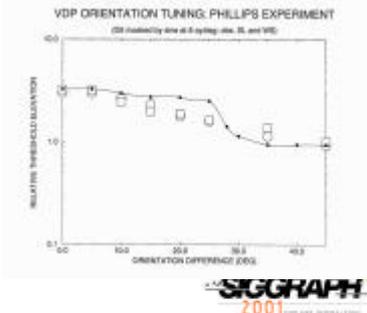
---

---

---

**Validation of Metrics: Orientation Channels**

- ✂ Tests shape of fan filters and their overlap
- ✂ Also interaction of radial fan cascade to create cortex filter
- ✂ Also tests off-frequency detection, and probability summation
- ✂ Note errors due to model's discrete channels



VDP ORIENTATION TUNING: PHILLIPS EXPERIMENT  
(200 cycles/sec at 0 cycles/sec, 30, and 90°)

RELATIVE THRESHOLD ELEVATION

ORIENTATION DIFFERENCE (DEG)

SIGGRAPH 2001

---

---

---

---

---

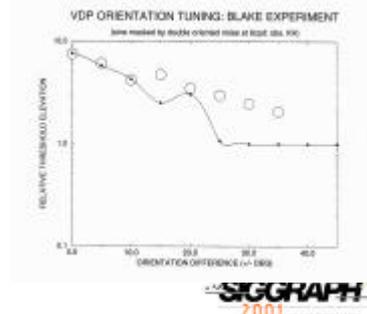
---

---

---

**Validation of Metrics: Orientation Channels**

- ✂ Here, the oriented noise has both orientations to minimize off-frequency looking
- ✂ Also tests similar model features as previous test



VDP ORIENTATION TUNING: BLAKE EXPERIMENT  
(noise masked by model's oriented noise at 90° and 180°)

RELATIVE THRESHOLD ELEVATION

ORIENTATION DIFFERENCE (°)

SIGGRAPH 2001

---

---

---

---

---

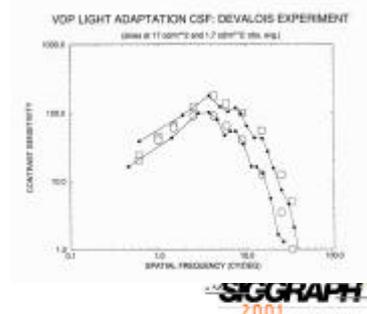
---

---

---

**Validation of Metrics: Spatial Frequency Response (CSF)**

- ✂ Tests global frequency response (CSF) and its light adaptation capabilities
- ✂ Also interaction of CSF with all post-CSF model components



VDP LIGHT ADAPTATION CSF: DEVALOIS EXPERIMENT  
(0.044 at 11 cd/m<sup>2</sup> and 1.7 cd/m<sup>2</sup> str. mag.)

CONTRAST SENSITIVITY

SPATIAL FREQUENCY (CYCLES/DEG)

SIGGRAPH 2001

---

---

---

---

---

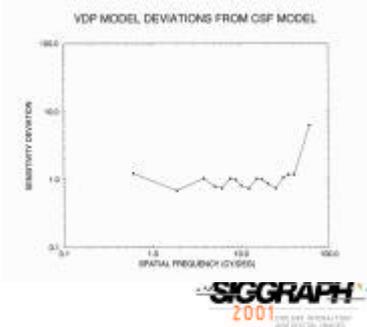
---

---

---

**Validation of Metrics: CSF, continued**

- ✂ Deviations from CSF model expectations
- ✂ Error shows underlying discrete channels, when true HVS has more of a continuum of channels



The graph shows 'VDP MODEL DEVIATIONS FROM CSF MODEL'. The y-axis is 'SENSITIVITY DEVIATION' on a log scale from 0.1 to 100.0. The x-axis is 'SPATIAL FREQUENCY (CYCLES)' on a log scale from 0.1 to 1000.0. The data points fluctuate around a value of 1.0, with a notable spike at the highest frequency.

**SIGGRAPH 2001**

---

---

---

---

---

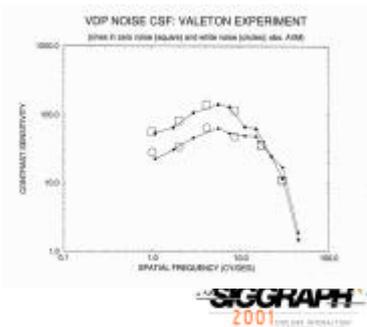
---

---

---

**Validation of Metrics: CSF in Noise**

- ✂ Tests ability to predict the effects of noise on the CSF
- ✂ Tests masking in conjunction with CSF
- ✂ Tests probability summation since some channels have higher masking due to preceding CSF



The graph shows 'VDP NOISE CSF: VALETON EXPERIMENT'. The y-axis is 'CONTRAST SENSITIVITY' on a log scale from 1.0 to 1000.0. The x-axis is 'SPATIAL FREQUENCY (CYCLES)' on a log scale from 0.1 to 100.0. Two data series are shown, both peaking around 100 cycles and then declining.

**SIGGRAPH 2001**

---

---

---

---

---

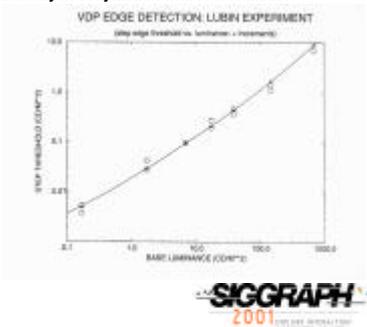
---

---

---

**Validation of Metrics: Luminance Nonlinearity**

- ✂ Tests amplitude nonlinearity in conjunction with rest of model
- ✂ Experiment is the detection of sharp luminance edge at different gray levels
- ✂ Directly relevant to practical problem of contouring artifacts throughout grayscale range



The graph shows 'VDP EDGE DETECTION: LUBIN EXPERIMENT'. The y-axis is 'STEP FREQUENCY (CYCLES/PI)' on a log scale from 0.01 to 10.0. The x-axis is 'BASE LUMINANCE (CD/MT²)' on a log scale from 0.1 to 100.0. The data points form a clear upward-sloping linear trend on the log-log scale.

**SIGGRAPH 2001**

---

---

---

---

---

---

---

---

### Validation of Metrics: Blur

- Practical test of ability to see blur as a function of contrast
- Tests entire model
- Note that masking at edges will affect visibility of blur signal, which lies close to edges

**SIGGRAPH 2001**

---

---

---

---

---

---

---

---

### Validation of Metrics: Vernier Acuity

- Vernier Acuity: ability to see lateral offset (break) in thin lines, as a function of line length
- Vernier acuity tests entire model, most important interacting mechanisms are the CSF and the orientation of channels
- Practical importance to visibility of jaggies, form various aliasing/antialiasing tradeoffs.
- HVS acuity initially seemed better than that allowed by cone sampling aperture

**SIGGRAPH 2001**

---

---

---

---

---

---

---

---

### Validation of Metrics: Current Status for Spatial Detection

- Modelfest: from Vision Science Community**
  - collect data that will challenge present models of spatial vision.
  - way to overcome vagaries due to different labs, methods, display calibration, observers, & undocumented details in previous psychophysics literature
- 1st year result: Spatial luminance detection**
  - key stimuli set shown to right:
  - 8 observers
  - standardized methods, display calibration, and display conditions
  - Watson has compared 5 basic models, with Gabor performing best
- Modelfest data set and modelling activities are example of the most rigorous building block that can be used for image quality
- 2nd year goal: Masking

**SIGGRAPH 2001**

---

---

---

---

---

---

---

---

**Validation of Metrics: Image Distortion Demos**

- ✂ **Testing of Models in actual image applications**
- ✂ **Primarily due to masking aspects, distortions are not present uniformly through image; they are localized**
- ✂ **Visualization of distortion by visual model:**
  - Location and Contrast Sign (Daly, since only detection model)
  - Location and Magnitude (Lubin, Heeger, most others)
- ✂ **Early Approach:**
  - Rather than observer study of ranking image quality, ask observers where they can see distortions in an image.
  - User feedback used to fine-tune model
  - Result is a collection of practical visible distortion localization demos (publication constraints)
- ✂ **Quantitative localization approach (Zhang, Setiawan, Wandell '97):**
  - Spatially quantize images,
  - Observers click where they can see distortions




---

---

---

---

---

---

---

---

**Validation of Metrics: Direct Assessment of Image Quality**

- ✂ **Distorted image sequences are rated for image quality on a quality scale**
- ✂ **Sometimes a reference sequence is shown**
- ✂ **Reference-free rating is referred to as open-ended**
- ✂ **Most techniques obtain single rating across all locations in image and across all frames**
  - Difficult for observers to perform well
  - Can only look at one location within each image per frame (eccentricity effect)
  - Theoretically must know where observer is looking
  - Psychological assessment of varying quality into single number not well understood
- ✂ **Continuous quality assessment (observer varies slider throughout sequence)**
  - Recency effect (Hamburg and DeRidder)
  - Minkowski summation pooling methods (Rohaly, et al)
  - Minkowski exponent varies (Keelan)
    - ✂ If large distortion exists, it dominates (exponent large, e.g., >2)
    - ✂ If image has many different small distortions, they sum linearly (exponent = 1)




---

---

---

---

---

---

---

---

**Validation of Metrics: Direct Assessment of Image Quality: VQEG**

- ✂ **VQEG (Video quality experts group) recent study**
  - Intention was to form a video image quality standard that could be used in product spec sheets
  - Well-controlled experiment using mostly MPEG2 compressed video
  - Various models predicted overall video image quality observer ratings for various sequences
  - Models performed well, but little better than weighted MSE
  - Key problems: No viewing distance changes were used
    - No display characterizations were made (spatial MTF, noise, tone= only ?)
    - No models could handle localized spatial displacement (via MPEG2)
    - DCT of MPEG2 agreeable with MSE if frequencies < 10 cy/deg
- ✂ **Overly Ambitious?**




---

---

---

---

---

---

---

---

**Summary**

- ✂ **Ability to predict:**
  - spatial still luminance threshold detection is very well established
  - spatial still luminance suprathreshold well tested for key applications, but not tested against key psychophysical experiments
  - spatiochromatic threshold detection not well tested
  - spatiochromatic appearance modeled and simulated for qualitative testing (Ferwerda)
  - spatiotemporal luminance detection tested for a few key psychophysical experiments, but mostly only tested for applications
  - spatiotemporal luminance suprathreshold tested for overall image quality along with color
  - spatiochromatic temporal appearance not well explored
- ✂ **One key problem: patience and funding**
  - Applied people using visual models have skipped basic psychophysical validation and many expect standardization for complete image quality models
  - Near-impossible to get funding for key groundwork validation
  - Vision science community is moving ahead but consolidation a la Modelfest is slow
- ✂ **Areas needing refinement:**
  - More rigorous testing of spatiochromatic
  - Individual variations
  - Eye movements and region of interest
- ✂ **Complexity of visual models surpasses often neglected display modeling**




---

---

---

---

---

---

---

---

---

---

**References:**

- ✂ **Visual Models:**
  - S. Daly (1993) The Visible Differences Predictor: An algorithm for the assessment of image fidelity. In A. B. Watson (ed.), Digital Images and Human Vision, MIT Press, 179-206.
  - S. Daly (1993) Quantitative performance assessment of an algorithm for the determination of image fidelity, SID 93 Digest, 317-320.
  - P. Teo and D. Heeger (1994) Perceptual image distortion, IEEE Proc. ICIP-94 V.2 982-986.
  - J. Lubin (1995) A visual system discrimination model for imaging system design and evaluation, in E. Peli (ed.), Visual Models for Target Detection and Recognition, WorldScientific Publishers.
  - J. Lubin, M. Brill, and R. Crane (1996), Vision model-based assessment of distortion magnitudes in digital video, presented at the Nov. meeting of the International Association of Broadcasters (IAB).
  - B. Li, G. Meyer, R. V. Klassen (1998) A comparison of two imagequality models, Proc. SPIE V. 3299, 98-109.
  - S. Pattanaik, M. Fairchild, J. Ferwerda, D. Greenburg (1998) Multiscale model of adaptation, spatial vision and color appearance, IS&T CIC Proceedings V. 6, 2-7.
  - M. D'Zmura, T. Shen, W. Wu, H. Chen, and M. Vassiliou (1998) Contrast gain control for color image quality, Proc. SPIE V. 3299 194-201.
  - S. Daly (1998) Engineering observations from spatiotemporal and spatiovelocity visual models, SPIE Proc. 3299, 180-191.




---

---

---

---

---

---

---

---

---

---

**References:**

- ✂ **Efficiency versions of visual models:**
  - M. Bolin and G. Meyer (1998) "A perceptually based adaptive sampling algorithm" Siggraph.
  - A. Bradley (1999) "A wavelet visible differences predictor" IEEE Trans. Image Processing, V. 8, 717-730.
- ✂ **Modelfest:**
  - <http://www.neurometrics.com/projects/Modelfest/stimulusList.htm>
  - Watson study: <http://www.opticsexpress.org/oearchive/source/14103.htm>
- ✂ **Image quality rating and other observer studies**
  - Zhang, Setiawan, Wandell, 5th Color Imaging Conference 1997).
  - R. Hamburg and H. de Ridder, (2000) Submitted manuscript
  - A. M. Rohaly (1999) SPIE Proc. V. 3644, 218-225.
  - B. Keelan (2000) IS&T PICS conference Proc., 197-203.
- ✂ **Psychophysical and physiological data**
  - A.J. van Doorn and J.J. Koenderink, "Spatiotemporal contrast detection threshold surface is bimodal," Optics Letters 4, 32-34 (1979).
  - G. Orban, H. Kennedy, and J. Bullier (1986) Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey. J. of Neurophysiology V. 56 #2 462-480.




---

---

---

---

---

---

---

---

---

---

# Seeing is Believing: Reality Perception in Modeling, Rendering & Animation

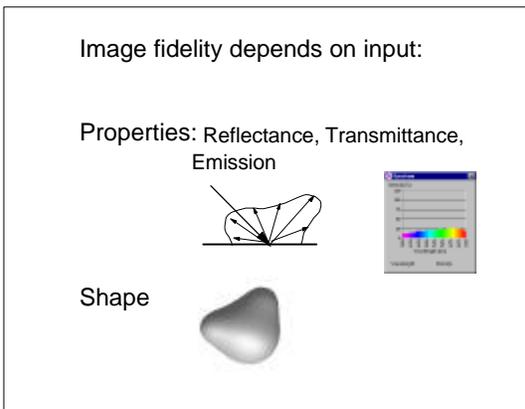
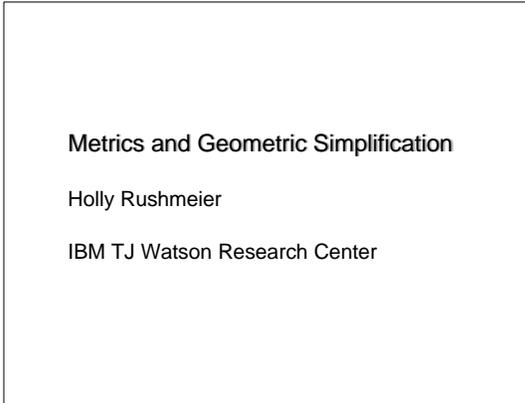


## V. Metrics and Geometric Simplification

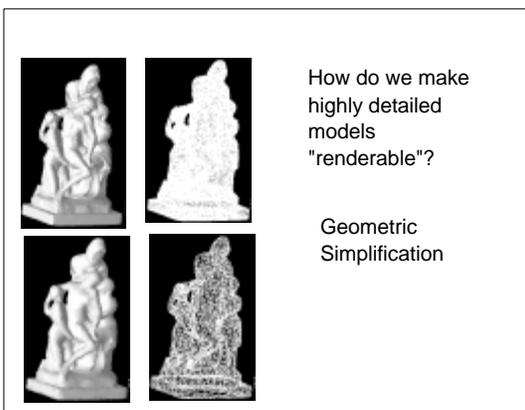
Alan Chalmers  
Scott Daly  
Ann McNamara  
Karol Myszkowski  
**Holly Rushmeier**  
Tom Troscianko

Course #21

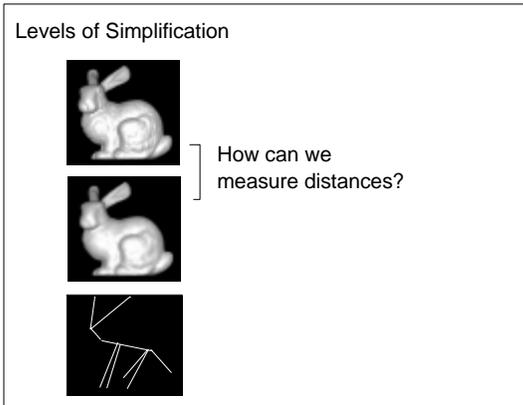




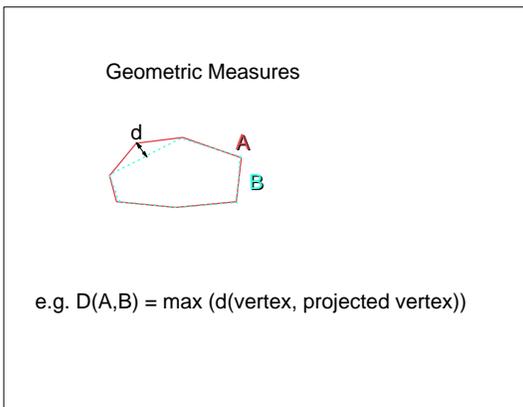
- ▶ The accuracy of a final image depends on input data, discretization error and computation errors\*. The input data includes the properties of how a surface interacts with light, and the shape of a surface. Given enough time, an image quality metric could be use to measure perturbations in the input data, as well as their effects on subsequent calculations. In practice however, we need some measure of what input data will be adequate. In this section we consider the problem of developing perceptual metrics for shape.
- ▶ \*Ref: Arvo, Torrance, Smits "A Framework for the Analysis of Error in Global Illumination Algorithms, SIGGRAPH 94,pp 75-84.



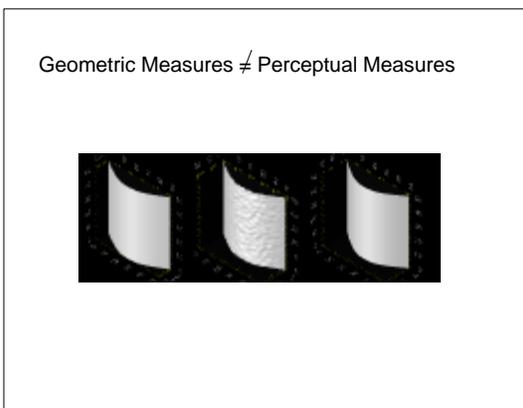
- ▶ 3D scanning systems have made it possible to capture very high spatial resolution models\*. Many models are obviously hugely over-sampled and consume a great deal of time and memory to render. Numerous methods have been developed to simplify geometries.\*\*
- ▶ \*Ref. Bernardini & Rushmeier "The 3D Model Acquisition Pipeline", Eurographics 2000, State of the Art Report.
- ▶ \*\*P. Cignoni, C. Montani and R. Scopigno, "A Comparison of Mesh Simplification Algorithms", Computers and Graphics, 2000.



- ▶ To find the best simplification we need a way to measure the difference between representations of the same object. It is possible to simplify all the way down to a near symbolic representation of an object. Here we focus on representations that attempt to maintain the appearance of the original object.



- ▶ Tools have been developed to measure the *geometric* distance between surfaces in a variety of ways.\*
- ▶ \*Ref: P. Cignoni, C. Rocchini and R. Scopigno "Metro: Measuring Error in Simplified Surfaces" Computer Graphics Forum, Volume 17, Number 2, 1998, pp. 167-174.



- ▶ Geometric measures are not perfect measures of how much surfaces will look alike. In this example, the original surface is on the far left. The RMS distance of the surface in the center to the surface at the left is smaller than the RMS distance for the surface on the right. However, the surface on the right looks more like the surface on the far left.

What we would like:

$P(A,B)$  where

$P(A,B) = 0 \Rightarrow$  all images generated using B look identical to images generated using A

if  $P(A,C) = 2P(A,E)$ , any image using C in place of A rather than E in place of A will look twice as bad

We don't have and won't get  $P( )$

Why don't we have  $P( )$  from human vision or computer vision??

-- lot's of work in shape perception, but it doesn't take this form

What can we do??

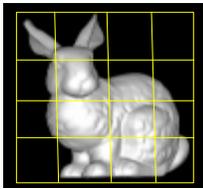
-- so far, use more general insights

- ▶ An example of work on shape perception in computer graphics: R. Browse, J. Rodger and R. Adderly "Perception of Object Shape in Computer Graphics Displays" Journal of Electronic Imaging 10(1) 181-187, January 2001

Applying insights about perception:

Limit the spatial extent of simplification:

SCROOGE

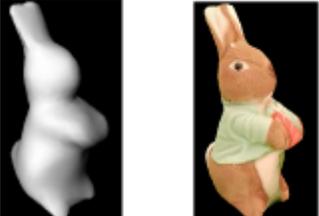


- ▶ Human vision sensitivity varies with spatial frequency. Some algorithms try to take advantage of this by limiting the spatial frequencies that can be impacted by a change caused by simplification.
- ▶ Ref. M. Reddy "SCROOGE: Perceptually-Driven Polygon Reduction", Computer Graphics Forum, Vol. 15, 1996, No. 4, pp. 191-203.

Applying insights about perception:

Texture detail can distract from simplification artifacts:

Masking, separate geometry and color



- ▶ "Visual masking" is the phenomenon in which one visual pattern affects the detectability of another. The content of texture map applied to a surface can obscure artifacts of simplification.\* This is commonly exploited, but not with specific metrics, in the simplification of colored-point data.\*\*
- ▶ \*Ref: Ferwerda, Pattanaik, Shirley and Greenberg, "A Model of Visual Masking for Computer Graphics", SIGGRAPH 97, pp. 143-152.
- ▶ \*\*Ref: Soucey, Godin and Rioux, "A texture-mapping approach for the compression of colored 3D triangulations" The Visual Computer, vol 12, pp. 503-513, 1996.

Applying insights about perception:

Only what is visible matters, and what is close matters most:

View dependent simplification



- ▶ Many algorithms store hierarchical representations of objects and only display visible and close portions of the model in detail.\*
- ▶ \*Ref: J.C. Xia and A. Varshney "Dynamic View-Dependent Simplification for Polygonal Models", Proceedings of IEEE Visualization 96, pp. 327-334, 498.
- ▶ \*Ref: R. Klein and A. Schilling "Efficient Rendering of Multiresolution Meshes with Guaranteed Image Quality", Visual Computer, 1999, Vol 15, pp. 443-452.

Applying insights about perception:

Only the interaction with light matters:

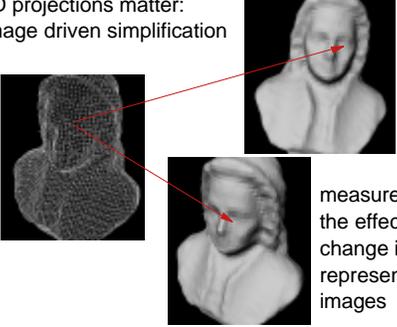
Normals maps



- ▶ Here a simplified geometry at the left is represented lit in two different ways with detail provided by normals mapped onto the surface at a higher resolution. This was originally suggested for simplification\* and can be used to acquire data in simplified form.\*\*
- ▶ \*Ref. J. Cohen, M. Olano, and D. Manocha, "Appearance Preserving Simplification", Proceedings of SIGGRAPH 98, pp. 155-122.
- ▶ \*\*Ref. H. Rushmeier, F. Bernardini, J. Mittleman and G. Taubin, "Acquiring Input for Rendering at Appropriate Levels of Detail: Digitizing a Pieta" Ninth Eurographics Rendering Workshop, June 1998, pp. 81-92.

Applying insights about perception:

2D projections matter:  
Image driven simplification

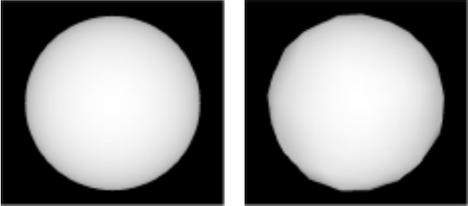


measure the effect of change in representative images

- ▶ Simplification can be driven by making incremental changes that have the smallest effect on a small set of representative images.\*
- ▶ \*Ref. P. Lindstrom and G. Turk, "Image-Driven Simplification", ACM Transactions on Graphics, Vol. 19, No. 3, July 2000, pp. 204-241.

Applying insights about perception:

Silhouettes matter:  
save silhouettes separately



- ▶ Since silhouettes are important, it improves the representation to save silhouettes separately at a higher resolution.\*
- ▶ \*Ref. P. Sander, S. Gortler, H. Hoppe and J. Snyder, "Silhouette Clipping", Proceedings of SIGGRAPH 2000, pp. 327-334.

Methods measure perceptually related parameters, but not the perceived quality of a representation, so:

- How can we judge which method is best for a storage/time budget?
- How can we compare the quality of representation of different objects?
- How can we design how to mix approaches?
- How can we tell when a simplification is adequate?

Experiments in Measuring Model Quality

Naming Times

Similarity Ratings

Naming Times

Ben Watson\*, Alinda Friedman, Aaron McGaffey  
Depts. of Comp. and Psych. U. Alberta  
\*now Northwestern

Time to name an object related to perceived quality.

► Ref: B. Watson, A. Freidman and A. McGaffey. "Using naming time to evaluate quality predictors for model simplification, Proceedings of ACM CHI 2000, Pages 113 - 120

Naming times:

Used simple models as "0%"  
Simplified 50% and 80%

Longer naming time => decreased quality

How did naming time difference between 0% and the two levels of simplification correlate with geometric and image measures?

► Naming times can be used as a measure of quality, but how to we transfer this evaluation in to an automated process? We need to find computable metrics that correlate well with the naming time data.

Correlations not significant for 50% models,  
with geometric MSE doing best.

Correlations better for 80% models,  
with perceptually based image metric  
having significant correlation.

- ▶ Neither geometric or image-based metrics do well.

Similarity Measures:

How effective is replacing geometry  
with normals?

- ▶ Ref. H. Rushmeier, B. Rogowitz and C. Piatko, "Perceptual Issues in Substituting Texture for Geometry", Proceedings of Human Vision and Electronic Imaging, SPIE Vol. 3959, pp. 372-383.

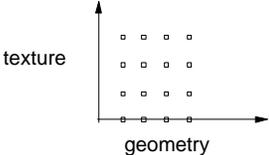
ISSUES

- what is the limit for replacing geometry for texture?
- is texture replacement always better?
- are certain types of geometry more suitable for texture replacement?
- are there different rules for different classes of objects, different viewing conditions?

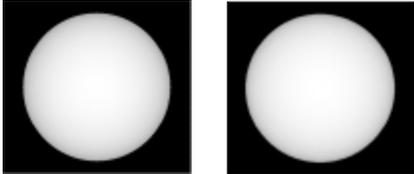
A Simple Experiment:

Two simple objects with white matte surface

Two lighting conditions (front, side)

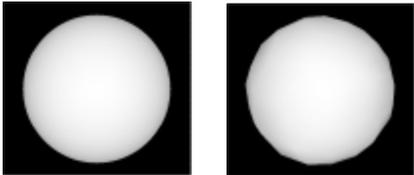


- Two objects: a smooth sphere, and a sphere with a crinkled surface were used. The appearance of approximations of the object with various levels of geometry and texture simplification were judge, to get insight into the trade-off between representation as pure geometry and texture in the form of normals maps.



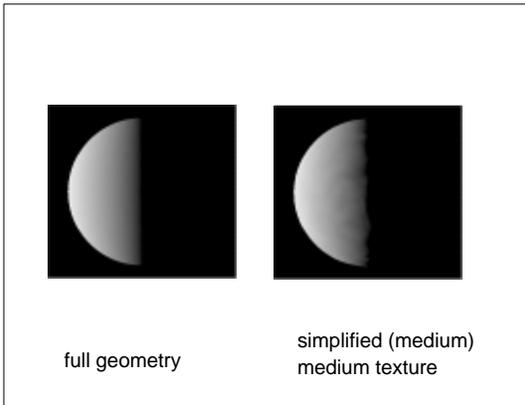
full geometry                  simplified (medium)  
no texture

- An example of the smooth sphere, with no texture mapping, front lit.

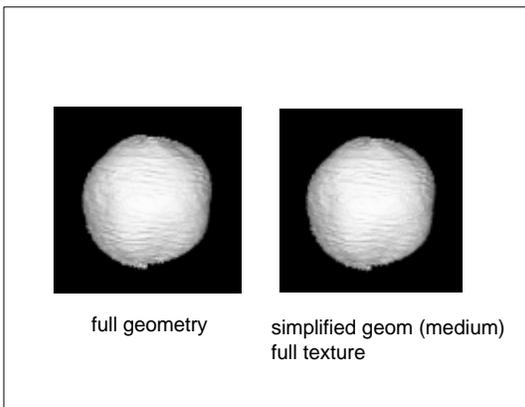


full geometry                  simplified (small)  
small texture

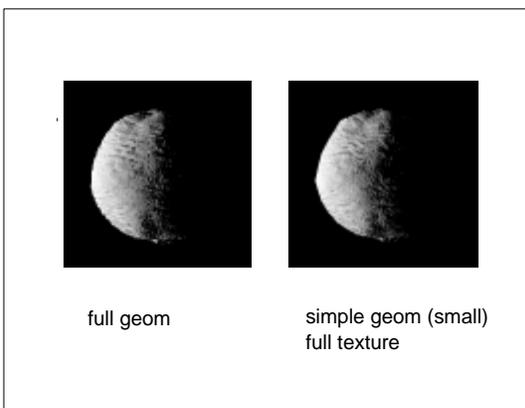
- An example of the smooth sphere using texturing, front lit.



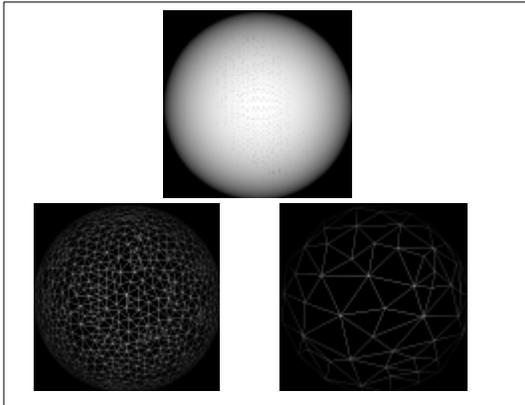
- ▶ An example to examine the effect of lighting direction.



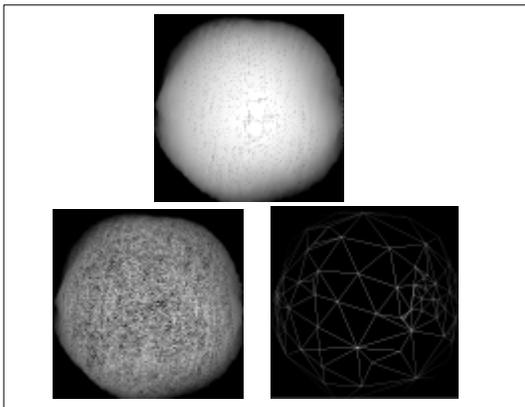
- ▶ The "crinkled" sphere case.



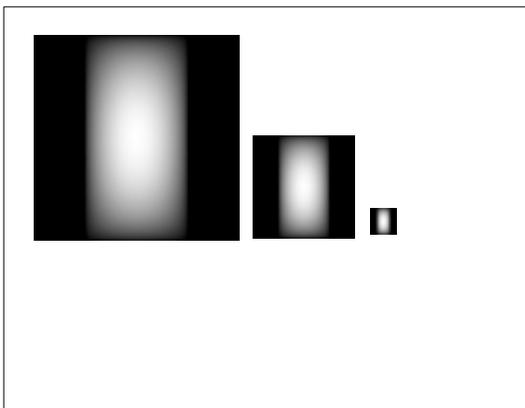
- ▶ The effect of lighting in the crinkled sphere case.



- ▶ The underlying geometry for the three levels of the smooth sphere.

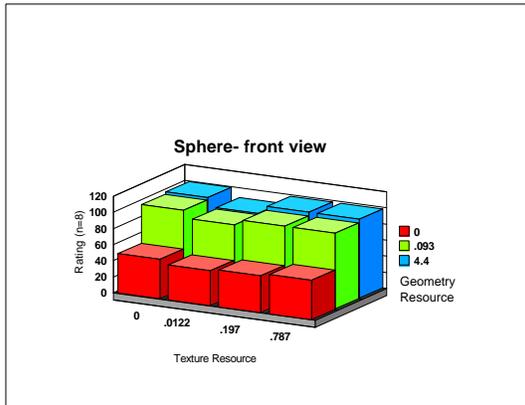


- ▶ The underlying geometry for the three levels of the crinkled sphere.

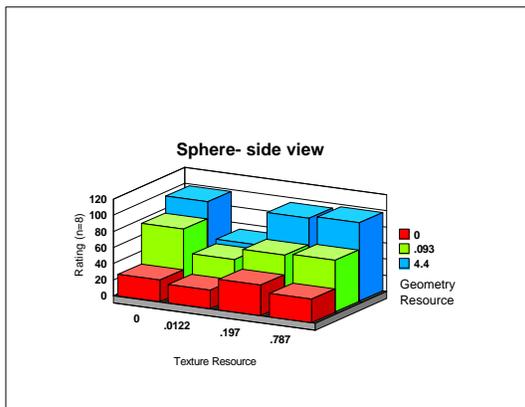


- ▶ Various levels of texture simplification -- the texture image resolution is just reduced.

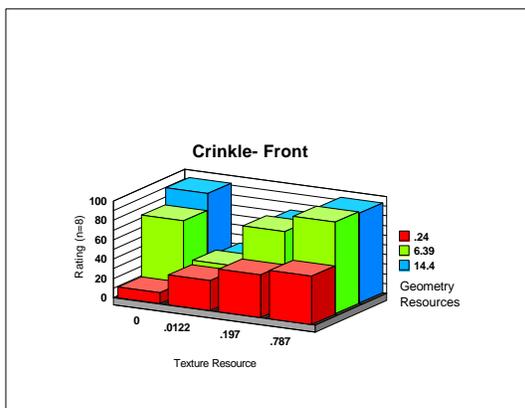
- Texturing didn't improve the simplified geometries

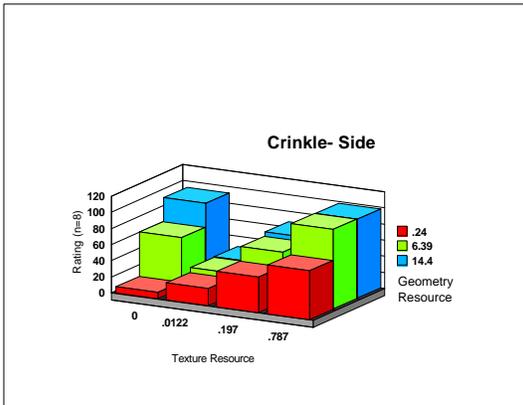


- The quality of the approximations depends on lighting direction.



- Applying texture for the crinkled case always improves the highly simplified object and the detailed texture improves the somewhat simplified geometry.

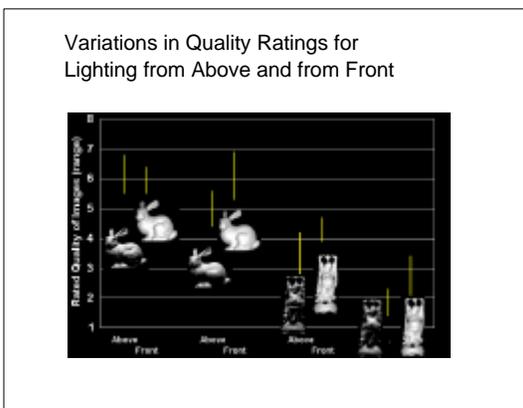




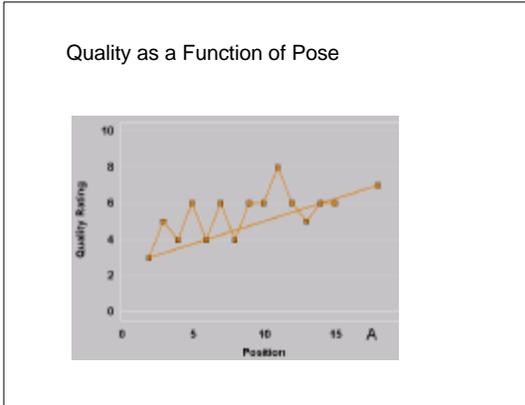
- Light direction also has an impact on the appearance of the crinkled surface.

More questions than answers:  
 Are experiments on static images adequate for evaluating geometric quality? Which image(s)?

- A single object can be viewed in many poses, under many different lighting conditions. If we are going to use image-based metrics, which images are important?



- We asked observers to rate the quality of different levels of simplifications for different lighting conditions. For the same level of simplification, the range of results was always different for the two conditions.



- ▶ Even for a single observer, the assessment of quality varied with the object pose. The quality of an animation composed of these poses is not the max, min or average of the quality of the individual poses.

Conclusion:

Evaluating the quality of a geometric object representation requires more than a good image metric.

# Seeing is Believing: Reality Perception in Modeling, Rendering & Animation



## VI. Applications in Rendering and Animation

Alan Chalmers  
Scott Daly  
Ann McNamara  
**Karol Myszkowski**  
Holly Rushmeier  
Tom Troscianko

Course #21



## Perceptually-based global illumination and rendering techniques

Karol Myszkowski  
Max-Planck-Institute for Computer Science  
Germany



---

---

---

---

---

---

---

---

## Outline

- ⌘ Questions of Appearance Preservation
- ⌘ Daly's Visible Differences Predictor and its human psychophysical validation for rendering tasks
- ⌘ Selected applications of the VDP to guide global illumination computation
- ⌘ Animation Quality Metric and its applications toward automatic keyframe selection in animation rendering



---

---

---

---

---

---

---

---

## Questions of Appearance Preservation

**The concern is not whether images are the same; rather the concern is whether images appear the same.**

- ⌘ How much computation is enough?  
How much reduction is too much?
- ⌘ An objective metric of image quality which takes into account basic characteristics of the human perception could be of some help to answer these questions without human assistance.



---

---

---

---

---

---

---

---

### Modeling important characteristics of the Human Visual System

- Temporal and spatial mechanisms (channels) which are used to represent the visual information at various scales and orientations as it is believed that primary visual cortex does.
- Contrast Sensitivity Function which specifies the detection threshold for a stimulus as a function of its spatial and temporal frequencies.
- Visual masking affecting the detection threshold of a stimulus as a function of the interfering background stimulus which is closely coupled in space and time.




---

---

---

---

---

---

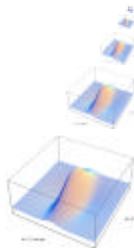
---

---

### Spatial and orientation mechanisms

The following filter banks are commonly used:

- Gabor functions (Marcelja80),
- Steerable pyramid transform (Simoncelli92),
- Discrete Cosine Transform (DCT),
- Difference of Gaussians (Laplacian) pyramids (Burt83, Wilson91),
- Cortex transform (Watson87, Daly93).




---

---

---

---

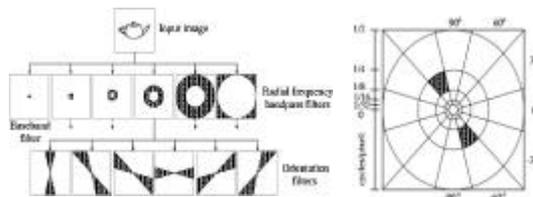
---

---

---

---

### Cortex transform: organization of the filter bank




---

---

---

---

---

---

---

---

### Cortex transform: orientation bands

The diagram illustrates the cortex transform process. In the center is an 'Input image' of a modern interior. Four red arrows point outwards from this image to four separate orientation bands. Each band shows a grayscale image where only features of a specific orientation are visible, while other orientations are suppressed. The bands are arranged in a 2x2 grid. A SIGGRAPH 2001 logo is in the bottom right corner.

---

---

---

---

---

---

---

---

### Spatiovelocity Contrast Sensitivity Function

Contrast sensitivity data for traveling gratings of various spatial frequencies were derived in Kelly's psychophysical experiments (1960).

The figure contains two plots. On the left is a 3D surface plot showing the contrast sensitivity function as a function of spatial frequency and temporal frequency. The surface is highest at low spatial and temporal frequencies and decreases as they increase. On the right is a 2D line graph with 'Spatial Frequency (cycles/deg)' on the x-axis and 'Contrast Sensitivity' on the y-axis. It shows two curves: a solid line for 'Spatial Frequency (cycles/deg)' and a dashed line for 'Temporal Frequency (Hz)'. Both curves show a peak at low frequencies and then decrease. A SIGGRAPH 2001 logo is in the bottom right corner.

Daly (1998) extended Kelly's model to account for target tracking by the eye movements.

---

---

---

---

---

---

---

---

### Visual masking

Masking is strongest between stimuli located in the same perceptual channel, and many vision models are limited to this intra-channel masking.

The following threshold elevation model is commonly used:

The graph shows a curve representing a threshold elevation model. The x-axis is labeled 'log 1 + (Mask Contrast / Threshold)^2' and ranges from -1.0 to 1.0. The y-axis is labeled 'log Threshold Elevation' and ranges from 0.0 to 2.0. The curve is flat at 0.0 for negative x-values and then rises exponentially for positive x-values. A SIGGRAPH 2001 logo is in the bottom right corner.

---

---

---

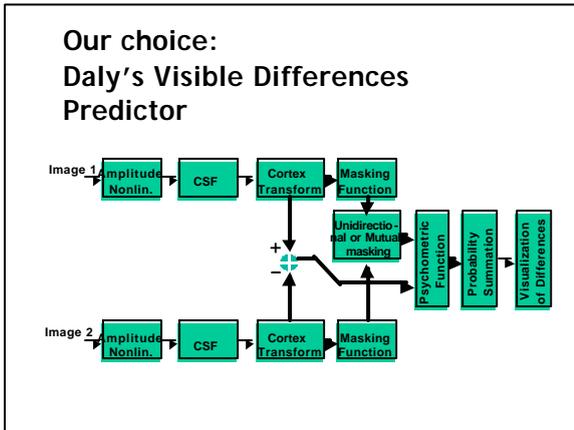
---

---

---

---

---




---

---

---

---

---

---

---

---

**Daly's VDP: outstanding features**

- ⌘ Predicts local differences between images
- ⌘ Takes into account important visual characteristics:
  - ⌘ Weber's law-like amplitude compression,
  - ⌘ advanced CSF model,
  - ⌘ masking (mutual or unidirectional)
- ⌘ Uses the Cortex transform, which is a pyramid-style, invertible, and computationally efficient image representation

---

---

---

---

---

---

---

---

**VDP validation in typical global illumination and rendering tasks**

<http://www.mpi-sb.mpg.de/resources/vdp>

---

---

---

---

---

---

---

---

### Human Psychophysical Validation of VDP

- Goal:
  - To determine whether VDP -based predictions match well with subjective reports of visible differences between images under conditions mimicking those in VDP applications
- Experiments:
  - Perceived quality of shadows cast upon textured surfaces for various texture scales
  - Perceived image quality for successive stages of global illumination solution



---

---

---

---

---

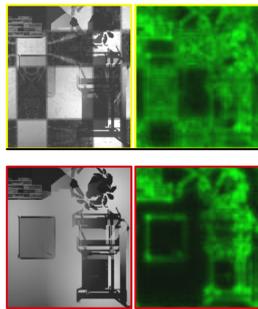
---

---

---

### Shadow masking

Visualization of the contrast threshold elevation due to masking. Stronger masking occurs when the target image contains a texture (top row). Bright green denotes more masking.



---

---

---

---

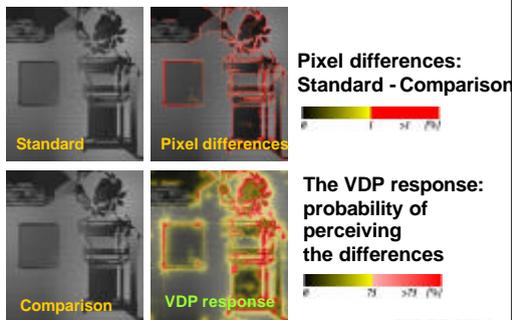
---

---

---

---

### Shadow masking vs. texture pattern size (1)



---

---

---

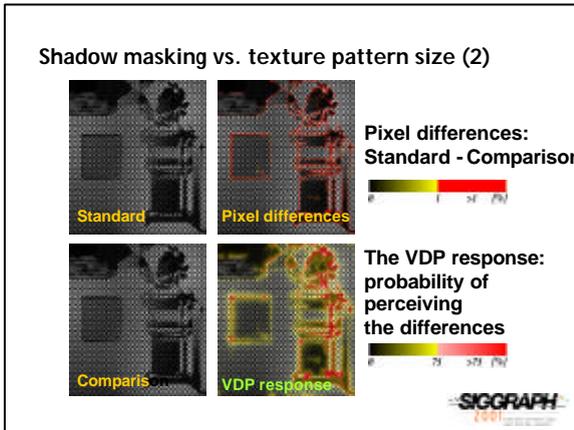
---

---

---

---

---




---

---

---

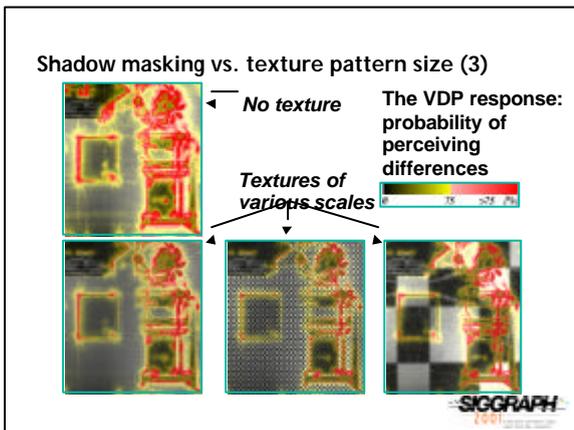
---

---

---

---

---




---

---

---

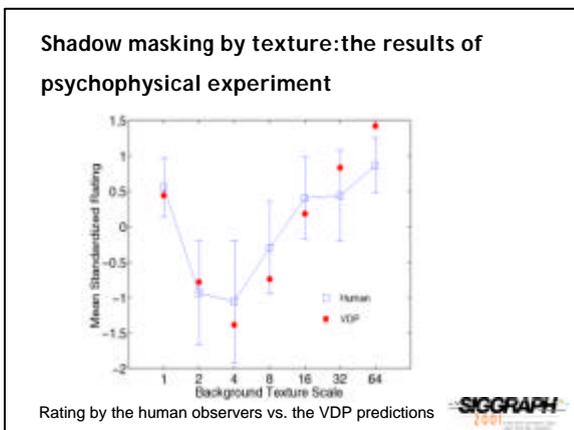
---

---

---

---

---




---

---

---

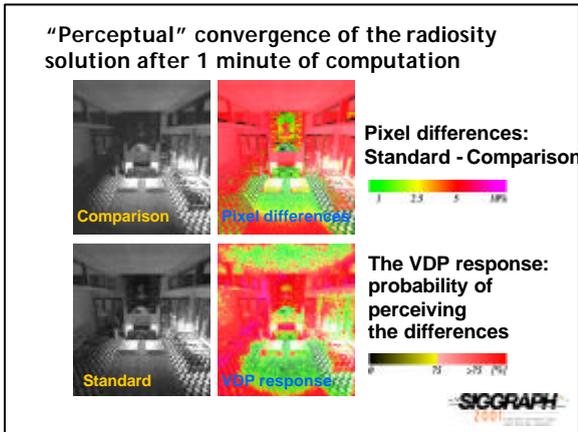
---

---

---

---

---



---

---

---

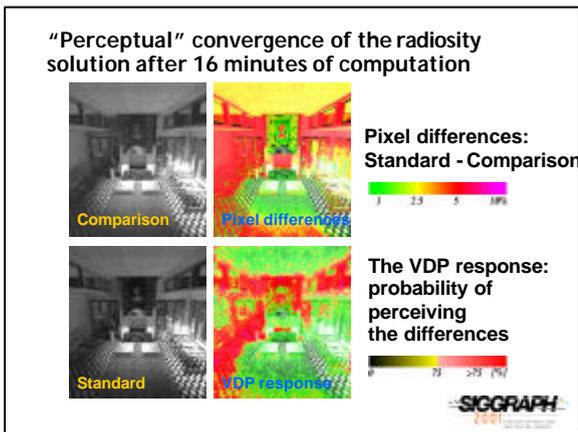
---

---

---

---

---



---

---

---

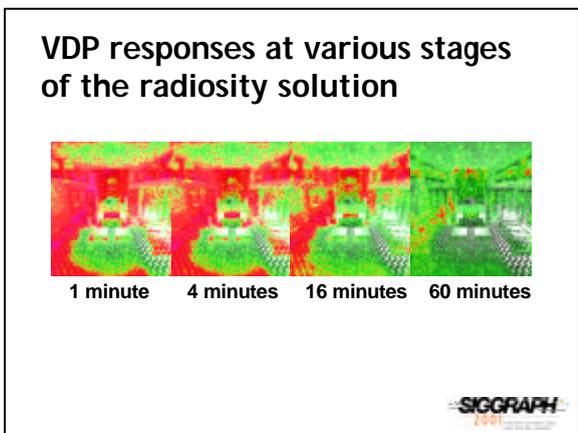
---

---

---

---

---



---

---

---

---

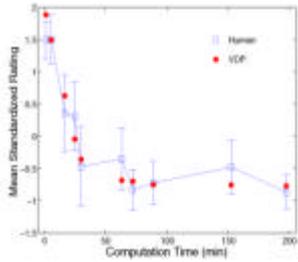
---

---

---

---

**“Perceptual” convergence of the radiosity solution:  
the results of psychophysical experiment**



Rating by the human observers vs. the VDP predictions



---

---

---

---

---

---

---

---

**A progressive  
global illumination solution**



---

---

---

---

---

---

---

---

**Algorithm selection  
procedure**

- ⌘ Goal: minimize the perceived differences between the intermediate and final images as a function of time by selecting the best algorithm at every stage of the global illumination solution.
- ⌘ Method: use the perception-based Visible Differences Predictor (VDP) developed by Daly to get quantitative measures of such differences. (Currently, VDP is used off-line to select the most efficient algorithms in terms of image quality progression.)



---

---

---

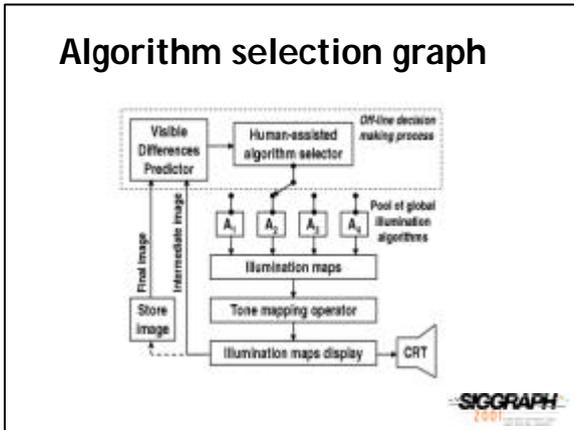
---

---

---

---

---




---

---

---

---

---

---

---

---

### Pool of algorithms

- ⌘ Deterministic Direct Lighting (DDL).
- ⌘ Indirect lighting computation using Hierarchical Progressive Radiosity with clustering (HPR).
- ⌘ Density Estimation Particle Tracing (DEPT) from light sources with lighting function filtering performed in 3D space. Photons for direct (DDEPT) and indirect (IDEPT) lighting are stored separately.

---

---

---

---

---

---

---

---

### Illumination maps filtering in the DEPT algorithm

- ⌘ Static, balanced kd-tree with mesh vertices is used to search for local illumination estimates
- ⌘ Adaptive selection of density estimation filter support based on mathematically-sound local statistic measures of illumination variation
- ⌘ As solution converges the local filter support shrinks reducing bias.

---

---

---

---

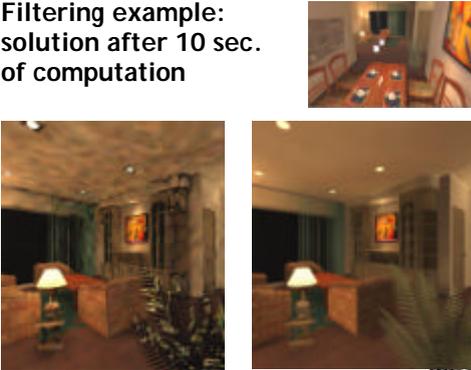
---

---

---

---

**Filtering example:  
solution after 10 sec.  
of computation**



Without filtering      With filtering



---

---

---

---

---

---

---

---

**Pool of algorithms (2)**

- ⌘ All three algorithms use mesh to store lighting simulation results.
- ⌘ Graphics hardware can be used to display mesh-reconstructed lighting, and walkthrough animation is possible at any stage of computation.
- ⌘ The final gather step is not required to obtain images of good quality.



---

---

---

---

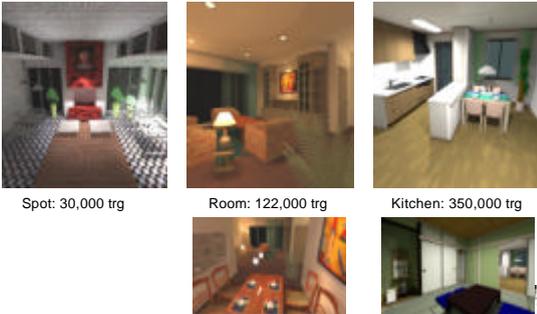
---

---

---

---

**Measuring basic algorithms  
performance: test scenes**



Spot: 30,000 trg      Room: 122,000 trg      Kitchen: 350,000 trg

---

---

---

---

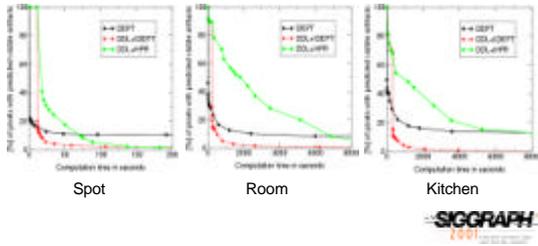
---

---

---

---

**Measuring basic algorithms performance: "perceptual" convergence rather than "physical" convergence**




---

---

---

---

---

---

---

---

**Our algorithm choice supported by the VDP responses**

- ⌘ Use at first a customized DEPT (compromise bias of the intermediate solution to get fast image feedback upon the user demand and reduce excessive noise).
- ⌘ At T1 switch to deterministic direct lighting computations to reconstruct fine details of the lighting function.
- ⌘ At T2 switch back to DEPT to reduce variance of the indirect lighting solution. This makes it possible to reduce the final solution bias by relaxing noise filtering.




---

---

---

---

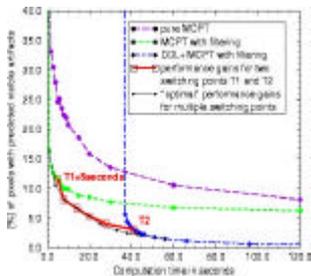
---

---

---

---

**Selection of switching points between algorithms:  $T_1$  and  $T_2$  versus  $T_1, \dots, T_N$**



Spot (old results on slower R10000 processor)




---

---

---

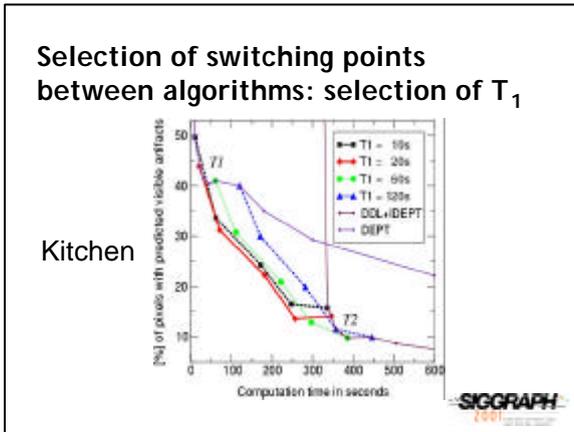
---

---

---

---

---




---

---

---

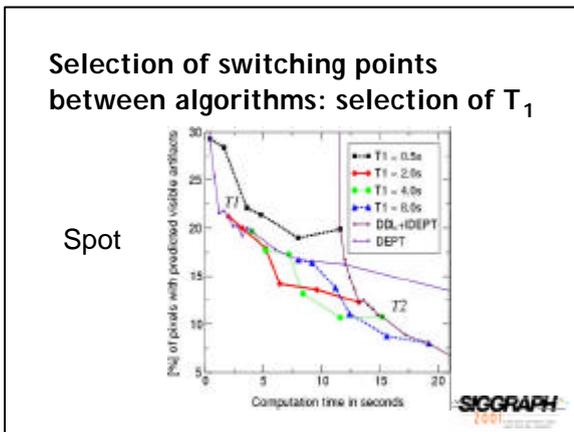
---

---

---

---

---




---

---

---

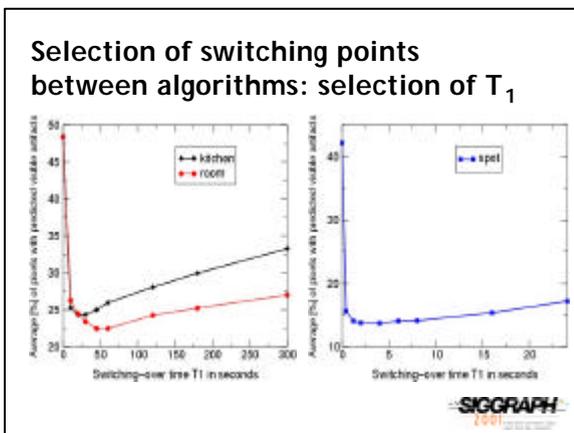
---

---

---

---

---




---

---

---

---

---

---

---

---

**Progressive rendering example (1)**



DEPT: 3 seconds



DEPT: 20 seconds



---

---

---

---

---

---

---

---

**Progressive rendering: example (2)**



DEPT: 20 s + DDL: 326s



Converged solution: 2 hours



---

---

---

---

---

---

---

---

**Rendering of high quality animation sequences**

<http://www.mpi-sb.mpg.de/resources/aqm>



---

---

---

---

---

---

---

---

### Motivation

- ⌘ In the traditional approach to rendering of high quality animation sequences every frame is considered separately. This precludes accounting for the visual sensitivity to temporal detail.
- ⌘ Our goal is to improve the performance of walkthrough sequences rendering by considering both the *spatial* and *temporal* aspects of *human perception*.
- ⌘ We want to focus computational efforts on those image details that can be readily perceived in the animated sequence.
- ⌘ Our goal is to improve efficiency of rendering of walkthrough sequences.




---

---

---

---

---

---

---

---

### Experimental findings on the human perception of animated sequences

- ⌘ The requirements imposed on the quality of still images must be higher than for images used in an animated sequence. The quality requirements can usually be relaxed as the velocity of the visual pattern increases.
- ⌘ The perceived sharpness of blurred visual patterns increases with their motion velocity, which is attributed to the higher level processing in the visual system.
- ⌘ The human eye is less sensitive to higher spatial frequencies than to lower frequencies.




---

---

---

---

---

---

---

---

### Video quality metrics

- ⌘ Virtually all state-of-the-art perception-based video quality metrics account for the discussed HVS characteristics.
- ⌘ A majority of the existing video quality metrics have been developed to evaluate performance of digital video coding and compression techniques, e.g., Lambrecht (1996), Lubin (1997), and Watson (1998).
- ⌘ The lack of comparative studies make it unclear which metric performs best.
- ⌘ We use our own metric that takes advantage of data readily available for synthetic images.




---

---

---

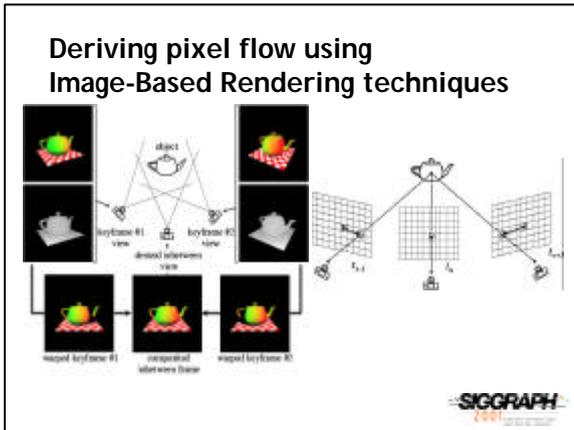
---

---

---

---

---




---

---

---

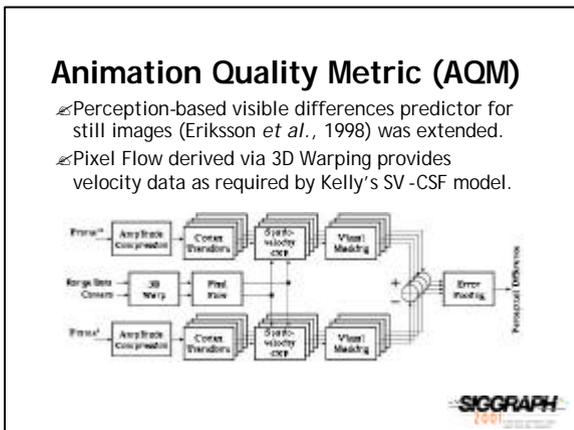
---

---

---

---

---




---

---

---

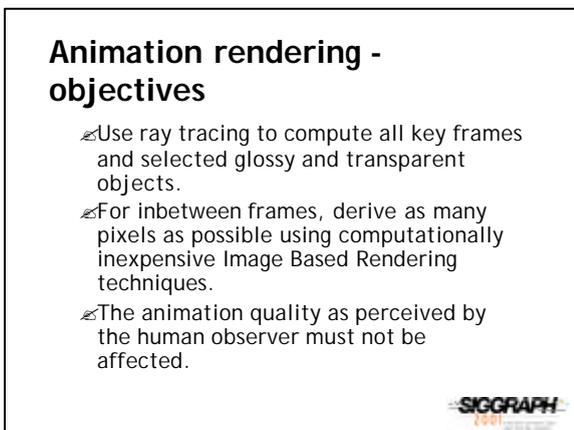
---

---

---

---

---




---

---

---

---

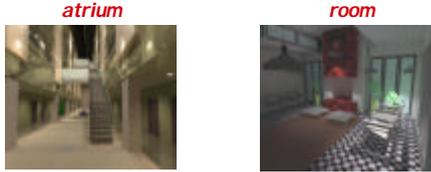
---

---

---

---

### Selected case study scenes



- ⌘ Interesting occlusion relationships between objects which are challenging for IBR.
- ⌘ Many specular objects for the *atrium* scene.
- ⌘ Animation path causing great variations of the pixel flow for the *room* scene.




---

---

---

---

---

---

---

---

### Keyframe placement

- ⌘ The selection of keyframes should be considered in the context of the inbetween frame computation technique.
- ⌘ In IBR techniques reference frames are usually placed:
  - ⌘ uniformly in space at the nodes of 2D or 3D grid (Chen95),
  - ⌘ uniformly along the animation path (Mark97),
  - ⌘ at manually selected locations (Darsa97).
- ⌘ A notable exception is work done by Nimeroff et al. 1996, who used a simple quality criterion.




---

---

---

---

---

---

---

---

### Keyframe placement - our approach

- ⌘ Our goal is to find inexpensive and automatic solution, which reduces animation artifacts which can be perceived by the human observer.
- ⌘ Our solution consists of two stages:
  - ⌘ initial keyframe placement which reduces the number of pixels which cannot be properly derived using IBR techniques due to occlusion problems,
  - ⌘ further refinement of keyframe placement which takes into account perceptual considerations, and is guided by AQM predictions.




---

---

---

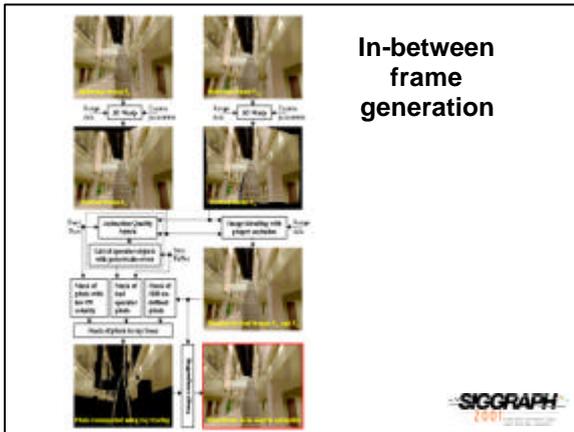
---

---

---

---

---




---

---

---

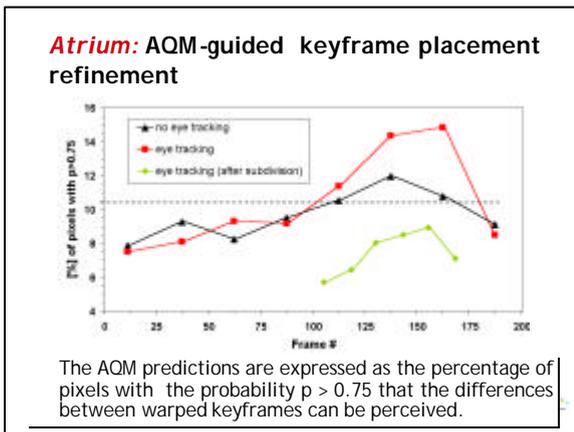
---

---

---

---

---




---

---

---

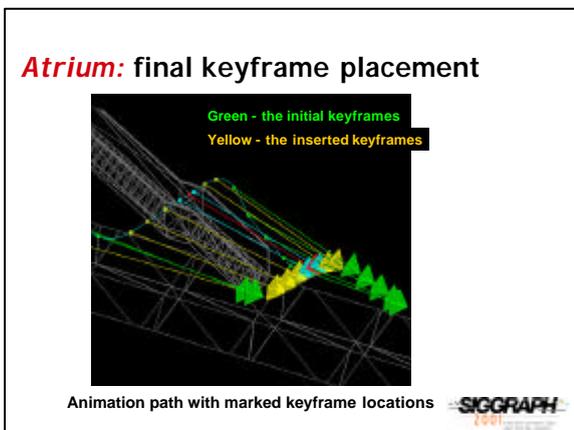
---

---

---

---

---




---

---

---

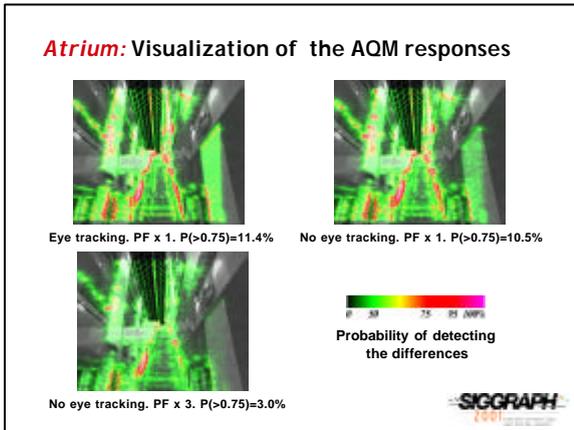
---

---

---

---

---




---

---

---

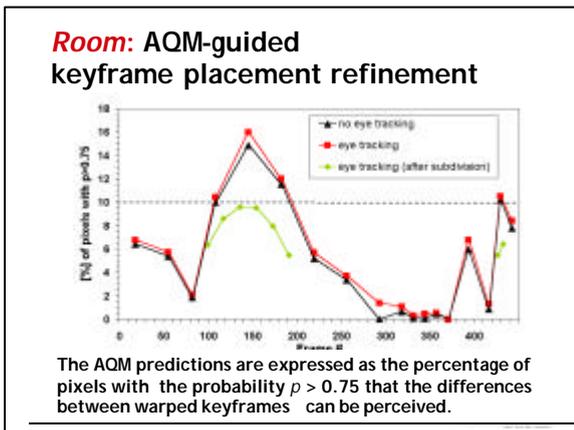
---

---

---

---

---




---

---

---

---

---

---

---

---

**Spatio-temporal antialiasing**

- ⌚ 3D low-pass filtering in the spatio-temporal domain is performed as a post-process on the complete animation sequence.
- ⌚ Motion-compensated filtering is performed in the temporal domain (this is another application of the Pixel Flow derived as a by-product of IBR computations).
- ⌚ To our experience, for moving visual patterns a single ray-traced sample per pixel is enough to produce an animation which is visually indistinguishable from its counterpart based on supersampled images.




---

---

---

---

---

---

---

---

### Examples of final frames

Supersampled frame used  
in traditional animations



Corresponding frame  
derived using our approach



*In both cases the perceived quality of animation appears to be similar!*




---

---

---

---

---

---

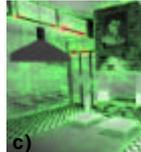
---

---

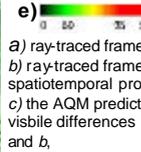
### Pixel Flow separation for occluding objects



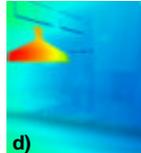
a)



b)



c)



d)



e)



f)

a) ray-traced frame,  
b) ray-traced frame with spatiotemporal processing,  
c) the AQM prediction of visible differences between a and b,  
d) the corresponding Pixel Flow velocity,  
e) color scales for the AQM response, and  
f) color scale for the Pixel Flow velocity [deg/sec].




---

---

---

---

---

---

---

---

### IBR-derived pixels to be ray traced

- ⌘ Pixels representing specular objects selected by the AQM predictions for recomputation.
- ⌘ Pixels with occlusion problems inherent to IBR techniques.
- ⌘ Pixels for slowly moving visual patterns, which are selected based on the Pixel Flow magnitude. The threshold velocity was found experimentally using subjective and objective (AQM) judgment of the resulting animation quality.




---

---

---

---

---

---

---

---

Experimental settings for estimating the upper threshold Pixel Flow velocity which is used to identify image regions that require ray-traced pixels to avoid degradation of the animation quality as perceived by the human observer.

---

---

---

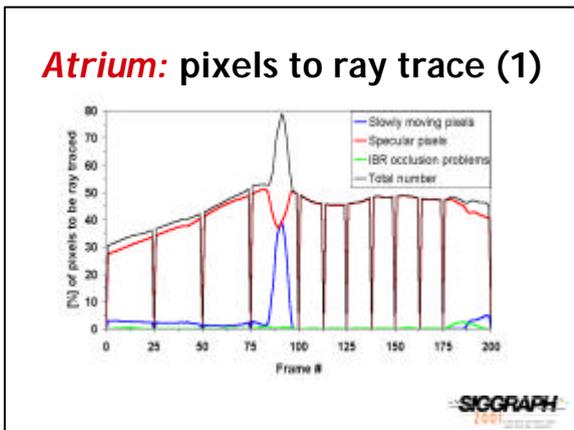
---

---

---

---

---




---

---

---

---

---

---

---

---

**Atrium: pixels to ray trace (2)**

- ⚡ The atrium scene is a very hard case for our technique because a majority of objects exhibits some glossy reflectance properties.
- ⚡ The percentage of pixels to be recalculated by ray tracing:

Specular pixels	40.8%
Slow motion	2.4%
IBR occlusions	0.3%
Keyframes	6.0%
-----	
Total	49.5%

---

---

---

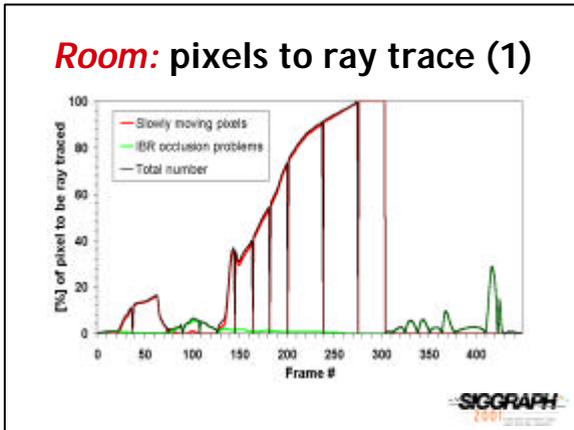
---

---

---

---

---




---

---

---

---

---

---

---

---

**Room: pixels to ray trace (2)**

- ✍ The room scene is a hard case for our technique because for many frames the camera motion is very slow which results in high sensitivity of the human visual system.
- ✍ The percentage of pixels to be recalculated by ray tracing:

Slow motion	28.1%
IBR occlusions	1.9%
Keyframes	5.1%
-----	
Total	35.1%

---

---

---

---

---

---

---

---

**Summary**

- ✍ We investigated the robustness of the VDP predictions in typical global illumination and rendering tasks.
- ✍ Based on good agreement of the VDP predictions with the human observer responses we used the VDP to tune parameters of global illumination techniques to improve their "perceptual" convergence.
- ✍ We proposed an Animation Quality Metric suitable for estimating quality of sequences of synthetic images.
- ✍ We developed a system for animation rendering featuring perception-based guidance of inbetween frames computation which reduces the rendering costs.

---

---

---

---

---

---

---

---

## Acknowledgments

☞The following colleagues and friends contributed to the presented work:

- ☞William Martens - the VDP validation experiments,
- ☞Vladimir Volevich, Andrei Khodulev and Edward Kopylov - development of the DEPT technique,
- ☞Przemyslaw Rokita and Takehiro Tawara - perception-based rendering and antialiasing of walkthrough sequences.



---

---

---

---

---

---

---

---

# Seeing is Believing: Reality Perception in Modeling, Rendering & Animation



## VII. Conclusion

Alan Chalmers  
Scott Daly  
Ann McNamara  
Karol Myszkowski  
Holly Rushmeier  
Tom Troscianko

Course #21



## Summary

- ✍ Course has shown
  - ✍ Image Quality Metrics are important
  - ✍ Number of subjective and objective approaches
  - ✍ Rich areas of application in both rendering and animation
  - ✍ Exciting topic for future research



---

---

---

---

---

---

---

---

## Related Talks

- ✍ A number of related events this week
  - ✍ Wednesday Morning: Human Factors Session
    - ✍ Katerina Mania
    - ✍ Carol OSullivan
  - ✍ Friday Morning : Models of Visual Representation Session
    - ✍ Kirsten Cater



---

---

---

---

---

---

---

---

# Seeing is Believing: Reality Perception in Modeling, Rendering & Animation



## Course Notes

Alan Chalmers  
Scott Daly  
Ann McNamara  
Karol Myszkowski  
Holly Rushmeier  
Tom Troscianko

Course #21



# Contents

<b>1</b>	<b>Illumination: Measurement &amp; Simulation</b>	<b>3</b>
1.1	Light and Materials . . . . .	4
1.1.1	Radiometry . . . . .	6
1.1.2	Photometry . . . . .	7
1.1.3	Characterising Surface Materials . . . . .	8
1.2	Illumination Models . . . . .	10
1.2.1	Raytracing . . . . .	11
1.2.2	Radiosity . . . . .	14
1.2.3	Radiance . . . . .	18
1.3	Summary . . . . .	21
<b>2</b>	<b>Visual Perception in Realistic Image Synthesis</b>	<b>23</b>
2.1	Visual Perception . . . . .	23
2.1.1	The Human Visual System . . . . .	23
2.1.2	Human Visual Perception . . . . .	27
2.1.3	Lightness Perception . . . . .	28
2.2	Perception and Graphics . . . . .	29
2.2.1	Perceptually driven rendering . . . . .	29
2.2.2	Perceptually Based Image Quality Metrics . . . . .	31
2.2.3	Comparing Real and Synthetic Scenes . . . . .	34
2.2.4	Tone Mapping . . . . .	36
2.2.5	Single Scale Tone Reproduction Operators . . . . .	38
2.2.6	Multi-Scale Tone Reproduction Operators . . . . .	40
2.3	Summary . . . . .	41
<b>3</b>	<b>A Psychophysical Investigation</b>	<b>43</b>
3.1	The Pilot Study . . . . .	43
3.1.1	Participants in the Pilot Study . . . . .	44
3.1.2	Apparatus . . . . .	44
3.1.3	The Graphical Representation . . . . .	45
3.1.4	Procedure . . . . .	46
3.2	Results and Discussion of Pilot Study . . . . .	47
3.2.1	Condition A: Real Versus Rendered . . . . .	49
3.2.2	Condition B: Real Versus 30% Blurred . . . . .	49
3.2.3	Condition C: Real Versus Partially Occluded . . . . .	50
3.2.4	Conclusion from Conditions of Pilot Experiment . . . . .	50
3.2.5	Compared to Numerical Techniques . . . . .	51
3.3	Conclusions and Problems with Pilot Study . . . . .	51
3.4	Experiment: The Physical Framework . . . . .	52
3.4.1	Ordering Effects . . . . .	52
3.4.2	Matching to Patches . . . . .	52
3.4.3	Custom Paints . . . . .	53
3.4.4	Three Dimensional Objects . . . . .	54
3.5	Modifications to the Original Apparatus . . . . .	54

3.6	Experimental Design . . . . .	55
3.6.1	Participants . . . . .	55
3.6.2	Randomisation and Counterbalancing . . . . .	55
3.7	Procedure . . . . .	56
3.8	Experiment . . . . .	56
3.8.1	Training on Munsell Chips . . . . .	57
3.8.2	Matching to Images . . . . .	58
3.8.3	The Graphical Representations . . . . .	58
3.8.4	Instructions . . . . .	60
3.9	Summary . . . . .	60

# Chapter 1

## Illumination: Measurement & Simulation

1

Since its conception, the pursuit of realistic image synthesis has been the creation of representative, high quality imagery [41, 38, 37, 86]. The production (rendering) of realistic imagery requires a precise treatment of lighting effects, which involves the simulation of physical phenomena including light emission, propagation, and reflection. To achieve this, the environment under consideration is first modelled as a collection of virtual lights, objects and a camera (or eye) point. *Physically-based* rendering algorithms [35, 3, 49, 81] focus on producing realistic images by simulating the light energy, or *radiance*, that is visible at every pixel of the image. The computed radiance values must then be mapped to values suitable for display on some display device [87, 50]. Figure 1.1 illustrates this process.

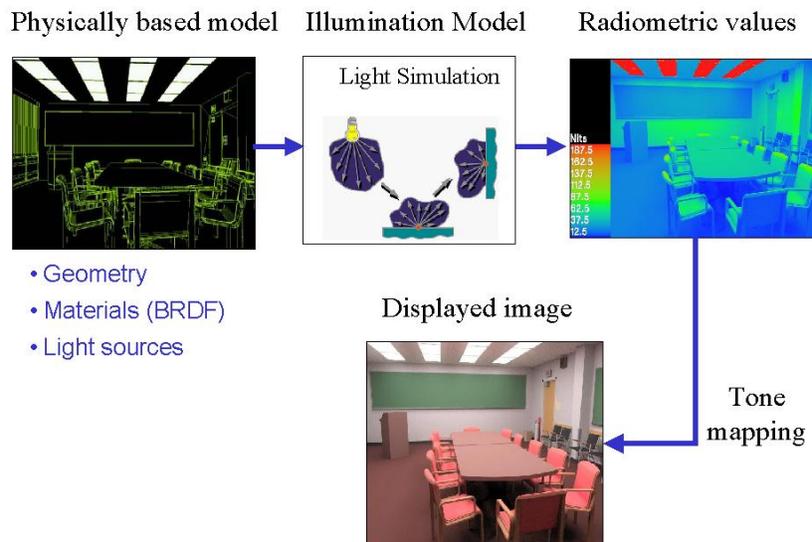


Figure 1.1: The rendering pipeline

Understanding the natural illumination process, and how to quantify illumination, provides the foundations for designing and controlling physically based image synthesis algorithms. A precise terminology exists to quantify illumination [12], from this terminology the underlying equations used to build the mathematical models for illumination simulation algorithms are derived.

<sup>1</sup>written by Ann McNamara ann.mcnamara@tcd.ie

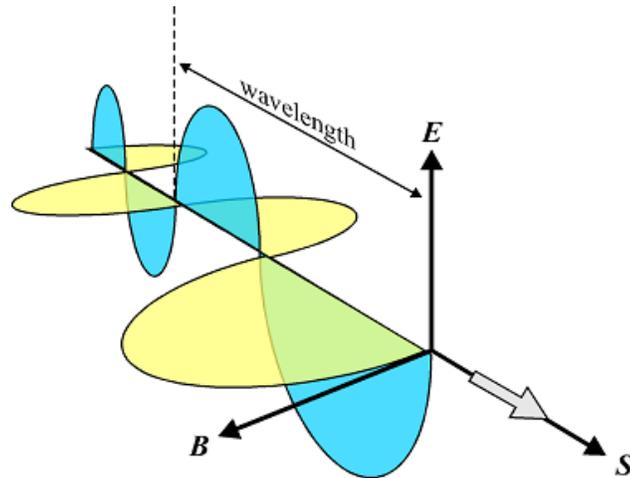


Figure 1.2: Mutually orthogonal E and B fields of an electromagnetic wave propagating in the x axis

## 1.1 Light and Materials

Light is a form of electro-magnetic energy comprising waves of coupled electric and magnetic fields perpendicular to each other and to the direction of propagation of the wave, Figure 1.2. The portion of light

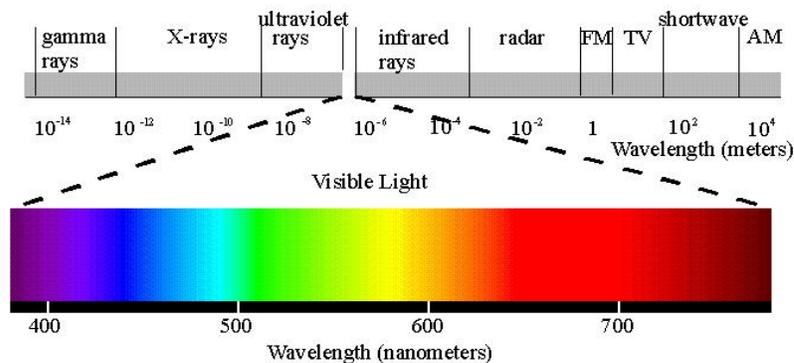


Figure 1.3: The visible portion of the electromagnetic spectrum

which can be seen by the human eye, *visible light*, is just a tiny fraction of the electromagnetic spectrum, which extends from very high frequency of radio waves through to low frequency microwaves, infra red and ultra violet light to x-ray and gamma rays. The range of visible light, which lies approximately between 380nm and 720nm, is shown in the context of the whole electromagnetic spectrum in Figure 1.3. The scenes which humans perceive are based on an integration over the visible spectrum of incoming radiation. The following definitions form an introduction to the measurement and perception of light. Most of the following definitions are taken from the Illumination Engineering Society Lighting Handbook, given by the IES [47].

**Illuminating hemisphere( $\Omega$ ):** The illuminating hemisphere, Figure 1.4, is a convenient notation to describe the illumination events above or below a surface. These events such as light sources, or other reflecting surfaces, are projected onto this hemisphere, which for convenience is usually of radius 1 (a unit hemisphere). Integrating over the hemisphere means considering all events above the surface weighted by the solid angles of their projections onto the hemisphere. Using this form, the illumination at a given point can be computed by considering **all** illumination events captured on the illumination hemisphere.

**Solid Angle:** The solid angle is the three-dimensional “angle” formed by the vertex of a cone. When this vertex is the centre of a sphere of radius  $r$  and the base of the cone cuts out an area  $A$  on the surface

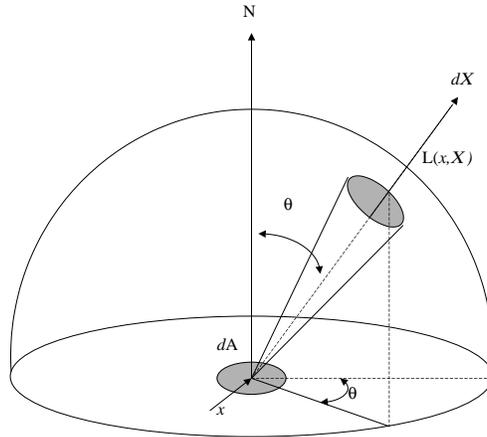


Figure 1.4: The illumination hemisphere

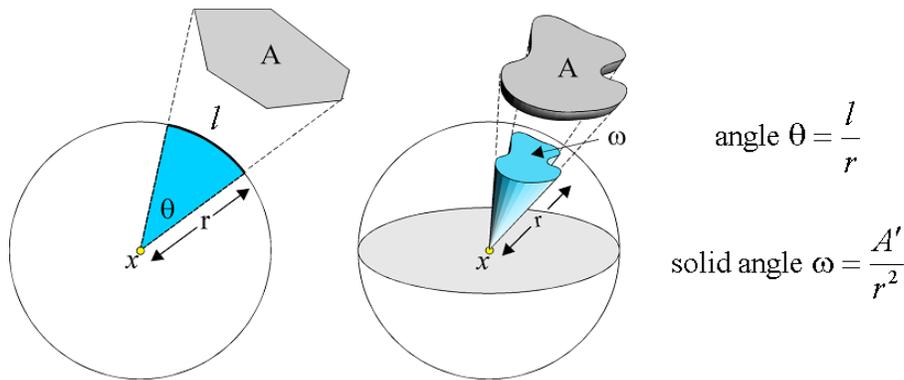


Figure 1.5: Solid Angles

of the sphere, the solid angle in steradians is defined as  $A/r^2$ . The solid angle of the entire sphere is  $4\pi sr$ , so solid angle of an entire hemisphere is  $2\pi sr$ . Figure 1.5 depicts the relationship between angle and solid angle.

**Projected Area:** This is the *apparent* area of an object as seen by an observer from a particular view direction. Projected area,  $dA_i$ , is the actual area  $dA$ , times the cosine of the angle,  $\theta$ , which is the angle between the surface normal and the view direction, Figure 1.6.

$$dA_i = dA \cos \theta$$

Clearly projected area varies according to viewing direction.

When simulating the propagation of light through an environment, two related methods of measuring and characterising light distributions are of interest to the computer graphics practitioner [46, 4]

**Radiometry** is the science of measuring radiant energy from any part of the electromagnetic spectrum. In general, the term usually applies to the measurement using optical instruments of light in the visible, infrared and ultraviolet wavelength regions. Radiometric terms and units have been standardised in the ANSI publication [44].

**Photometry** is the science of measuring light *within* the visible portion of the electromagnetic spectrum in units *weighted* in accordance with the sensitivity of the human visual system [92]. Photometry

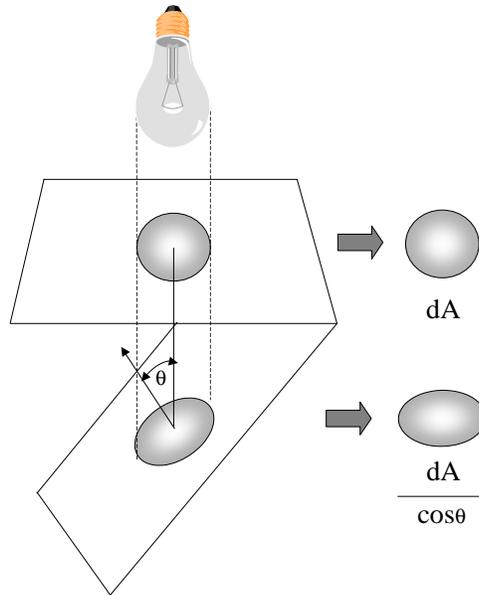


Figure 1.6: The greater the angle, the greater the area over which light is distributed, so energy at a given point will be proportionally less

deals with perceptual issues; if a surface radiates a given amount of energy, then how bright does that surface *appear* to an average viewer? In 1924 the *Commission Internationale d’Eclairage* or CIE, attempted to standardise the *luminous efficiency* of the human visual system by performing empirical tests with over one hundred observers [12].

### 1.1.1 Radiometry

Radiometry is the science of measuring radiant energy, in any portion of the electromagnetic spectrum. As light is a form of radiometric energy, radiometry is used in graphics to provide the basis for illumination calculations.

**Radiant Energy(Q):** *measured in Joules (J).* Light is radiant energy, photons of a certain frequency have a specific quantum of energy, defined by  $E = hf$ , where  $h$  is Planck’s Constant <sup>2</sup> and  $f$  is the frequency.

**Radiant Flux( $\Phi$ ):** *Measured in Watts(W).* This is simply the radiant energy flowing through an area per unit time,  $dQ/dt$ .

**Radiant Flux Density ( $d\Phi/dA$ ):** *Measured in Watts per square metre ( $Wm^{-2}$ ).* The quotient of the radiant flux incident on or emitted by a surface element surrounding the point and area of the element. *Emittance* is radiant flux density emitted from a surface, and *irradiance* is the term for radiant flux density incident on a surface.

**Radiant Exitance(M):** *Watts per square metre ( $Wm^{-2}$ ).* The radiant flux leaving the surface per unit area of the surface. (emittance)

**Irradiance(E):** *Measured in Watts per square metre ( $Wm^{-2}$ ).* The radiant flux incident on the receiver per unit area of the receiver.

<sup>2</sup>Planck discovered that light energy is carried by photons, he found that the energy of a photon is equal to the frequency of its electromagnetic wave multiplied by a constant,  $h$ , or Planck’s Constant which is equal to  $6.626 \times 10^{-20}$  Joules per second

**Radiant Intensity(I):** *Measured in watts per steradian ( $Wsr^{-1}$ ).* Radiant Intensity represents the radiant flux from a point source in a particular direction. Thus it is the flux per unit solid angle.  $d\Phi/d\omega$ .

**Radiance(L):** *Measured in watts per steradian per metre squared ( $Wsr^{-1}m^{-2}$ ).* Radiance is radiant flux arriving at or leaving from a surface, per unit solid angle per unit projected area. It is defined as  $L = d^{-2}/(\cos\theta dA d\omega)$  for a given direction  $\Theta$ . Radiance does not attenuate with distance. It is the quantity to which most light receivers, including the human eye, are sensitive.

### 1.1.2 Photometry

Photometry is the science of measuring light within the visible portion of the electromagnetic spectrum in units that are weighted according to the sensitivity of the human eye. It is a quantitative science based on a statistical model of the human visual response to light. Photometry attempts to measure the subjective impression produced by stimulating the human visual system with radiant energy. This is a complex task, nevertheless the subjective impression of a scene can be quantified for “normal” viewing conditions. In 1924, CIE asked over one hundred observers to visually match the brightness of monochromatic light sources with different wavelengths, under controlled conditions. The results from those experiments show the *Photopic Luminous Efficiency Curve* of the Human Visual System as a function of wavelength. It provides a weighting function that can be used to convert radiometric units into photometric measurements. Radiant flux is a physical quantity, whereas the *light* due to radiant flux is not, the amount of light is dependent on the ability of the radiation to stimulate the eye. The conversion of radiant flux to light involves a factor that depends on the physiological and psychological processes of seeing. Photometric terms are

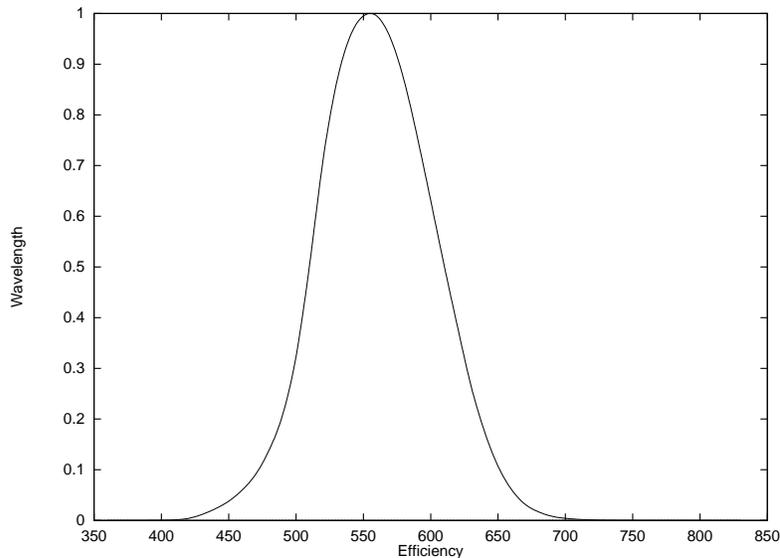


Figure 1.7: Luminous Efficiency Curve

equivalent to radiometric terms weighted by  $V(\lambda)$ , the *photopic spectral luminous efficiency* curve, shown in Figure 1.7. Radiation outside the visible spectrum does not play a role in photometry. The photopic quantities relevant to computer graphics imagery are the following:

**Light:** Light is radiant energy, evaluated according to its capacity to produce a visual sensation.

**Luminous Flux ( $\Phi_v$ ):** *Measured in Lumens.* The rate of flow of light with respect to time. The lumen is defined as the luminous flux of monochromatic radiation of wavelength of  $555nm$  whose radiant flux is  $(1/683)W$ . As this wavelength generates the maximal sensation in the eye, larger radiant flux at other visible wavelengths will correspond to 1 *lumen* of luminous flux. The quantity can be expressed as a factor  $f$  times  $(1/683)W$  where  $f$  is the reciprocal of the sensitivity of the corresponding wavelength, relative to the the sensitivity of  $555nm$ .

**Luminous Factor or Luminous Efficacy:** *Measured in lumen/watt.* The sensitivity of the human eye to the visible wavelengths is expressed by *luminous efficacy*. Luminous efficacy of a particular wavelength is the ratio of the luminous flux at that wavelength to the corresponding radiant flux.

**Luminous Intensity:** *Measured in candelas.* Luminous intensity,  $I_v$ , is the solid angular flux density of a point light source in a particular direction,  $\frac{d\Phi}{d\omega}$ . The *candela* is the unit of luminous intensity, one candela is one lumen per steradian. Since the total solid angle about a point is  $4\pi$  steradians it follows that a point source having a uniform intensity of 1 *candela* has a luminous flux of  $4\pi$  *lumens*.

**Illuminance ( $E_v$ ):** *Measured in Lux.* Illuminance,  $E_v$ , or *illumination*, is the area density of the luminous flux incident on a surface  $\frac{d\Phi}{dA}$ .

**Luminous Exitance(M):** *Measured in Lux/Nit.* Luminous exitance,  $M$ , is the area density of luminous flux leaving a surface at a point. This is the total luminous flux emitted, reflected and transmitted from a surface independent of direction.

**Luminance( $L_v$ ):** *Measured in Candelas per square metre.* Luminance,  $L_v$ , is the radiometric equivalent of radiance and is hence a useful quantity to represent directional luminous flux for an area light source. *Luminance*,  $L_v$ , along a direction  $(\theta, \phi)$ , is the luminous flux per projected surface area per unit solid angle centred around that direction.

Physics	Radiometry	Radiometric Units
	Radiant Energy	joules [ $J = kgm^2/s^2$ ]
Flux	Radiant Power	watts [ $W = joules/s$ ]
Angular Flux Density	Radiance	$[W/m^2sr]$
Flux Density	Irradiance	$[W/m^2]$
Flux Density	Radiosity	$[W/m^2]$
	Radiant Intensity	$[W/sr]$
Physics	Photometry	Photometric Units
	Luminous Energy	talbot
Flux	Luminous Power	lumens [ $talbots/second$ ]
Angular Flux Density	Luminance	Nit [ $lumens/m^2sr$ ]
Flux Density	Illuminance	Lux [ $lumens/m^2sr$ ]
Flux Density	Luminosity	Lux [ $lumens/m^2sr$ ]
	Luminous Intensity	Candela [ $lumens/sr$ ]

Table 1.1: Radiometric and Photometric Quantities

Radiometric and corresponding photometric quantities are summarised in Table 1.1, along with their units.

### 1.1.3 Characterising Surface Materials

The next key problem to be addressed in the simulation of light distribution involves characterising the reflections of light from surfaces. Various materials reflect light in different ways, for example a matt house paint reflects light very differently than the often highly specular paint often used on sports cars. *Reflection* is the process whereby light of a specific wavelength is (at least partially) propagated outward by a material without change in wavelength, or more precisely, “*reflection is the process by which electromagnetic flux (power), incident on a stationary surface or medium, leaves that surface or medium from the incident side without change in frequency; reflectance is the fraction of the incident flux that is reflected*” [69].

The effect of reflection depends on the directional properties of the surface involved. The reflective behaviour of a surface is described by its *Bi-Directional Reflectance Distribution Function(BRDF)*. The

BRDF expresses the probability that the light coming from a given direction will be reflected in another direction [16, 33]. Hence, the BRDF is the ratio of outgoing intensity to incoming energy, Figure 1.8. Generally we define BRDF as:

$$R_{bd}(\lambda_i, \phi_i, \theta_i, \phi_r, \theta_r)$$

where  $\lambda_i$  = wavelength of the incoming light  
 $\phi_i, \theta_i$  = incoming direction  
 $\phi_r, \theta_r$  = outgoing direction

This relates the light in the direction  $(\phi_i, \theta_i)$  to outgoing light in the direction  $(\phi_r, \theta_r)$ . BRDF is a function of wavelength.

$$R_{bd}(\lambda_i, \phi_i, \theta_i, \phi_r, \theta_r) = \frac{I_r(\phi_r, \theta_r, \phi_i, \theta_i)}{E_i(\phi_i, \theta_i)}$$

Incoming energy is related to outgoing intensity by

$$E_i(\phi_i, \theta_i) = I_i(\theta_i, \phi_i) \cos \theta_i d\omega$$

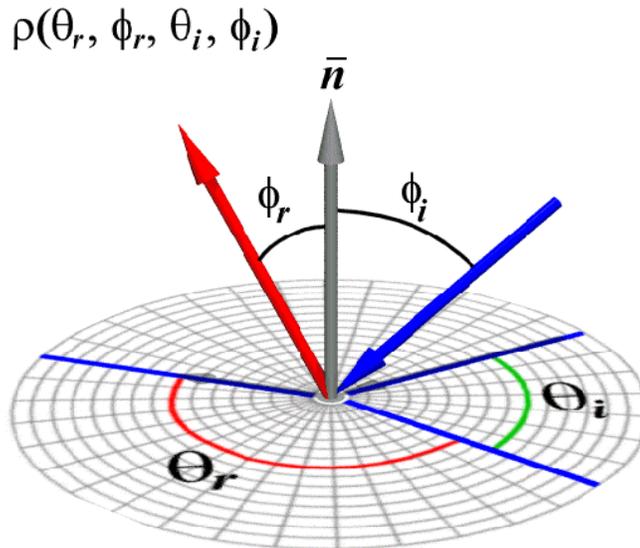


Figure 1.8: Geometry of the BRDF

Figure 1.9 shows different types of material behaviour, which are defined as follows [33]:

**Specular (mirror):** Specular materials reflect light in one direction only, the mirror direction. The outgoing direction is in the incident plane and the angle of reflection is equal to the angle of incidence.

**Diffuse:** Diffuse, or Lambertian materials reflect light equally in all directions. Reflection of light from a diffuse surface is independent of incoming direction. The reflected light is the same in all directions and does not change with viewing angle.

**Mixed:** Reflection is a combination of specular and diffuse reflection. Overall reflectance is given by a weighted combination of diffuse and specular components.

**Retro-Reflection:** Retro-Reflection occurs when the light is reflected back on itself, that is the outgoing direction is equal, or close to the incident direction. Retro-reflective devices are widely used in the areas of nighttime transportation and safety.

**Gloss:** Glossy materials exhibit the property that involves mixed reflection and is responsible for a mirror like appearance of a rough surface.

Most materials do not fall exactly into one of the idealised material categories described above, but instead exhibit a combination of specular and diffuse characteristics. Real materials generally have a more complex behaviour, with a directional character resulting from surface finish and sub-surface scattering.

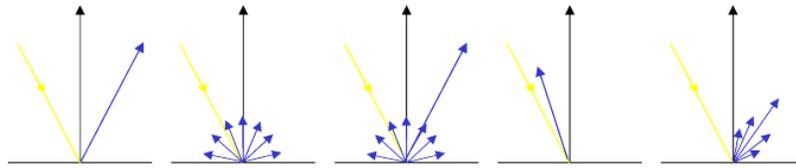


Figure 1.9: Types of reflection, from left to right, specular, diffuse, mixed, retro-reflection, gloss

## 1.2 Illumination Models

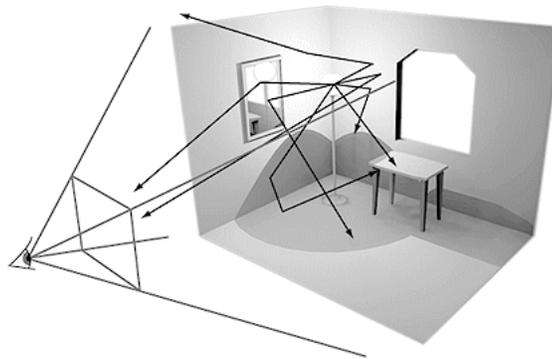


Figure 1.10: Light behaviour in an environment

The purpose of an *illumination model* is to model the distribution of light in an environment. Typically this is achieved by using the laws of physics to compute the trajectory of light energy through the scene being modelled. *Local illumination models* calculate the distribution of reflected light as a function of the incoming energy from the light source(s). Local is used to emphasise the fact that the illumination of a surface is determined solely by the characteristics of the surface itself and those of the light source. The Phong illumination model [74] was one of the earliest local reflection models in computer graphics. Light interaction is considered as reflecting in terms of three separate components, a diffuse, a specular and an ambient term. The linear combination of these three can then be used to model light intensity at each point on a surface (or at certain points on a surface, then the appearance of the entire surface can be calculated using interpolation of the values at these points).

$$I = I_a I_k + \sum I_i [k_d \cos \theta + k_s \cos^n \phi]$$

where  $I$ , the intensity leaving a point, is calculated as the accumulation of contributions from  $N$  light sources, each of intensity  $I_i$ . The wavelength dependent diffuse reflectivity,  $k_d$ , gives the *diffuse* term. This is the fraction of light scattered in all directions. The specular coefficient,  $k_s$  is used to model light reflected in the mirror direction. If a surface faces away from the light source it will not receive any light, hence will appear black. In reality, direct light and reflected light combine to give the illumination of each surface in an environment, so such surfaces would receive light indirectly via interreflections from other surfaces, to account for this, local illumination models include a constant *ambient term*,  $I_a k_a$ .

However, the interreflection of light can account for a high proportion of the total illumination in a scene. This is especially true for indoor scenes where light cannot “escape” the scene but instead is always

reflected back into the scene by some surface, as in Figure 1.10 [5]. To account for such interreflections, all objects must be considered a potential source of illumination for all other objects in the scene. This constitutes a *global illumination model*. Global illumination models endeavour to include all of the light interaction in a scene, to give rise to effects such as indirect illumination, soft shadows and colour bleeding, all of which influence the perception of the resulting imagery, and hence the quality of the image. The complexities of the interaction of light and surfaces in an environment can be neatly described in a compact form by the *rendering equation* [45]:

$$L_r(\theta_r, \phi_r) = L_e + \int L_i(\theta_i, \phi_i) f_r(\theta_i, \phi_i, \theta_r, \phi_r) |\cos \theta_i| \sin \theta_i d\theta_i d\phi_i$$

where:

$$\begin{aligned} L_e &= \text{emitted radiation} \\ \int L_i(\theta_i, \phi_i) &= \text{The incoming luminance distribution} \\ f_r(\theta_i, \phi_i, \theta_r, \phi_r) &= \text{BRDF} \end{aligned}$$

The problem of global illumination can be seen as solving the rendering equation for each point in an environment. The rendering equation is a complex integral equation<sup>3</sup>. In all but the simplest case, there is no closed form solution for such an equation so it must be solved using numerical techniques. Numerical techniques imply approximation. For this reason most illumination computations are approximate solutions to the rendering equation.

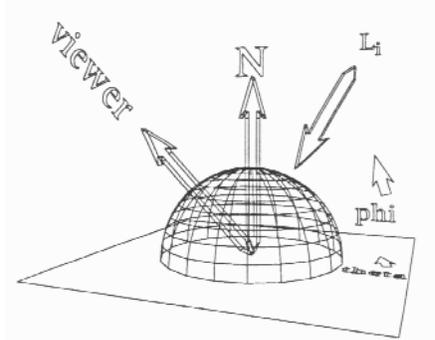


Figure 1.11: Illustration of the Rendering Equation which determines radiance by summing self emitting radiance and reflected radiance.

## 1.2.1 Raytracing

Raytracing is a versatile technique for computing images by tracing individual paths of light through a scene. Raytracing algorithms attempt to capture *view-dependent* specular effects as well as reflections and transmissions [2, 91]. Raytracing unifies the processes of hidden surface removal, shading, reflection, refraction and shadowing. In raytracing, it is recognised that although millions of photons travel through an environment, only those photons striking the eye are needed to compute the image. Hence, raytracing proceeds by tracing a number of rays starting at the eye point or camera into the scene, this way only the necessary information is computed. The disadvantage of this is that the result of raytracing is a single image, making it a view-dependent technique. Initially one ray is passed through (the centre of) each pixel, this is called *the primary ray*. Each primary ray is tested for intersection with all objects in the scene to determine the object closest to the eye. A *shadow ray* is then traced toward each light source in the scene. If this ray does not intersect any other objects, that is there is a clear path from the point of intersection to the light source, then a local illumination model is applied to determine the contribution of the light source(s) to that surface point. If the light source(s) is occluded then the point under consideration is in shadow.

In the case of reflective or transparent surfaces, the direction in which light arrives by reflection or transmission is also needed. Reflected rays are easily computed since the angle of reflection is equal to the

<sup>3</sup>The rendering equation is a linear inhomogeneous Fredholm integral equation of the second kind, which exhibits a recursive nature making it difficult to evaluate.

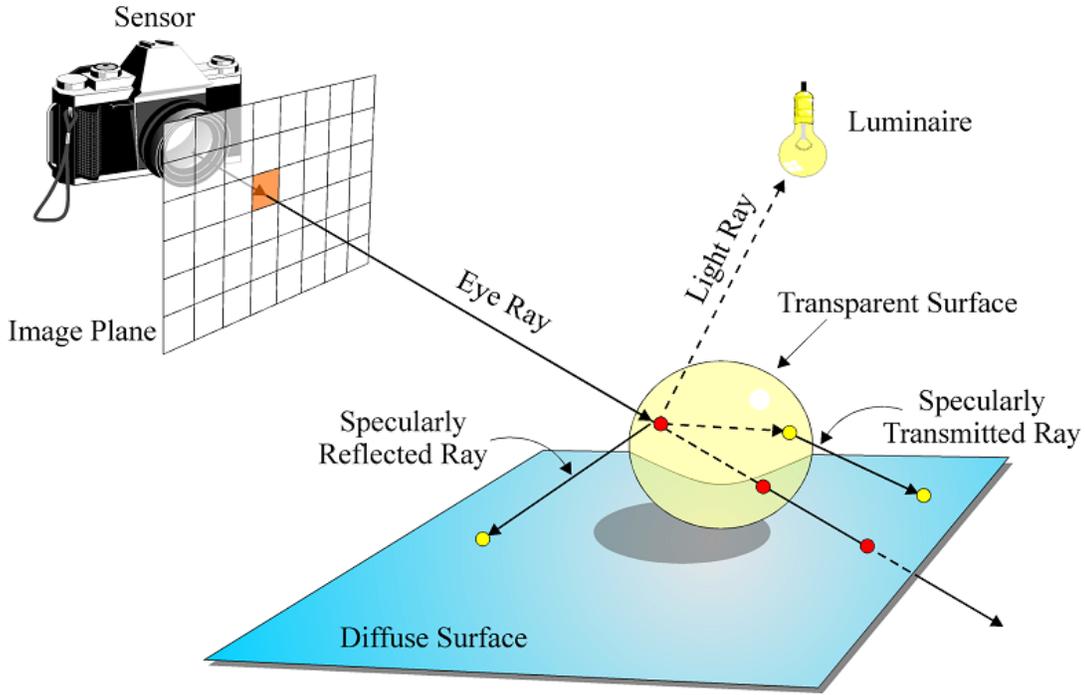


Figure 1.12: Raytracing: Rays are traced from the eye into the scene in an attempt to capture specular reflection, transparency effects and shadowing

angle of incidence, Figure 1.13. Transmitted rays are computed according to *Snell's Law*, which describes the relationship between the angle of incidence,  $\theta_i$ , and the angle of transmission,  $\theta_t$ :

$$\frac{\sin \theta_i}{\sin \theta_t} = \frac{\eta_i}{\eta_t}$$

where  $\eta_i$  and  $\eta_t$  are the *indices of refraction* of the materials through which the ray travels. Snell's Law states that the ratio of the sine of the angle of the incident ray in one medium to the sine of the angle it makes in another medium is constant, Figure 1.13.

A *recursive* evaluation must be employed, at each surface, Figure 1.14 [5]. By recursively tracing rays through the scene, until no further objects are encountered or some maximum number of levels has been reached, colour contributions for each pixel are calculated. A weakness of raytracing is the manner in which diffuse interreflections are handled. Surfaces receiving no direct illumination appear black. To overcome this an indirect illumination term, referred to as *ambient light*, is accounted for by a constant *ambient term*, which is usually assigned an arbitrary value. Figure 2.15 gives pseudo code to illustrate the recursive raytracing procedure.

Raytracing can model a large range of lighting effects accurately accounting for the global illumination characteristics of direct illumination, shadows, specular reflection and transparency. The main drawback of raytracing is that it can prove to be computationally expensive and time consuming, even for moderate environments. Intersection tests dominate the cost of raytracing algorithms. Typically in raytracing several intersections per pixel are computed. Performing intersection tests with *all* objects in an environment is inefficient. Several algorithms, such as *spatial subdivision* [23, 32], have been developed which attempt to minimise the number of ray object intersections. By enclosing a scene in a cube, that cube can be successively subdivided until each sub-region (voxel or cell) contains no more than a preset maximum number of objects. This subdivision can then be stored in an *octree* to establish a hierarchical description of the occupancy of voxels. Subdivision can be *uniform*, the cube is divided into eight equal sized octants at each step, or *adaptive* where only regions of the cube containing objects are subdivided. Using such a framework allows spatial coherence to be exploited. Rays are traced through individual voxels, with intersection tests performed only for the objects contained within. The ray is then processed through the voxels by determining the entry and exit points for each voxel traversed by the ray until an object is intersected or the scene

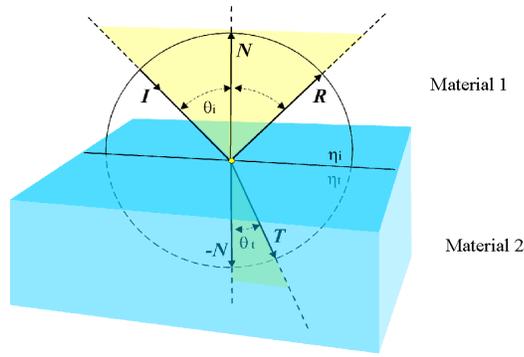


Figure 1.13: Behaviour of light ray incident on a surface

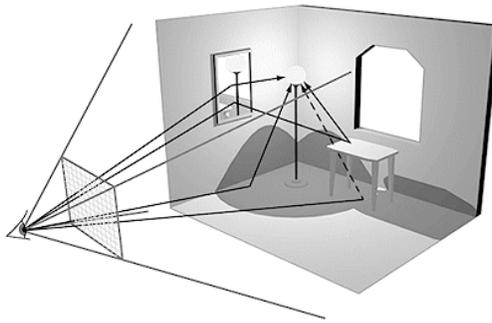


Figure 1.14: Raytracing

boundary is reached. In a *Spatially Enumerated Auxiliary Data Structure (SEADS)* space is divided into equally sized voxels regardless of object position, resulting in more voxels than an octree division. Using this strategy many rays can be traced with increased speed from region to region using a 3D-DDA (Digital Differential Analyser), speed can be further augmented by implementing this in hardware.

*Aliasing* effects, (Figure 1.16), occur when attempting to represent a continuous phenomena (radiance) with discrete samples (pixel values). *Spatial aliasing* effects appear as a consequence of the spatial resolution of the pixels in the image plane. Figure 1.16 illustrates this concept, attempting to represent a curved surface on a square grid, the resulting “blockiness” is referred to as aliasing, or “jaggies”. Due to the digital nature of computers, it is not possible to completely eliminate aliasing. Fortunately, many *anti-aliasing* techniques exist to *minimise* the effect. *Supersampling* takes the average radiance produced by shooting several rays through each pixel, this reduces aliasing but increases the cost of raytracing. An alternative is to use *adaptive sampling* focusing extra rays where they are required. Initially a low number of rays are traced per pixel, only if there are sufficient differences in the values returned are subsequent rays traced for that pixel.

In traditional raytracing only one ray is traced in each of the shadow and reflection directions. As a result the images generated often contain unnaturally sharp shadows and sharp mirror reflections. *Distributed Raytracing* [17, 18] extends classical recursive raytracing to include stochastic methods to simulate an array of optical effects including gloss, translucency, shadow penumbrae, depth of field and motion blur. This is achieved by distributing rays over several domains (pixel positions, lens position, area sampling position etc). In distributed raytracing several shadow or reflection rays are cast, each in a slightly different direction and the result is averaged over the number of rays cast.

Further details of the raytracing method can be found in [33, 22, 34].

```

For each pixel, p, in an image
Set I = ray starting at eye through pixel p
rad = Trace(I);
DrawPixel(p, I);
float Trace(Ray I){
    radiance = 0;
    Intersect I with all objects in the scene
    to determine o, the closest object.
    Compute P, the point of intersection of I with o
    DO LOCAL SHADING
    for each light source in the scene {
        trace a ray from P to L;
        If L is visible at P
            radiance += LocalShade(L, P);
        else
            P is in shadow, do nothing;
    }
}
DO GLOBAL SHADING
ReflectedRay
TransmittedRay
return(radiance);

```

Figure 1.15: Pseudo Code for the Raytracing approach

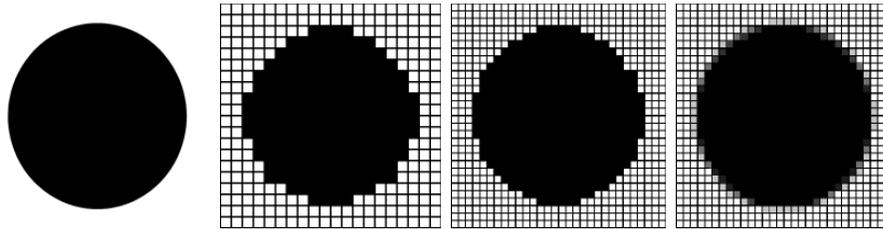


Figure 1.16: Antialiasing: a) A circle b) Strongly aliased circle c) Aliased Circle at Higher Resolution d) Antialiased Circle

## 1.2.2 Radiosity

Radiosity methods [35, 70, 15] attempt to capture *view-independent* diffuse interreflections in a scene, Figure 1.17 [5]. Techniques originally developed to compute the radiant interchange between surfaces, were first applied to the global illumination problem in the mid 1980s. *Radiosity*<sup>4</sup> methods are applicable to solving for the interreflection of light between ideal (Lambertian) diffuse surfaces. Radiosity assumes ideal diffuse reflections. The algorithm achieves global illumination by explicitly creating a global system of equations to capture interreflections of light in a scene and automatically accounting for the effects of multiple reflections. To accomplish this the surfaces of a scene are first divided into a *mesh of patches*. The radiance of these patches is computed by solving a system of equations, Figure 1.18 [9]. The result of a radiosity solution is not just a single image but a full three dimensional representation of the distribution of light energy in an environment, making it a view independent method.

The amount of light leaving each patch can be expressed as a combination of its emitted light and its reflected light.

<sup>4</sup>The term radiosity refers to a measure of radiant energy, specifically the energy leaving a surface per unit area per unit time. Now, radiosity has also come to mean a set of computational techniques for computing global illumination.

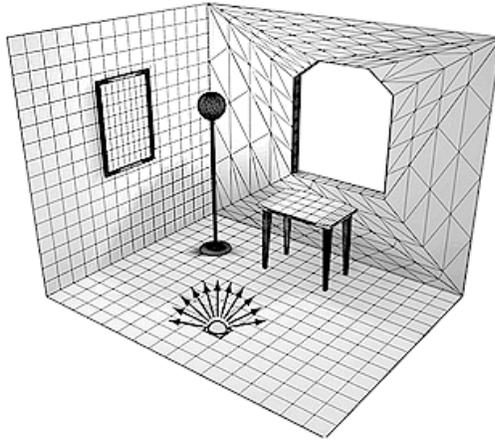


Figure 1.17: Radiosity



Figure 1.18: Radiosity: An image on the left, meshed representation on the right

$$B_i = E_i + \rho_i \sum_{j=1}^n F_{ij} B_j$$

$B_i$  is the exitance radiosity of patch  $i$ . (impident energy per unit area per unit time per unit surface)

$E_i$  is the radiosity emitted from patch  $i$ . (energy per unit area per unit time)

$F_{ij}$  is the form factor from  $i$  to  $j$ , the fraction of energy leaving patch  $i$  that arrives at patch  $j$ .

$\rho_i$  is the reffectivity of patch  $i$ .

$n$  is the number of patches in the environment.

The *form-factor*, Figure 1.19,  $F_{ij}$  is the fraction of energy transferred from patch  $i$  to patch  $j$ . The *reciprocity relationship* [80] states:

$$A_j F_{ji} = A_i F_{ij}$$

For all patches in a scene we get a *linear system of equations*:

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

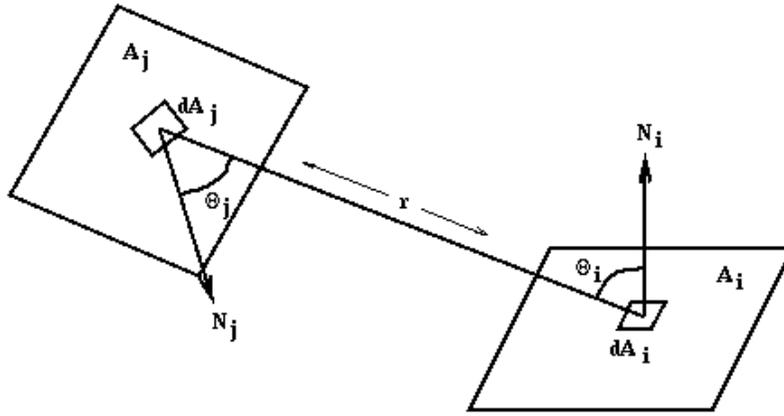


Figure 1.19: The relationship between two patches

A patch can contribute to its own reflected energy (in the case of concave objects) so this must be taken into account; so in general, terms along the diagonal are not merely 1. Due to the wavelength dependency of the  $\rho_i$  and  $E_i$  the matrix must be solved for each band of wavelengths to be considered, in computer graphics this usually includes a band for each of red, green and blue wave bands. However, the form factors are solely dependent on geometry, and not wavelength dependent and so do not need to be recomputed if the lighting or surface reflectivity changes.

This system of equations can be solved for the radiosity values by using iterative methods, for example Gauss-Seidel Iteration. Once the values for each pass have been obtained then the values at the vertices of the patches are calculated and the patches can then be passed to a standard polygon rendering pipeline that implements Gouraud shading. The value at a vertex can be calculated by averaging the radiosity values of the surrounding patches [36].

*Form Factor Computation* The form-factor, from differential area  $dA_i$  to differential area  $dA_j$  is:

$$dF_{di-dj} = \frac{\cos \theta_i \cos \theta_j}{\pi r^2} H_{ij} dA_j$$

As shown in Figure 1.19, for the ray between differential areas  $dA_i$  and  $dA_j$ ;  $\theta_i$  is the angle between the ray and the surface normal of  $A_i$ ,  $\theta_j$  is the angle between the ray and the surface normal of  $A_j$ ,  $r$  is the length of the ray,  $H_{ij}$  takes the value of 1 or 0 depending on whether or not  $dA_i$  is visible from  $dA_j$ . To calculate the form factor,  $F_{ij}$  from differential area  $dA_i$  to finite area  $A_j$  integrate over the area of patch  $j$ :

$$F_{dij} = \int \frac{\cos \theta_i \cos \theta_j}{\pi r^2} H_{ij} dA_j dA_i$$

So the form-factor from  $A_i$  to  $A_j$  is computed as the area average of the above equation over patch  $i$ :

$$F_{ij} = \frac{1}{A_i} \int_{A_i} \int_{A_j} \frac{\cos \theta_i \cos \theta_j}{\pi r^2} H_{ij} dA_j dA_i$$

By assuming that the centre of a patch typifies other points on that patch, then  $F_{ij}$  can be approximated by  $F_{dij}$  calculated for  $dA_i$  at the centre of patch  $i$ .

An equivalent to computing form-factors, Nusselt [71] projected parts of  $A_j$  visible from  $dA_i$  onto a unit hemisphere, this projected area is then projected orthographically down onto the hemisphere's unit circle base, then dividing by the area of the circle, Figure 1.20. Projecting onto the unit hemisphere accounts for  $\cos \theta_j / r^2$ , the projection to the base accounts for the multiplication by  $\cos \theta_i$ , and dividing by the area of the base accounts for the division by  $\pi$ .

An alternative algorithm, proposed by Cohen and Greenberg [15], projects onto the upper half of a cube, *hemicube*, centred about  $dA_i$ , with the cube's top parallel to the surface, Figure 1.21. The hemicube

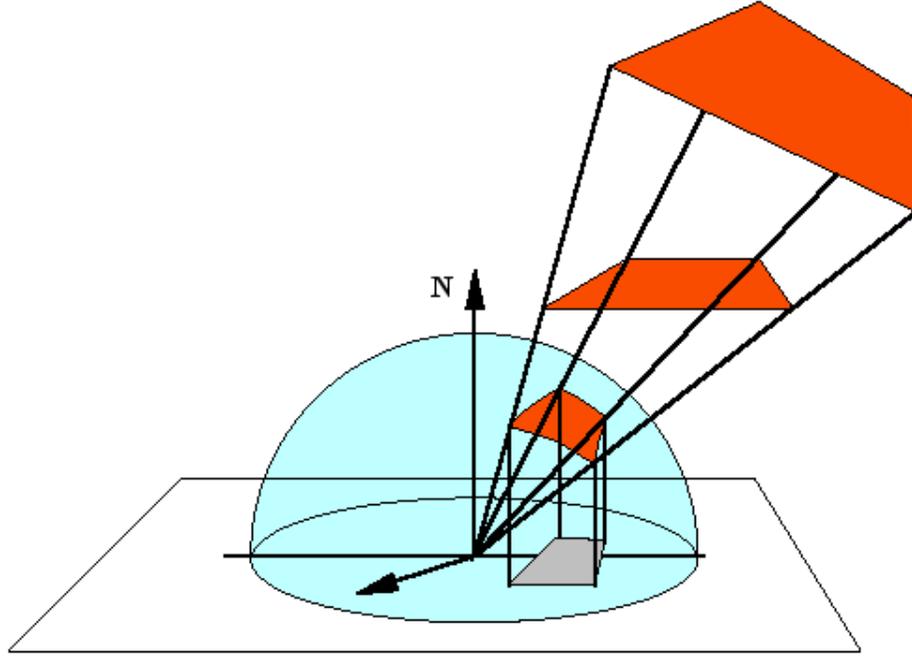


Figure 1.20: The Nusselt Analog

is divided into a uniform grid. All patches in the environment are clipped to the view-volume frusta defined by the centre of the cube and each of its five faces, then each of the clipped patches is projected onto the appropriate face of the hemispherical dome.

Each cell,  $p$ , which Cohen [15] refers to as a pixel, of the hemispherical dome has a precomputed *delta form factor* associated with it:

$$\Delta F_p = \frac{\cos \theta_i \cos \theta_j}{\pi r^2} \Delta A$$

$\theta_p$  is the angle between the surface normal of cell  $p$  and the vector between  $dA_i$  and  $p$ ,  $r$  is the length of the vector. Assigning the hemispherical dome a  $(x, y, z)$  co-ordinate system, with the origin at the centre of the bottom face, then for the top face:

$$r = \sqrt{x_p^2 + y_p^2 + 1}$$

$$\cos \theta_i = \cos \theta_p = \frac{1}{r}$$

$x_p$  and  $y_p$  are the co-ordinates of the hemispherical dome.

The approximate form factor,  $F_{dij}$  for any patch  $j$  can be found by summing the values of  $\Delta F_p$  associated with each cell  $p$  in  $A_j$ 's hemispherical dome projections. The values of  $\Delta F_p$  for *all* the hemispherical dome cells sum to 1. Assuming that the distance between the patches is large relative to the size of the patch, these values for  $F_{dij}$  can be used as the values of  $F_{ij}$  to compute patch radiosities.

The full matrix algorithm solves each  $B_i$  value one at a time by “gathering” light contributions from all other patches in the scene. One of the disadvantages of this method is only *after* all radiosities have been computed is the resultant image displayed. For complex environments the time taken to produce a solution can be extensive. This means that the user is unable to alter any of the parameters of the environment until the entire computation is complete. Then once the alteration is made, the user must once again wait until the full solution is recomputed. To alleviate this Cohen et al. [14] proposed the *progressive refinement radiosity* which uses the notion of *adaptive refinement* of images, to provide the user as soon as possible

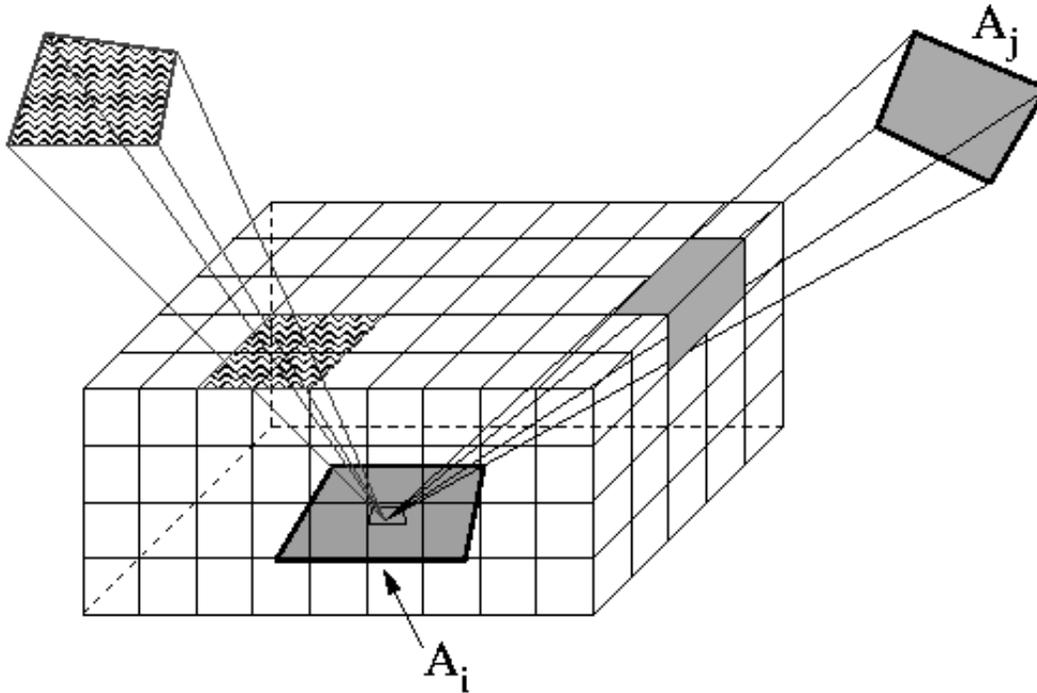


Figure 1.21: The Hemicube

with an approximation of the full solution. Rather than evaluating the effect that all other radiosities have on a particular patch, progressive refinement examines the effect that a patch has on all other patches in the environment.

With early radiosity techniques it was necessary to build the complete matrix of form-factors *before* solving the radiosity method. By re-ordering computation so that the complete form-factor doesn't need to be stored *progressive refinement* radiosity allows partial solutions to be displayed, Figure 2.22. The progressive refinement approach simultaneously solves all patch radiosities by repeatedly choosing a patch to “shoot” and distributing that patch's energy to all other patches. This is attractive as it provides a very good approximation to the final solution after only a few iterations.

The main advantage of radiosity methods lies in the view independence of the solution, and the ability to accurately simulate lighting effects.

More details of the radiosity method may be found in, for example [4, 16, 82]

### 1.2.3 Radiance

Radiance is a physically based lighting simulation tool for visualising lighting in virtual environments [89, 51]. This system employs a raytracing strategy, with significant extensions to traditional raytracing, to achieve accurate simulation of the propagation of light through an environment. The approach encompasses a hybrid deterministic/stochastic raytracing approach to efficiently solve the rendering equation, while maintaining an optimum balance between speed of computation and accuracy of the solution. This is achieved by applying a recursive algorithm to solve the rendering equation, reformulated below, with the notion of energy transfer between two points replaced by energy passing through a point in a specific direction (i.e. the definition of radiance):

$$L_r(\theta_r, \phi_r) = L_e + \int \int L_i(\theta_i, \phi_i) f(\theta_r, \phi_r, \theta_i, \phi_i) |\cos \theta_i| \sin \theta_i d\theta_i d\phi_i$$

```

BEGIN
  initially  $b_i = \Delta b_i = 0$  for non light sources
  initially  $b_i = \Delta b_i = e_i$  for light sources
  select patch  $i$ 
  WHILE (  $\Delta b_i \geq$  TOLERANCE)
    calculate form factors at patch  $i$  using the hemi-cube method
    FOR  $j:= 1$  TO all other patches (except  $i$ ) DO
      /* Include  $i$  if it is concave */
      DO Updates
         $\Delta RAD := (\rho_j \Delta b_i F_{ij} A_i) \div A_j$ 
        /* update change since last patch  $j$  shot light */
         $\Delta b_i = \Delta b_i + \Delta Rad$ 
        /* update total radiosity of patch  $j$  */
         $b_j = b_j + \Delta Rad$ 
      END FOR
    /* patch  $i$  has just shot - reset unshot radiosity to 0 */
     $\Delta b_i := 0$ 
    select next patch  $i$ 
  END WHILE
END

```

Figure 1.22: Pseudo Code for the Progressive Refinement Radiosity Solution

To accelerate the solution, the integral is separated into those parts which can be computed deterministically, and those which are better solved using stochastic methods. The *direct component* is computed by tracing rays to random locations on the light sources. The *specular indirect component* is computed by distributing rays about the mirror and transmission directions using uniform Monte Carlo sampling. Once these two prominent components are calculated, the *diffusely interreflected component* is computed by occasional evaluation of the integral at dynamically selected locations.

### The Direct Component

Light sources are responsible for a large proportion of the illumination in a scene. Therefore it makes sense to determine which objects are sources and compute them separately in a "direct calculation". This involves identifying those objects that contribute to the illumination more significantly than others, mostly these are light sources but in some cases may include other reflecting objects.

Light source testing, or *shadow testing*, has been used since the introduction of raytracing, and is one of the best strategies for improving efficiency and reducing noise. However there are a number of problems associated with conventional shadow testing algorithms:

- The time taken to perform shadow testing increases linearly according to the number of light sources in the scene. So it would take twice as long to render an image containing two sources as it would to render an image containing a single source. Clearly for complex environments containing many light sources this become prohibitive.
- Subdivision of large light sources<sup>5</sup> is necessary if inaccurate solutions containing excessive noise are to be avoided. If subdivision is done unconditionally, this leads to oversampling of sources at points that are far enough away that a single shadow ray would be sufficient.
- Most direct calculations fail to include light reflection from mirrors. Determining such significant sources with undirected ray samples is usually impractical, and results in missing illumination.

Radiance remedies the above shortcomings by:

<sup>5</sup>large relative to the distance to the illuminated surface

**Selective Shadow Testing:** A prioritised list of potential source contributions is created at each evaluation of the rendering equation. The largest potential contributors are tested for shadows first, testing is halted when the remainder of the potential contributions has fallen below some specified fraction of the accumulated total. The total estimate of the direct component is the sum of the tested light sources and a statistical estimate of the remainder.

**Adaptive Source subdivision:** The simplest approach for sources that are large relative to their distance is to send multiple sample rays. Unfortunately, breaking a source into smaller pieces and sending multiple rays is unsatisfactory for distant points - an adaptive sampling technique is more practical. Multiple rays are sent to sources if they are large relative to the distance of the evaluation point. Sources are recursively divided into smaller portions until each portion satisfies some size/distance criterion.

**Virtual Light Source Calculation:** A mirror surface may reflect sunlight on a diffuse or semi-specular surface. Although the diffuse interreflection calculation could in principle handle this effect, it is possible that insufficient sampling of an intense light source will occur. A small source reflected specularly is still too small to find in a practical number of naive Monte Carlo samples, the algorithm must know where to look. Virtual light sources, which do not exist in reality, are introduced to combat this problem. Virtual light sources direct the shadow rays in the appropriate directions to find reflected or otherwise transferred light sources.

### The Indirect Component

The calculation of the indirect component computes all sources of illumination not accounted for during the direct component computation. This includes light reflected and transmitted in specular directions and light bouncing diffusely between surfaces. Thus the indirect component is computed separately as specular sampling and diffuse interreflections. To achieve specular sampling a single ray is sent in the designated specular direction. In the case of ideal reflection or transmission, the direction is completely determined. In the case of rough-specular Monte Carlo importance sampling is used to determine the actual sample direction, and the light source(s) contribution is computed as part of the direct component.

Despite the considerable savings achieved by removing the direct lighting and specular reflections components from the integral, the diffuse indirect contributions remaining prove too expensive to recalculate at every pixel. Such a process would involve tracing hundreds of rays per pixel to ensure variance within tolerable levels. This is the reason why many conventional ray tracing approaches account for these indirect contributions using an arbitrary constant *ambient term*. This has been successful because the illumination changes gradually across a surface and the eye is more sensitive to contrast and therefore the resulting surfaces appear smooth. Radiance however takes a more accurate approach by spreading out this influence over a number of pixels which results in a smooth, accurate result at a modest sampling cost.

The basic idea is to perform a full evaluation of the rendering equation for indirect diffuse contributions *only as needed*, caching and interpolating these values over each surface. Direct and specular components are still computed on a pixel by pixel basis, but hemispherical sampling occurs less frequently. The result is a good estimate of the indirect diffuse contribution when required by sending more samples than would be possible for a pixel-independent calculation. Computing the indirect irradiance at a point in a scene, involves sending a few hundred uniformly distributed rays over the projected hemisphere. If any of these rays hits a light source its contribution is ignored, as direct component is handled separately. This sampling process is applied recursively for multiple reflections, and does not grow exponentially as each level has its own cache of indirect values.

The hemisphere samples not only give information about the indirect illumination, but also information about the locations and brightness of surfaces visible from the evaluation point. Using this gradient information in addition to the value of the function, a higher order interpolation can be used to get a better irradiance estimate between the calculated points. In effect, a smoother and more accurate result is achieved without having to do any additional sampling, and with very little overhead.

To summarise Radiance uses a raytracing method with extensions to handle specular, diffuse and directional diffuse reflection, and transmission to generate high quality lighting simulations. Studies have shown that Radiance is capable of producing highly realistic and accurate imagery [48], for this reason we have

chosen to use the Radiance suite of lighting simulation programs to generate the images used in the study presented in this thesis.

### 1.3 Summary

“Realistic representation in computer graphics requires quantifying light and materials, developing rules that describe their interaction (illumination models) and presenting the results through some display medium so the correct perceptions are created” [40].

This chapter opened with a brief introduction to light followed by a description of the fundamental terms, definitions and nomenclature associated with light energy including explanations of relevant radiometric terms, along with their photometric counterparts. Materials in nature interact with light in various ways, and materials can be characterised according to such behaviour. Diffuse materials reflect light uniformly in all directions, specular materials reflect light in the mirror direction only, materials of a glossy nature reflect light, and those materials that exhibit retro reflection reflect light back in the direction of the incident light. Illumination models take a description of a set of light sources and a set of surfaces which includes geometric, emissive and reflective properties, and attempt to calculate the interaction of the light with surfaces in the environment to produce an image. Early illumination models performed such calculations by considering the distribution of reflected light as function of incoming energy from the light source(s) to produce a *local illumination model*. Local is used to emphasise the fact that the illumination of a surface is determined by, and only by, the characteristics of the surface itself and those of the light source.

The interreflection of light can account for a high proportion of the total illumination in a scene. This is especially true for indoor scenes where light cannot “escape” the scene but instead is always reflected back into the scene by some surface. To account for such interreflections, all objects must be considered a potential source of illumination for all other objects in the scene. This constitutes a *global illumination model*. Global illumination models attempt to include all of the light interaction in a scene, giving rise to effects such as indirect illumination, soft shadows and colour bleeding, all of which have an impact on the perception of the resulting imagery, and hence the quality of the image.

The *rendering equation* which describes the fundamentals of global light exchanges in a scene is presented. Most image synthesis methods involve attempting to solve the rendering equation for each point in a scene.

Some of the more popular existing solution strategies for global illumination, *Raytracing* and *Radiosity*, were briefly reviewed. Raytracing provides an accurate method for obtaining global, specular reflection and transmission effects. Pixel rays are recursively traced through a scene, accumulating intensity contributions through repeated bounces between objects. By constructing a ray-tree for each pixel in the image plane, contributions are added from the terminal nodes to determine the intensity value for the pixel. A number of spatial subdivision techniques have been introduced to minimise the number of ray object intersections. Using Monte Carlo sampling to trace multiple rays per pixel, *distributed raytracing* provides an accurate method for modelling effects such as gloss, shadow effects and motion blur. Radiosity methods provide accurate methods for modelling diffuse reflection effects. This is achieved by calculating radiant energy transfer between surface patches in an environment. The solution of radiosity techniques are greatly accelerated using a number of extensions including the hemicube, progressive radiosity and hierarchical techniques. Finally Radiance, a physically based lighting simulation, which uses an extended raytracing methodology was introduced. Radiance accelerates its solution of the rendering equation by treating the direct component, the specular component and indirect interreflections separately.

To give an idea of differences between the three approaches, raytracing, radiosity and radiance, Figure 1.23 shows from left to right, a raytraced image, an image generated using radiosity and finally an image computed with the Radiance lighting simulation package.

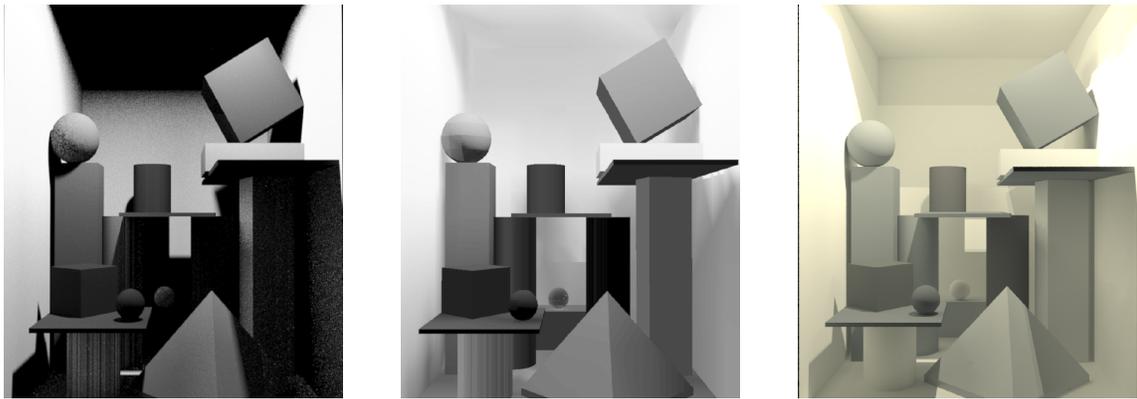


Figure 1.23: Renderings of a simple environment. Raytraced Solution (left), Radiosity Solution (center), & Radiance Solution (right)

## Chapter 2

# Visual Perception in Realistic Image Synthesis

1

Realism is often a primary goal in computer graphics imagery. We strive to create images that are perceptually indistinguishable from an actual scene. Rendering systems can now closely approximate the physical distribution of light in an environment. However, physical accuracy does not guarantee that the displayed images will have an authentic visual appearance. In recent years the emphasis in realistic image synthesis has begun to shift from the simulation of light in an environment to images that *look* as real as the physical environment they portray. In other words the computer image should be not only physically correct but also perceptually equivalent to the scene it represents. This implies aspects of the Human Visual System (HVS) must be considered if realism is required. Visual perception is employed in many different guises in graphics to achieve authenticity. Certain aspects of the HVS must be considered to identify the perceptual effects that a realistic rendering system must achieve in order to effectively reproduce a similar visual response to a real scene. This chapter outlines the main characteristics of the HVS and the manner in which knowledge about visual perception is increasingly appearing in state-of-the-art realistic image synthesis. It is organised into three Sections, each exploring the use of perception in realistic image synthesis, each with slightly different emphasis and application. First, perception driven rendering algorithms are described, which focus on embedding models of the HVS directly into global illumination computations in order to improve their efficiency. Then perception based image quality metrics, which aim to compare images on a perceptual rather than a physical basis, are presented. These metrics can be used to evaluate, validate and compare imagery. Finally, Tone Reproduction Operators, which attempt to map the vast range of computed radiance values to the limited range of display values, are discussed.

## 2.1 Visual Perception

Perception is the process by which humans, and other organisms, interpret and organise sensation in order to understand their surrounding environment. Sensation refers to the immediate, relatively unprocessed result of stimulation of sensory receptors. Perception, on the other hand, is used to describe the ultimate experience and interpretation of the world and usually involves further processing of sensory input. Sensory organs translate physical energy from the environment into electrical impulses processed by the brain. In the case of vision light, in the form of electromagnetic radiation, activates receptor cells in the eye triggering signals to the brain. These signals are not understood as pure energy, rather, perception allows them to be interpreted as objects, events, people and situations.

### 2.1.1 The Human Visual System

Vision is a complex process that requires numerous components of the human eye and brain to work together. Vision is defined as the ability to see the features of objects we look at, such as colour, shape,

---

<sup>1</sup>written by Ann McNamara    ann.mcnamara@tcd.ie

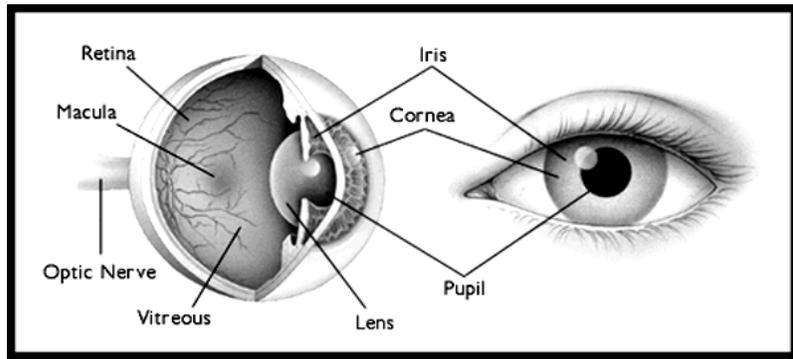


Figure 2.1: Cross section of the human eye

size, details, depth, and contrast. Vision begins with light rays bouncing off the surface of objects. These reflected light rays enter the eye and are transformed into electrical signals. Millions of signals per second leave the eye via the optic nerve and travel to the visual area of the brain. Brain cells then decode the signals providing us with sight.

The response of the human eye to light is a complex, still not well understood process. It is difficult to quantify due to the high level of interaction between the visual system and complex brain functions. A sketch of the anatomical components of the human eye is shown in Figure 2.2 [43]. The main structures are the iris, lens, pupil, cornea, retina, vitreous humor, optic disk and optic nerve.

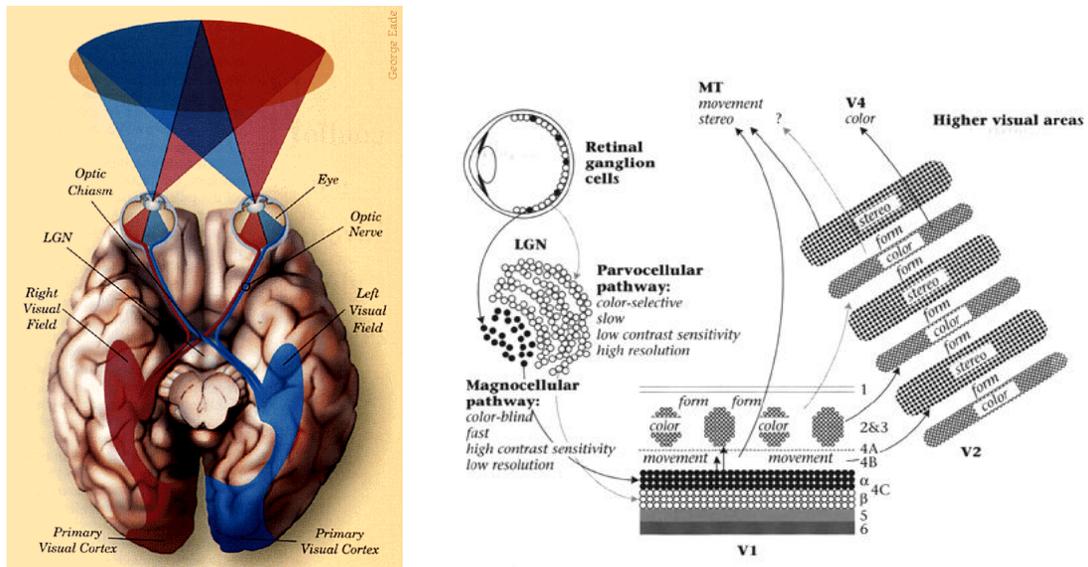


Figure 2.2: The components of the HVS

The path of light through the visual system begins at the *pupil*, is focused by the *lens*, then passes onto the *retina*, Figure 2.3 [39], which covers the back surface of the eye. The retina is a mesh of *photoreceptors*, which receive light and pass the stimulus on to the brain. Figure 2.1 [43] shows the internal structure of the human eye, a sphere, typically 12mm in radius, enclosed by a protective membrane, the *sclera*. At the front of the sclera lies the *cornea*, a protruding opening, and an optical system comprising the *lens* and *ciliary muscles* which change the shape of the lens providing variable focus. Light enters the eye through the lens and proceeds through the *vitreous humor*, a transparent substance, to the rear wall of the eye, the *retina*. The retina has photoreceptors coupled to nerve cells, which intercept incoming photons and output neural signals. These signals are transmitted to the brain through the *optic nerve*, connected to the retina at the *optic disk* or *papilla*, more commonly known as the *blind spot*. The retina is composed of two major classes of receptor cells known as *rods* and *cones*. The rods are extremely sensitive to light and provide achromatic

vision at low (*scotopic*) levels of illumination. The cones are less sensitive than the rods but provide colour vision at high (*photopic*) levels of illumination. A schematic drawing of rod and cone cells is shown in Figure 2.3. Cones are nerve cells that are sensitive to light, detail, and colour. Millions of cone cells are packed into the macula, aiding it in providing the visual detail needed to scan the letters on an eye chart, see a street sign, or read the words in a newspaper.

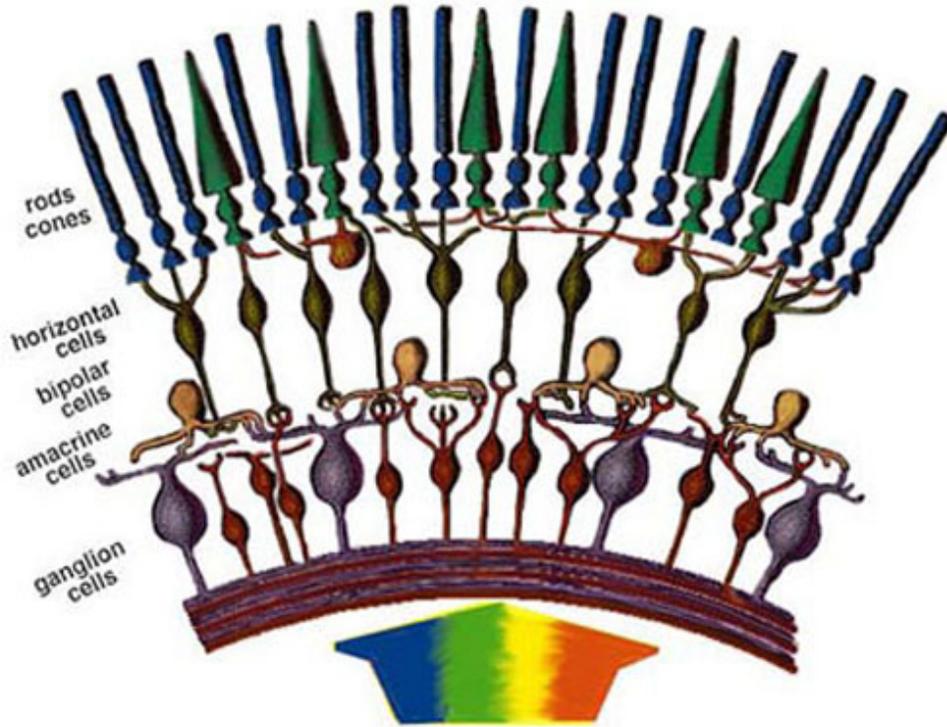


Figure 2.3: Retinal structure

Rods are designed for night vision. They also provide peripheral vision, but they do not see as acutely as cones. Rods are insensitive to colour. When a person passes from a brightly lit place to one that is dimly illuminated, such as entering a movie theatre during the day, the interior seems very dark. After some minutes this impression passes and vision becomes more distinct. In this period of *adaptation* to the dark, the eye becomes almost entirely dependent on the rods for vision, which operate best at very low light levels. Since the rods do not distinguish colour, vision in dim light is almost colourless.

Cones provide both luminance and colour vision in daylight. They contain three different pigments, which respond either to blue, red, or green wavelengths of light. A person who is missing one or more of the pigments is said to be colour-blind and has difficulty distinguishing between certain colours, such as red from green.

These photoreceptor cells are connected to each other and the *ganglion cells* which transmit signals to and from the optic nerve. Connections are achieved via two layers, the first and second *synaptic layers*. The interconnections between the rods and cones are mainly horizontal links, indicating a preferential processing of signals in the horizontal plane.

Normal daytime vision, where the cones predominate visual processing, is termed *photopic*, whereas low light levels where the rods are principally responsible for perception is termed *scotopic vision*. When both rods and cones are equally involved then vision is termed *mesopic*. Figure 2.4 [73], shows the range of luminance encountered by a typical human observer in a natural environment along with associated visual parameters.

*Visual acuity* is the ability of the Human Visual System (HVS) to resolve detail in an image. The human eye is less sensitive to gradual and sudden changes in brightness in the image plane but has higher sensitivity to intermediate changes. Acuity decreases with increase in distance. Visual acuity can be measured using

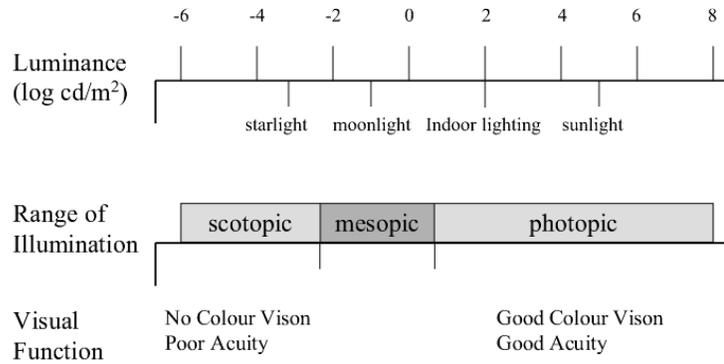


Figure 2.4: The range of luminance in the natural environment and associated visual parameters

a *Snellen Chart*, a standardised chart of symbols and letters. *Visual field* indicates the ability of each eye to perceive objects to the side of the central area of vision. A normal field of vision is  $180^\circ$ .

Contrast is defined as:

$$\frac{l_{max} - l_{min}}{l_{max} + l_{min}}$$

where  $l_{max}$  and  $l_{min}$  are the maximum and minimum luminance. Human brightness sensitivity is logarithmic, so it follows that for the same perception, higher brightness requires higher contrast. Apparent brightness is dependent on background brightness. This phenomenon, termed simultaneous contrast, is illustrated in 2.5. Despite the fact that *all* centre squares are the same brightness, they are perceived as different due to the different background brightness.



Figure 2.5: Simultaneous contrast: the internal squares all have the same luminance but the changes in luminance in the surrounding areas change the *perceived* luminance of the internal squares

*Depth Perception* is the ability to see the world in three dimensions and to perceive distance. Images projected onto the retina are two-dimensional, and from these flat images vivid three dimensional worlds are constructed. *Binocular Disparity* and *monocular cues* provide information for depth perception. Binocular disparity is the difference between the images projected onto the left and right eye. The brain integrates these two images into a single three dimensional image to allow depth and distance perception. Monocular cues are cues to depth that are effective when viewed with only one eye, including interposition, atmospheric perspective, texture gradient, linear perspective, size cues, height cues and motion parallax.

*Perceptual Constancy* is a phenomenon which enables the same perception of an object despite changes in the actual pattern of light falling on the retina. Psychologists have identified a number of perceptual constancies including lightness constancy, colour constancy, size constancy and shape constancy.

- **Lightness Constancy:** The term lightness constancy describes the ability of the visual system to perceive surface lightness correctly despite changes in the level of illumination.

- **Colour Constancy:** Closely related to lightness constancy, this is the ability of the HVS to perceive the correct colour of an object despite changes in illumination.
- **Shape Constancy:** Objects are perceived as having the same shape regardless of changes in their orientation. - example with cube, from front and side
- **Size Constancy:** This is the tendency to perceive objects as staying the same size despite changes in viewing distance.

## 2.1.2 Human Visual Perception

A number of psychophysical experimental studies have demonstrated many features of how the HVS works. However, problems arise when trying to generalise these results for use in computer graphics. This is because, often, experiments are conducted under limited laboratory conditions and are typically designed to explore a single dimension of the HVS. As described earlier, the HVS comprises complex mechanisms, which rather than working independently, often features work together, and therefore it makes sense to examine the HVS as a whole. Instead of reusing information from previous psychophysical experiments, new experiments are needed. Some examples will support this.

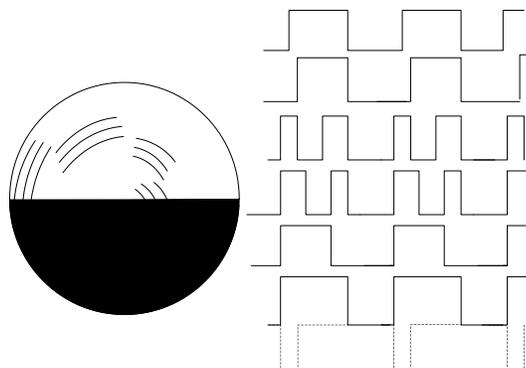


Figure 2.6: When a black and white patterned top shown on the left is rotated at 5-10 revolutions per second, coloured rings can be seen. The light intensity distribution of the rotating pattern as a function of time is shown on the right. Spatiotemporal interactions between antagonistic, spectrally opponent colour mechanisms account for this phenomenon.

A Benham's disk is a flat disc, half of which is black and the other half has three sets of lines like the grooves on a record but more spaced out, Figure 2.6. When the disk is spun a human observer sees red, yellow and green rings, despite the fact that there are no colours in the pattern. The curves on the right of the pattern begin to explain what happens. Each curve plots the temporal light intensity distribution at the different radii from the centre, created when the top is spun. These changing light patterns produce spatiotemporal interaction in the HVS that unbalance antagonistic, spectrally-opponent mechanisms to create the appearance of coloured rings. This illusion demonstrates that, although it may be convenient to model the HVS in terms of unidimensional responses to motion, pattern and colour, human percepts are in fact the product of complex multidimensional response.

A second example, Figure 2.7 [1], shows the panels in checkerboard block on the left and a flat pattern on the right, which have the same reflectance, but differences in their three-dimensional organisation means they are perceived differently. The two panels marked with X's have the same reflectance, but on the block they appear to have different reflectance under different levels of illumination. Conversely, the two panels marked with O's have different reflectance values but on the block appear to be the same colour due to the different illumination conditions. This demonstrates the complexity of interactions between apparent reflectance, apparent illumination and apparent shape that can dramatically affect human perception.

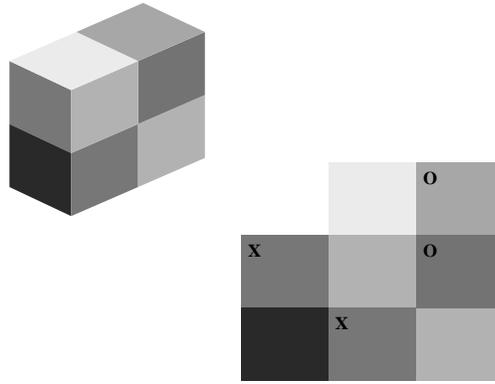


Figure 2.7: Interaction between apparent reflection, apparent illumination and apparent three-dimensional shape. Corresponding panels in the two patterns have the same physical reflectance. Differences in the perceived spatial organisation of the patterns produces differing interpretations in terms of lightness (apparent reflectance) and brightness (apparent illumination)

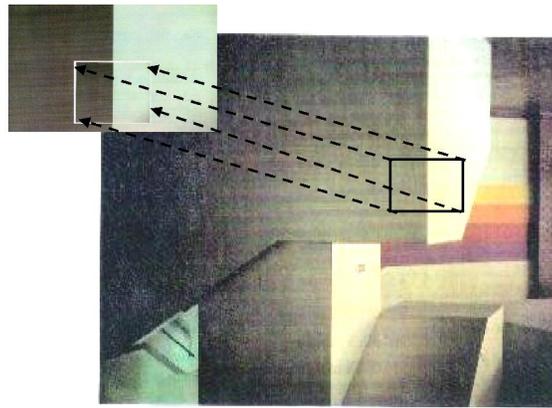


Figure 2.8: Importance of depth perception for lightness constancy

### 2.1.3 Lightness Perception

Gilchrist [31, 30, 29] justified the systematic study of lightness error as an understanding of the HVS. He found that there are always errors when judging lightness, and these errors are not random, but *systematic*. The pattern of these systematic errors therefore provide a signature of the visual system. He defines a lightness error as “any difference between the actual reflectance of a target surface and the reflectance of the matching chip selected from a Munsell chart”. The task defined for the psychophysical experiments described later in this thesis involves asking human observers to match the reflectance of real world objects to a Munsell chart, which gives a measure of errors in lightness matching. The observer is then asked to match the reflectance of simulated objects (in a computer generated rendition of the real world) to the same Munsell chart. This gives a measure of lightness errors with respect to the computer image. There are limitations on the HVS, so there will be errors (systematic errors) in both cases. For the rendered image to be deemed a faithful representation, both sets of lightness errors should be close to each other.

Gilchrist (1977) showed that the perception of the degree of “lightness” of a surface patch (i.e. whether it is white, gray or black) is greatly affected by the perceived distance and orientation of the surface in question, as well as the perceived illumination falling on the surface - where the latter was experimentally manipulated through a variety of cues such as occlusion, or perspective.

Perception of the lightness of patches varying in reflectance may thus be a suitable candidate for the

choice of visual task. It is simple to perform, and it is known that lightness constancy depends on the successful perception of lighting and the 3D structure of a scene, for example Figure 2.8. When viewed in isolation, the patches on the top left hand corner appear to be of different luminance. However, when examined in the context of the entire scene, it can be seen that the patches have been cut from the edge of the stairwell, and are perceived as an edge where the entire stairwell has the same luminance. Eliminating the depth cues means that the patches are perceived as different, demonstrating the dependency of lightness perception on the correct perception of three-dimensional structure [31]. As the key features of any scene are illumination, geometry and depth, the task of lightness matching encapsulates all three key characteristics into one task. This task is particularly suited to an experimental framework, as apart from being simple to perform it also allows excellent control over experimental stimuli. Subsequent chapters describe an experimental framework, with such a lightness matching task at the core, to allow human observers to compare real and synthetic scenes, but first we look at how knowledge of the HVS can be applied for the advancement of computer graphics techniques.

## 2.2 Perception and Graphics

Recent years have seen an increase in the application of visual perception to computer graphics. As mentioned earlier, in certain applications it is important that computer images should not only be physically correct but also perceptually equivalent to the scene it is intended to represent. Realism implies computational expense, and research is beginning to emerge to investigate how knowledge of the human visual system can be used to “cut corners” and minimise rendering times by guiding algorithms to compute only what is necessary to satisfy the observer. Perception based image quality metrics, which can be used to evaluate, validate and compare imagery have been presented. Also, tone reproduction operators, have been introduced to map the vast range of computed radiance values to the limited range of the display device.

### 2.2.1 Perceptually driven rendering

Even for realistic image synthesis there may be little point spending time or resources to compute detail in an image that would not be detected by a human observer. By eliminating any computation spent on calculating image features which lie below the threshold of visibility, rendering times can be shortened leading to more efficient processing. Because the chief objective of physically based rendering is *realism*, incorporating models of HVS behaviour into rendering algorithms can improve performance, as well as improving the quality of the imagery produced. So by taking advantage of the limitations of the human eye, just enough detail to satisfy the observer can be computed without sacrificing image quality. Several attempts have been made to develop image synthesis algorithms that detect threshold visual differences and direct the algorithm to work on those parts of an image that are in most need of refinement.

Raytracing produces an image by computing samples of radiance, one for each pixel in the image plane. Producing an anti-aliased image is difficult unless very high sampling densities are used. Mitchell [65] realised that deciding where to do extra sampling can be guided by knowledge of how the eye perceives *noise as a function of contrast and colour*. Studies have shown that the eye is most sensitive to noise in intermediate frequencies [78]. While frequencies of up to 60 cycles per degree (cpd) can be visible, the maximum response to noise is at approximately 4.5 cpd, so sampling in regions with frequency above this threshold can be minimised, without affecting the visual quality of the image. Mitchell begins by sampling the entire image at low frequency then uses an adaptive sample strategy on the image according to the frequency content. This results in a non uniform sampling of the image, which enables aliasing noise to be channelled into high frequencies where artefacts are less conspicuous. However, non-uniform sampling alone doesn’t eliminate aliasing, just changes its characteristics to make it less noticeable. Mitchell applies two levels of sampling. To decide whether the high sampling density should be invoked the variance of samples could be used [52], but this is a poor measure of visual perception of local variation. Instead Mitchell chooses to use contrast to model the non-linear response of the eye to rapid variations in light intensity:

$$C = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

As each sample consists of three separate intensities for red, green and blue, three separate contrasts can be computed for each of them. These three contrasts are tested against separate thresholds, 0.4, 0.3 and

0.6 for red, green and blue respectively, and super-sampling is done if any one exceeds the threshold. The contrast metric is then used to determine when the high sampling density should be invoked. This test is most sensitive to green in accordance with the human eye's response to noise as a function of colour. Multi stage filters are then used to reconstruct the non-uniform samples into a digital image. Although this idea has the beginnings of a perceptual approach, it is at most a crude approximation to the HVS. Only two levels of sampling are used and it doesn't account for visual masking <sup>2</sup>.

The HVS exhibits different spatial acuities in response to different colours. Evidence exists that colour spatial acuity is less than monochrome spatial acuity. Exploiting this *poor colour spatial acuity* of the HVS, Meyer and Liu [62] developed an adaptive image synthesis algorithm which uses an opponents processing model of colour vision [61] comprising chromatic and achromatic colour channels. Using a Painter and Sloan [72] adaptive subdivision, a k-D tree representation <sup>3</sup> of the image is generated. Areas of the image containing high frequency information are stored at the lower levels of the tree. They then modified a screen subdivision raytracer to limit the depth to which the k-D tree must be descended to compute the chromatic colour channels. The limit is determined by psychophysical results describing the colour spatial frequency. They achieved a modest saving in computational effort and showed, using a psychophysical experiment, that decreasing the number of rays used to produce the chromatic channels had less of an effect on image quality than reducing the number of rays used to create the achromatic channels. This was the first work to attempt to minimise the computation of colour calculations, as opposed to just decreasing costly object intersection calculations.

Bolin and Meyer [6] took a frequency based approach to raytracing, which uses a simple vision model, making it possible for them to control how rays are cast in a scene. Their algorithm accounts for the *contrast sensitivity*, *spatial frequency* and *masking properties* of the HVS. The contrast sensitivity response of the eye is non-linear. So, when deciding where rays should be cast, the algorithm deems a luminance difference at low intensity to be of greater importance than the same luminance difference at high intensity. The spatial response of the HVS is known to be less for patterns of pure colour than for patterns that include luminance differences. This means that it is possible to cast fewer rays into regions with colour spatial variations than are cast in regions with spatial frequency variations in luminance. Finally, it is known that the presence of high spatial frequency can mask the presence of other high frequency information (masking). When used in conjunction with a Monte Carlo raytracer, more rays are spawned when low frequency terms are being determined than when high frequency terms are being found. Using this strategy, the artefacts that are most visible in the scene can be eliminated from the image first, then noise can be channelled into areas of the image where artefacts are less conspicuous. This technique is an improvement on Mitchell's method because the vision model employed accounts for *contrast sensitivity*, *spatial frequency* and *masking*.

Despite the simplicity of the vision models used in these approaches, the results are promising, especially as they demonstrate the feasibility of embedding HVS models into the rendering systems to produce more economical systems without forfeiting image quality. Fuelled by the notion that more sophisticated models of the HVS would yield even greater speedup, several researchers began to introduce more complex models of the HVS into their global illumination computations.

Myszkowski [67] applied a more sophisticated vision model to steer computation of a Monte Carlo based raytracer. Aiming to take maximum advantage of the limitations of the HVS, his model included *threshold sensitivity*, *spatial frequency sensitivity* and *contrast masking*. A perceptual error metric is built into the rendering engine allowing adaptive allocation of computation effort into areas where errors remain above perceivable thresholds and allowing computation to be halted in all other areas (i.e. those areas where errors are below the perceivable threshold and thus not visible to a human observer). This perceptual error metric takes the form of Daly's [20] **Visible Difference Predictor (VDP)**, discussed in Section 2.2.2.

Bolin and Meyer [7] devised a similar scheme, also using a sophisticated vision model, in an attempt to make use of all HVS limitations. They integrated a simplified version of the Sarnoff Visible Discrimination Model (VDM) into an image synthesis algorithm to detect threshold visible differences and, based on those differences direct subsequent computational effort to regions of the image in most need of refinement. The VDM takes two images, specified in CIE XYZ colour space, as input. Output of the model is a Just Noticeable Difference (JND) map. One JND corresponds to a 75% probability that an observer viewing the two images would detect a difference [54]. They use the upper and lower bound images from the computation results at intermediate stages and used the predictor to get an error estimate for that stage.

---

<sup>2</sup>The presence of high spatial frequency in an image can mask the presence of other high frequency information

<sup>3</sup>A KD Tree is a data structure that is used in computer science during orthogonal range searching

The image quality model is used to control where to take samples in the image, and also to decide when enough samples have been taken across the entire image, providing a visual stopping condition. A more comprehensive description of the VDM is given in Section 2.2.2.

Applying a complex vision model at each consecutive time step of image generation requires repeated evaluation of the embedded vision model. The VDP can be expensive to process due to the multi-scale spatial processing involved in some of its components. This means that in some cases the cost of re-computing the vision model may cancel the savings gained by employing the perceptual error metric to speed up the rendering algorithm. To combat this, Ramasubramanian [76] introduced a metric that handles luminance-dependent processing and spatially-dependent processing independently, allowing the expensive spatially-dependent component to be *precomputed*. Ramasubramanian developed a physical error metric that predicts the *perceptual* threshold for detecting artefacts in the image. This metric is then used to predict the sensitivity of the HVS to noise in the indirect lighting component. This enables a reduction in the number of samples needed in areas of an image with high frequency texture patterns, geometric details, and direct lighting variations, giving a significant speedup in computation.

Using *validated* image models that predict image fidelity, programmers can work toward achieving greater efficiencies in the knowledge that resulting images will still be faithful visual representations. Also in situations where time or resources are limited and fidelity must be traded off against performance, perceptually based error metrics could be used to provide insights into where computation could be economised with least visual impact.

## 2.2.2 Perceptually Based Image Quality Metrics



Figure 2.9: Photograph of a conference room (left) & photo-realistic rendering (right)

Reliable image quality assessments are necessary for the evaluation of realistic image synthesis algorithms. Typically the quality of the image synthesis method is evaluated using image to image comparisons. Often comparisons are made with a photograph of the scene that the image depicts, as shown in Figure 2.9 [51].

Several image fidelity metrics have been developed whose goals are to predict the amount of differences that would be visible to a human observer. It is well established that simple approaches like mean squared error do not provide meaningful measures of image fidelity, Figure 2.10. The image on the left has been slightly blurred, while the image on the right has deliberate scribbles. The Root Mean Square Error (RMSE) value for blurred image is markedly higher than the RMSE for the image on the right. However, a human observer might indicate a higher correlation between the two images. This illustrates that the use of RMSE is not sufficient [75]. Clearly more sophisticated measures which incorporate a representation of the HVS are needed. It is generally recognised that more meaningful measures of image quality are obtained using techniques based on visual (and therefore subjective) assessment of images, as after all most final uses of computer generated images will be viewed by human observers.

In 1998, Li and Meyer conducted a comprehensive study that compared two of the more successful image quality models, outlined here:

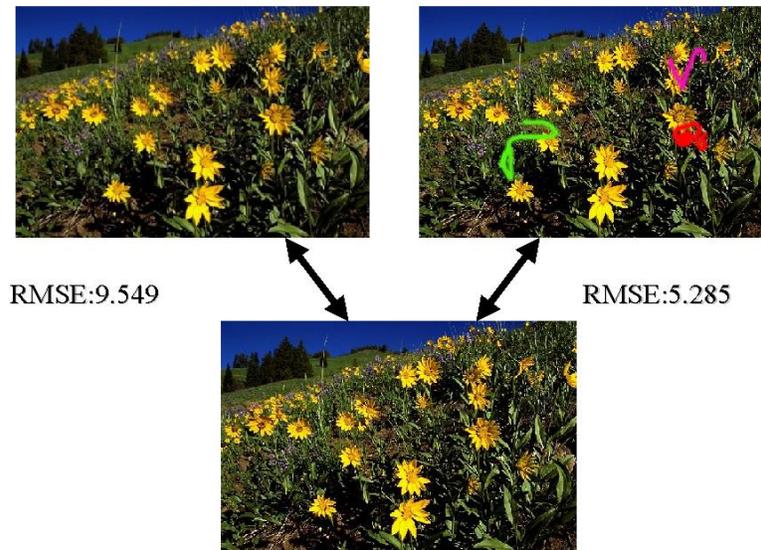


Figure 2.10: Comparing top images to the image on the bottom using RMSE

### Daly's Visible Differences Predictor

The **Visible Difference Predictor (VDP)** is a perceptually based image quality metric proposed by Daly [20]. The VDP takes a psychophysically based approach to construct a model of human vision. Two images serve as input to the VDP, and a difference map is produced as output. This difference map predicts the probability of detection of differences between the two images. Figure 2.11 gives a block diagram of the components of the predictor. The main stages are an initial non-linearity, frequency domain weighting with the human contrast sensitivity function CSF, and a series of detection mechanisms.

To account for adaptation and the non-linear response of retinal neurons, a non-linear response function is applied to each image. Daly assumed that adaptation is a function of each pixel individually. The model used for adaptation estimates the relationship between brightness sensation and luminance. At low levels of luminance a cube-root power law is applied, while at higher luminance levels it approximates the logarithmic dependence.

The next stage involves converting the image to the frequency domain. The transformed data is weighted with the CSF i.e. the scaled amplitude for each frequency is multiplied by the CSF for that spatial frequency. This data is then normalised (by dividing each point by the original image mean) to give local contrast information.

The image is then divided into 31 independent streams. It is known that the HVS has specific selectivities based on orientation (6 channels) and spatial frequency (approximately one octave per channel). Each of the five overlapping spatial frequency bands is combined with each of the six overlapping orientation bands to split the image into thirty channels. Along with the orientation-independent base band this gives a total of 31 channels. At this point the individual channels are transformed back into the spatial domain.

A mask, which is a function of image location in the image, is associated with each channel. The presence of masking information at a specific location, spatial frequency and orientation increases the threshold of detectability for a signal with those characteristics. A threshold elevation map for each channel is computed as a function of the mask contrast. Finally, mutual masking is applied between the two sets of threshold elevation maps from both input images to produce a single threshold elevation map per channel.

Contrasts of corresponding channels in one image are subtracted from those of the other images, and the difference is scaled down by threshold elevation. The scaled contrast differences are used as the argument to a psychometric function to compute a detection probability. The psychometric function yields a probability of detection of a difference for each location in the image, for each of the 31 channels. The detection

probabilities for all of the channels are combined using the assumption of independent probabilities, giving an overall signed detection probability for each location in the image.

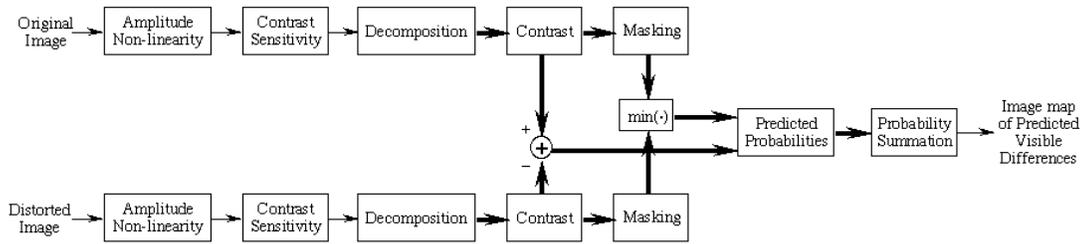


Figure 2.11: A block diagram of the Visible Differences Predictor (VDP)

### Sarnoff Visual Discrimination Model

The Sarnoff VDM [53] focuses more attention on modelling the physiology of the visual pathway. Therefore the VDM operates in the spatial domain (as opposed to the frequency domain approach of VDP). The main components of the VDM include spatial resampling, wavelet-like pyramid channelling, a transducer for JND calculations and a final refinement step to account for CSF normalisation and dipper effect simulation. The VDM also takes as input two images along with a set of parameters for viewing conditions, and here the output is a map of JND's. The overall structure of the VDM is shown in Figure 2.12.

To account for the optics of the eye and mosaic structure of the retina, a single point spread function (PSF) is used to predict the foveal performance of the two dimensional optics of the eye (it is assumed the PSF is circularly symmetric). The effect of the PSF convolution is blurring of the input images. A spatial resampling, at a rate of 120 pixels per degree, is then applied to account for the fixed density of the cones in the fovea. This resampling is essential in a spatial domain approach since the extraction of the different frequency bands is dependent on the resampling kernels and the resampling rates. If the original image is too big, and the local image quality cannot be assessed in a single glance, then the image can be subdivided into smaller blocks.

A Laplacian pyramid stores a wavelet representation of the resampled input images and a quadrature mirrored pair of convolution kernels records information along each of the four orientations. On completion of this stage, the raw luminance signal has been converted into units of local contrast. Due to the use of a spatial domain convolution approach, the peak frequency of each level has to be a power of two. The seven bandpass levels have peak frequencies from 32 to 0.5 cycles/degree, where each level is separated from its neighbours by one octave. A steerable pyramid is used to perform the decomposition, to increase performance. This is a multi-scale, multi-orientation, image transform with both frequency and orientation components. The last step in the decomposition process is computation of a phase-independent energy response by squaring and summing odd phase and even phase coefficients. They are determined by convolving the quadrature mirror pair filters with a certain frequency band.

The energy measure is normalised by the square of the reciprocal of the CSF, then a transducer is used to refine the JND map by taking the spatial masking dipper effect into account. The dipper shape reflects on characteristic of the contrast discrimination function. This stage involves the transformation by a sigmoid non-linearity. Finally the model includes a pooling stage in which transducer outputs are averaged over a small region by convolving with a disc-shaped kernel.

Once the JND difference map has been computed for each channel, the final stage involves putting together the contributions from each channel. This leads to the concept of a space of multiple dimensions. There are 28 channels involved in the summation, seven pyramid levels times four different orientations. For each spatial position the final JND distance can be regarded as the distance between the 28-dimensional vectors.

Meyer and Li concluded that although both methods performed comparably, the Sarnoff VDM was deemed slightly more robust producing better JND maps and required less re-calibration than the Daly VDP. Despite this both have been successfully incorporated into global illumination algorithms to produce favourable results [66, 68, 8].

The main contribution of Meyer and Li's study was the independent verification of the major features of each model. Meyer and Li do agree however, that psychophysical experiments involving a large set of

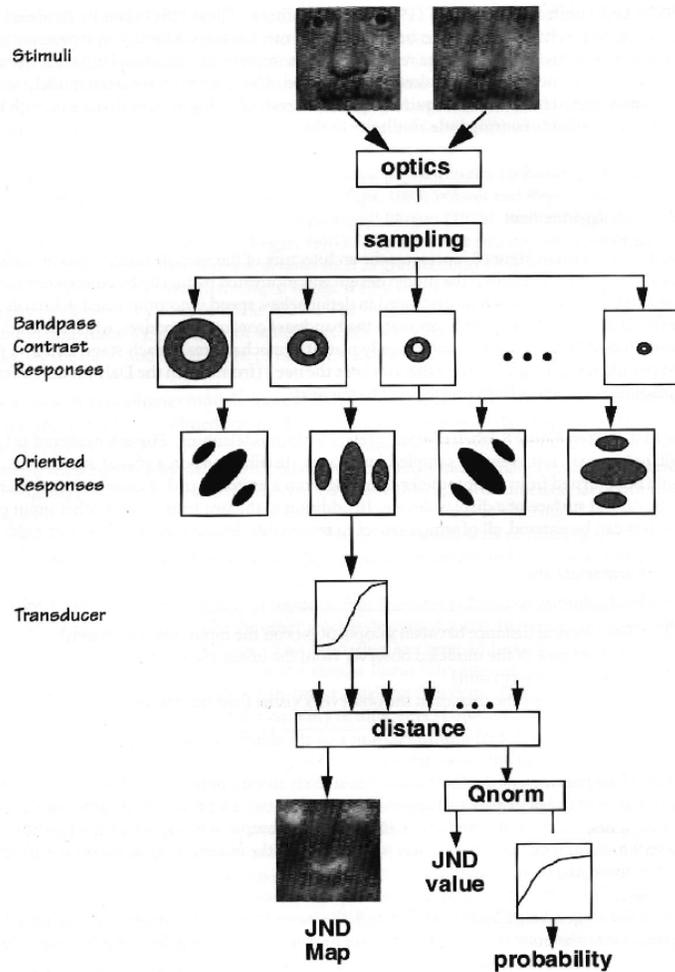


Figure 2.12: A block diagram of the Visual Discrimination Model (VDM)

images would be needed for a complete evaluation, to investigate the performance of models under a wider range of conditions.

### 2.2.3 Comparing Real and Synthetic Scenes

While image quality metrics have been successfully incorporated into global illumination algorithms to guide computations more efficiently, metrics can also be useful to validate and compare rendering techniques. As the goal of realistic image synthesis is to generate representations of a physical scene, simulations should therefore be compared to the real world scenes.

Using a simple five sided cube as their test environment, Meyer et al [63] presented an approach to image synthesis comprising separate physical and perceptual modules. They chose diffusely reflecting materials to build a physical test model. Each module is verified using experimental techniques. The test environment was placed in a small dark room. Radiometric values predicted using a radiosity lighting simulation of a basic scene are compared to physical measurements of radiant flux densities in the real scene. Then the results of the radiosity calculations are transformed to the RGB values for display, following the principles of colour science. Measurements of irradiation were made at 25 locations in the plane of the open face for comparison with the simulations. Results show that irradiation is greatest near the centre of the open side of the cube. This area provides the best view of the light source and other walls. In summary, there is good agreement between the radiometric measurements and the predictions of the lighting model.

Meyer et al. then proceeded by transforming the validated simulated value to values displayable on

a television monitor. Twenty participants were asked to differentiate between a real environment and the displayed image, both of which were viewed through the back of a view camera. They were asked which of the images was the real scene. Nine out of the twenty participants (45%) indicated that the simulated image was actually the real scene, i.e. selected the wrong answer, revealing that observers would have done just as well by simple guessing. Although participants considered the overall match and colour match to be good, some weaknesses were noticed in the sharpness of the shadows (a consequence of the discretisation in the simulation) and in the brightness of the ceiling panel (a consequence of the directional characteristics of the light source). The overall agreement lends strong support to the perceptual validity of the simulation and display process. This was the first attempt to compare real and simulated scenes side by side, using human observers.

Although the results of the study are encouraging, there are some drawbacks with this approach: The scene under examination was very simple, the methodology for comparison itself was not inherently controlled, and the results suggest that the participants could have simply guessed. To really investigate the differences between a real environment and its synthetic representation, a more robust approach is required.

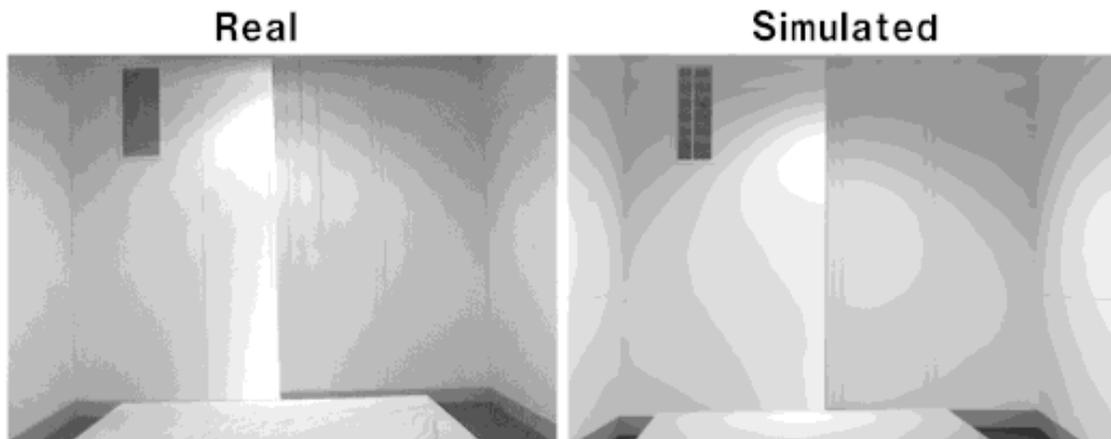


Figure 2.13: NIST Comparison using a Conference Room

Another approach to comparing real and simulated scenes takes a *captured* image of the real scene in question and uses numerical techniques to determine the perceptual differences between the two. Rushmeier et al. [77] explored using perceptually based metrics, based on image appearance, to compare the image quality of a rendered scene to a captured image of the scene being represented, Figure 2.13. The following image comparison metrics were derived from [20, 24, 55] in a study which compared real and synthetic images by Rushmeier *et al* [77]. Each is based on ideas taken from image compression techniques. Image compression techniques seek to minimise storage space by saving only what will be visible in an image (similar to the goal of perceptually driven rendering where the aim is to minimise rendering times by computing only what will be visible in the image). The goal of Rushmeier’s study was to obtain results from comparing two images using these models that were large if large differences between the images exist, and small when they are almost the same. These suggested metrics include some basic characteristics of human vision described in image compression literature. First, within a broad band of luminance, the eye senses relative rather than absolute luminances. For this reason a metric should account for luminance variations, not absolute values. Second, the response of the eye is non-linear. The perceived “brightness” or “lightness” is a non-linear function of luminance. The particular non-linear relationship is not well established and is likely to depend on complex issues such as perceived lighting and 3-D geometry. Third, the sensitivity of the eye depends on the spatial frequency of luminance variations. The following methods attempt to model these three effects. Each model uses a different Contrast Sensitivity Function (CSF) to model the sensitivity to spatial frequencies.

**Model 1 After Mannos and Sakrison:** [55].

This model is adapted from a study in image compression which attempted to derive a numerically based measure of distortion which corresponds to the *subjective* evaluation of the image by a human observer, in order to simulate the optimum encoding technique. First, all the luminance values are normalised by the mean luminance. The non-linearity in perception is accounted for by taking the

cubed root of each normalised luminance. A Fast Fourier Transform (FFT) is computed of the resulting values, and the magnitude of the resulting values are filtered with a CSF to an array of values. Mannos and Sakrison [MaSa74] proposed a model of the human contrast sensitivity function. The contrast sensitivity function tells us how sensitive we are to the various frequencies of visual stimuli. If the frequency of visual stimuli is too high we will not be able to recognise the stimulus pattern any more. Imagine an image consisting of vertical black and white stripes. If the stripes are very thin (i.e. a few thousand per millimetre) humans will be unable to see individual stripes. All that we will see is a gray image. If the stripes then become wider and wider, there is a threshold width, after which humans are able to distinguish the stripes. The contrast sensitivity function proposed by Mannos and Sakrison is:

$$A(f) = 2.6 \cdot [0.0192 + 0.114\sqrt{f}]e^{-(0.114\sqrt{f})^{1.1}}.$$

where  $f$  is the spatial frequency of the visual stimuli given in cycles/degree. Finally, the distance between the two images is computed by finding the Mean Square Error (MSE) of the values for each of the two images. This technique therefore measures similarity in Fourier amplitude between images.

**Model 2 After Gervais et al:** [24].

The original purpose of this model was to identify confusion among letters of the alphabet. Even though this problem is quite different to image comparison, Rushmeier et al. justify using this model as it includes the effect of phase as well as magnitude in the frequency space representation of the image. Once again the luminances are normalised by dividing by the mean luminance. A Fast Fourier Transform (FFT) is computed, producing an array of phases and magnitudes. These magnitudes are then filtered with an anisotropic CSF filter function constructed by fitting splines to psychophysical data. The distance between two images is computed using methods described in [24].

**Model 3 After Daly:** adapted from [20].

Described in more detail in Section 2.2.2, this model combines the effects of adaptation and non-linearity into a single transformation, which acts on each pixel individually. In the first two models each pixel has significant global effect in the normalisation by contributing to the image mean. Each luminance is transformed by an amplitude non-linearity value. A FFT is applied to each transformed luminance and then they are filtered by a CSF (computed for a level of 50 cd/m<sup>2</sup>). The distance between the two images is then computed using MSE as in model 1.

Myszkowski [67] realised the VDP had many potential applications in realistic image synthesis. He completed a comprehensive validation and calibration of VDP response via human psychophysical experiments. He subsequently used the VDP local error metric to steer decision making in adaptive mesh subdivision, and in isolating regions of interest for more intensive global illumination computations, Figures 2.14, 2.15. The VDP was tested to determine how close VDP predictions come to subjective reports of visible differences between images by designing two human psychophysical experiments. Results from these experiments showed a good correspondence between human observations and VDP results.

These perception based image quality metrics have demonstrated the success of implementing a visual model, in spite of the fact that knowledge of the visual process is as yet incomplete. However, there is a fundamental problem with all these methods from the point of view of *validation*. Although these methods are capable of producing images based on models of the HVS, there is no standard way of telling if the images “capture the visual appearance” of scenes in a meaningful way. One approach to validation could compare observers’ perception and performance in real scenes against the predictions of the models. This would enable calibration and validation of the models to assess the level of fidelity of the images produced.

## 2.2.4 Tone Mapping

The range of luminance we encounter in natural environments (and hence the range of luminances that can be computed by a physically based rendering algorithm) is vast. Over the course of the day the absolute level of illumination can vary by more than a 100,000,000 to 1 from bright sunlight down to starlight. The dynamic range of light energy in a single environment can also be large, in the order of 10,000 to 1 from highlights to shadows. However, typical display media have useful luminance ranges of approximately 100 to 1. This means some mapping function must be used to translate real world values into values displayable

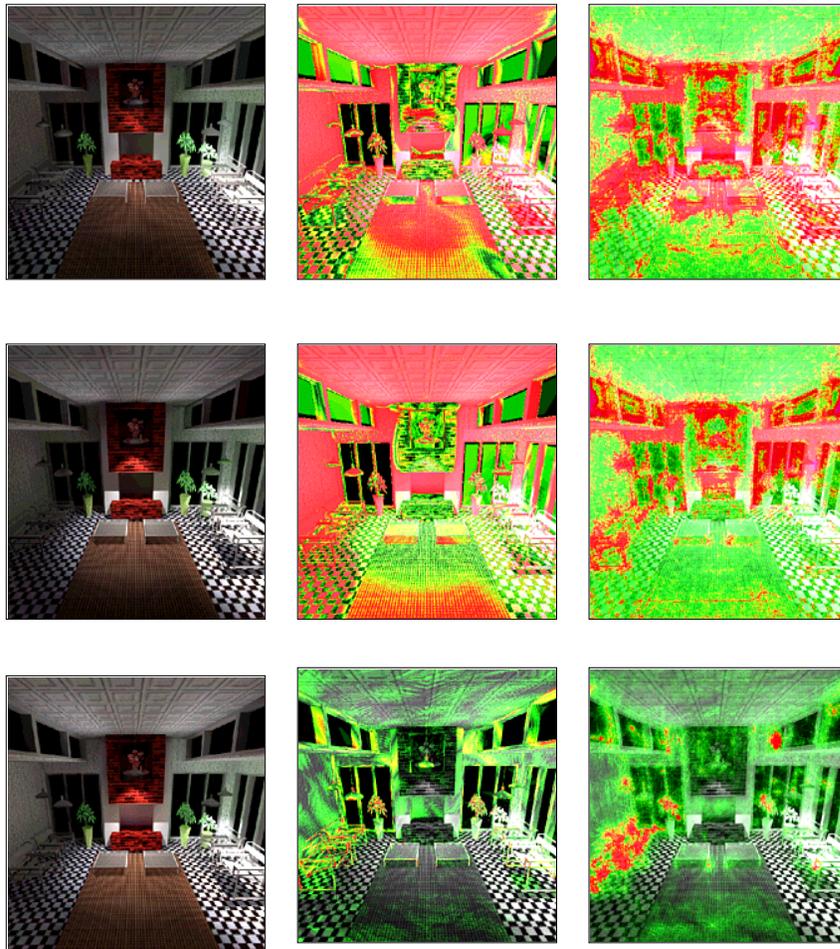


Figure 2.14: "Perceptual" convergence of the lighting solution. The first column shows the current solution, then the absolute and VDP predicted difference with respect to the fully converged solution shown in 2.15

by the device in question, be it electronic (CRT) or print media. Initial attempts to develop such a mapping were simple *ad-hoc* methods which failed miserably for high dynamic range scenes. These ad-hoc methods proceeded by employing a linear arbitrary scaling, either mapping the average of a luminance in the real world to the average of the display, or the maximum non-light source luminance to the maximum displayable value. While such a scaling proved appropriate for scenes with similar dynamic range to the display media, it failed to preserve visibility in scenes with high dynamic ranges of luminance. This is due to the fact that very bright or very dim values must be clipped to fall within the range of displayable values. Also, using this method all images are mapped in the same manner irrespective of *absolute* value. This means a room illuminated by a single candle could be mapped to the same image as a room illuminated by a search light, resulting in loss of the overall impression of brightness and so losing the subjective correspondence between real and displayed scene. It follows that more sophisticated mappings were required.

*Tone Mapping*, originally developed for use in photography and television, addresses the problem of mapping to a display, and is an attempt to recreate the same *perceptual* response in the viewer of a synthetic image as they would have if looking at the real scene. Taking advantage of HVS sensitivity to *relative* luminances rather than *absolute* luminances allows the overall subjective impression of a real environment to be replicated on some display media, despite the fact that the range of real world luminances often dwarfs the displayable range.

Tone reproduction operators can be classified according to the manner in which values are transformed. *Single-scale* operators proceed by applying the *same* scaling transformation for each pixel in the image, and that scaling only depends on the current level of adaptation, and not on the real-world luminances. *Multi-*

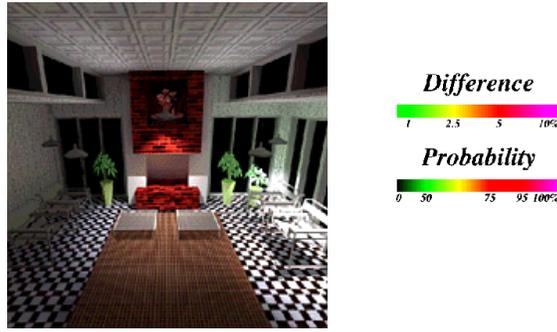


Figure 2.15: Fully Converged Image, and Perceptual Scales

*scale* operators take a differing approach and may apply a different scale to each pixel in the image, this time the scaling is influenced by many factors.

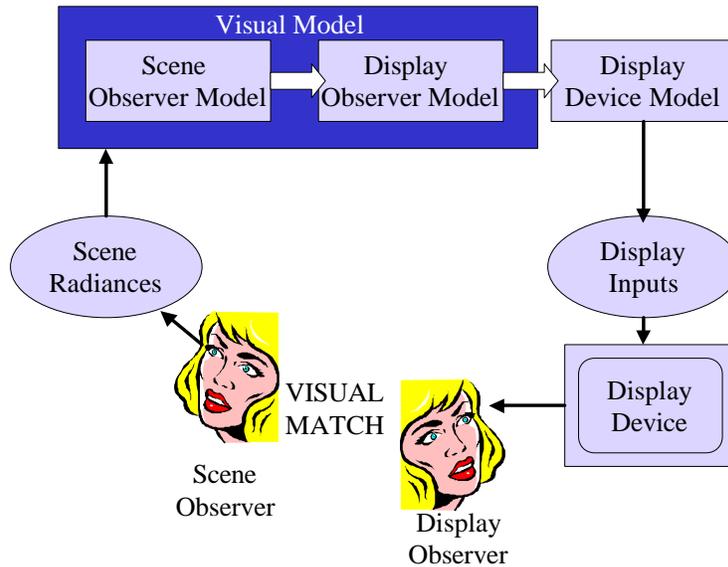


Figure 2.16: A block diagram of tone reproduction

### 2.2.5 Single Scale Tone Reproduction Operators

Tumblin and Rushmeier were the first to apply the dynamics of tone reproduction to the domain of realistic image synthesis [88]. Using a psychophysical model of brightness perception first developed by Stevens and Stevens [83], they produced a tone reproduction operator that attempted to match the brightness of the real scene to the brightness of the computed image displayed on a CRT. To achieve this an *observer model* is built which describes how real world and display luminances are perceived, and a *display model* that describes how a frame-buffer value is converted into displayed luminance, Figure 2.16 [88]. The image is presented to a hypothetical real world observer, who adapts to a luminance  $L_{a(w)}$ . Applying Stevens' equation, which relates brightness to target luminance, the perceived value of a real world luminance,  $L_w$ , is computed as:

$$B_w = 10^{\beta(L_{a(w)})} (\pi \times 10^{-4} L_w)^{\alpha(L_{a(w)})}$$

where  $\beta(L_{a(w)})$  and  $\alpha(L_{a(w)})$  are functions of the real world adaptation level:

$$\alpha(L_{a(w)}) = 0.4 \log_{10}(L_{a(w)}) + 1.519$$

$$\beta(L_{a(w)}) = -0.4(\log_{10}(L_{a(w)}))^2 - 0.218 \log_{10}(L_{a(w)}) + 6.1642$$

Luminances are in  $cd/m^{-2}$ . If it is assumed that a display observer viewing a CRT screen adapts to a luminance,  $L_{a(d)}$ , the brightness of a displayed luminance value can be similarly expressed:

$$B_d = 10^{\beta(L_{a(d)})} (\pi \times 10^{-4} L_d)^{\alpha(L_{a(d)})}$$

where  $\beta(L_{a(d)})$  and  $\alpha(L_{a(d)})$  are as before. To match the brightness of a real world luminance to the brightness of a display luminance,  $B_w$  must equal  $B_d$ . The luminance required to satisfy this can be determined:

$$L_d = \frac{1}{\pi \times 10^{-4}} 10^{\frac{\beta_{a(w)} - \beta_{a(d)}}{\alpha_{a(d)}}} (\pi \times 10^{-4} L_w)^{\frac{\alpha_{a(w)}}{\alpha_{a(d)}}}$$

This represents the concatenation of the real-world observer and the inverse display observer model. To determine  $n$ , the frame buffer value, the inverse display system model is applied to give:

$$n = \left[ \frac{L_d - L_{amb}}{L_{dmax}} \right]^{\frac{1}{7}}$$

giving

$$\tau_{TUMB}(L_w) = \left[ \frac{10^{\frac{\beta_{a(w)} - \beta_{a(d)}}{\alpha_{a(d)}}} (\pi \times 10^{-4} L_w)^{\frac{\alpha_{a(w)}}{\alpha_{a(d)}}}}{\pi \times 10^{-4}} \right]$$

Taking a slightly different approach, Ward [89] searched for a linear transform to give a similar result, while keeping computational expense to a minimum. He proposed transforming real world luminances,  $L_w$ , to display luminances,  $L_d$ , through  $m$ , a scaling factor:

$$L_d = mL_w$$

The consequence of adaptation can be thought of as a shift in the absolute difference in luminance required in order for a human observer to notice a variation. Based on psychophysical data collected by Blackwell [13], Ward defines a relationship that states that if the eye is adapted to luminance level  $L_a$ , the smallest alteration in luminance that can be seen satisfies:

$$\Delta(L_a) = 0.0594(1.219 + L_a^{0.4})^{2.5}$$

Real world luminances are mapped to the display luminances so the smallest discernible differences in luminance can also be mapped, using:

$$\Delta L(L_{a(d)}) = m \Delta L(L_{a(w)})$$

Where  $L_{aw}$  and  $L_{ad}$  are the adaptation levels to the real world scene and display device respectively. The scaling factor,  $m$ , dictates how to map luminances from the world to the display such that a Just Noticeable Difference (JND) in world luminances maps to a JND in display luminances :

$$m = \frac{\Delta L(L_{a(d)})}{\Delta L(L_{a(w)})} = \left( \frac{1.219 + L_{a(d)}^{0.4}}{1.219 + L_{a(w)}^{0.4}} \right)^{2.5}$$

To estimate the adaptation levels,  $L_{aw}$  to  $L_{ad}$ , Ward assumes that the adaptation level is approximately half the average radiance of the image, ( $L_{a(d)} = L_{dmax}/2$ ). Substituting in to equation (above) results in values

from 0 to  $L_{dmax}$ , and dividing by  $L_{dmax}$  then gives values in the required range from [0..1]. The scaling factor is then given by:

$$m = \frac{1}{L_{dmax}} \left[ \frac{1.219 + (L_{dmax}/2)^0 \cdot 4}{1.219 + (L_{a(w)}^0 \cdot 4)} \right]^{2.5}$$

where  $L_{dmax}$  is typically set to  $100\text{cdm}^{-2}$ .

In 1996, Ferwerda et al. [21] developed a model conceptually similar to Ward's, but in addition to preserving threshold visibility, this model also accounted for changes in colour appearance, visual acuity, and temporal sensitivity. Different tone reproduction operators are applied depending on the level of adaptation of the real world observer. A *threshold sensitivity function* is constructed for both the real world and display observers given their level of adaptation. A linear scale factor is then computed to relate real world luminance to photopic display luminance. The required display luminance is calculated by combining the photopic and scotopic display luminances using a parametric constant,  $k$ , which varies between 1 and 0 as the real world adaptation level goes from top to bottom of the mesopic range.

To account for loss of visual acuity, Ferwerda et al. used data obtained from experiments that related the detectability of square wave gratings of different spatial frequencies to changes in background luminance. By applying a Gaussian convolution filter, frequencies in the real world image which could not be resolved when adapted to the real world adaptation level are removed. Light and dark adaptation are also considered by Ferwerda, by adding a parametric constant,  $b$ , to the display luminance, the value of which changes over time.

A critical and underdeveloped aspect of all this work is the visual model on which the algorithms are based. As we move through different environments or look from place to place within a single environment, our eyes adapt to the prevailing conditions of illumination both globally and within local regions of the visual field. These adaptation processes may have dramatic effects on the visibility and appearance of objects and on our visual performance. In order to produce realistic displayed images of synthesised or captured scenes, a more complete visual model of adaptation needs to be developed. This model will be especially important for immersive display systems that occupy the whole visual field and therefore determine the viewer's visual state.

## 2.2.6 Multi-Scale Tone Reproduction Operators

After careful investigation of the effect tone mapping had on a small test scene illuminated only by a single incandescent bulb, Chiu et al [11] believed it was incorrect to apply the same mapping to each pixel. By uniformly applying any tone mapping operator across the pixel of an image, incorrect results are likely. They noted that the mapping applied to a pixel should be dependent on the spatial position in the image of that pixel. This means that some pixels having the same intensities in the original images may have differing intensity values in the displayed image. Using the fact that the human visual system is more sensitive to *relative* changes in luminance rather than *absolute* levels, they developed a spatially non-uniform scaling function for high contrast images. First the image is blurred to remove all the high frequencies, and then the result is inverted. This approach was capable of reproducing all the detail in the original image, but reverse intensity gradients appeared in the image when very bright and very dark areas were close to each other. Schlick [79] proposed a similar transformation based on a rational tone reproduction operator rather than a logarithmic one. Neither of these methods accounted for differing levels of adaptation. Their solutions are based purely on experimental results, and no attempt is made to employ psychophysical models of the HVS.

Larson et al. [50] developed a histogram equalisation technique that used a spatial varying map of foveal adaptation to transform a histogram of image luminances in such away that the resulting image lay within the dynamic range of the display device and image contrast and visibility were preserved. First a histogram of brightness (approximated as a logarithm of real-world luminances) is created for a filtered image in which each pixel corresponds to approximately  $1^\circ$  of visual field. A histogram and a cumulative distribution function are then obtained for this reduced image. Using threshold visibility data from Ferwerda, an automatic adjustment algorithm is applied to create an image with the dynamic range of the original scene compressed into the range available on the display device, subject to certain restrictions regarding limits of contrast sensitivity of the human eye.

In addition to tone reproduction operators being useful for rendering calculated luminance to the screen, they are also useful for giving a measure of the perceptible difference between two luminances at a given

level of adaptation. This function can then be used to guide algorithms, such as discontinuity meshing, where there is a need to determine whether some process would be noticeable or not to the end user.

Gibson and Hubbard [26] have used features of the *threshold sensitivity* displayed by the HVS to accelerate the computation of radiosity solutions. A perceptually based measure controls the generation of view independent radiosity solutions. This is achieved with an *a-priori* estimate of real-world adaptation luminance, and uses a tone reproduction operator to transform luminance values to display colours and is then used as a numerical measure of their perceived difference. The model stops patch refinement once the difference between successive levels of elements becomes perceptually unnoticeable. The perceived importance of any potential shadow falling across a surface can be determined, this can be used to control the number of rays cast during visibility computations. Finally, they use perceptual knowledge to optimise the element mesh for faster interactive display and save memory during computations. This technique was used on the adaptive element refinement, shadow detection, and mesh optimisation portions of the radiosity algorithm.

Discontinuity meshing is an established technique used to model shadows in radiosity meshes. It is computationally expensive, but produces meshes which are far more accurate and which also contain fewer elements. Hedley [42] used a perceptually informed error metric to optimise adaptive mesh subdivision for radiosity solutions, the goal being to develop scaleable discontinuity meshing methods by considering visual perception. Meshes were minimised by discarding discontinuities which had a negligible *perceptible* effect on a mesh. They demonstrated that a perception-based approach results in a greater reduction in mesh complexity, without introducing more visual artefacts than a purely radiometrically-based approach.

## 2.3 Summary

The beginning of this chapter gave a brief introduction to the physiological and psychological workings of the Human Visual System, with particular attention paid to the perception of lightness. It has been established that the correct perception of lightness is critically dependent on the correct perception of the illumination in a scene, as well as the correct perception of depth. As illumination and three dimensional geometry are the key features of any scene, this makes the task of lightness perception an ideal task to investigate the perception of imagery. In subsequent chapters, an experimental framework is presented based on the task of lightness perception which allows comparison of computer generated imagery to real physical scenes.

Later in the chapter, the applications of visual perception in computer graphics were explored. For many applications, computer imagery should not only be physically correct but also perceptually equivalent to the scene it represents. Knowledge of the HVS can be employed to greatly benefit the synthesis of realistic images at various stages of production. Global illumination computations are costly in terms of computation. There is a great deal of potential to improve the efficiency of such algorithms by focusing computation on the features of a scene which are more conspicuous to the human observer. Those features that are below perceptual visibility thresholds have no impact on the final solution, and therefore can be omitted from the computation, increasing efficiency without causing any perceivable difference to the final image. Perceptual metrics involving advanced HVS models can be used to determine the visible differences between a pair of images. These metrics can then be used to compare and evaluate image quality. They can also be used within the rendering framework to steer computation into regions of an image which are in most need of refinement, and to halt computation when differences in successive iterations of the solution become imperceptible.

Future applications will require perceptual accuracy in addition to physical accuracy. Without perceptual accuracy it is impossible to assure users of computer graphics that the generated imagery is anything like the scene it depicts. Imagine a visualisation of an architectural design. Without perceptual accuracy it is difficult to guarantee the architect that the visualisation sufficiently represents their design, and that the completed building will look anything like the computer representation. This chapter discussed how knowledge of the HVS is being incorporated at various stages in the image synthesis pipeline. The problem is that much of the data used has been obtained from specific psychophysical experiments which have been conducted in specialised laboratory environments under reductionistic conditions. These experiments are designed to examine a single dimension of human vision, however, evidence exists to indicate that features of the HVS do not operate individually, but rather functions overlap and should be examined as a whole rather than in isolation. Tone reproduction operators map computed radiance values to display values in a manner

that preserves perception of the original scene. Tone reproduction operators produce a perceptual match between the scene and the image in the hopes that the image may be used predictively.

There is a strong need for the models of human vision currently used in image synthesis computations to be *validated* to demonstrate that their performance is comparable to the actual performance of the HVS. Subsequent chapters introduce a framework that provides a first step toward such validation.

## Chapter 3

# A Psychophysical Investigation

<sup>1</sup> This chapter outlines the steps involved in designing and building a psychophysical experiment to facilitate easy comparison of real scenes and synthetic images by a human observer [60, 59, 85, 56, 57, 58]. The apparatus includes a calibrated light source and a well articulated scene containing three dimensional objects placed within a custom built environment which were placed to introduce shadows into the scene, paired with various synthetic images of that scene. Measurements of virtual environments are often approximate. However, for the application described in this thesis an accurate description of the environment is essential to avoid introducing errors at such an early stage of the rendering pipeline. Also, once an image has been rendered, it is important to display this image in the correct manner taking into account the limitations of the display device. The measurements required for this study and the equipment used to record them are described herein. Also described is the process involved in designing the experiment to ensure a robust set-up which yields valid results. This is achieved by employing psychophysical techniques for the study of the HVS. Psychophysics comprises a collection of methods used to conduct non-invasive experiments on humans, the purpose of which is to study mappings between events in an environment and levels of sensory responses [19, 25]. This thesis is concerned with the levels of human visual responses, the goal being to examine perceptual behaviour in response to real and computer imagery.

Use of a perceptual lightness matching procedure is chosen because it is sensitive to errors in perceived depths, as well as the more obvious dependence on the array of luminance values in the scene. Lightness constancy depends on a correct representation of the three-dimensional structure of the scene [28, 27] and the illumination in the scene. Any errors in depth/illumination perception, when viewing the computer model, will result in errors of constancy, and thus poor psychophysical matching performance. This makes the lightness matching task a good candidate for comparing scene content in real and synthetic images.

This chapter provides a detailed description of the construction of an experimental framework which enables human observers to perform the light-matching task in real scenes and computer generated representations. Task performance in each case (real versus rendered) can then be compared to give a measure of perceptual equivalence. To ensure correctness in such a study it is necessary to begin with a smaller *pilot study* to identify any shortcomings of the framework, which can then be corrected *before* a fuller study is undertaken.

### 3.1 The Pilot Study

When conducting experiments it is common to conduct a relatively small, preliminary study designed to put the experimenter in a better position to conduct a fuller investigation. Such studies, *pilot studies*, are useful for working through the practical details that are difficult to anticipate, and also help to familiarise the experimenter with logical and theoretical facets of the experiment that might not be apparent from just thinking about the situation. Often during the pilot study, the experimenter recognises needed controls, flaws in logic and so on [25]. For these reasons a small preliminary study preceded the main experiments. For the pilot study a simple test scene was constructed that allows implementation and testing of various conditions. The main function of this section is to describe precisely how the pilot study was conducted,

---

<sup>1</sup>written by Ann McNamara ann.mcnamara@tcd.ie

discuss the results obtained and the modifications necessary to eliminate some unwanted influences present in the pilot method.

### 3.1.1 Participants in the Pilot Study

Fifteen observers participated in the pilot study. In each condition participants were naive as to the purpose of the experiment. All reported to have normal or corrected-to-normal vision. The average age of participants was twenty-five, the group was made up of eleven males and four females.

### 3.1.2 Apparatus

This study required an experimental set-up comprised of a real environment and a computer representation of that environment. On face value this may appear to be a simple task, however *accurate* modelling of even simple scenes can prove difficult. For this study it was important to take great care during both the modelling and display stages of the rendering pipeline. The study seeks to evaluate the *lighting simulations* used. In order to draw reliable conclusions from the study, errors at the modelling and display stages must be avoided, otherwise errors are likely to arise from any one of the three stages involved in generating the image. Here we describe the equipment used to construct the real world test environment, along with the physical measurements performed to attain the necessary input for the synthetic representations.

#### The Real Scene

The test environment was a five sided box of 557 mm high, 408 mm wide and 507 mm deep, with an opening on one side. All interior surfaces of the box were painted with white matt house paint. To the right of this enclosure a chart showing thirty grey level patches, labelled as in Figure 3.2, were positioned on the wall to act as reference. The thirty patches were chosen to provide perceptually spaced levels of reflectance from black to white, according to the Munsell Renotation System [92]. A series of fifteen of these grey level patches were chosen at random, reshaped, and placed in no particular order within the physical environment. A small front-silvered, optical mirror was incorporated into the set up to facilitate

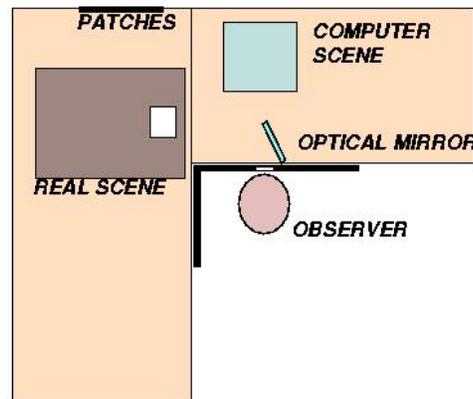


Figure 3.1: Experimental set up

alternation between the two settings, viewing of the original scene or viewing of the rendered scene on the computer monitor. When the optical mirror was in position, participants viewed the original scene. In the absence of the optical mirror the computer representation of the original scene was viewed. The angular subtenses of the two displays were equalised, and the fact that the display monitor had to be closer to the participant for this to occur, was allowed for by the inclusion of a +2 diopter lens in its optical path; the lens equated the optical distances of the two displays, Figure 3.1.

#### Illumination

The light source consisted of a 24-volt quartz halogen bulb mounted on optical bench fittings at the top of the test environment. This was supplied by a stabilised 10 amp DC power supply, stable to 30 parts per million in current. The light shone through a 70 mm by 115 mm opening at the top of the enclosure. Black masks, constructed of matt cardboard sheeting, were placed framing the screen and the open wall of the

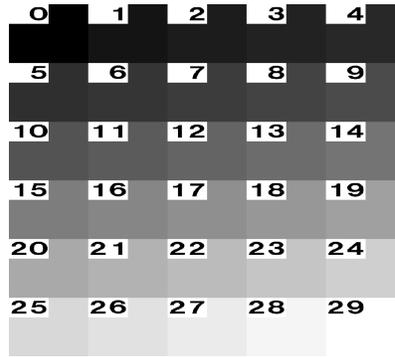


Figure 3.2: Reference patches

enclosure, a separate black cardboard sheet was used to define the eye position. An aperture in this mask was used to enforce monocular vision, since the VDU display did not permit stereoscopic viewing, and it was felt that therefore monocular viewing would give a fair comparison of the two scenes.

### 3.1.3 The Graphical Representation

The geometric model of the real environment was created using Alias Wavefront [90]. To describe a physical environment, a lighting simulation program takes as input the geometry of the environment and objects in the environment, the properties of the light source(s), and the material characteristics of the surfaces and objects. To obtain input to the lighting simulation software the following measurements were required.



Figure 3.3: The Minolta CS-100 chroma meter

**Geometry:** A tape measure was used to measure the geometry of the test environment. Length measurements were made with an accuracy of the order of one millimetre.

**Illumination:** The photometric instrument used throughout the course of the pilot experiments was the Minolta Spot Chroma meter CS-100. The Minolta chroma meter is a compact, tristimulus colourimeter for non contact measurements of light sources or reflective surfaces, Figure 3.3 [64]. The one degree acceptance angle and through the lens viewing system enables accurate targeting of the subject. The chroma meter was used to measure the chromaticity and luminance values of the materials in the original scene and from the screen simulation. The luminance meter was also used to take similar readings of the thirty reference patches. The illuminant was measured by illuminating an Eastman Kodak Standard White powder, pressed into a circular cavity, which reflects 99% of incident light in a diffuse manner. The chroma meter was then used to determine the illuminant tristimulus values. Measured chromaticity values were converted to RGB triplets by applying a matrix based on the chromaticity co-ordinates of the monitor phosphors and a monitor white point [84].

**Materials:** The chroma meter was used for material chromaticity measurements. To ensure accuracy of the measurements five measurements were recorded for each material, the highest and lowest luminance magnitude recorded for each material discarded and an average was taken of the remaining three values. The CIE (1931) xy chromaticity co-ordinates of each primary were obtained and the relative luminance for each phosphor were recorded using the chroma meter, measurements were transformed to RGB triplets using the same method applied for the illumination measurements.

Using the required measurements as input the *rendered* image was then created using the radiance lighting simulation package [89] to generate the graphical representation of the real scene. Radiance is a physically based lighting simulation package, which means that physically meaningful results may be expected, provided the input to the renderer is meaningful. Radiance is described in detail in section 1.2.3. Radiance uses RGB tristimulus values to describe surface characteristics, so the values obtained for the illuminant and surfaces in the scene need to be transformed from xy chromaticity co-ordinates to values usable by radiance, namely tristimulus RGB values.

The CIE (1931) xy chromaticity co-ordinates of each primary were obtained using the Minolta chroma meter [64]. Then these values were transformed to screen RGB tristimulus values as input to Radiance lighting simulation program using the following method [84]. Using measured xy values the z for each primary can be calculated using the relationship

$$x + y + z = 1$$

For each phosphor the relative luminances are recorded using the chroma meter. These are normalised to sum to 1. The resulting values are the *Y tristimulus values*. From the Y tristimulus values and the chromaticity co-ordinates for each primary we compute the X and Z tristimulus values using the formulas

$$\begin{aligned} X_r &= Y_r \times \frac{x_r}{y_r} \\ Z_r &= Y_r \times \frac{z_r}{y_r} \end{aligned}$$

Similarly for green and blue phosphor. By this method a matrix,  $T$ , is constructed.

$$T = \begin{vmatrix} X_r & X_g & X_b \\ Y_r & Y_g & Y_b \\ Z_r & Z_g & Z_b \end{vmatrix}$$

Now to compute the tristimulus (RGB), first the X, Y and Z tristimulus values need to be calculated:

$$\begin{aligned} X &= x \times \frac{Y}{y} \\ Y &= Y \\ Z &= (1 - x - y) \times \frac{Y}{y} \end{aligned}$$

Then applying the following matrix gives the RGB values which can then be used as input to radiance.

$$\begin{vmatrix} R \\ G \\ B \end{vmatrix} = \begin{vmatrix} T^{-1} \end{vmatrix} \begin{vmatrix} X \\ Y \\ Z \end{vmatrix}$$

The entire experimental set-up resided in an enclosed dark laboratory in which the only light sources are the DC bulb (shielded from direct view) or illumination from the monitor. As described earlier, Gilchrist [10, 28, 30] has shown that such an experimental environment is sufficient for the purposes of this experiment.

### 3.1.4 Procedure

The participants' task was to match grey level patches within the physical environment, to a set of control patches, Figure 3.4. Then participants were asked to repeat the same task with the original environment replaced by its computer representation, and in addition some slight variations of the computer representation, such as changes in Fourier composition (blurring), see Figure 3.5. The reason for deliberate distortion of the original rendered image is due to the fact that many image comparison techniques make assumptions about what is important in an image, i.e. spatial frequency. What really is important may depend on *task*. So while slightly disturbing spatial frequency (for example) in an image will cause a numerical



Figure 3.4: Patches within the physical environment(left) are compared with a set of control patches (right)

technique to produce large differences, the difference may not be detectable by a human observer. The HVS is not a photometer, as discussed in section 3.2, while pixel by pixel comparisons might be sufficient, they can give misleading results as the outcome does not always correspond to the *perceptible* differences between images. In the **Original Scene**, physical stimuli were presented in the test environment, described



Figure 3.5: Rendered image, left, with blurring, right

in the previous section. Participants viewed the scene monocularly through a fixed viewing position. The experiment was undertaken in a darkened room under constant, controlled illumination conditions.

While viewing the **Computer Simulated Scene**, representation of the stimuli, rendered using Radiance, were presented on the monitor of a Silicon Graphics 02 machine. Again, participants viewed the screen monocularly through a fixed viewing position.

## 3.2 Results and Discussion of Pilot Study

A number of different conditions were investigated during the complete pilot study, ranging from a rendered image to its real counterpart, to comparing distorted, deliberately incorrect images to the physical scene. For the pilot study data were obtained for fifteen participants. Participants had either normal or corrected-to-normal vision. Each participant performed a number of conditions, in random order, and within each condition the participant's task was to match the fifteen grey test patches to a reference chart on the wall. Each patch was matched only once. Although each participant started on a randomly chosen patch, they

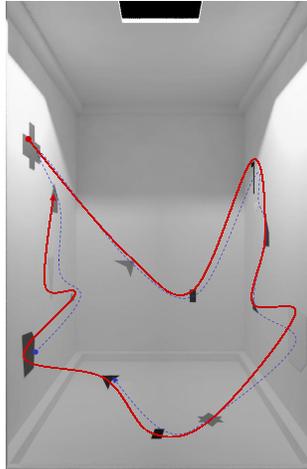


Figure 3.6: Participants started at different patches but followed the same path

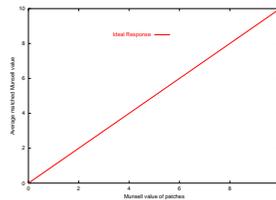


Figure 3.7: Ideal response: obtained if each participant matched each object to its correct corresponding Munsell Value

followed the same path around the patches. As illustrated in Figure 3.6, one participant may have followed the red track while another followed the blue track, but the sequence in which they were visited was the same. This can lead to problems due to the order in which patches are viewed, (section 3.4.1).

Figure 3.7 shows the results that *would* be obtained if each participant matched each patch *exactly* to its corresponding patch on the chart. A perfect set of data would lie along a 45° diagonal line. The lightness matching technique is used as a measure of perceptual response to a physical scene. The HVS is not perfect, and so some errors are made in the lightness matching task. These “errors”, or deviations from what would be an *ideal* response are shown in Figure 3.8.

The blue line shows the *actual* response to the real scene. Now this response becomes the “ground-truth” against which all other responses are measured. The closer a response is to this line, the closer the perceptual match between the scene and the image. The experimental data for the real environment lie close to this line, with some small but systematic deviations for specific test patches. These deviations show that lightness constancy is not perfect for the original scene. What this means is as follows: when observing a given scene, small (but significant) errors of lightness perception are likely to occur. A perceptually-perfect reconstruction of the scene should produce a very similar pattern of errors if it is perceptually similar to the original. So for high quality renderings the response should follow the ground-truth closely in order to conclude the image is a faithful representation.

The main idea behind the pilot study was first to establish the framework and ensure the chosen task was sufficient as a comparison measure. For this reason a number of different conditions were run to evaluate the feasibility of the framework. This is a pilot study so only simple statistics were applied for a basic analysis of the results.

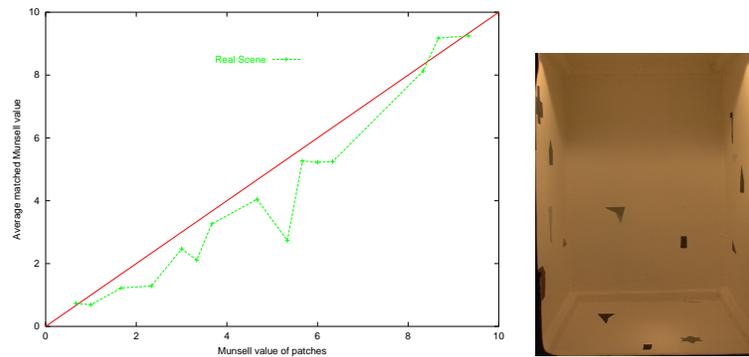


Figure 3.8: Response of average matching of Munsell values to REAL

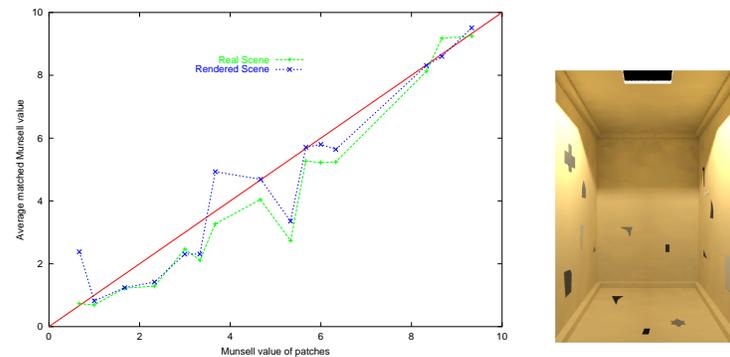


Figure 3.9: Response of average matching of Munsell values real with rendered, with the rendered image shown on the right

### 3.2.1 Condition A: Real Versus Rendered

The purpose of the first experiment was to determine how similar perceptual response to a rendered scene was to that of the real scene it represents. Participants performed the lightness matching task on both the real scene and the rendered image. Results are averaged over all participants to give the average match per patch.

Figure 3.9 shows the results obtained for the comparison. The x-axis gives the *actual* Munsell value of each patch, the y-axis gives the *matched* Munsell value, averaged across the participants. (Results for following conditions are presented on the same axes)

In general, it can be seen that the matched values are very similar to those of the original scene, in other words, the same (small) failures of constancy apply both to the real scene and the rendered image. This, in turn suggests that there is no significant perceptual difference between the original scene and the rendered version. This is in spite of the fact that the mean luminance of the rendered versions was lower by a factor of about 30 compared to the original.

### 3.2.2 Condition B: Real Versus 30% Blurred

In condition B, participants were presented with a 30% blurred version of the rendered image used in condition 1, Figure 3.10. Blurring was achieved by passing a gaussian filter over each pixel in the image. Again the matched values are very similar to those chosen in the original scene. This suggests that the HVS is not very sensitive to slight variations in spatial frequency.

Under these conditions the blurred version looked very different subjectively, due to visible softening of edges, but again similar data were obtained.

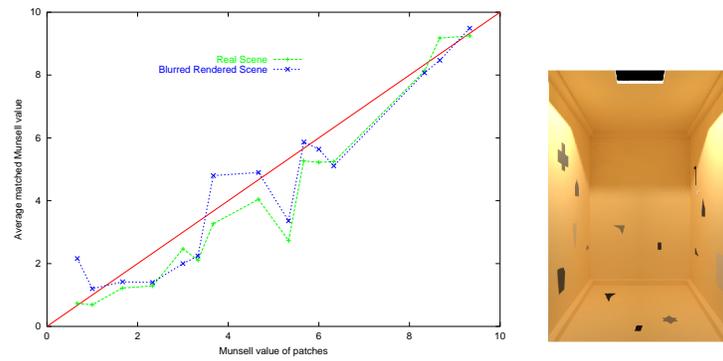


Figure 3.10: Response of average matching of Munsell values real with blurred rendered

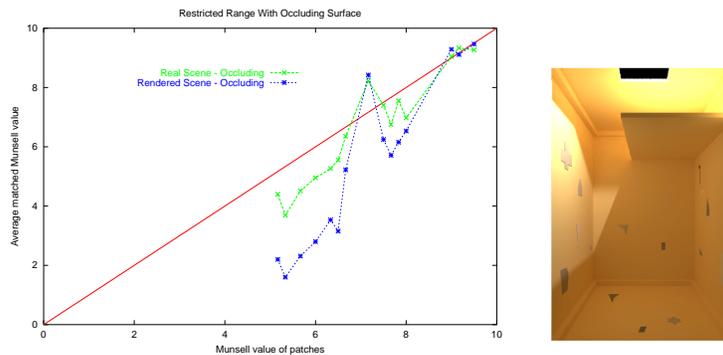


Figure 3.11: Comparison of average matching of Munsell values with real with partial occlusion

### 3.2.3 Condition C: Real Versus Partially Occluded

As a first attempt at introducing shadows into the environment, a “shelf” like structure was placed at the top of the test scene to limit the amount of light falling on the regions, Figure 3.11. This was done in both real and virtual scenes. The results are plotted as for the previous conditions, and show that despite the shadow, the response is again similar in both cases, suggesting little perceptual difference between the two.

### 3.2.4 Conclusion from Conditions of Pilot Experiment

The pilot experiment shows for simple scenes that lightness constancy is extremely robust against changes to the rendered image.

For each condition in the pilot experiments the biggest deviations occur at regions 9 and 16. This is interesting as both these regions are partially in shadow. At this stage however, no conclusive evidence exists as to the reason for such deviation, due to the manner in which the pilot study was conducted. For instance, participants followed the *same* sequence when performing the task, so regions were tested in an *orderly* manner, a more dependable approach should examine regions in *random* order to avoid *ordering effects*. Also, the control chart was separate from the scenes/images under consideration, forcing participants to shift their gaze between the two. *Adaptation* is affected when moving between scene and control chart, it would be more satisfactory if the viewer was allowed to adapt fully to the scene in question before attempting evaluations. So any one of a number of factors may contribute to these mis-matches. To correctly conclude which factor contributes to these fluctuations these issues need to be addressed before the comprehensive study is made. This will allow conclusions to be drawn in the absence of such errors.

### 3.2.5 Compared to Numerical Techniques

In an effort to compare this method to currently available techniques it is possible to reduce the pattern of results to a single value as follows :

- taking the matches to the original scene as reference, calculate the mean signed deviation for the rendered, blurred rendered and partial occlusion functions.
- Compute the mean and standard deviation of these

Table 3.1 shows the results obtained. A value of zero in this table would indicate *perceptually perfect* match; the actual values given come close to this and are statistically not significantly different from zero. This, therefore, again indicates high perceptual fidelity in both versions of the rendered scene.

Compared to Real	Mean Munsell Value Deviation
Rendered Scene	-0.37 ( $\sigma = 0.44$ )
Blurred Scene	-0.23 ( $\sigma = 0.57$ )
Partial Occlusion Scene	-1.16 ( $\sigma = 0.93$ )

Table 3.1: Comparison of test images to real environment

How do these values compare to other methods? Using the algorithm of Daly [20] a 5.04% difference between the rendered and blurred rendered images is returned. As a comparison, a left-right reversal of the image gives a difference value of 3.71%; and a comparison of the image with a white noise grey level image results in a difference value of 72%. Thus, the algorithm suggests that there is a marked difference between the rendered image and blurred rendered image; for example this is a 36% greater difference than that with a left-right reversed image. (This difference increases for less symmetrical images). However, our method suggests that these two scenes are perceptually equivalent in terms of our task. It *may* therefore be that there is a dissociation between our method and that of the Daly technique. In addition, the algorithmic method cannot give a direct comparison between the original scene and the rendered version; this could only be achieved by frame grabbing the original which is a process likely to introduce errors due to the non-linear nature of the capture process. This idea of comparing the psychophysical approach to image quality metrics is explored further in [56].

**In Summary:** the results of the pilot study show that the rendered scenes used in this study have high perceptual fidelity compared to the original scene, and that other methods of assessing image fidelity yield results which are markedly different from ours. The results also imply that a rendered image can convey albedo.

## 3.3 Conclusions and Problems with Pilot Study

The main purpose of the pilot study was to test the feasibility of using psychophysical techniques to give an index of image fidelity. Shadows are important for the correct perception of a scene [29]. Although the pilot study gave confidence in the lightness-matching task as a basis for scene-image comparison, the environment used was very basic, consisting of flat polygonal patches placed in an empty five sided cube.

The pilot study demonstrated the potential of the lightness matching based experiment to measure the perceptual correspondence between real and synthetic images. As previously mentioned the purpose of a pilot study is to reveal problems with the experiment before a full study is made. The following difficulties arose during the pilot study and needed to be addressed before a more comprehensive study was made. The main problems to be addressed were:

**Ordering Effects:** Each participant matched each patch in the same sequence, so, for example, each participant examined patch 1 *directly* after patch 6. This can lead to *ordering effects* [19] which means the perception of a patch may be influenced by the perception of the patch examined immediately previous. Consider looking at a light object immediately after viewing a dark object. This may cause the lightness of the light object to be artificially exaggerated, and so will be perceived as *lighter* than it really is. To avoid the *ordering effect* from the experiment, it is better to allow participants to examine the objects in a *random* manner rather than a predefined sequence

**Patch Selection:** As described the experiment involved asking participants to match each patch in the scene to one of a set of control patches hanging on the wall. In order to view this control chart and make a selection, the participant was forced to move his gaze *away* from the scene. So the viewer switched between adaptation to the scene, and adaptation to the control chart. As the process of adaptation can take several minutes, it is not a good idea to allow alternation between views. A better approach is to first *train* each participant to *learn* the control chart, thus allowing selection of a match from memory *without* moving their gaze from the scene in question, guaranteeing that the adaptation level remains constant.

**Measuring Equipment:** The device used to measure the reflectance and emissive properties for the pilot study outputs Y, xy tristimulus values. A more sophisticated device allows more precise measurement, resulting in more accurate readings. For this reason a *spectroradiometer* is used for measuring input in the full experiment.

This suite of pilot experiments instilled confidence in the methodology, while establishing some common methods and conditions. So while the results of the pilot study look promising, strong claims cannot be based on the results of this set of experiments due to the shortcomings of the experimental design. By taking action to combat the shortcomings a more complete study can be undertaken. Results from an improved framework can then be used with a higher level of confidence to draw conclusions about HVS perception in real and synthetic images and the differences between the two.

The actions taken to remedy these shortcomings are presented in the next section.

## 3.4 Experiment: The Physical Framework

Although the pilot study demonstrated the usefulness of the technique, more importantly it highlighted some of the flaws in the framework which may otherwise have escaped unnoticed. These flaws and the actions taken to remedy them are addressed here before moving on to the discussion of the main set of experiments which form the foundations for the new image comparison framework. To introduce more complexity into the environment, the idea of shadowing used in condition C of the pilot study is extended by replacing two dimensional patches by three dimensional objects to allow the exploration of effects such as shadowing and depth perception.

### 3.4.1 Ordering Effects

In the pilot experiments, participants were asked to match patches in the physical scene to patches on the Munsell Chart. Each participant started on a different (randomly selected) patch, but then followed the same path as before, for example, patch 4 was always examined directly after patch 15 and directly before patch 6. This leads to what is known in experimental psychology as *ordering effects*. To explain this phenomenon consider how observing a dark object immediately after a brighter object may influence perception of the dark object. As an extreme example bear in mind the experience of matinee cinema goers, when on emerging from the dark cinema theatre find themselves temporarily “blinded” by their bright environment. Ordering effects are perhaps the reason for such sharp “spikes” in the data collected during the pilot experiments. To eliminate any doubts and error introduced by ordering effects, participants were asked to examine objects in the new set up in a random order. Each participant began by examining a different randomly selected object, followed by another randomly selected object, and so on, examining randomly selected objects until each object in the scene had been tested. In addition to randomisation of object examination, the order of presentation of images was conducted in a completely random manner. For example, if a high quality image was presented first to every participant, this may affect their perception of lower quality images. This may artificially exaggerate the “goodness” or “badness” of an image. To avoid this scenario images are presented randomly, including presentation of the real environment.

### 3.4.2 Matching to Patches

Through the course of the pilot study it became apparent that moving the eye between the screen and the control patches was unacceptable. In addition to adding to time taken to complete each experiment this procedure introduced possible errors due to adaptation effects. Adaptation occurs over time, quickly

switching gaze between scene and control patches did not allow the viewer to fully adapt to *either* the scene or the patches, and so potential errors of judgement are likely when testing in this manner. A new method of matching to patches was devised for the main experiment, as described in Section 4.8.1.

### 3.4.3 Custom Paints

Patch Number	Patch Reflectance	Paint Reflectance	Paint Name
0	0.049429567	0.04708, 0.048301	Black, 9:1
1	0.054809032	-	-
2	0.066896445	0.063563	4:1
3	0.060829058	-	-
4	0.083818697	0.077891	2:1
5	0.087025778	-	-
6	0.101228624	0.096187	3:1
7	0.112001329	0.113279	1:1
8	0.123494339	-	-
9	0.122389474	0.138329	2:3
10	0.159463761	-	-
11	0.151914931	-	-
12	0.15708198	-	-
13	0.174353557	-	-
14	0.168015403	0.161078	3:7
15	0.22593297	0.200236	1:4
16	0.270090158	-	-
17	0.291413257	-	-
18	0.295541683	-	-
19	0.339228031	0.328583	1:9
20	0.367078689	-	-
21	0.37551392	-	-
22	0.394189794	-	-
23	0.434945453	0.420213	1/2:9 1/2
24	0.464873337	-	-
25	0.506370125	0.529190123	1/4:9 7/4
26	0.551247036	0.531189	1/8:9 7/8
27	0.610832586	-	-
28	0.682142642	-	-
29	0.755641895	0.879516	White

Table 3.2: Paint reflectance along with reflectance of corresponding patch

Due to the three dimensional nature of objects in the new scene, simple two dimensional patches were no longer appropriate. To accommodate the three dimensional objects, custom paints were mixed, using precise ratios to serve as the basis for materials in the scene. To ensure correct, accurate ratios were achieved 30ml syringes were used to mix paint in ratios of black to white as shown in Table 4.3. Although the ratios of each paint sample was known it was still necessary to determine the *reflectance* values for each paint sample. This was done using the same method used to determine the illumination and materials during the pilot study. The paint samples were measured against a standard. This standard was set by illuminating an Eastman Kodak Standard White powder, pressed into a circular cavity, which reflects 99% of incident light in a diffuse manner. This circular cavity was housed in a black oblong box, coated with black velvet (to avoid interreflections). At one end of the box was an opening to allow the light to shine through, and an opening on the side to allow the measurements to be taken as shown in Figure 3.12. The custom paints were used to paint fifteen samples which were cut such that they could be slotted into the oblong box, and measured under controlled conditions.

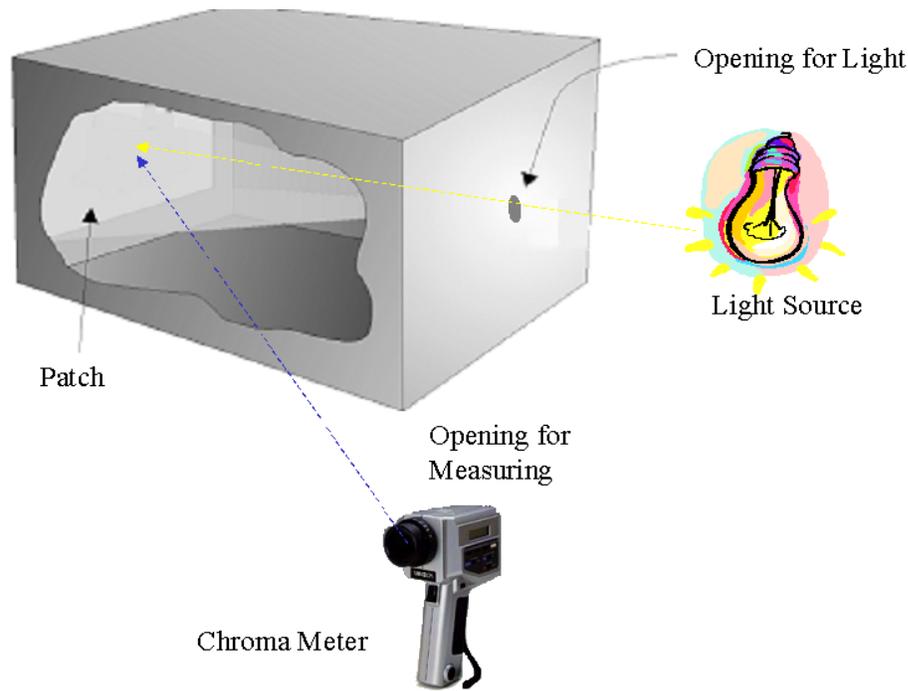


Figure 3.12: Sketched representation of measuring environment

The relationship between the patches that the participants were trained on and the custom paints used on the objects in the scene is shown in Figure 3.13.

### 3.4.4 Three Dimensional Objects

While the pilot study gave confidence in the method, it became obvious that a full investigation would require a more complex scene, showing shadows and three dimensional objects. Several objects were chosen ranging from household objects, to custom made oblong wooden pillars. The objects and their dimensions are given in Table 4.3. Each object was painted with a randomly chosen custom paint from the samples described in Section 4.3.

## 3.5 Modifications to the Original Apparatus

Extending the environment to introduce complexity meant some additional measurements were needed. In the pilot study the patches were generated using known reflectance, then verified using the Minolta CS-100 Chroma Meter. For the main experiment, although the *ratios* of the paint were known, their *reflectance* needed to be measured. More comprehensive measurements can be achieved using a spectroradiometer to obtain measurements across the visible spectrum. A *spectroradiometer* is an instrument used for detecting and measuring the intensity of radiant thermal energy. The radiometer is essentially a partially evacuated tube within which is mounted a shaft with four light vanes. One side of the vanes is blackened and the other is of polished metal. Upon receiving external radiation, the blackened side absorbs more heat than the polished side, and the free molecules in the bulb react more strongly on the dark side, pushing the dark side away from the source of radiation. The spectroradiometer used for these measurements was a TOPCON spectroradiometer (model sr-1)<sup>2</sup>. The sr-1 outputs the spectral radiance of the sample under examination, in 5nm increments.

For the actual experiments more complete *spectral reflectances* of the paints used were measured using a TOPCON-100 spectroradiometer. This data is recorded in Appendix A.

<sup>2</sup>kindly loaned for this study by DERA

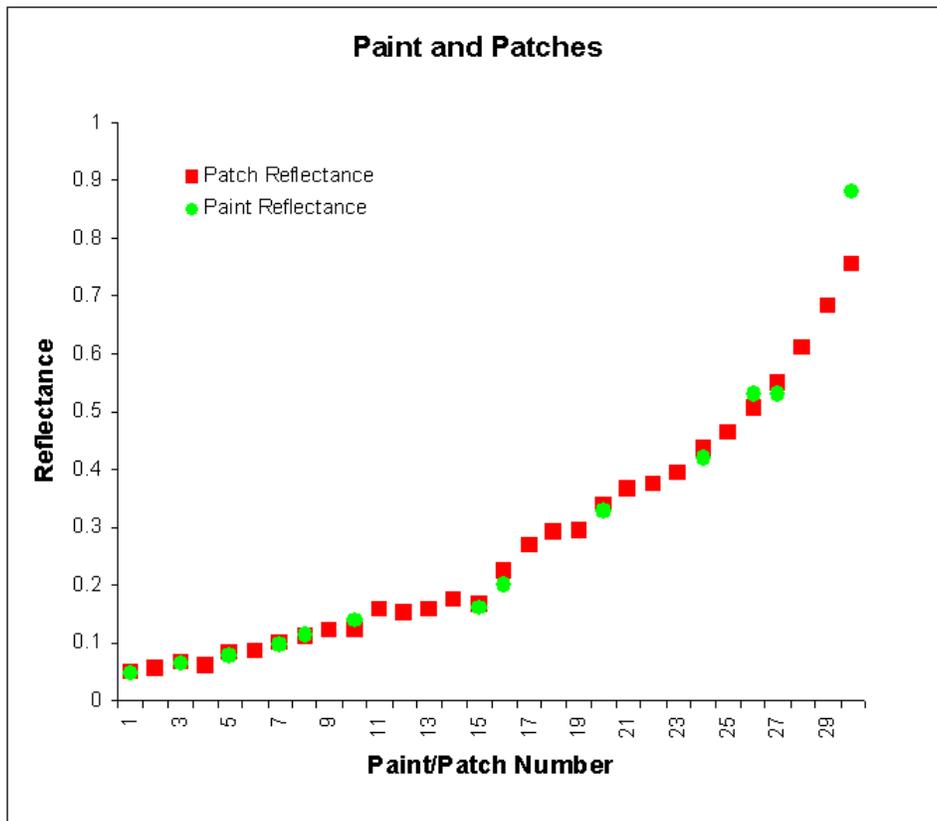


Figure 3.13: Corresponding paint and patch reflectance

## 3.6 Experimental Design

### 3.6.1 Participants

Eighteen observers participated in each experiment. In each condition participants were naive as to the purpose of the experiment. All reported to have normal or corrected-to-normal vision. Participants are assigned to groups in such a way that groups are approximately equivalent, this is achieved through **randomisation**, a term used extensively throughout the remainder of this chapter.

### 3.6.2 Randomisation and Counterbalancing

Experiments are designed to detect the human visual response to a certain condition, and to detect the response to **only** that condition. It may seem the best manner to control for other effects would be to identify them and eliminate their effects. However, to identify all variables might conceivably influence an experimental outcome. By *randomising*, an experiment is arranged so that extraneous features tend to be equally represented in experimental groups. Random assignment to conditions in an experiment is inclined to produce equal representation of variables requiring control. Randomisation of order of presentation means conditions are as likely to occur in one order as another. It also means that presenting a condition in one position for a given participant, say light environment first, has no influence on whether the same condition is presented in any other position. If order of presentation is completely randomised this would mean no “balancing” occurs. It is assumed a truly random process will eventually result in a fairly even balance of various orders of presentation. Randomisation has the distinct disadvantage that imbalances in order of presentation may occur simply on a chance basis. This is especially true if the number of conditions is small. Randomisation will even things out in the *long run* but only if the experiment is extensive. It is even possible that the same condition will be presented in the same manner each and every time just as

Object	Dimensions	Paint
Pyramid	17 x 14 x 14	1:9
Small Cylinder	7.6 x 7.6 x 9.7	WHITE
Ledge on Small Cylinder	15 x 25 x 7.0	3:2
Small Sphere at Front	radius 2.2	4:1
Small Cube at Front	7 x 7 x 7	9:1
Tall Rectangle on Right	7 x 7 x 23	3:7
Large Sphere	radius 8.8	2:3
Tall Cylinder on Right	7.6 x 7.6 x 24.1	1:4
Ledge on Tall Cylinders	15 x 15 x .6	1:1
Small Cylinder	7.6 x 7.6 x 9.7	BLACK
Tall Cylinder on Left	7.6 x 7.6 x 24.1	BLACK
Tilted Box	11.5 x 11.5 x 11.5	2:1
Box Under Tilted Box	15 x 25.5 x .6	$\frac{1}{4} : 9\frac{3}{4}$
Ledge on Rectangle on Right	41 x 15 x .6	3:7
Tall Rectangle on Right	7 x 7 x 32.3	$\frac{1}{8} : 9\frac{7}{8}$
Walls		WHITE

Table 3.3: Objects in the mixed environment, their placement and assigned paint

it is possible to draw four aces from a deck of cards without cheating [19]. To avoid such imbalances **counterbalancing** is often used instead of randomisation. Counterbalancing means that the experimenter ensures that various possible presentation orders occur equally often.

In this study there are three distinct conditions, the design of the experiment is counter balanced by ensuring each condition is presented first one third of the times, second one third of the time and last one third of the time. By counterbalancing the effect of either of the three conditions being presented first will be present equally in each condition. By examining results when a treatment comes first and comparing results when the same treatment comes second or third, effects of ordering can be seen. Many variables have effects that need to be taken into account. Fatigue or hunger for example can be present depending on the time of day the experiment is conducted. This condition must therefore be counterbalanced to avoid unwanted influences on the data. For this experiment, time was divided into three zones, namely morning, middle of day and afternoon. This division worked out neatly resulting in eighteen different combinations of time of day/condition. Using eighteen participants, one for each combination counterbalances the experiment, thus removing any time of day or ordering effects. Experimentation time for each condition was approximately 45 minutes, with 54 conditions meant the experiments ran over 50 hours.

### 3.7 Procedure

Each object was matched once only and the order in which each participant performed the matches was varied between participants and conditions. The experimental conditions were kept constant over each participant, and the instructions given were the same in each case. To avoid data contamination it is critical to keep treatments as similar as possible across participants. In general, such explanations were given when the question was raised by an observer, the task being clear to most observers. The following steps outline a single experiment.

### 3.8 Experiment

### 3.8.1 Training on Munsell Chips



Figure 3.14: Patch arrangement used to train participants with reference chart

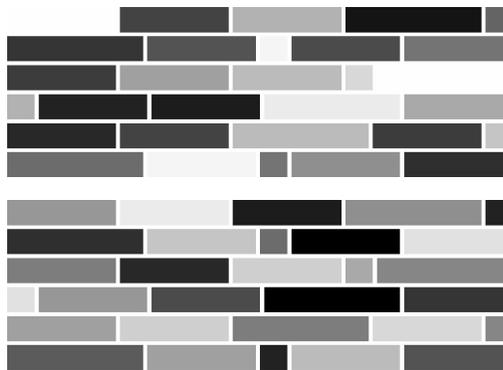


Figure 3.15: Patch arrangement used to train participants without the reference chart)

In [60], the task involved matching regions to a control chart which meant observers had to look away from the scene under examination to choose a match. Moving between scene and chart may affect adaptation to the scene in question, also the view point is not fixed, for this reason we decided to *train* participants on the control patches first. Once trained on the patches participants could then recall the match from memory. Training was conducted as follows. Observers were asked to select, from a numbered grid of 30 achromatic Munsell chips presented on a white background, a sample to match to a second unnumbered grid (Figure 3.14) which was simultaneously displayed on the same background, under constant illumination. The unnumbered grid comprised 60 chips. At the start of each experiment participants were presented with two grids, one an ordered numbered regular grid the other an unordered unnumbered irregular grid comprising one or more of the chips from the numbered grid.

Both charts were hung on the wall approximately one meter from the participant. Each participant was asked to match the chips on the unnumbered grid to one of the chips on the numbered grid on the left. In other words they were to pick a numbered square on the left and place it right next to the grid on the right which in the grid would match it exactly. This is done in a random manner, a laser pointer<sup>3</sup> was used to point to the unnumbered chip under examination. Then the numbered chart was removed, and the unnumbered chart replaced by a similar chart but one where the chips had a different order. Participants repeated the task, this time working from memory to recall the number each chip would match to. The results are shown in Figure 3.16. The graph on the left shows the average match across 18 participants, both with the reference chart and without the reference chart. The graph on the right shows the average correlation. This correlation gives an indication of the extent to which two sets of data are linearly related. A value close to 1 indicates a strong relationship, while a value of 0 signifies there is no linear relationship. A correlation of 1 would result if the participant matched each unnumbered patch to its corresponding numbered patch, in reality

<sup>3</sup>non-invasive medium

this is not the case and some small errors are made, what we need to determine is if the errors made when matching from memory i.e. without the chart are about the same size as the errors made with the reference chart in place. The correlation value when matching the patches with the chart in place is 0.96, and when matching from memory the result is 0.92, indicating a very small difference between the two conditions. From this small difference we can conclude that participants are *approximately as good* at matching the patches without the reference chart in place. Thus, this training paradigm proved to be reliable and stable. This has the dual benefit of speeding up the time taken per condition, as well as ensuring participants do not need to move their gaze from image to chart, thus eliminating any influence due to adaptation.

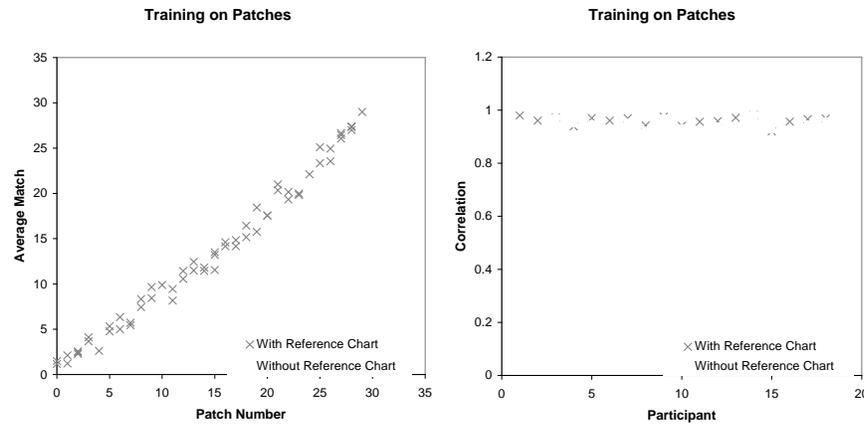


Figure 3.16: Results from matching patches with and without the reference chart

### 3.8.2 Matching to Images

Each participant was presented with a series of images, in a random order, one of which was the real environment. Participants were not explicitly informed which image was the physical environment. Each object was matched only once, however for certain paint samples there was more than one object painted with that paint in the environment.

The media used for stimulus presentation was a gamma corrected 20-inch monitor with the following phosphor chromaticity co-ordinates:

$$\begin{aligned} x_r &= 0.6044 & x_g &= 0.2808 & x_b &= .1520 & x_w &= 0.2786 \\ y_r &= 0.3434 & y_g &= 0.6016 & y_b &= .0660 & y_w &= 0.3020 \end{aligned}$$

### 3.8.3 The Graphical Representations

Eleven images (conditions) were presented to each participant, the real environment and ten synthetic images. Conditions are listed here along with the aims that we hoped to achieve from the comparison.

- A. **The Real Environment:** Human judgements of lightness are not perfect. This was illustrated during the pilot study when participants made small errors when comparing patches in the real scene to the set of control patches on the Munsell Chart. The pattern of these errors thus becomes the "ground truth" against which the pattern of errors in the synthetic images are compared. If similar patterns occur in both cases, it can be concluded that the synthetic representation has high fidelity to the real environment.

- B. **Photograph:** Comparison to a photograph is needed to enable us to evaluate our method to more traditional image comparison metrics. The reasoning behind this is that most current techniques compare to “reality” by comparing to a captured image. We wanted to see if this is equivalent to comparing to a real physical environment and so included a photograph, taken with a digital camera, as one of our test images.
- C. **Radiance: 2 Ambient Bounces:** A Radiance [89] image generated using 2 ambient bounces is generally considered to be a high quality image. Here we wanted to determine if 2 ambient bounces gives a similar perceptual impression to an 8 ambient bounce image which is more compute intensive.
- D. **Radiance: 8 Ambient Bounces:** We wanted to investigate if there was a marked difference using a Radiance image generated using 8 ambient bounces, as this involves considerably more compute time, and might not be necessary i.e. may not provide any more perceptual information than an image rendered using 2 ambient bounces.
- E. **Radiance: 8 Ambient Bounces BRIGHT:** This image had its brightness increased manually to see if this affected perception. The brightness was doubled (i.e. the intensity of each pixel was multiplied by 2) to see what, if any, effect this had on the perception of the image.
- F. **Radiance: Default:** Image generated with the default Radiance parameters. This would determine whether extra compute time makes a significant difference. The default image renders in a very short time, however ambient bounces of light are absent, we wanted to compare this to imagery where interreflections were catered for.
- G. **Radiance: Controlled Errors in Estimate Reflectance Values:** The RGB values for the materials were set to equal values to see what difference, if any, this made compared to using measured values. A poor perceptual response to this image would confirm our suspicion that material properties must be carefully quantified if an accurate result is required. This comparison, and the next, was to demonstrate the importance of using exact measurements rather than estimations for material values.
- H. **Radiance: Controlled Errors in Estimate of Light Source:** The RGB values for the light source were set to equal values to see what difference this made compared to using measured values. This experiment will show the necessity of measuring emission properties of sources in an environment if an accuracy is the aim.
- I. **Radiance: Tone Mapped:** We wanted to investigate the difference tone mapping would make to our test image. Tone mapping transforms the radiance values computed by the rendering engine to values displayable on a display device in a manner that preserves the *subjective* impression of the scene. The Tone Mapping Operator (TMO) used here was introduced by Ferwerda et al. [21]. Although the image examined does not have a very high dynamic range, we were interested to see the effects tone mapping would have on image perception.
- J. **Renderpark: Raytraced:** This was a very noisy image generated using stochastic raytracing. This experiment was designed to see how under-sampling would affect perception. Here the effect of under-sampling is exaggerated but might give insights in to how much undersampling a rendering engine can “get away with” without affecting perceptual performance.
- K. **Renderpark: Radiosity:** Finally, to investigate the effects of meshing in a radiosity solution, a poorly meshed radiosity image was used. We wanted to demonstrate the importance of using an accurate meshing strategy when employing radiosity techniques.

The images used are shown in Figures 3.17, 3.18 and 3.19.

Three full experiments were run, in each of the three, ten images were considered. The three environments are shown in Figure 4.20. Building on condition 5 of the pilot study, two new environments were introduced. First an environment containing entirely dark objects was constructed. Then a replica of this environment was also built with the same objects but they were painted only with light paint. All conditions for each experiment are shown in Figure 3.17 the original environment, Figure 3.18 shows conditions examined using the light environment, and the dark environment conditions are shown in Figure 3.19.

	Mixed Environment	Dark Environment	Light Environment
A. Real Environment	✓	✓	✓
B. Photograph	✓	✓	✓
C. 2 ambient bounces	✓	✓	✓
D. 8 ambient bounces	✓	✓	✓
E. 8 ambient bounces - Bright	✓	✓	✓
F. Default	✓	✓	✓
G. Guessed Materials	✓		✓
H. Guessed Illumination	✓	✓	✓
I. Tone Mapped	✓	✓	✓
J. Raytracing	✓		✓
K. Radiosity	✓		✓

Table 3.4: Experimental conditions: note some of the dark Environments were too dark to use so were not considered

### 3.8.4 Instructions

Each observer was presented with eleven images, one of which was the real scene. Images were presented in a random order to avoid any influences that might arise from a certain presentation sequence. Participants were not explicitly informed which scene they were viewing. After training on the patches, participants could recall from memory the number corresponding to each Munsell chip. Participants were asked to examine each target object in the scene, in a completely random order, and recall from memory the number of the Munsell chip they felt closely matched the paint on the target object. Such an explanation was given when the question was raised by an observer, the task being clear to each participant.

## 3.9 Summary

We have introduced a method for measuring the perceptual equivalence between a real scene and a computer simulation of the same scene. Because this model is based on psychophysical experiments, results are produced through study of vision from a human rather than a machine vision point of view. We have presented a method for modelling a real scene, then validated that model using the response of the human visual system. By conducting a series of experiments, based on the psychophysics of lightness perception, we can estimate how much alike a rendered image is to the original scene. We conduct a series of psychophysical experiments to assess the fidelity of graphical reconstruction of real scenes. Methods developed for the study of human visual perception are used to provide evidence for a perceptual, rather than a mere physical, match between the original scene and its computer representation. Results show that the rendered scene has high perceptual fidelity compared to the original scene, which implies that a rendered image can convey albedo<sup>4</sup>. This enables us to evaluate the quality of photo-realistic rendering software, and develop techniques to improve such renderer’s ability to produce high fidelity image. Because the complexity of human perception and the computational expensive rendering algorithms that exist today, future work should focus on developing efficient methods from which resultant graphical representations of scenes yield the same perceptual effects as the original scene. To achieve this the full gamut of colour perception, as opposed to simply lightness, must be considered by introducing scenes of increasing complexity.

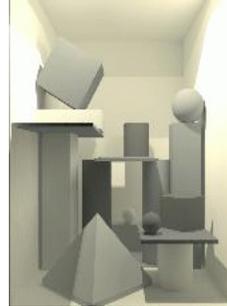
<sup>4</sup>albedo is the proportion of light or radiation reflected by a surface



A. Photograph



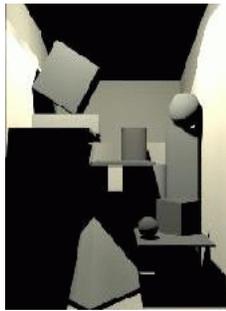
B. 2 Ambient Bounces



C. 8 Ambient Bounces



D. 8 Ambient Bounces  
Bright



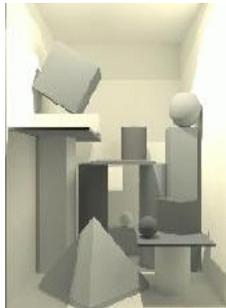
E. Default



F. Estimated Materials



G. Estimated Illumination



H. Tone Mapped



I. Raytraced



J. Radiosity

Figure 3.17: Images (Conditions) examined



B. Photograph



C. 2 Ambient Bounces



D. 8 Ambient Bounces



E. 8 Ambient Bounces  
Bright



F. Default



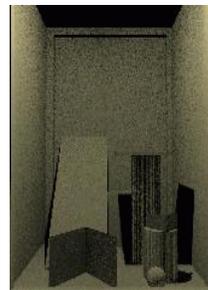
G. Estimated Materials



H. Estimated Illumination



I. Tone Mapped



J. Raytraced



K. Radiosity

Figure 3.18: Images examined - The Light Environment



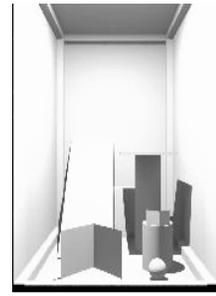
B. Photograph



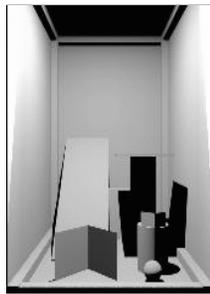
C. 2 Ambient Bounces



D. 8 Ambient Bounces



E. 8 Ambient Bounces  
Bright



F. Default



H. Estimated Illumination



I. Tone Mapped

Figure 3.19: Images examined - The Dark Environment

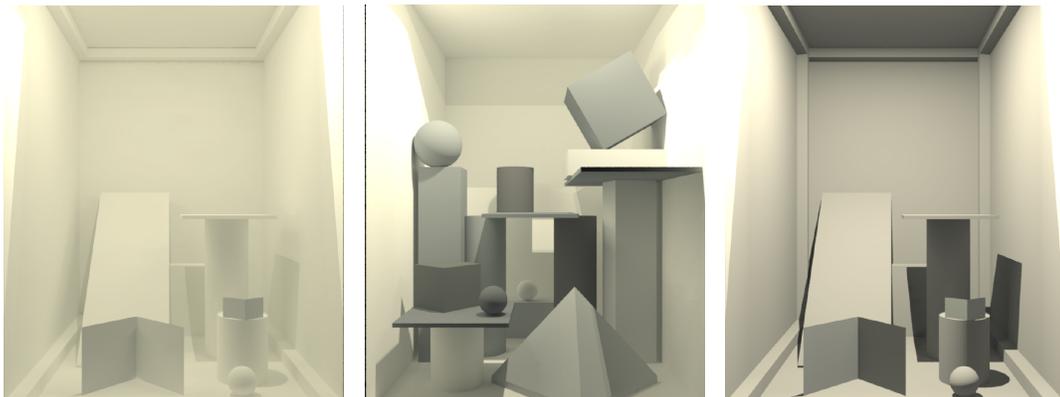


Figure 3.20: The three test scenes, a light environment, a mixed environment, and a dark environment

# Bibliography

- [1] E. H. Adelson. *Lightness Perception and Lightness Illusions*, pages 339–351. MIT Press, 1999.
- [2] A. Appel. Some techniques for shading machine renderings of solids. In *AFIPS 1968 Spring Joint Computer Conf.*, volume 32, pages 37–45, 1968.
- [3] J. Arvo and D. Kirk. Particle transport and image synthesis. volume 24, pages 63–66, Aug. 1990.
- [4] I. Ashdown. *Radiosity A Programmers Perspective*. John Wiley and Sons, New York, NY, 1994.
- [5] I. Ashdown, 1999. <http://www.ledalite.com>.
- [6] M. Bolin and G. Meyer. A frequency based ray tracer. In *ACM SIGGRAPH '95 Conference Proceedings*, pages 409–418, 1995.
- [7] M. Bolin and G. Meyer. A perceptually based adaptive sampling algorithm. In *ACM SIGGRAPH '98 Conference Proceedings*, pages 299–310, 1998.
- [8] M. R. Bolin and G. W. Meyer. A perceptually based adaptive sampling algorithm. *Computer Graphics*, 32(Annual Conference Series):299–309, Aug. 1998.
- [9] M. M. C. Ehrlich, 1994. <http://radsite.lbl.gov>.
- [10] J. Cataliotti and A. Gilchrist. Local and global processes in lightness perception. In *Perception and Psychophysics*, volume 57(2), pages 125–135, 1995.
- [11] K. Chiu, M. Herf, P. Shirley, S. Swamy, C. Wang, and K. Zimmerman. Spatially Nonuniform Scaling Functions for High Contrast Images. In *Proceedings of Graphics Interface '93*, pages 245–253, San Francisco, CA, May 1993. Morgan Kaufmann.
- [12] CIE. *CIE Proceedings*. Cambridge University Press, Cambridge, 1924.
- [13] T. C. . CIE 19/2.1, 1981. An Analytical Model for Describing the Influence of Lighting Parameters upon Visual Performance.
- [14] M. F. Cohen, S. E. Chen, J. R. Wallace, and D. P. Greenberg. A progressive refinement approach to fast radiosity image generation. *Symposium on Computational Geometry*, 22(4):75–84, 1–5 Aug. 1988.
- [15] M. F. Cohen and D. P. Greenberg. The hemi-cube: A radiosity for complex environments. volume 19, pages 31–40, July 1985.
- [16] M. F. Cohen and J. R. Wallace. *Radiosity and Realistic Image Synthesis*. Academic Press Professional, Boston, MA, 1993.
- [17] R. L. Cook. Stochastic sampling in computer graphics. *ACM Transactions on Graphics*, 5(1):51–72, January 1986.
- [18] R. L. Cook, T. Porter, and L. Carpenter. Distributed ray tracing. *Computer Graphics (SIGGRAPH '84 Proceedings)*, 18(3):137–145, July 1984.
- [19] H. Coolican. *Research Methods and Statistics in Psychology*. Hodder and Stoughton, Oxford, 1999.

- [20] S. Daly. The Visible Differences Predictor: An algorithm for the assessment of image fidelity. In A. Watson, editor, *Digital Image and Human Vision*, pages 179–206. Cambridge, MA: MIT Press, 1993.
- [21] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg. A model of visual adaptation for realistic image synthesis. *Computer Graphics*, 30(Annual Conference Series):249–258, 1996.
- [22] J. Foley, A. van Dam, S. Feiner, and J. Hughes. *Computer Graphics: Principles and Practice*. Addison-Wesley Publishing Co., ISBN 0-201-12110-7, 1990.
- [23] A. Fujimoto, T. Tanaka, and K. Iwata. ARTS: Accelerated ray tracing system. *IEEE Computer Graphics and Applications*, 6(4):16–26, 1986.
- [24] J. Gervais, J. L.O. Harvey, and J. R.s. Identification confusions among letters of the alphabet. In *Journal of Experimental Psychology: Human Perception and Performance*, volume 10(5), pages 655–666, 1984.
- [25] G. Gescheider. *Psychophysics: Method, Theory and Application*. Lawrence Erlbaum Associates, New Jersey, 1885.
- [26] S. Gibson and R. J. Hubbard. Perceptually-driven radiosity. *Computer Graphics Forum*, 16(2):129–141, 1997.
- [27] A. Gilchrist. Lightness contrast and illusions of lightness constancy: a common explanation. In *Perception and Psychophysics*, volume 43(5), pages 125–135, 1988.
- [28] A. Gilchrist. *Lightness, Brightness and Transparency*. Hillsdale: Lawrence Erlbaum Associates, 1996.
- [29] A. Gilchrist, S. Delman, and A. Jacobsen. The classification and integration of edges as critical to the perception of reflectance and illumination. *Perception and Psychophysics*, 33(5):425–436, 1983.
- [30] A. Gilchrist and A. Jacobsen. Perception of lightness and illumination in a world of one reflectance. *Perception*, 13:5–19, 1984.
- [31] A. L. Gilchrist. The perception of surface blacks and whites. *Scientific American*, 240(3):88–97, Mar. 1979.
- [32] A. S. Glassner. Space sub-division for fast ray tracing. *IEEE Computer Graphics and Applications*, October 1984, 4:15–22, 1984.
- [33] A. S. Glassner. *Principles of Digital Image Synthesis*. Morgan Kaufmann Publishers, San Francisco, California, 1995.
- [34] A. S. Glassner, R. L. Cook, E. Haines, P. Hanrahan, P. Heckbert, and L. R. Speer. Introduction to ray tracing. In *SIGGRAPH '87 Introduction to Ray Tracing*. Academic Press, July 1987.
- [35] C. M. Goral, K. Torrance, D. P. Greenberg, and B. Battaile. Modelling the interaction of light between diffuse surfaces. *Computer Graphics*, 18(3):213–222, July 1984.
- [36] H. Gourard. Continuous shading for curved surfaces. *IEEE Transactions on Computers*, c-20(6):623–629, 1971.
- [37] D. P. Greenberg. A framework for realistic image synthesis. *Communications of the ACM*, 42(8):44–53, Aug. 1999.
- [38] D. P. Greenberg, K. E. Torrance, P. Shirley, J. Arvo, J. A. Ferwerda, S. Pattanaik, E. P. F. Lafortune, B. Walter, S. Foo, and B. Trumbore. A framework for realistic image synthesis. In T. Whitted, editor, *SIGGRAPH 97 Conference Proceedings*, Annual Conference Series, pages 477–494. ACM SIGGRAPH, Addison Wesley, Aug. 1997. ISBN 0-89791-896-7.
- [39] R. N. H. Kolb, E. Fernandez, 2000. <http://www.insight.med.utah.edu/Webvision/>.

- [40] R. Hall. *Illumination and Color in Computer Generated Imagery*. Springer-Verlag, New York, NY, 1989.
- [41] R. A. Hall and D. P. Greenberg. A testbed for realistic image synthesis. *IEEE Computer Graphics and Applications*, 3(8):10–20, Nov. 1983.
- [42] D. Hedley. *Discontinuity Meshing for Complex Environments*. PhD thesis, Department of Computer Science, University of Bristol, Bristol, UK, Aug. 1998. Available from <http://www.cs.bris.ac.uk/Tools/Reports/Abstracts/1998-hedley.html>.
- [43] <http://www.hhmi.org>, 1999. <http://www.hhmi.org/senses/front/110.htm>.
- [44] A. N. S. Institute. ANSI standard nomenclature and definitions for illuminating engineering. ANSI/IES RP-16-1986, Illuminating Engineering Society, 345 East 47th Street, New York, NY 10017, June 1986.
- [45] J. T. Kajiya. The rendering equation. In D. C. Evans and R. J. Athay, editors, *Computer Graphics (SIGGRAPH '86 Proceedings)*, volume 20(4), pages 143–150, Aug. 1986.
- [46] J. T. Kajiya. Radiometry and photometry for computer graphics. In *SIGGRAPH '90 Advanced Topics in Ray Tracing course notes*. ACM SIGGRAPH, Aug. 1990.
- [47] J. E. Kaufman and H. Haynes. *IES Lighting Handbook*. Illuminating Engineering Society of North America, 1981.
- [48] A. Khodulev and E. Kopylov. Physically accurate lighting simulation in computer graphics software. In *GraphiCon '96 - The Sixth International Conference on Computer Graphics and Visualization*, volume 2, pages 111–119, St. Petersburg, Russia, July 1–5 1996.
- [49] E. Lafortune. *Mathematical Models and Monte Carlo Algorithms for Physically Based Rendering*. Ph.D. thesis, Department of Computer Science, Katholieke Universiteit Leuven, Leuven, Belgium, Feb. 1996.
- [50] G. W. Larson, H. Rushmeier, and C. Piatko. A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes. *IEEE Transactions on Visualization and Computer Graphics*, 3(4):291–306, Oct. 1997.
- [51] G. W. Larson and R. Shakespeare. *Rendering with Radiance: The Art and Science of Lighting Visualization*. Morgan Kaufmann, San Francisco, CA, 1998. ISBN 1-55860-499-5.
- [52] M. E. Lee, R. A. Redner, and S. P. Uselton. Statistically optimized sampling for distributed ray tracing. volume 19, pages 61–67, July 1985.
- [53] J. Lubin. A visual discrimination model for imaging system design and development. In P. E., editor, *Vision models for target detection and recognition*, pages 245–283. World Scientific, 1995.
- [54] J. Lubin. A human vision model for objective picture quality measurements. In *Conference Publication No. 447*, pages 498–503. IEE International Broadcasting Convention, 1997.
- [55] J. L. Mannos and D. J. Sakrison. The effects of a visual criterion on the encoding of images. *IEEE Transactions on Information Theory*, IT-20(4):525–536, July 1974.
- [56] A. McNamara. Evaluating image quality metrics v human evaluation. In *Visual Proceedings, Technical Sketch at ACM SIGGRAPH 2000*, page 249, 2000.
- [57] A. McNamara. Image quality metrics. In *SIGGRAPH 2000 Image Quality Metrics Course Notes*. ACM SIGGRAPH, July 2000.
- [58] A. McNamara. Visual perception in realistic image synthesis. In *State of the Art Report at EUROGRAPHICS 2000*, 2000.

- [59] A. McNamara, A. Chalmers, T. Troschianko, and I. Gilchrist. Evaluating images using human lightness judgements. In *Proceedings of the 11th Eurographics Rendering Workshop*, pages 207–219. Springer Verlag, June 2000.
- [60] A. McNamara, A. Chalmers, T. Troschianko, and E. Reinhard. Fidelity of graphics reconstructions: A psychophysical investigation. In *Proceedings of the 9th Eurographics Rendering Workshop*, pages 237–246. Springer Verlag, June 1998.
- [61] G. W. Meyer. Wavelength selection for synthetic image generation. *Computer Vision, Graphics, and Image Processing*, 41(1):57–79, Jan. 1988.
- [62] G. W. Meyer and A. Liu. Color spatial acuity control of a screen subdivision image synthesis algorithm. *Human Vision, Visual Processing, and Digital Display*, 1666(3):387–399, 1992.
- [63] G. W. Meyer, H. E. Rushmeier, M. F. Cohen, D. P. Greenberg, and K. E. Torrance. An Experimental Evaluation of Computer Graphics Imagery. *ACM Transactions on Graphics*, 5(1):30–50, Jan. 1986.
- [64] Minolta, 2000. <http://www.minolta.co.uk>.
- [65] D. P. Mitchell. Generating antialiased images at low sampling densities. *Computer Graphics*, 21(4):65–72, July 1987.
- [66] K. Myszkowski. The visible differences predictor: Applications to global illumination problems. In G. Drettakis and N. Max, editors, *Rendering Techniques '98 (Proceedings of Eurographics Rendering Workshop '98)*, pages 233–236, New York, NY, 1998. Springer Wien.
- [67] K. Myszkowski. The Visible Differences Predictor: Applications to global illumination problems. *Eurographics Rendering Workshop 1998*, pages 223–236, June 1998.
- [68] K. Myszkowski, A. B. Khodulev, and E. A. Kopylov. Validating global illumination algorithms and software. In *Visual Proceedings, Technical Sketch at ACM Siggraph'97*, page 156, 1997.
- [69] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. Geometric considerations and nomenclature for reflectance. Monograph 161, National Bureau of Standards (US), Oct. 1977.
- [70] T. Nishita and E. Nakamae. Continuous tone representation of three-dimensional objects taking account of shadows and interreflection. volume 19, pages 23–30, July 1985.
- [71] W. Nusselt. Grapische Bestimmung des Winkelverhältnisses bei der Warmestrahlung. *Zeitschrift des Vereines Deutscher Ingenieure*, 72(20):673, 1928.
- [72] J. Painter and K. Sloan. Antialiased ray tracing by adaptive progressive refinement. In J. Lane, editor, *Computer Graphics (SIGGRAPH '89 Proceedings)*, volume 23,3, pages 281–288, July 1989.
- [73] S. N. Pattanaik, J. A. Ferwerda, D. A. Greenberg, and M. D. Fairchild. A multiscale model of adaptation and spatial vision for realistic imaging. In *Computer Graphics (ACM SIGGRAPH '98 Proceedings)*, pages 287–298, 1998.
- [74] B. Phong. Illumination for computer generated pictures. *CACM June 1975*, 18(6):311–317, 1975.
- [75] J. Prikryl, 1999. <http://www.cs.kuleuven.ac.be/graphics/SEMINAR/program.html>.
- [76] M. Ramasubramanian, S. Pattanaik, and D. Greenberg. A perceptually based physical error metric for realistic image synthesis. *Proceedings of SIGGRAPH 99*, pages 73–82, August 1999.
- [77] H. Rushmeier, G. Ward, C. Piatko, P. Sanders, and B. Rust. Comparing real and synthetic images: Some ideas about metrics. In *Eurographics Rendering Workshop 1995*. Eurographics, June 1995.
- [78] D. J. Sakrison. On the role of the observer and a distortion measure in image transmission. *IEEE Trans. Commun.*, 25(11):1251–1267, 1977.
- [79] C. Schlick. An inexpensive BRDF model for physically-based rendering. *Computer Graphics Forum*, 13(3):C/233–C/246, 1994.

- [80] R. Siegel and J. R. Howell. *Thermal Radiation Heat Transfer*. Hemisphere Publishing Corporation, Washington, D.C., 3rd edition, 1992.
- [81] F. Sillion. The State of the Art in Physically-based Rendering and its Impact on Future Applications. In P. Brunet and F. W. Jansen, editors, *Photorealistic Rendering in Computer Graphics (Proceedings of the Second Eurographics Workshop on Rendering)*, pages 1–10, New York, NY, 1994. Springer-Verlag.
- [82] F. Sillion and C. Puech. *Radiosity and Global Illumination*. Morgan Kaufmann, San Francisco, 1994.
- [83] S. Stevens and J. Stevens. Brightness function: Effects of adaptation. In *Journal of the Optical Society of America*, volume 53, pages 375–385, March 1963.
- [84] D. Travis. *Effective Color Displays*. Academic Press, 1991.
- [85] T. Troscianko, A. McNamara, and A. Chalmers. Measures of lightness constancy as an index of the perceptual fidelity of computer graphics. In *European Conference on Visual Perception 1998, Perception Vol 27 Supplement*, pages 25–25. Pion Ltd, Aug. 1998.
- [86] B. Trumbore, W. Lytle, and D. P. Greenberg. A testbed for image synthesis. In W. Purgathofer, editor, *Eurographics '91*, pages 467–480. North-Holland, Sept. 1991.
- [87] J. Tumblin and H. E. Rushmeier. Tone Reproduction for Realistic Images. *IEEE Computer Graphics and Applications*, 13(6):42–48, Nov. 1993.
- [88] J. Tumblin and H. E. Rushmeier. Tone reproduction for realistic images. *IEEE Computer Graphics and Applications*, 13(6):42–48, Nov. 1993.
- [89] G. J. Ward. The RADIANCE lighting simulation and rendering system. In A. Glassner, editor, *Proceedings of SIGGRAPH '94 (Orlando, Florida)*, Computer Graphics Proceedings, Annual Conference Series, pages 459–472, July 1994.
- [90] A. Wavefront. *Alias Wavefront User Manual*. Alias Wavefront, 1997.
- [91] T. Whitted. An improved illumination model for shaded display. *Computer Graphics (Special SIGGRAPH '79 Issue)*, 13(3):1–14, Aug. 1979.
- [92] Wyszecki and Stiles. *Colour Science: concepts and methods, quantitative data and formulae (2nd edition)*. New York: Wiley, 1986.

# Perception-Guided Rendering and Animation Techniques

Karol Myszkowski\*  
Max-Planck-Institut für Informatik

## Abstract

Synthesis of realistic images which predict the appearance of the real world has many applications including architecture and interior design, illumination engineering, environmental assessment, special effects and film production, along with many others. Due to costly global illumination computation, which is required for the prediction of appearance, physically-based rendering still remains the domain of research laboratories, and is rarely used in industrial practice. In this overview special attention is paid to the solutions which use perception-guided algorithms to improve their performance. This makes it possible to focus the computation on readily visible scene details, and to stop it when further improvement of the image quality cannot be perceived by the human observer. Also, by better use of perception-motivated physically-based partial solutions, meaningful images can be presented to the user at the early stages of computation. Since many algorithms make simplifying assumptions about the underlying physical model in order to achieve gains in rendering performance, a validation procedure for testing lighting simulation accuracy and image quality is proposed. To check the requirement of appearance predictability imposed on the developed algorithms, the rendered images are compared against the corresponding real-world views.

---

\*The author is a senior researcher at Max-Planck-Institut für Informatik, Stuhlsatzenhausweg 85, 66-123 Saarbrücken, Germany, email: [myszkowski@mpi-sb.mpg.de](mailto:myszkowski@mpi-sb.mpg.de)

---

## Contents

<b>1</b>	<b>Perception-Driven Global Illumination and Rendering Computation</b>	<b>3</b>
1.1	Low-Level Perception-Based Error Metrics . . . . .	5
1.2	Advanced Perception-Based Error Metrics . . . . .	6
1.3	Visible Differences Predictor . . . . .	7
1.3.1	VDP Integrity . . . . .	11
1.3.2	Psychophysical Validation of the VDP . . . . .	12
1.4	VDP Applications in Global Illumination Computation . . . . .	13
1.4.1	Evaluating Progression of Global Illumination Computation . . . . .	13
1.4.2	Optimizing Progression of Global Illumination Computation . . . . .	16
1.4.3	Stopping Conditions for Global Illumination Computation . . . . .	21
<b>2</b>	<b>Perception-Driven Rendering of High-Quality Walkthrough Animations</b>	<b>23</b>
2.1	Image-Based Rendering Techniques . . . . .	24
2.2	Animation Quality Metric . . . . .	25
2.2.1	Spatiovelocity CSF Model . . . . .	26
2.2.2	Animation Quality Metric Algorithm . . . . .	27
2.3	Rendering of the Animation . . . . .	28
2.3.1	Adaptive Refinement of Keyframe Placement . . . . .	30
2.3.2	A Case Study Walkthrough Animation . . . . .	31
<b>3</b>	<b>Validation of Global Illumination and Rendering Solutions</b>	<b>35</b>
3.1	Experiments with the Atrium Scene . . . . .	36
3.1.1	Simplified Model . . . . .	36
3.1.2	Complete Model . . . . .	38
3.2	Multi-Stage Validation Procedure . . . . .	43

# 1 Perception-Driven Global Illumination and Rendering Computation

One of the basic goals of realistic rendering is to create images which are perceptually indistinguishable from real scenes. Since the fidelity and quality of the resulting images are judged by the human observer, the perceivable differences between the appearance of a virtual world (reconstructed on a computer) and its real world counterpart should be minimized. Thus, perception issues are clearly involved in realistic rendering, and should be considered at various stages of computation such as global illumination computation, rendering, and image display.

The most common approach is to consider the visual characteristics of the human observer at the stage of image display. Such considerations are mostly driven by the need to overcome numerous physical limitations of the display device. Let us recall that as a result of view-independent global illumination computation, a distribution of the radiometric (photometric) values over the scene surfaces is obtained. Then, for given viewing parameters, those values are projected on the image plane, and are presented on the display device in the form of a 2D array of discrete values for every pixel. Usually, current display technologies cannot directly reproduce these values due to their limited absolute and dynamic luminance ranges, and color gamut. However, the visual system of humans has limited capabilities in detecting differences in absolute luminance levels, and concentrates more on spatial aspects when comparing two images [79]. Thus, it might be possible to obtain the perceptual match between a real scene and a displayed image even though the display device is not able to reproduce the actual luminance values. To achieve this goal, Tumblin and Rushmeier [71] introduced the Tone Mapping Operator (TMO) that maps the simulated luminance values to the corresponding display luminance values taking into account the limitations of the display device, such that the impression of brightness is similar for the scene and display observer. The perception of brightness varies with the state of adaptation, and therefore the adaptation states of the scene and display observer are determined. The development of efficient TMOs that better utilize the low contrast display devices to reproduce the appearance of high contrast scenes without loss of shading and texture details, is an active research topic in computer graphics. A good survey of recent results can be found in [70]. The problem of color reproduction on the limited gamut of a display device has attracted much less attention in the computer graphics literature. However, this problem is well addressed in the more specialized literature of imaging and color science. An example of an advanced color reproduction model, which predicts a wide range of visual phenomena, is the Hunt model [26] developed in the Kodak Research Laboratories. A comprehensive survey of the color appearance models can be found in [15].

In this chapter we focus on embedding the characteristics of the Human Visual System (HVS) directly into global illumination and rendering algorithms to improve their efficiency. This research direction has recently gained much attention within the computer graphics community [3, 21, 22, 50, 61]. Since global illumination solutions are costly in terms of computation, significant efficiency improvements can be made by focusing computation on those scene features which can be readily perceived by the human observer under given viewing conditions. This means that those features that are below perceptual visibility thresholds, can be simply

omitted from the computation without causing any perceivable difference in the final image appearance.

Current global illumination algorithms usually rely on energy-based metrics of solution errors, which do not necessarily correspond to the visible improvements of the image quality [34]. Ideally, one may advocate the development of perceptually-based error metrics which can control the accuracy of every light interaction between surfaces. This can be done by predicting the visual impact those errors may have on the perceived fidelity of the rendered images. In practice, there is a trade-off between the robustness of such low-level error metrics and their computational costs. In Section 1.1 we give some examples of such low-level metrics applied in the context of hierarchical radiosity and adaptive meshing computations.

Another approach is to develop a perceptual metric which operates directly on the rendered images. If the goal of rendering is just a still frame, then the image-based error metric is adequate. In the case of (view-independent) solutions for interactive viewing, the application of the metric becomes more complex because a number of “representative” views should be chosen. In practice, instead of measuring the image quality in absolute terms, it is much easier to derive a relative metric which predicts the perceived differences between a pair of images [62]. (It is well-known that the common mean-squared error metric usually fails in such a task [8, 18, 62, 68].) A single numeric value might be adequate for some applications; however, for more specific guiding of computation, a local metric operating at the pixel level is required. In Section 1.2 we give a brief overview of the application of such local metrics to guide the global illumination and rendering solutions. Such metrics usually involve advanced HVS models, which may incur non-negligible computation costs. An important issue becomes whether the savings in computation that are obtained through the usage of such metrics can compensate these costs.

A representative example of such an advanced image fidelity metric is the Visible Differences Predictor (VDP) developed by Daly [8]. In Section 1.3 we overview briefly the VDP, which we use extensively in this work. The VDP metric, when applied in global illumination computation, provides a summary of the algorithm performance as a whole rather than giving a detailed insight into the workings of its particular elements. However, *a priori* knowledge of the current stage of computation can be used to obtain more specific measures for such tasks as adaptive meshing performance, accuracy of shadow reconstruction, convergence of the solution for indirect lighting, and so on. Since the VDP is a general purpose image fidelity metric, we validate its performance in these tasks. In Section 1.3.1 we report the results of comparisons of the VDP predictions when the model incorporates a variety of contrast definitions, spatial and orientation channel decomposition methods, and Contrast Sensitivity Functions (CSFs) derived from different psychophysical experiments. The goal of these experiments was to test the VDP integrity and sensitivity to differing models of visual mechanisms, which were derived by different authors and for different tasks than those which have been originally used by Daly. Also, we conducted psychophysical experiments with human subjects to validate the VDP performance in typical global illumination tasks (Section 1.3.2). An additional goal of these experiments was to test our implementation of the complex VDP model.

When our rigorous validation procedure of the VDP performance was successfully completed, we could then apply the metric to our actual global illumination applications. We used the VDP to monitor the progression of computation as a function of time for hierarchical radiosity and Monte Carlo solutions (Section 1.4.1). Based on the results obtained, we propose

a novel global illumination algorithm which is a hybrid of stochastic (density estimation) and deterministic (adaptive mesh refinement) techniques used in an optimized sequence to reduce the differences between the intermediate and final images as perceived by the human observer in the course of lighting computation (Section 1.4.2). The VDP responses are used to support selection of the best component algorithms from a pool of global illumination solutions, and to enhance the selected algorithms for even better progressive refinement of the image quality. The VDP is used to determine the optimal sequential order of component-algorithm execution, and to choose the points at which switchover between algorithms should take place. Also, we used the VDP to decide upon stopping conditions for global illumination simulation, when further continuation of computation does not contribute to perceivable changes in the quality of the resulting images (Section 1.4.3).

## 1.1 Low-Level Perception-Based Error Metrics

One of the research directions towards perception-driven improvement of global illumination computation performance relies on direct embedding of some simple error metrics to find the adequate level of light interactions between surfaces. Gibson and Hubbold [19] proposed a perception-driven hierarchical algorithm in which a TMO and the perceptually uniform color space CIE  $L^*u^*v^*$  are used to decide when to stop the hierarchy refinement. Links between patches are not further refined once the difference between successive levels of elements becomes unlikely to be detected perceptually. Gibson and Hubbold applied a similar error metric to measure the perceptual impact of the energy transfer between two interacting patches, and to decide upon the number of shadow rays that should be used in a visibility test for these patches. A similar strategy was assumed by Martin et al. [43], whose oracle of patch refinement operates directly in the image space and tries to improve the radiosity-based image quality for a given view. More detailed analysis of these and other similar techniques can be found in [60].

Perceptually-informed error metrics were also successfully introduced to control adaptive mesh subdivision [19, 24, 57] and mesh simplification [78] in order to minimize the number of mesh elements used to reconstruct the lighting function without introducing visible shading artifacts. The quality of lighting reconstruction is judged by the human observer, so it is not a surprise that purely energy-based criteria used in the discontinuity meshing [12, 36] and adaptive mesh subdivision [5, 35, 74] methods are far from optimal. These methods drive meshing refinement based on the measures of lighting differences between sample points, which are expressed as radiometric or photometric quantities. However, the same absolute values of such differences might have a different impact on the final image appearance, depending on the scene illumination and observation conditions (which determine the eye sensitivity). To make things even more complicated, a TMO must also be taken into account because it determines the mapping of simulated radiometric or photometric values into the corresponding values of the display device.

Myszkowski et al. [57] noticed that mesh refinement can be driven by some metrics which measure quantitatively visual sensation such as brightness instead of commonly used radiometric or photometric quantities. Myszkowski et al. transformed the stimulus luminance values to predicted perceived brightness using Stevens' power law [71] and a decision on the edge splitting was made based on the local differences in brightness. The threshold differences of brightness, which triggered such subdivision, corresponded to the Just Noticeable Difference (JND) values that were selected experimentally and had different values depending on the local illumination

level. For darker regions of the displayed image the eye is more sensitive and smaller values of the thresholds are chosen. Conversely, for bright regions that are close to the image saturation the threshold values are significantly larger. For the optimal threshold selection the global illumination should be known. However, in the radiosity technique [53] only direct illumination is known at the stage of mesh refinement, which might result in an overly conservative threshold selection. In such conditions, some lighting discontinuities predicted as perceivable could be washed out in the regions of significant indirect lighting. Obviously, this could lead to excessive mesh refinement, which is a drawback of the technique presented in [57].

Gibson and Hubbard [19] showed that the meshing performance can be improved even if some crude approximation of global illumination such as the ambient correction term [6] is used. Also, Gibson and Hubbard improved the method from [57] further by introducing color considerations into their mesh subdivision criteria.

Further improvement of meshing performance was reported in [75] by using a lighting simulation algorithm (discussed in more detail in Section 1.4.2) which provides local estimates of global illumination quickly. These estimates are available at the mesh refinement stage, which makes a more reliable evaluation of the contrast at lighting discontinuities possible. Thus, the prediction of discontinuity perceivability also becomes more robust and excessive mesh subdivision can be avoided. In the example given in [75], the uniform mesh built of 30,200 triangles was subdivided into 121,000, 97,000, and 86,000 elements using techniques proposed in [57], [19], and [75] respectively, without any noticeable difference in the resulting image quality.

Perception-based criteria have also been used to remove superfluous mesh elements in the discontinuity meshing approach [24]. Also, a similar perception-driven mesh simplification was performed as a post-process to a density estimation solution applying a dense, uniform mesh [78].

All techniques discussed so far used perceptual error metrics on the atomic level (e.g., every light interaction between patches, every mesh element subdivision), causing a significant overhead on procedures that are repeated thousands of times in the course of the radiosity solution. This imposes severe limitations on the complexity of human spatial vision models, which in practice are restricted to models of brightness and contrast perception. Recently, more complete (and costly) vision models have been used in rendering to develop higher level perceptual error metrics which operate on the complete images. In the following section we briefly overview applications of such metrics to global illumination and rendering solutions.

## 1.2 Advanced Perception-Based Error Metrics

Embedding advanced HVS models into global illumination and rendering algorithms is very attractive scenario, which enables computation to be perception-driven specifically for a given scene. Bolin and Meyer [3] have developed an efficient approximation of the Sarnoff Visual Discrimination Model (VDM) [38], which made it possible to use this model to guide the placement of samples in a rendered image. Because samples were only taken in areas where there were visible artifacts, some savings in rendering time compared to the traditional uniform or adaptive sampling were reported. Myszkowski [50] has shown some applications of the VDP to drive adaptive mesh subdivision taking into account visual masking of the mesh-reconstructed lighting function by textures. Ramasubramanian *et al.* [61] have developed their own image quality metric, which they applied to predict the sensitivity of the human observer to noise in

the indirect lighting component. This made a more efficient distribution of indirect lighting samples possible by reducing their number for pixels with higher spatial masking (in areas of images with high frequency texture patterns, geometric details, and direct lighting variations). All computations were performed within the framework of the costly path tracing algorithm [28] and a significant speedup of computation was reported compared to the sample distribution based on purely stochastic error measures.

A practical problem arises due to the fact that the computational costs incurred by the HVS models introduce an overhead to the actual lighting computation, which may become more significant the more rapid the lighting computation becomes. The potential gains of such perception-driven computation can be easily canceled out by this overhead, depending on many factors such as scene complexity, performance of a given lighting simulation algorithm for a given type of scene, image resolution, and so on. The HVS models can be simplified to reduce the overhead, e.g., Ramasubramanian *et al.* [61] ignore spatial orientation channels in their visual masking model, but then underestimation of visible image artifacts becomes more likely. To prevent such problems and to compensate for ignored perceptual mechanisms, more conservative (sensitive) settings of the HVS models should be applied, which may also reduce gains in the lighting computation driven by such models.

It seems that keeping the HVS models at some high level of sophistication and embedding them into rendering algorithms, which are supposed to provide a meaningful response rapidly, e.g., in tens of seconds or single minutes, may be a difficult task. For example, full processing of the difference map between a pair of images at a resolution of  $256 \times 256$  pixels using the VDP model [8] takes about 20 seconds on a R10000, 195 MHz processor and such processing should be repeated a number of times to get reasonable monitoring of progress in image quality. In this work we explore two different approaches, in which the advanced HVS models are used off-line or on-line. In the former case, the VDP is used only at the design stage of the global illumination algorithms and the tuning of their parameters. Thus, the resulting algorithms can spend 100% of their computation time in lighting simulation, and the costs of HVS processing (which is performed off-line) are of secondary importance (Section 1.4.2). In the latter case, the VDP processing is performed along with the time-consuming global illumination computation to decide upon its stopping condition. However, in this application the VDP computation is performed exclusively at later stages of computation, and involves only a small fraction of the overall computation costs (Section 1.4.3).

In the following section we briefly describe the VDP as a representative example of an advanced image fidelity metric that is strongly backed by findings in physiology and psychophysics.

### 1.3 Visible Differences Predictor

Although substantial progress in the study of physiology and psychophysics has been made in recent years, the HVS as a whole and the higher order cognitive mechanisms in particular, are not yet fully understood. Only the early stages of the visual pathway beginning with the retina and ending with the visual cortex are considered as mostly explored [11]. It is believed that the internal representation of an image by cells in the visual cortex is based on spatial frequency and orientation channels [40, 82, 88]. The channel model explains well such visual characteristics as:

- the overall behavioral Contrast Sensitivity Function (CSF) - visual system sensitivity is a function of the spatial frequency and orientation content of the stimulus pattern;
- spatial masking - detectability of a particular pattern is reduced by the presence of a second pattern of similar frequency content;
- sub-threshold summation - adding two patterns of sub-threshold contrast together can improve detectability within a common channel;
- contrast adaptation - sensitivity to selected spatial frequencies is temporarily lost after observing high contrast patterns of the same frequencies; and,
- the spatial frequency after-effects - as result of the eye's adaptation to a certain grating pattern, other nearby spatial frequencies appear to be shifted.

Because of these favorable characteristics, the channel model provides the core of the most recent HVS models that attempt to describe spatial vision. Our application of the HVS model is concerned with how to predict whether a visible difference will be observed between two images. Therefore, we were most interested in the HVS models developed for similar tasks [7,8,16,18,37,44,66,68,86,92], which arise from studying lossy image compression, evaluating dithering algorithms, designing CRT and flat-panel displays, and generating computer graphics. Let us now describe briefly the Visible Differences Predictor (VDP) developed by Daly [8] as a representative example, which was selected by us for our experiments on global illumination algorithms.

The VDP is considered one of the leading computational models for predicting the differences between images that can be perceived by the human observer [32]. The VDP receives as input a pair of images, and as output it generates a map of probability values, which characterize perceivability of the differences. The input target and mask images undergo identical initial processing (Figure 1.1). At first, the original pixel intensities are converted to physical luminance values in the display device. If the exact range of luminance values is not known for a given CRT display it is usually assumed that the maximum luminance value is about  $100 \text{ cd/m}^2$ . Weber's law-like behavior is applied to derive brightness sensation for every pixel based on the corresponding luminance values. The non-linear response of retinal neurons and their adaptation characteristics are taken into account. For the sake of simplicity it is assumed that the HVS adapts separately to each pixel. Then the resulting image is converted into the frequency domain and processing of the CSF is performed. The resulting data is decomposed into the spatial frequency and orientation channels using the Cortex transform, which is a pyramid-style, invertible, and computationally efficient image representation. In Figure 1.2a we show organization of the filter bank in the Cortex transform, which models the combined radial frequency and orientational selectivity of cortical neurons. After decomposing the input image into six frequency bands, each of these bands (except the lowest-frequency baseband) undergoes identical orientational selectivity processing. The resulting decomposition of the image frequency plane into 31 radial frequency and orientation channels is shown in Figure 1.2b. Then the individual channels are transformed back to the spatial domain, in which visual masking is processed.

As the result of CSF computation the contrast values in all channels are normalized by the corresponding values of detection thresholds. Due to visual masking characteristics of the HVS those threshold values can be further elevated with increases in the contrast of image

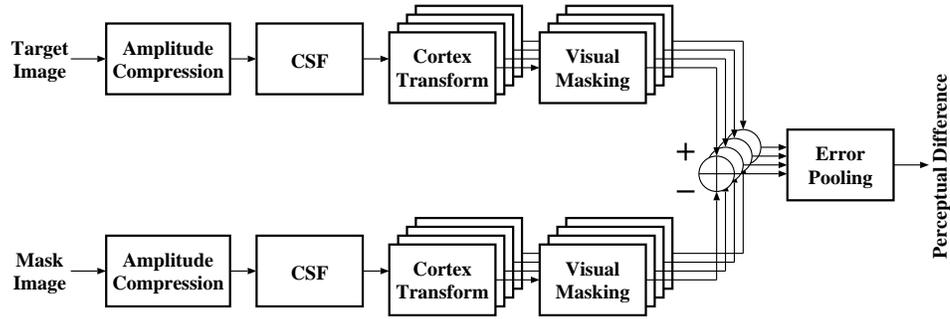


Figure 1.1: Block diagram of the Visible Differences Predictor (multiple arrows indicate parallel processing of the spatial frequency and orientation channels).

(mask) pattern. For every channel and for every pixel, the elevation of the detection threshold is calculated based on the mask contrast for that channel and that pixel. It is usually assumed that the threshold elevation is computed for the mask image. Also, more conservative approach can be chosen in which mutual masking is considered by taking the minimal threshold elevation value for the corresponding channels and pixels of the two input images. The resulting threshold elevation maps are then used to normalize the contrast differences between target and mask images. The normalized differences are input to the psychometric function which estimates the probability of detecting the differences for a given channel. This estimated probability value is summed across all channels for every pixel. Finally, the probability values are used to visualize visible differences between the target and mask images. It is assumed that the difference can be perceived for a given pixel when the probability value is greater than 0.75, which is the standard threshold value for discrimination tasks [88]. When a single numeric value is needed to characterize the differences between images, the percentage of pixels with probability greater than this threshold value is reported. The former measure is suitable to estimate the differences locally, while the latter measure provides global information on the differences for the whole image.

The main advantage of the VDP is a prediction of local differences between images (on the pixel level). The Daly model also takes into account the visual characteristics that we think are extremely important in our application: a Weber's law-like amplitude compression, advanced CSF model, and visual masking function.

The original Daly model also has some disadvantages, for example, it does not process chromatic channels in input images. However, in global illumination applications many important effects such as the solution convergence or the quality of shadow reconstruction can be relatively well captured by the achromatic mechanism, which is far more sensitive than its chromatic counterparts.

The VDP seems to be one of the best existing choices for our applications involving prediction of image quality for various settings of global illumination solutions. This claim is supported by our extensive VDP integrity checking, and validation in psychophysical experiments that we briefly summarize in the following two sections. More extensive documentation of these tests is provided on the VDP project Web pages [1].

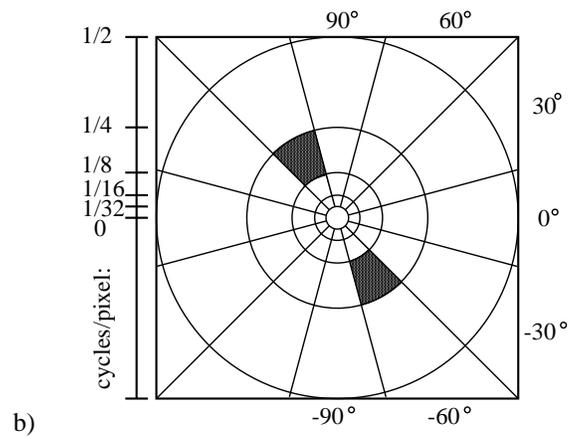
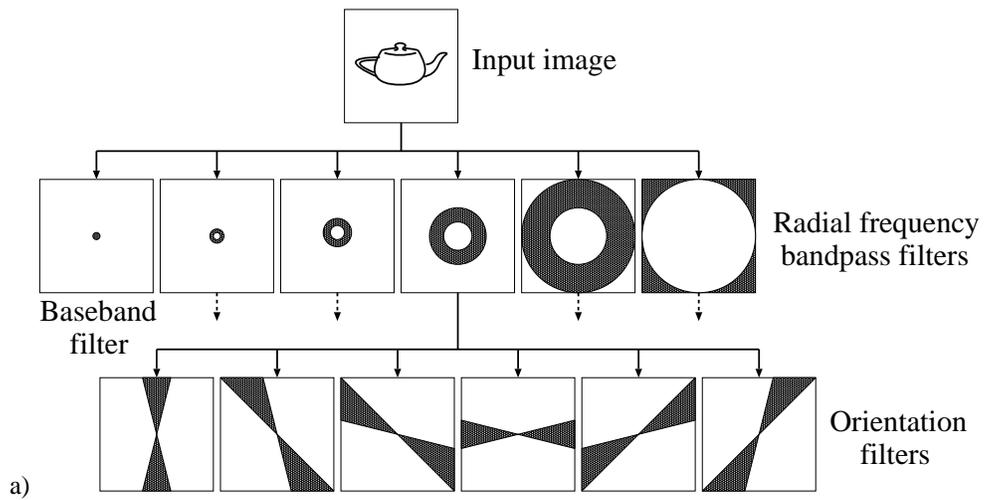


Figure 1.2: Cortex transform: a) organization of the filter bank and b) decomposition of the image frequency plane into the radial and orientation selectivity channels. The filled regions show an example of the spatial frequencies allocated to a single channel. At the left side of b) the spatial frequency scale in cycles per pixel is shown, which makes possible estimation of the bandpass frequencies of every radial channel. Also, the orientation of the band center in degrees is shown for every orientation channel.

### 1.3.1 VDP Integrity

The VDP model predicts many characteristics of human perception. However, the computational models of these characteristics were often derived from the results of various unrelated experiments, which were conducted using completely different tasks. As pointed out by Taylor et al. [66] this is a potential threat for the VDP integrity. The approach promoted in [66, 89] was to execute psychophysical experiments that directly determined the model parameters. However, such experiments usually cover significantly less visual mechanisms, for example, the model proposed by Taylor et al. does not support visual masking. In this respect, the strategy taken by Daly results in a more complete model, although, perhaps at the expense of its integrity.

We decided to examine the integrity of Daly's model to understand how critical are its major components in maintaining a reasonable output. We replaced some model components by functionally similar components, which we obtained from well-established research results published in the literature. We investigated how the VDP responses will be affected by such replacements.

We experimented with three types of CSF used in the following HVS models: [8], [16, 44], and [18]. The response of the VDP was very similar in the former two cases, while for the latter one discrepancies were more significant. A possible reason for such discrepancies is that the CSF used in [18] does not take into account luminance adaptation for our test, which could differ from the conditions under which the CSF was originally measured.

Also, we experimented with different spatial and orientation channel decomposition methods. We compared the Cortex transform [8] with 6 spatial and 6 orientation channels (a typical output of every channel for our standard test image is shown on the VDP project Web pages [1]) and the band-pass (Laplacian) pyramid proposed by Burt [4] with 6 spatial frequency channels, and extended to include 4 orientation channels. While the quantitative results are different, the distribution of probabilities of detection differences between images corresponds quite well. The quantitative differences can be reduced by an appropriate scaling of the VDP responses.

Daly's original VDP model used an average image mean to compute the global contrast for every channel of the Cortex transform. We experimented with the local contrast using a lowpass filter on the input image to provide an estimate of luminance adaptation for every pixel. This made the VDP more sensitive to differences in dark image regions, and we found that in many cases the VDP responses better matched our subjective impressions.

In experiments we performed, we found that the VDP prediction was quite robust across the tasks we examined and variations in the configuration of VDP modules. While the quantitative results we obtained were different in many cases (i.e., the probability values for perceiving a difference which are reported for every pixel), the distribution of predicted perceivable differences over the image surface usually matched quite well. A comparison of the VDP output for all experiments discussed in this section is provided on the VDP project Web pages [1].

In [50] we report representative results of more specialized VDP experiments, which were focused on prediction of the perceived shadow quality as a function of the visual masking by a texture, the CRT device observation distance, and the global illumination solution convergence. In all cases tested we obtained predictions that matched well our subjective judgments. On the VDP project Web pages [1] we provide input images along with the VDP predictions for the full set of experiments we performed. We disseminate this material on the Internet so that it can be used for testing other metrics of differences between images.

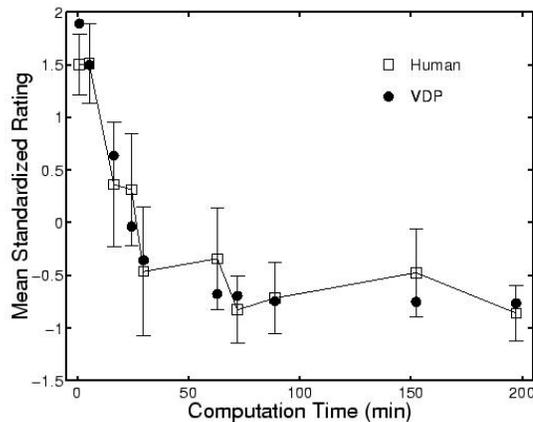


Figure 1.3: The standardized mean ratings (squares) at each of 10 cumulative computation times are shown along with corresponding VDP predictions (filled circles).

### 1.3.2 Psychophysical Validation of the VDP

Since the VDP is a general purpose predictor of the differences between images, it can be used to evaluate sets of images from a wide range of applications. In our experiments we chose to test its performance in global illumination tasks, which correspond to our intended use of the VDP. In this work we discuss one selected experiment in which we compared VDP responses with those obtained from human subjects for a series of image pairs resulting from successive refinement in a progressive hierarchical radiosity solution. We chose this experiment because it validates the VDP role in the development of our novel global illumination algorithm described in Section 1.4.2. The description of our other psychophysical experiments with subjects concerning visual masking of shadows by textures, and image fidelity following JPEG compression can be found in [42]. As postulated in [22] the experiments were performed in cooperation with an experimental psychologist.

In the experiment reported here, subjective judgments from 11 human observers were collected for pairs of images presented on a high-quality CRT display under controlled viewing conditions. The experimental subjects were requested to rank on a scale from 1 to 9 the perceived global difference between each of a pair of images. In every pair, the final image for the fully converged radiosity solution was presented side-by-side with an image generated at an intermediate stage of radiosity computation. In total ten intermediate images taken at different stages of computation were considered, and presented to subjects in a random order. We used the HTML forms to present stimuli, and the subjects could freely scroll the display and adjust their ranking (we include examples of our HTML forms on the VDP project Web pages [1]). The prediction of differences for the same pairs of images was computed using the VDP, and compared against the subjects' judgments. Figure 1.3 summarizes the results obtained. A good agreement was observed between VDP results and subjective ratings. This means that as the progressive radiosity solution converged, close agreement between the VDP predictions and the

subjective judgments was maintained.

The results of our psychophysical experiment suggest that the VDP can be used to estimate what might be termed “perceptual” convergence in image quality rather than “physical” convergence. Such an image quality measure can be used to compare the performance of various rendering techniques as a function of the computation time, and to decide when the computation can be finished because further improvement of the image quality cannot be perceived by the human observer [50].

Encouraged by the positive results of VDP validation in psychophysical experiments and integrity tests, we used the VDP in actual applications where the main goal was to improve the performance of global illumination computation. In the following section we discuss a number of examples of such applications.

## 1.4 VDP Applications in Global Illumination Computation

A common measure of the physical convergence of a global illumination solution is the Root Mean Square (RMS) error computed for differences between pixel values of the intermediate and final images. The RMS error is not suitable to monitor the progress of computation because it poorly predicts the differences as perceived by the human observer [8, 18, 62, 68]. In Section 1.3.2 a new metric of the perceptual convergence in image quality was discussed, and we used this metric to compare the performance of selected global illumination techniques (Section 1.4.1). As the result of such a comparison, a hybrid global illumination solution has been proposed in which the technique that performs best in terms of the perceptual convergence is selected at every stage of computation [75]. We discuss this hybrid technique in Section 1.4.2.

As can be seen in Figure 1.3 the ranking for the final stages of the radiosity solution (70–200 minutes) was considerably more difficult because the corresponding images were very similar. This suggests a novel application of the VDP (and other similar metrics) to decide upon the computation stopping conditions, when further computation will not result in noticeable changes in the image quality as perceived by the human observer. We discuss this topic more in detail in Section 1.4.3.

### 1.4.1 Evaluating Progression of Global Illumination Computation

In many practical applications it is important to obtain the intermediate images which correspond well to the final image at possibly early stages of solution. A practical problem arises how to measure the solution progression, which could lead to the selection of an optimal global illumination technique for a given task. Clearly, since the human observer ultimately judges the image quality, basic characteristics of the HVS must be involved in such a measure of the solution progression. We used the VDP to provide the quantitative measures of the perceptual convergence by predicting the perceivable differences between the intermediate and final images [50].

We investigated the perceptual convergence of the following view-independent algorithms:

- Deterministic Direct Lighting (DDL) computation with perceptually-based adaptive mesh subdivision [57]. We provide more details on the DDL algorithm in [53, 54].

- Shooting iteration Hierarchical (link-less and cluster-based) Radiosity (SHR) [53, 54] for indirect lighting computation. By default, a pre-calculated fixed mesh is used to store the resulting lighting.
- Density Estimation Photon Tracing (DEPT) from light sources with photons bucketed into a non-adaptive mesh [75]. By Direct DEPT (DDEPT) we denote buckets with photons coming directly from light sources, and by Indirect DEPT (IDEPT) we denote a different set of buckets with photons coming via at least one reflection. We provide more details on the DEPT algorithm in [75].

The DDL and SHR techniques are deterministic, while the DEPT algorithm is stochastic. Obviously direct (DDL and DDEPT) and indirect (SHR and IDEPT) lighting computation techniques are complementary, but in practice the following combinations of these basic algorithms are used: DDL+SHR, DDL+IDEPT, and DDEPT+IDEPT (DEPT for short).

We measured the performance of these basic techniques in terms of perceived differences between the intermediate and final images using the VDP responses. As we discussed in Section 1.3, the VDP response provides the probability of difference detection between a pair of images, which is estimated for every pixel. We measured the difference between images as the percentage of pixels for which the probability of difference detection is over 0.75, which is the standard threshold value for discrimination tasks [88]. In all tests performed, we used images of resolution  $512 \times 512$ . The diagonal of the images displayed on our CRT device was 0.2 meters, and we assumed that images were observed from the distance of 0.5 meters.

We assumed that the final images used for the VDP computation are based on the DDL+SHR and DDL+IDEPT global illumination solutions, which converge within some negligible error tolerance. The final images obtained using these methods are usually only slightly different. Minor discrepancies can be explained by various approximations assumed by each of these completely different algorithms, e.g., different handling of the visibility problem, the lighting function discretization during computation used by the SHR technique. To eliminate the influence of these differences on the VDP response, for a given method we considered the final image generated using this particular method. The only exception is the DDEPT+IDEPT method, for which we use the final image generated using the DDL+IDEPT technique because it provides more accurate direct lighting reconstruction for a given mesh/bucket density.

In this work we report results obtained for a scene, which we will refer to as the POINT (in [75] we consider three different scenes of various complexity of geometry and lighting). Both direct and indirect lighting play a significant role in the illumination of the POINT scene. The scene is built of about 5,000 polygons, and the original scene geometry was tessellated into 30,200 mesh elements using the DDL technique.

The graphs in Figure 1.5 show that the perceptual convergence of the indirect lighting solution for the SHR technique is slower than the IDEPT approach (direct lighting is computed using the same DDL method). In our experience, the difference in performance between the IDEPT over SHR methods is far more significant for complex scenes. The SHR technique shows better performance for simple scenes only. Based on these results, we use the DDL+SHR technique for scenes built of fewer than 500 polygons. For scenes of more practical complexity, we consider the DDL, DDEPT and IDEPT techniques to optimize the progressive refinement of image quality.

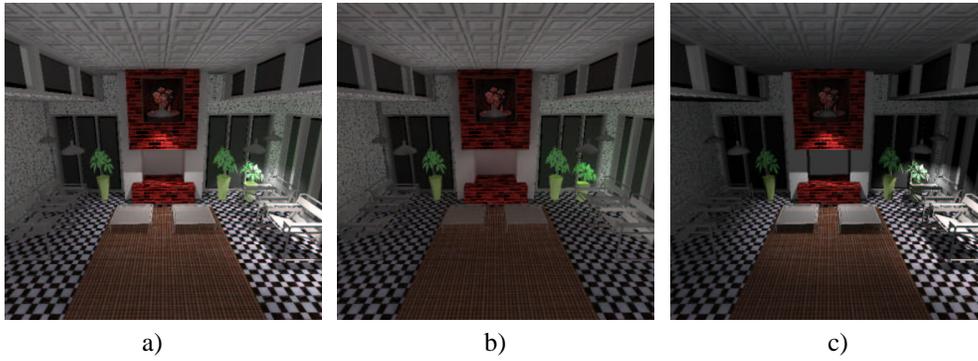


Figure 1.4: Test scene POINT: a) full global illumination solution, b) indirect lighting only, c) direct lighting only.

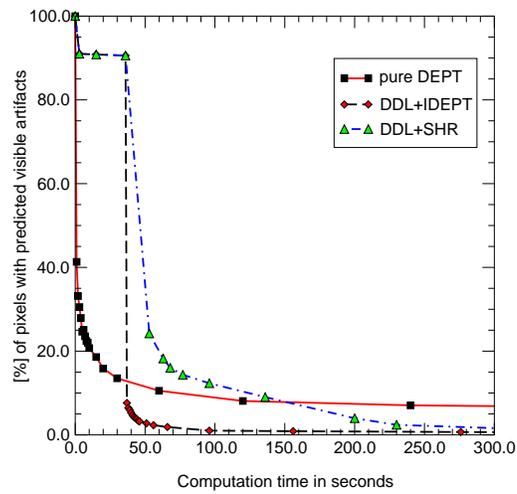


Figure 1.5: Plots of the VDP results (predicted global differences between the intermediate and final images) measuring the performance of the DEPT, DDL+IDEPT, and DDL+SHR algorithms for the POINT scene.

The graphs in Figure 1.5 show that at the initial stages of computation the DEPT technique provides the best performance, and rapidly gives meaningful feedback to the user. At later stages, the DDL+IDEPT hybrid shows faster perceptual convergence to the final image. In both cases, we used the same fixed mesh to bucket photons. Due to the basic mesh-element granularity, many subtle details of direct lighting distribution could not be captured well using the DDEPT technique. For example, small and/or narrow lighting patterns may be completely washed out. Also, when shadows are somehow reconstructed, they can be distorted and shifted with respect to their original appearance, and their boundaries can be excessively smooth. The problem of excessive discretization error, which is inherent in our DDEPT method, is reduced by the adaptive mesh subdivision used by the DDL technique.

The graphs in Figure 1.5 show that the algorithms examined have different performance at different stages of computation. This makes possible the development of a hybrid (composite) algorithm which uses the best candidate algorithm at a given stage of computation. This idea is further investigated in the following section.

### 1.4.2 Optimizing Progression of Global Illumination Computation

Based on the results of experiments measuring the perceptual convergence which were presented in the previous section for the POINT scene, and similar results obtained for different scenes we investigated (e.g., refer to [75]), a new hybrid technique that uses DDEPT, IDEPT and DDL can be proposed:

1. First, stochastic computations of direct and indirect lighting should be performed.
2. Second, the stochastically computed direct component should be gradually replaced by its deterministically computed counterpart to reconstruct the fine details of the lighting function.
3. Finally, stochastic indirect computation should be continued until some stopping criterion is reached, e.g., a criterion that is energy-based in terms of the solution variance (some engineering applications may require precise illumination values), or perception-based in terms of perceivable differences between the intermediate and final images [50].

All algorithms discussed use mesh vertices to store the results of direct and indirect lighting computations separately, so switching between them can easily be performed. The mesh is adaptively refined to fit the lighting distribution better in the case of the DDL technique only (refer to [53, 54]), but then indirect lighting computed using the IDEPT can be interpolated at the new vertices.

While the obtained ordering of the basic algorithms was the same across all tested scenes (refer also to [75]), the optimal selection of switchover points between the sequentially executed algorithms depended on the given scene characteristics. Ideally, the switchover points should be selected automatically based on the performance of component algorithms for a given scene, which could be measured by the on-line VDP computation. However, performing the VDP computation at the runtime of the composite algorithm computation is not acceptable because of the high costs of the VDP processing (Section 1.2). To overcome this problem we decided to elaborate a robust heuristic of the switchover points selection which provides good progression of the image quality for a wide range of indoor scenes. For this purpose, we designed another

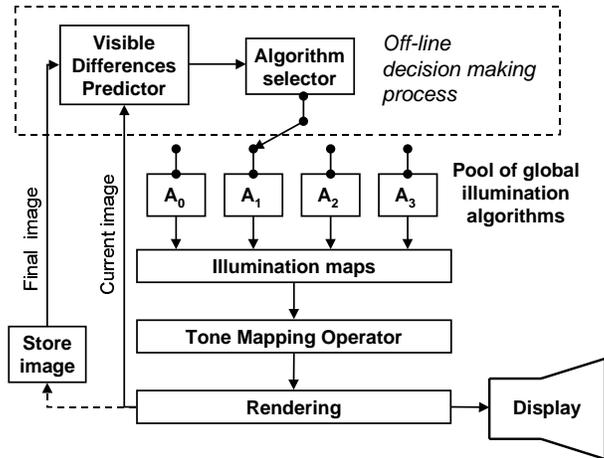


Figure 1.6: Experimental setting for evaluation of the image quality progression and selection of the switchover points between global illumination algorithms (the human-assisted selection is based on minimizing the perceptual distance between the intermediate and final images).

experiment involving the VDP off-line, and our experimental setting is shown in Figure 1.6.<sup>1</sup> Within this framework we applied the VDP to get quantitative measures of the image quality progression as a function of time points  $T_i$  at which switching between our basic algorithms DEPT (DDEPT+IDEPT), DDL, and IDEPT was performed.

The results of our experiments for the POINT test scene are summarized in Figure 1.7a. The thick line between two switchover points  $T_1$  and  $T_2$  depicts possible performance gains if DEPT is replaced by DDL at  $T_1$ , and then DDL is replaced by IDEPT at  $T_2$ . Also, we tried a different switching strategy, in which after switching from DEPT to DDL at  $T_1$ , we performed switching back and forth between the DDL and IDEPT algorithms. We refer to this strategy as  $T_1, \dots, T_N$ , where  $N$  stands for the number of switchover points. We investigated various choices of  $T_i$  ( $i > 1$ ), which controlled switching between the DDL and IDEPT algorithms. For example, we performed the switching after completion of every single iteration of the DDL computation, or every two such iterations and so on. Also, we changed  $T_1$ , which effectively controls the initial DEPT computation time. The thin line in Figure 1.7a shows an envelope of all graphs depicting our composite algorithm performance for all combinations of switchover points investigated by us. This envelope approximates the best expected performance of our composite technique assuming an “optimal” switching strategy between the DDL and IDEPT algorithms with multiple switchover points  $T_1, \dots, T_N$ . As can be seen, gains in performance achieved using the  $T_1, \dots, T_N$  strategy were negligible compared to the strategy based on well-chosen switchover points  $T_1$  and  $T_2$ . This observation was confirmed for other tests we performed [75].

<sup>1</sup>This setting is of general use and can be easily applied to any set of global illumination algorithms to select the best basic algorithm for a given task and computation stage.

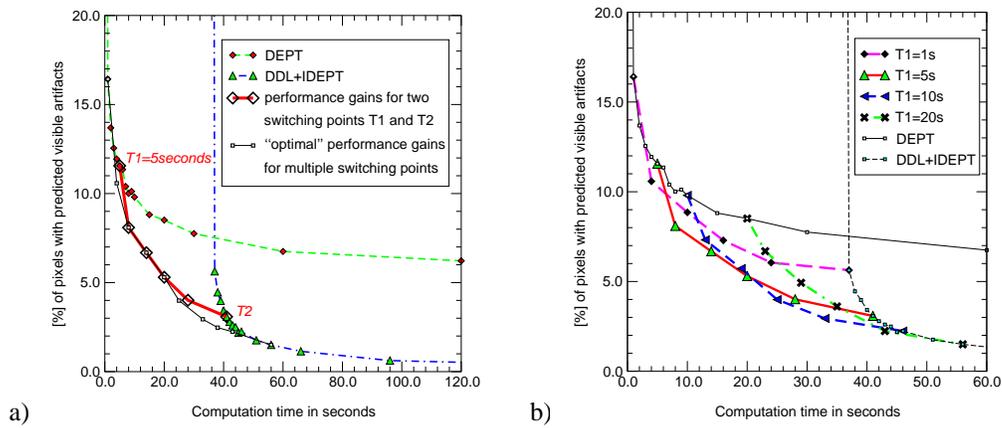


Figure 1.7: Plots of the VDP results (magnified from Figure 1.5) measuring the performance of DEPT and DDL+DEPT algorithms for the POINT test. a) The thick line between two switchover points  $T_1$  and  $T_2$  depicts possible performance gains if the DEPT is replaced by the DDL at  $T_1$ , and then the IDEPT is activated at  $T_2$ . The thin line depicts an “optimal” switching strategy between the DDL and IDEPT algorithms with multiple switchover points  $T_1, \dots, T_N$ . b) Performance gains for various choices of switching time  $T_1$ .

For the sake of simplicity of our composite algorithm, we decided to use just two switchover points  $T_1$  and  $T_2$ . We investigated various choices of  $T_1$ , which measures the duration of the initial DEPT computation. We assumed that  $T_2$  is decided automatically when the DDL computation is completed. The composite algorithm performance for various  $T_1$  is shown in Figure 1.7b. As can be seen our composite algorithm performs much better than the standalone DDL+IDEPT or DEPT methods for all choices of  $T_1$  which are considered in Figure 1.7b. In [75] we show that the choice of  $T_1$  is not extremely critical in terms of the image quality progressive refinement. However, a too short  $T_1$  may result in a poor quality of indirect lighting, which cannot be improved during the DDL computation. On the other hand, a too long  $T_1$  may result in an undesirable delay in reconstruction of shadows and other shading details. Because of this, the upper bound for  $T_1$  should be comparable to the computation time of the first iteration  $T_{i0}$  in the DDL processing, after which the first rendering of a complete direct lighting distribution becomes possible. We can estimate  $T_{i0}$  well by measuring the timings of pilot photons tracing and by knowing the number of initial mesh vertices, the number of light sources, and estimating the average number of shadow feelers (i.e., rays traced to obtain visibility information) for area and linear light sources.

Our heuristic for the  $T_1$  selection proceeds as follows. At first, we run the DEPT computation for time  $T_\alpha = \alpha T_{i0}$  (where  $\alpha = 0.1$ , and  $T_\alpha \geq 0.5$  seconds, since in our implementation we assumed that 0.5 seconds is the minimal interval for sampling DEPT solution errors). We then estimate the *RMS* error  $\tilde{E}$  of the indirect lighting simulation (we provide a derivation of the *RMS* error measure for the DEPT algorithm in [76]). Based on the results of DEPT computation for multiple scenes, we assume that a reasonable approximation of indirect lighting can usually be obtained for the *RMS* error threshold value  $E_{thr} \approx 15\%$ . Taking into account the basic properties of stochastic solution convergence [65], we estimate the required computation time  $T_{thr}$  to reach the accuracy level  $E_{thr}$  as

$$T_{thr} = T_\alpha \frac{\tilde{E}^2}{E_{thr}^2},$$

and finally, we set  $T_1$  as

$$T_1 = \min(T_{thr}, T_{i0}).$$

For simplicity, our heuristic relies on the energy-based criterion of indirect lighting accuracy. Obviously, in the perceptual sense this criterion does not guarantee the optimal  $T_1$  switchover point selection. However, we found that this heuristic provided stable progressive refinement of rendered image quality for all tests performed with multiple scenes. The robust behavior of our heuristic can be explained by the relative insensitivity of our composite algorithm to the selection of  $T_1$  [75], and the strong lowpass filtering properties of our lighting reconstruction method at the initial stages of computation.

Figure 1.8 shows an example of fast perceptual convergence of the intermediate solutions in terms of the perceived quality of the corresponding images. The THEATER scene is built of 17,300 polygons (tessellated into 22,300 mesh elements) and is illuminated by 581 light sources. Figures 1.8 depict nonfiltered (a) and filtered (b) illumination maps, obtained after 30 seconds of the DEPT computation. Figure 1.8b closely resembles the corresponding image in Figure 1.8c, which took 20 and 68 minutes of the DEPT and DDL computations, respectively. The final antialiased image (Figure 1.8d) was rendered using ray tracing, which took 234 minutes (the

image resolution was  $960 \times 740$  pixels). In the ray tracing computation, direct lighting was recomputed for every image sample. This solution is typical for multipass approaches, e.g., [27]. The indirect lighting was interpolated based on the results of the IDEPT computation, which are stored at mesh vertices. Since all surfaces of the scene in Figure 1.8 exhibit the Lambertian properties of light reflection, the illumination maps (Figures 1.8b and c) are of similar quality to that obtained using the ray tracing computation (Figure 1.8d). Obviously, once calculated, illumination maps make possible walkthroughs of adequate image quality almost immediately, while the ray tracing approach requires many hours of computation if the viewing parameters are changed. This example shows the advantages of high quality view-independent solutions for rendering environments with prevailing Lambertian properties.

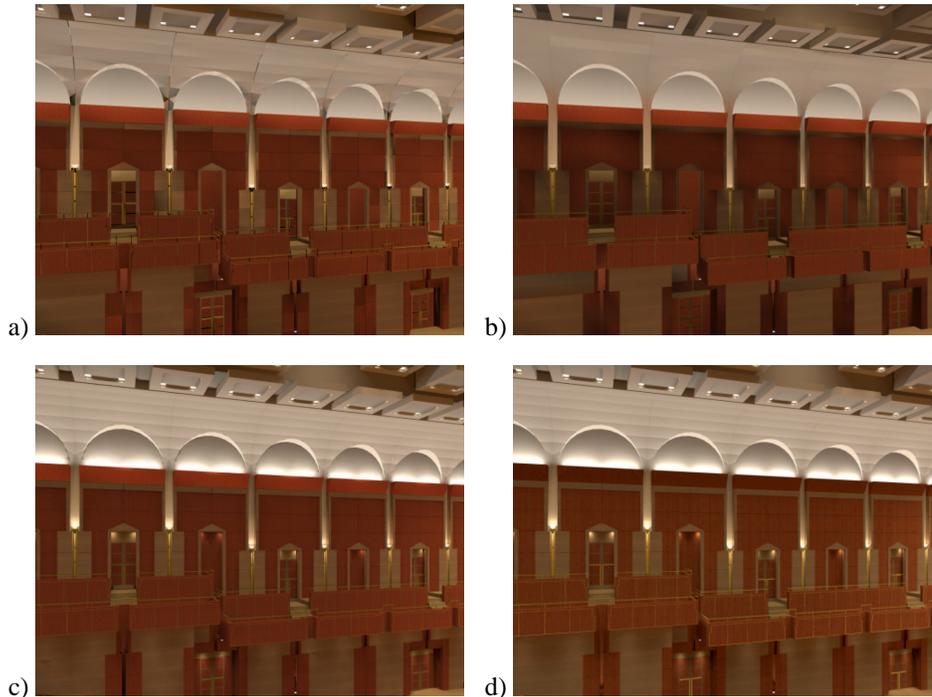


Figure 1.8: Comparison of various renderings for the THEATER scene: a) photon tracing without illumination map filtering (30 seconds), and b) photon tracing with filtering (30 seconds), c) enhanced accuracy of direct illumination (88 minutes), d) ray traced image (234 minutes).

It was impractical to use the VDP on-line (because of its computational costs) in algorithms that produce some intermediate results (images) rapidly, which was the case of our composite global illumination solution. However, for applications which require substantial computation time, embedding advanced HVS models might be profitable. In the following section we discuss an example of using the VDP on-line to decide upon the stopping conditions for global illumination computation which often requires many hours to be completed.

### 1.4.3 Stopping Conditions for Global Illumination Computation

Global illumination computation may be performed just to generate realistic images, or for some more demanding engineering applications. In both cases, quite different criteria to stop computation proved to be useful [50]. In the former case, computation should be stopped immediately when the image quality becomes indistinguishable from that of the fully converged solution for the human observer. A practical problem here is that the final solution is not known, because it is actually the goal of the computation. In the latter case, stopping conditions usually involve estimates of the simulation error in terms of energy, which is provided by the lighting simulation algorithm, and compared against a threshold value imposed by the user. For some algorithms such as radiosity it might be difficult to obtain a reliable estimate of simulation accuracy, while it is a relatively easy task for Monte Carlo techniques [73, 76, 77].

A common practice is to use energy-based error metrics to stop computation in realistic rendering applications. In our observation, such error metrics are usually too conservative, and lead to excessive computation times. For example, significant differences of radiance between the intermediate and final stages of solution which may appear in some scene regions, can lead to negligible differences in the resulting images due to the compressive power of the TMO used to convert radiance to displayable RGB. Occasionally, energy-based metrics prove to be unreliable and visible image artifacts may appear even though the error threshold value is set very low. Since the error is measured globally, it may achieve a low value for the whole scene but locally it can be still very high.

Clearly, some perception-informed metrics, which capture well local errors are needed to stop global illumination computation without affecting the final image quality. We decided to use the VDP for this purpose, encouraged by positive results of psychophysical experiments in similar tasks that we reported in Section 1.3.2. We assume that computation can be stopped if the VDP does not report significant differences between intermediate images. A practical problem is to select an appropriate intermediate image which should be compared against the current image to get robust stopping conditions.

We attempt to find a heuristic solution for this problem through experiments with the DDL+IDEPT technique which we described in Section 1.4.1. In this work we discuss the results obtained for the POINT test scene shown in Figure 1.4. However, we achieved similar results for other scenes we tested.

Let us assume that the current image  $\mathfrak{S}_T$  is obtained after the computation time  $T$ , and let us denote by  $VDP(\mathfrak{S}_T, \mathfrak{S}_{\alpha T})$  the VDP response for a pair of images  $\mathfrak{S}_T$  and  $\mathfrak{S}_{\alpha T}$  where  $0 < \alpha < 1$ . We should find an  $\alpha$  to get a reasonable match between  $VDP(\mathfrak{S}_T, \mathfrak{S}_{\alpha T})$  and  $VDP(\mathfrak{S}_C, \mathfrak{S}_T)$ , where  $\mathfrak{S}_C$  is an image for the fully converged solution. Figure 1.9 shows the numerical values of  $VDP(\mathfrak{S}_C, \mathfrak{S}_T)$  and  $VDP(\mathfrak{S}_T, \mathfrak{S}_{\alpha T})$  for  $T = \{100, 400, 1600\}$  seconds and various  $\alpha$ , for scene shown in Figure 1.4. While the numerical values of  $VDP(\mathfrak{S}_T, \mathfrak{S}_{0.5T})$  provide the upper bound for  $VDP(\mathfrak{S}_C, \mathfrak{S}_T)$  over all investigated  $T$ , it is even more important that the image regions with the perceivable differences are similar in both cases (refer to the VDP project Web pages [1] for color images with  $VDP(\mathfrak{S}_C, \mathfrak{S}_T)$  and  $VDP(\mathfrak{S}_T, \mathfrak{S}_{0.5T})$ ). This means that for certain regions of  $\mathfrak{S}_{0.5T}$  and  $\mathfrak{S}_T$  the variance of the luminance estimate is very small (below the perceived level), and it is likely that it will be so for  $\mathfrak{S}_C$ . For other regions such variance is high, and it is likely that luminance estimates for  $\mathfrak{S}_{0.5T}$  and  $\mathfrak{S}_T$  which fluctuate around the converged values for  $\mathfrak{S}_C$  will be different, and can be captured by the VDP. Thus, the

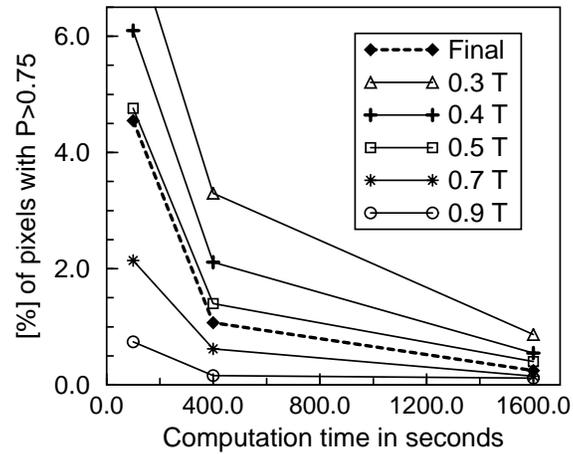


Figure 1.9: The VDP predicted differences between  $\mathfrak{S}_C$  and  $\mathfrak{S}_T$ , and  $\mathfrak{S}_T$  and  $\mathfrak{S}_{\alpha T}$  images.

choice of  $\alpha$  is a trade-off. The  $\alpha$  should be small enough to capture such perceivable fluctuations. However, it should not be too small because  $\mathfrak{S}_{\alpha T}$  may exhibit high variance in the regions in which the solution for  $\mathfrak{S}_T$  converged to that of  $\mathfrak{S}_C$ , with luminance differences below the noticeable level. In our experiments with stopping conditions for the DEPT technique for various scenes we found that  $\alpha = 0.5$  (50% of photons are the same for  $\mathfrak{S}_T$  and  $\mathfrak{S}_{0.5T}$ ) is such a reasonable trade-off.

## 2 Perception-Driven Rendering of High-Quality Walkthrough Animations

Rendering of animated sequences proves to be a very computation intensive task. In professional production this involves specialized rendering farms designed specifically for this purpose. Data revealed by major animation companies show that rendering times for the final antialiased frames are still counted in tens of minutes or hours [2], so shortening this time becomes very important. A serious drawback of traditional approaches to animation rendering is that error metrics controlling the quality of frames (which are computed separately one by one) are too conservative, and do not take advantage of various limitations of the HVS.

It is well-known in the video community that the human eye is less sensitive to higher spatial frequencies than to lower frequencies, and this knowledge was used in designing video equipment [72]. It is also conventional wisdom that the requirements imposed on the quality of still images must be higher than for images used in an animated sequence. Another intuitive point is that the quality of rendering can usually be relaxed as the velocity of the moving object (visual pattern) increases. These observations are confirmed by systematic psychophysical experiments investigating the sensitivity of the human eye for various spatiotemporal patterns [29, 81]. For example, the perceived sharpness of moving low resolution (or blurred) patterns increases with velocity, which is attributed to the higher level processing in the HVS [87]. This means that all techniques attempting to speed up the rendering of every single frame separately cannot account for the eye sensitivity variations resulting from temporal considerations. Effectively, computational efforts can be easily wasted on processing image details which cannot be perceived in the animated sequence. In this context, a global approach involving both spatial and temporal dimensions appears promising and is a relatively unexplored research direction.

In this work we present a framework for the perceptually-based accelerated rendering of animated sequences [55]. In our approach, computation is focused on those selected frames (keyframes) and frame fragments (inbetween frames), which strongly affect the appearance of the entire animation by depicting image details readily perceivable by the human observer. All pixels related to these frames and frame fragments are computed using a costly rendering method (we use ray tracing as the final pass of our SHR global illumination solution), which provides images of high quality. The remaining pixels are derived using inexpensive Image-Based Rendering (IBR) techniques [41, 45, 63]. Ideally, the differences between pixels computed using the slower and faster methods should not be perceived in animated sequences, even though such differences can be readily seen when the corresponding frames are observed as still images. A spatiotemporal perception-based quality metric for animated sequences is used to guide frame computation in a fully automatic and recursive manner.

In the following section we briefly introduce the basics of IBR techniques, and we show their non-standard applications in the context of animation walkthroughs. Then we propose our animation quality metric, and show its application to improve efficiency of rendering animation walkthrough sequences.

## 2.1 Image-Based Rendering Techniques

In recent years, Image-Based Rendering (IBR) techniques became an active research direction. The main idea behind the IBR is to derive new views of an object based on a limited number of reference views. The IBR solutions are especially appealing in the context of photographs of the real-world, because a high level of realism of the derived frames can be obtained while tedious geometric modeling required by the traditional (geometry-based) rendering can be avoided. A practical problem with the IBR techniques is that depth (range) data registered with every image are required to properly solve occlusions which arise when the camera translational motion is involved. For the real-world scenes this problem can be addressed using costly range scanning devices, or using computer vision methods [30], e.g., the stereo-pair method, which are usually far less accurate and robust.

The IBR approach is also used for generated images, in which case the geometrical model is available, so depth data of high accuracy can be easily obtained. The main motivation of using IBR techniques for synthetic scenes is rendering efficiency (it is relatively easy to achieve the rendering speed of 10 or even more frames per second on an ordinary PC without any graphics accelerator [63]). Figure 2.1 depicts the process of acquiring an image for the desired view based on two reference images (keyframes), and corresponding depth maps (the distance to the object is encoded in grey scale). At first, 3D warping [45] and reprojection of every pixel in the reference image to its new location in the desired image (inbetween frame) is performed. Usually a single reference image does not depict all scene regions that are visible from the desired view, which results in holes in the warped reference images. Such holes can be removed by combining information from multiple reference images during the compositing step (in the example shown in Figure 2.1 just two images are composited), which complements the desired image rendering. This requires a careful selection of the reference images to cover all scene regions which might be visible from desired views. For a walkthrough animation along a predefined path a proper selection of keyframes is usually easier because of the constraints imposed on the camera locations for desired views. We exploit this observation to improve the efficiency of high-quality rendering of walkthrough animations, which we discuss in Section 2.3.

The 3D warping technique [45] has one more application in the context of walkthrough animation sequences. As a result of the 3D warping of a selected frame to the previous (following) frame in the sequence, the displacement vector between positions of the corresponding pixels which represent the same scene detail is derived (refer to Figure 2.2). Because the time span between the subsequent animation frames is known (e.g., in the PAL composite video standard 25 frames per second are displayed), it is easy to compute the velocity vector based on the corresponding displacement vector. A vector field of pixel velocities defined for every image in the animation sequence is called the Pixel Flow (PF) which is the well-known notion in the digital video and computer vision communities [67]. In this work we focus on walkthrough animation sequences that deal exclusively with changes of camera parameters,<sup>2</sup> in which case a PF of good accuracy can be derived using the computationally efficient 3D warping technique. In Section 2.2.1 we show an application of the PF to estimate the human eye sensitivity to spatial patterns moving across the image plane.

<sup>2</sup>In the more general case of scene animation the PF can be computed based on the scripts describing motion of characters, changes of their shape and so on [64]. For the natural image sequences sufficient spatial image gradients must exist to detect pixel displacements, in which case so called optical flow can be computed [67]. The optical flow computation is usually far less accurate and more costly than the PF computation for synthetic sequences.

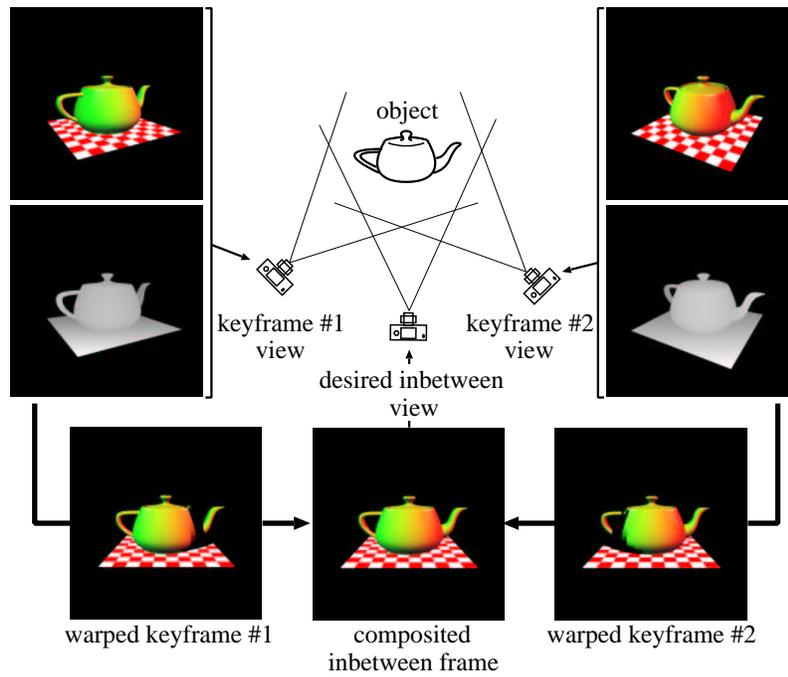


Figure 2.1: IBR: derivation of an image for the desired view based on two reference images.

## 2.2 Animation Quality Metric

Assessment of video quality in terms of artifacts visible to the human observer is becoming very important in various applications dealing with digital video encoding, transmission, and compression techniques. Subjective video quality measurement usually is costly and time-consuming, and requires many human viewers to obtain statistically meaningful results [69]. In recent years, a number of automatic video quality metrics, based on the computational models of human vision, has been proposed. Some of these metrics were designed for video [72,90], and are often specifically tuned (refer to [91]) for the assessment of perceivability of typical distortions arising in lossy video compression such as blocking artifacts, blurring, color shifts, and fragmentation. Also, some well-established still image quality metrics were extended into the time domain [39, 69, 83].

A lack of comparative studies makes it difficult to evaluate the actual performance of the discussed metrics. It seems that the Sarnoff model [39] is the most developed.<sup>3</sup> Also, Watson [83] proposed a metric based on the Discrete Cosine Transform which is computationally efficient and retains many basic characteristics of the Sarnoff model [84]. In this work we decided to use our own metric of animated sequence quality, which is specifically tuned for synthetic animation sequences.

<sup>3</sup>The product of Tektronix, Inc. called PQA-200 Picture Quality Analyzer test instrument includes the so called JNDmatrix which is based on this technology.

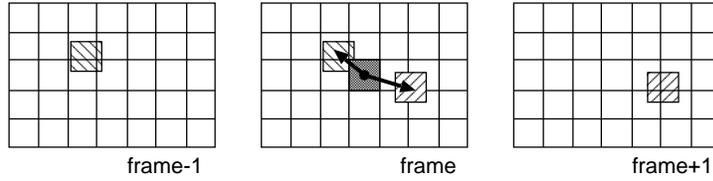


Figure 2.2: Displacement vectors for a pixel of the current frame in respect to the previous (frame-1) and following (frame+1) frames in an animation sequence. All marked pixels depict the same scene detail.

Before we move on to the description of our metric, we recall basic facts on the spatiotemporal Contrast Sensitivity Function (CSF) which is an important component of virtually all advanced video quality metrics. We show that in our application it is far more convenient to use the spatiovelocity CSF, which is a dual representation of the commonly used spatiotemporal CSF.

### 2.2.1 Spatiovelocity CSF Model

Spatiotemporal sensitivity to contrast, which varies with the spatial and temporal frequencies is an important characteristic of the HVS. The sensitivity is characterized by the so called spatiotemporal CSF, which defines the detection threshold for a stimulus as a function of its spatial and temporal frequencies. One of the most commonly used analytical approximations of the spatiotemporal CSF are the formulas derived experimentally by Kelly [29]. Instead of experimenting with flickering spatial patterns, Kelly measured contrast sensitivity at several fixed velocities for traveling waves of various spatial frequencies. Kelly used the well-known relationship of equivalence between the visual patterns flickering with temporal frequency  $\tau$ , and the corresponding steady patterns moving along the image plane with velocity  $\vec{v}$  such that [81]:

$$\tau = v_x \varrho_x + v_y \varrho_y = \vec{v} \cdot \vec{\varrho} \quad (2.1)$$

where  $v_x$  and  $v_y$  denote the horizontal and vertical components of the velocity vector  $\vec{v}$ , which is defined in the image plane  $xy$ , and  $\varrho_x$  and  $\varrho_y$  are the corresponding components of the spatial frequency  $\vec{\varrho}$ . Kelly found that the constant velocity CSF curves have a very regular shape at any velocity greater than about 0.1 degree/second. This made it easy to fit an analytical approximation to the contrast sensitivity data derived by Kelly in the psychophysical experiment. As a result, Kelly obtained the spatiovelocity CSF, which he was able to convert into the spatiotemporal CSF using equation (2.1).

We use the spatiovelocity CSF model provided by Daly [9], who extended Kelly's model to accommodate for the requirements of current CRT display devices (characterized by the maximum luminance levels of about  $100 \text{ cd/m}^2$ ), and obtained the following formula:

$$CSF(\varrho, v) = c_0 (6.1 + 7.3 \left| \log\left(\frac{c_2 v}{3}\right) \right|^3) c_2 v (2\pi c_1 \varrho)^2 \exp\left(-\frac{4\pi c_1 \varrho (c_2 v + 2)}{45.9}\right) \quad (2.2)$$

where  $\varrho = \|\vec{\varrho}\|$  is the spatial frequency in cycles per degree,  $v = \|\vec{v}\|$  is the retinal velocity in degrees per second, and  $c_0 = 1.14$ ,  $c_1 = 0.67$ ,  $c_2 = 1.7$  are coefficients introduced by Daly.

In [9, 55] a more extended discussion on estimates of the retinal velocity is available, which takes into account the eye natural drift, smooth pursuit, and saccadic movements. Figure 2.3 depicts the spatiovelocity CSF model specified in equation (2.2).

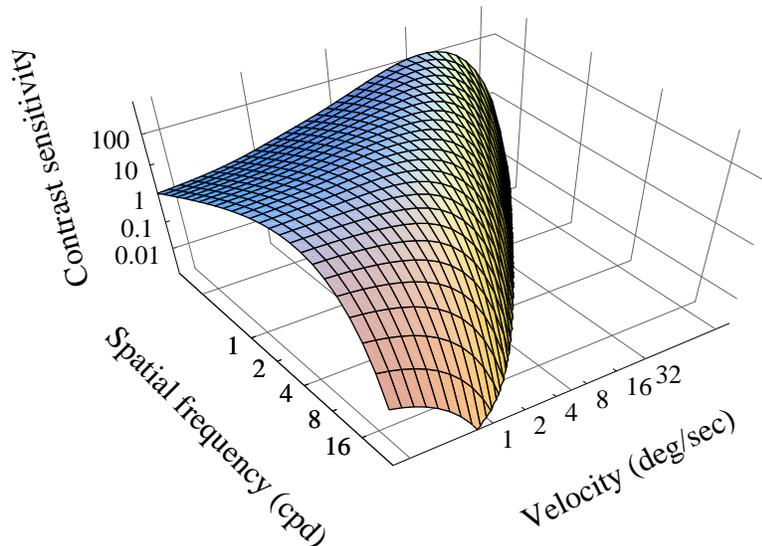


Figure 2.3: Spatiovelocity Contrast Sensitivity Function.

Although the spatiotemporal CSF is used by widely known video quality metrics, we chose to include the spatiovelocity CSF into our animation quality metric. Our design decision was encouraged by the observation that it is not clear whether the vision channels are better described as spatiotemporal (e.g., Hess and Snowden [25], and many other results in psychophysics) or spatiovelocity (e.g., Movshon et al. [49], and many other results especially in physiology). Also, accounting for the eye movements is more straightforward for a spatiovelocity CSF than for a spatiotemporal CSF [9]. Finally, the widely used spatiotemporal CSF was in fact derived from Kelly's spatiovelocity CSF, which was measured for moving stimuli (traveling waves). However, the main reason behind our choice of the spatiovelocity CSF is that in our application we deal with synthetic animation sequences for which it is relatively easy to derive the PF (as it was shown in Section 2.1). Based on the PF and using the spatiovelocity CSF, the eye sensitivity can be efficiently estimated for a given image pattern in the context of its motion across the image space. The spatiovelocity CSF is an important component of the Animation Quality Metric (AQM) developed by the author, which we describe in the following section.

### 2.2.2 Animation Quality Metric Algorithm

As the framework of our AQM we decided to expand the perception-based visible differences predictor for static images proposed by Eriksson et al. [14]. The architecture of this predictor was validated by Eriksson et al. through psychophysical experiments, and its integrity was

shown for various contrast and visual masking models [14]. Furthermore, we found that the responses of this predictor are very robust, and its architecture was suitable for an incorporation of the spatiovelocity CSF.

Figure 2.4 illustrates the processing flow of the AQM. Two comparison animation sequences are provided as input. For every pair of input frames the probability map  $P_{Map}$  of perceiving the differences between these frames is generated as output.  $P_{Map}$  provides for all pixels the probability values, which are calibrated in such a way that 1 Just Noticeable Differences (JND) unit [8, 39] corresponds to a 75% probability that an observer can perceive the difference between the corresponding image regions. While  $P_{Map}$  provides local information on the differences, for some applications it is more convenient to use just a single value which measures the differences globally. We assumed that the percentage of pixels in  $P_{Map}$  with the predicted differences over the 1 JND unit is a good measure of such global differences.

In the AQM computation each input frame undergoes the identical initial processing. At first, the original pixel intensities are compressed by the amplitude non-linearity and normalized to the luminance levels of the CRT display. Then the resulting images are converted into the frequency domain, and decomposition into spatial and orientation channels is performed using the Cortex transform which was developed by Daly [8] for the VDP. Then, the individual channels are transformed back to the spatial domain, and contrast in every channel is computed (the global contrast definition [14] with respect to the mean luminance value of the whole image was assumed).

In the next stage, the spatiovelocity CSF is computed according to the model of Kelly. The contrast sensitivity values are calculated using equation (2.2) for the center frequency  $\rho$  of each frequency band of the Cortex transform. The visual pattern velocity is estimated based on the average PF magnitude between the currently considered frame, and the previous and following frames (refer to Figure 2.2). As we discussed in Section 2.1, the PF can be estimated rapidly using the 3D warping technique, which requires access to the range data of the current frame and the camera parameters for all three involved frames. This means that the access to well localized data in the animation sequence is required. Since the visual pattern is maximally blurred in the direction of retinal motion, and spatial acuity is retained in the direction orthogonal to the retinal motion direction [13], we project the retinal velocity vector onto the direction of the filter band orientation. The contrast sensitivity values function are used to normalize the contrasts in every spatial frequency-orientation channel into the JND units. Next the visual masking is modeled using the threshold elevation approach [14]. The final stage is error pooling across all channels.

The AQM is well suited to computer graphics applications, and can be used to determine when a lower image quality will be not perceived for a given frame, and its local regions. As pointed out in [10] our AQM has a strong potential in efficient guiding of the video compression to determine the level of perceived details locally per frame. In this work we apply the AQM to guide inbetween frame computation, which we discuss in the following section.

### 2.3 Rendering of the Animation

For animation techniques relying on keyframing the rendering costs depend heavily upon the efficiency of inbetween frame computation because the inbetween frames usually significantly outnumber the keyframes. We use IBR techniques [41, 45] described in Section 2.1 to derive the inbetween frames. Our goal is to maximize the number of pixels computed using the IBR

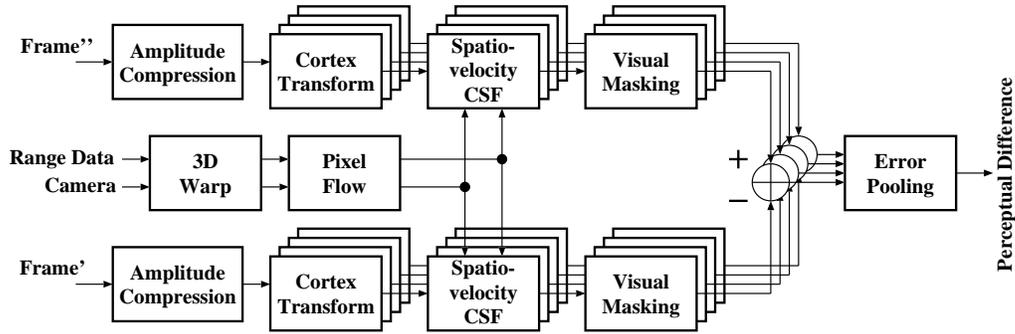


Figure 2.4: Animation Quality Metric. The spatiovelocity CSF requires the velocity value for every pixel, which is acquired from the PF. The PF is computed for the previous and following frames along the animation path in respect to the input frame' (or frame'' which should closely correspond to frame').

approach without deteriorating the animation quality as perceived by the human observer.

The quality of pixels derived using IBR techniques is usually lower than ray-traced pixels, e.g., in the regions of derived frames which are expanded in respect to the reference frames. The HVS sensitivity is especially high for such quality degradations when the PF values are low. We replace IBR-derived pixels in such regions with ray-traced pixels. The replacement is performed when the PF velocity is below a specified threshold value, which we estimated in subjective and objective (using the AQM) experiments [56]. In typical animations, usually only a few percent of the pixels are replaced, unless the camera motion is very slow. Those pixels are usually grouped around a so called focus of expansion [67] which represents the position in the image corresponding to the point towards which the camera is moving.

Since specular effects are usually of high contrast and they attract the viewers attention when looking at a video sequence [58], special care is taken to process them properly. Existing IBR methods require costly preprocessing to obtain specular effects of good quality [20, 31, 33, 48]. For example, a huge number of pre-calculated images is needed to obtain crisp mirror reflections. Because of these problems we decided to use ray tracing for pixels depicting objects with strong specular properties. We use our AQM to decide for which objects with glossy reflectance or transparent properties such computations are required.

Pixels representing objects in the inbetween frames which are not visible in the keyframes cannot be properly derived using the IBR techniques, and we apply ray tracing to fill the resulting holes in frames (refer to Section 2.1). An appropriate selection of keyframes is an important factor in reducing the number of pixels which must be ray traced. This issue is discussed in more details in [56]. In particular, it is shown that the accumulated PF along the animation path can be used to improve the performance of IBR computation. It turns out, that by reducing the variance of the accumulated PF between the animation segments, the number of pixels in inbetween frames which depict objects invisible in the keyframes can usually be reduced.

In this work for the sake of simplicity we assume that initially the keyframes are placed sparsely and uniformly along the animation path which is known in advance. Then adaptive keyframe selection is performed, which is guided by the AQM predictions. We provide a de-

tailed description of this solution in the following section. Then, we discuss the performance of our approach for a case study walkthrough animation.

### 2.3.1 Adaptive Refinement of Keyframe Placement

At first the initial keyframe placement is decided by choosing the constant length of  $N+1$  frames for all animation segments  $S$ . Then every  $S$  is processed separately applying the following recursive procedure:

1. Generate the first frame  $k_0$  and the last frame  $k_N$  in  $S$  using ray tracing. The keyframes that are shared by two neighboring segments are computed only once.
  2. Apply 3D warping to keyframes  $k_0$  and  $k_N$  to derive two instances  $k'_{[N/2]}$  and  $k''_{[N/2]}$  of an inbetween frame  $[N/2]$ .
  3. Use the AQM to compute the probability map  $P_{Map}$  with perceivable differences between  $k'_{[N/2]}$  and  $k''_{[N/2]}$ .
  4. Mask out from  $P_{Map}$  all pixels that must be ray traced because of the IBR deficiencies (discussed in Sections 2.1 and 2.3). The following order for masking out pixels is taken:
    - (a) Mask out from  $P_{Map}$  pixels with low PF values (in [56] we discuss experimental derivation of the PF threshold value used for such masking).
    - (b) Mask out from  $P_{Map}$  pixels depicting objects with strong specular properties (i.e., mirrors, transparent and glossy objects). The item buffer [85] of frame  $k_{[N/2]}$  is used to identify pixels representing objects with such properties. Only those specular objects are masked out for which the differences between  $k'_{[N/2]}$  and  $k''_{[N/2]}$  as reported in  $P_{Map}$  can be readily perceived by the human observer. In Section 2.3.2 we provide details on setting the thresholds of the AQM response, which are used by us to discriminate between the perceivable and imperceivable differences.
    - (c) Mask out from  $P_{Map}$  holes composed of pixels that could not be derived from keyframes  $k_0$  and  $k_N$  using 3D warping.
  5. If masked-out  $P_{Map}$  shows the differences between  $k'_{[N/2]}$  and  $k''_{[N/2]}$  for a bigger percentage of pixels than the assumed threshold value:
    - (a) Split  $S$  at frame  $k_{[N/2]}$  into two subsegments  $S_1$  ( $k_0, \dots, k_{[N/2]}$ ) and  $S_2$  ( $k_{[N/2]}, \dots, k_N$ ).
    - (b) Process recursively  $S_1$  and  $S_2$ , starting this procedure from the beginning for each of them.
- Else
- (a) Composite  $k'_{[N/2]}$  and  $k''_{[N/2]}$  with correct processing of object occlusions [41, 63] to derive  $k_{[N/2]}$ .
  - (b) Ray trace all pixels which were masked out in the step 4 of this procedure, and composite these pixels with  $k_{[N/2]}$ .

- (c) Repeat the two latter steps for all remaining inbetween frames, i.e.,  $k_1, \dots, k_{[N/2]-1}$  and  $k_{[N/2]+1}, \dots, k_{N-1}$  in  $S$ .

To avoid image quality degradation resulting from multiple resamplings, the fully ray-traced reference frames  $k_0$  and  $k_N$  are always warped in step 5c to derive all inbetween frames in  $S$ . Pixels to be ray traced, i.e., pixels with low PF values, pixels depicting specular objects with visible differences (such objects are selected once for the whole  $S$  in step 4b), and pixels with holes resulting from the IBR processing must be identified for every inbetween frame separately. We evaluate the AQM response only for frame  $k_{[N/2]}$ . We assume that derivation of  $k_{[N/2]}$  applying the IBR techniques is the most error-prone in the whole segment  $S$  because its arclength distance along the animation path to either the  $k_0$  or  $k_N$  frames is the longest one. This assumption is a trade off between the time spent for rendering and for the control of its quality (we discuss the AQM costs in Section 2.3.2), but in practice, it holds well for typical animation paths.

Figure 2.5 summarizes the computation and compositing of an inbetween frame. We used a dotted line to mark those processing stages that are performed only once for segment  $S$ . All other processing stages are repeated for all inbetween frames.

As a final step, we apply a spatiotemporal antialiasing technique, which utilizes the PF to perform motion-compensated filtering (refer to [55] for more details).

### 2.3.2 A Case Study Walkthrough Animation

In this work we choose as a case study a walkthrough animation for the ATRIUM scene shown in Figure 3.1b (more details on this scene are provided in Chapter 3). The main motivation for this choice were the interesting occlusion relationships between objects which are challenging for IBR. Also, a vast majority of the surfaces exhibit some view-dependent reflection properties, including the mirror-like and transparent surfaces, which made the computation of inbetween frames more difficult. Under such conditions, the AQM guided selection of keyframes and glossy objects within inbetween frames to be recomputed was more critical, and wrong decisions concerning these issues could be easy to perceive. In [56] we discuss another case study in which a walkthrough within the POINT scene (shown in Figures 1.4 and 2.5) is analyzed with an emphasis on strong variations of the PF along the animation path.

For our experiments we selected a walkthrough sequence of 200 frames. At the initial keyframe selection step, we assumed the length  $N + 1 = 25$  frames for each animation segment  $S$ . Figure 2.6a illustrates adaptive refinement of the initial keyframe placement guided by the AQM predictions. We use the global measure (refer to Section 2.2.2) of the differences between frames, i.e., the percentage of pixels in  $P_{Map}$  for which the differences are over 1 JND. Note that only pixels to be derived using the IBR approach are considered, while pixels to be ray traced are masked out (refer to Section 2.3.1). The filled squares in Figure 2.6a show the global AQM predictions of the differences between the subsequent keyframe pairs:  $k_0$  warped to  $k'_{[N/2]}$ , and  $k_N$  warped to  $k''_{[N/2]}$  for every initial segment  $S$ . Segments with global predictions over 10% are split, and the filled diamonds show the corresponding reduction of the predicted perceivable differences between the newly inserted frames. The 10% threshold was chosen experimentally, and can be justified by the fact that for an animated sequence the observer can only fixate at one location per frame. For such a location and its surround of approximately 1 visual degree

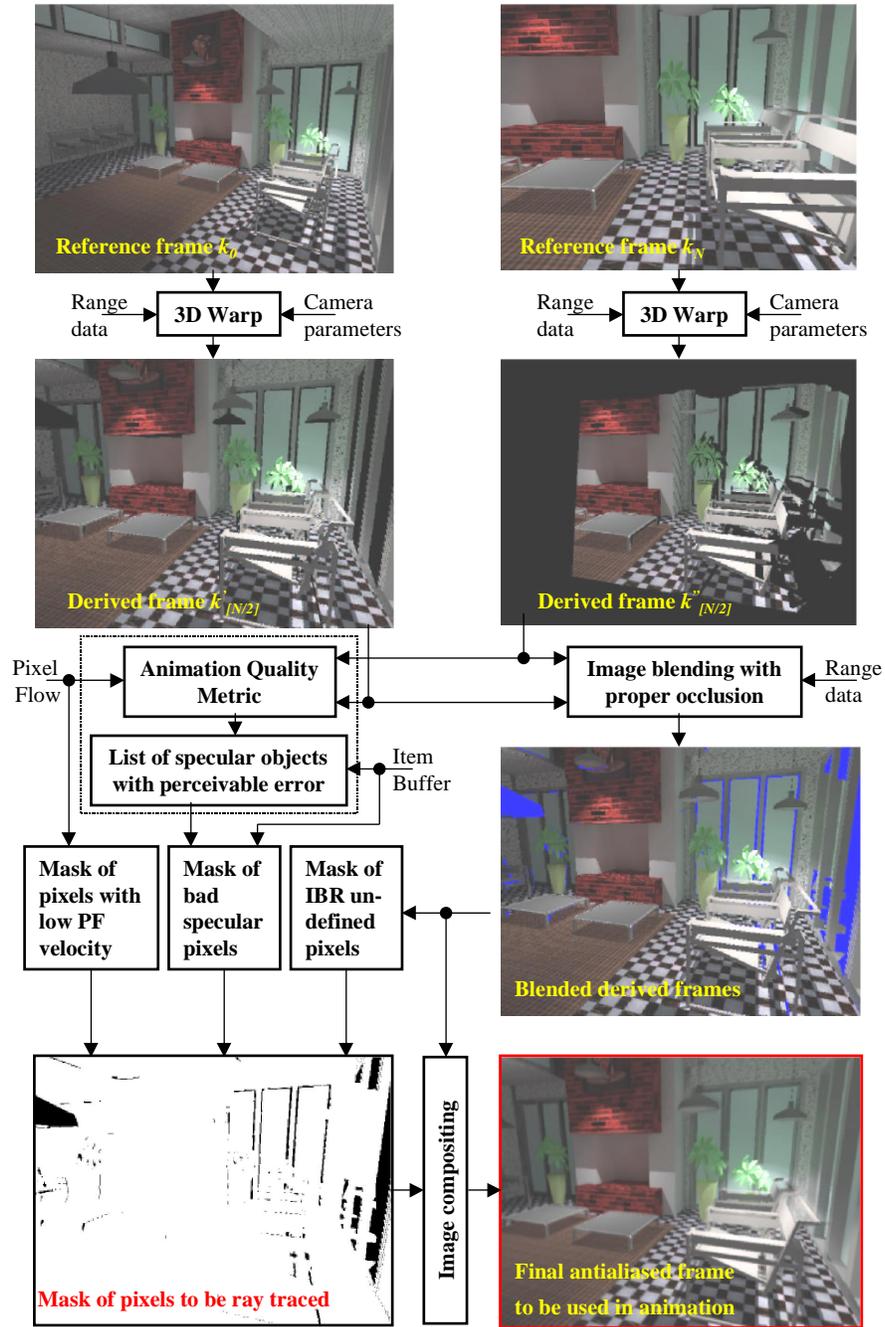
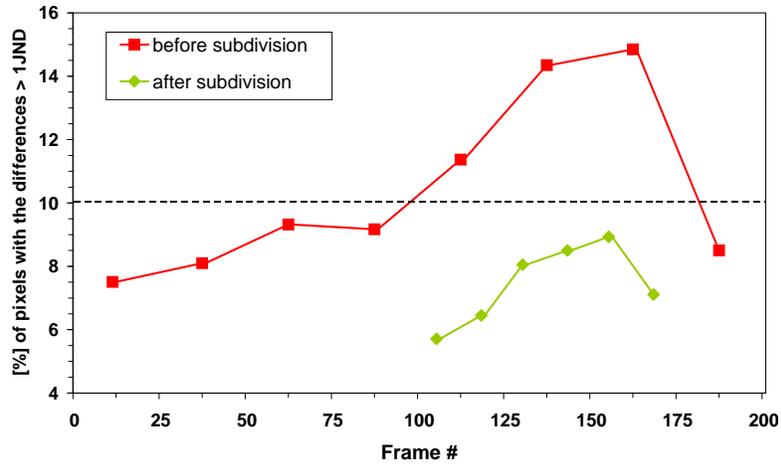
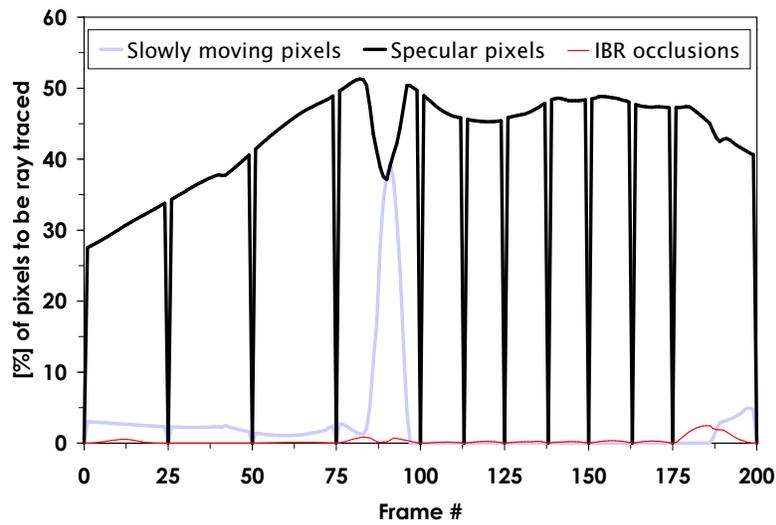


Figure 2.5: The processing flow for inbetween frames computation.



a)



b)

Figure 2.6: ATRIUM walkthrough statistics: a) the AQM prediction of the perceived differences between the warped images of two neighboring reference frames, b) the percentage of pixels to be recalculated by ray tracing. In a) lines connecting the symbols were added for the figure readability and they do not have any meaning for unmarked frames.

the eye sensitivity is high due to the foveal vision [79], while it decreases significantly for the remaining image regions which are perceived by means of the peripheral vision (eccentricity effect). The AQM sensitivity is tuned for the foveal vision because at the stage of animation rendering it is not known where the observer will be looking. This means that the AQM predictions might be too conservative for many image regions, and the degradation of quality usually cannot be perceived unless the observer attention is specifically attracted to these regions. To improve the rendering performance, we chose a trade-off solution in which visible differences between warped keyframes are allowed for a small number of pixels (up to 10%). Although some perceivable quality problems may arise for these pixels, it is most likely that the observer will not notice them at all.

The overall costs of the computation of inbetween frames are strongly affected by the number pixels of that must be ray traced. As we discussed in Section 2.3, we replace IBR-derived pixels by ray traced pixels for image patterns moving with low velocity. The graph in Figure 2.6b shows the percentage of such pixels for which the PF values are below the experimentally derived threshold velocity 0.5 degree/second (for details concerning the derivation of this threshold value refer to [56]). Also, we use ray tracing to derive pixels depicting specular objects for which the IBR technique leads to the AQM predicted degradation of the animation quality. As described in Section 2.3.1, for every segment  $S$  we run the AQM once to decide upon the specular objects which require recomputation. If a group of connected pixels representing an object (or a part of an object) exhibits differences greater than 2 JND units (a 93.75% probability of the difference discrimination), we select such an object for recalculation. If differences below 2 JND units are reported for an object by the AQM then we estimate the ratio of pixels exhibiting such differences to all pixels depicting this object. If the ratio is bigger than 25%, we select such an object for recomputation - 25% is an experimentally selected trade-off value, which makes a reduction in the number of specular objects requiring recomputation possible, at the expense of some potentially perceivable image artifacts. The graph in Figure 2.6b shows the percentage of specular pixels selected for recomputation. Finally, the percentage of pixels that are ray traced due to IBR occlusion problems is included in this graph. Table 2.1 summarizes the results depicted in Figure 2.6b by providing the average percentage of pixels per frame to be ray traced. Note that the keyframe pixels, which are always ray traced, are included in the average.

Although the AQM processing costs are relatively high (it takes 243 seconds<sup>4</sup> to process a pair of frames), the overall computation cost per frame was reduced from 40 minutes to 20.5 minutes without noticeable differences in the animation quality.<sup>5</sup> Even better performance can be expected for environments in which specular objects are depicted by a smaller percentage of pixels, and camera motion is faster.

---

<sup>4</sup>We had to consider images of resolution  $1024 \times 512$  for the Fast Fourier Transform processing of our frames of resolution  $640 \times 480$ .

<sup>5</sup>All timings reported in this section were measured on a MIPS 195 MHz processor.

Table 2.1: Statistics of the ray traced pixels in the ATRIUM walkthrough.

Slow motion [%]	Specular objects [%]	IBR occlusions [%]	Keyframes [%]	Total [%]
2.4	40.8	0.3	6.0	49.5

### 3 Validation of Global Illumination and Rendering Solutions

It is relatively easy to use commodity rendering techniques to create great looking images; however, it is much more difficult to create images that match the appearance of a real environment [46, 54]. The basic precondition to achieving this goal is physically-based lighting simulation, which is a computationally demanding problem. To make computation tractable in practical applications, many simplifying assumptions are usually introduced to underlying physical models. Because analytic evaluation of such simplifications and interactions between them is generally impractical, the correctness of a given technique must be checked experimentally by a comparison of simulation results to some reference data [52]. For example, the distribution of illumination at some predefined points derived analytically or measured experimentally can be used to validate the lighting simulation part of a rendering algorithm. An effective way to test complete rendering algorithms, including Tone Mapping Operators (TMO) used for displaying images on the CRT device [71], is a direct comparison of the appearance of virtual and real-world images as seen by the human observer [46, 47]. Unfortunately, such experimental validation was almost never performed for existing global illumination solutions, which makes it difficult to compare their efficiency, or even test their implementation correctness.

One of the reasons such validation experiments are rarely performed in practice is lack of standardized, robust, non-trivial, and easily accessible test data. Ideally, a standard set of diversified tests should be available, such that the performance of global illumination solutions could be measured in terms of the achieved lighting simulation accuracy, image fidelity, and computation time. The existence of such tests would make possible the comparison of particular elements of rendering algorithms as well. Other areas of research such as computer vision, and augmented reality could also benefit from such test scenes in their studies.

There have been some isolated attempts to develop such standard tests. A well-known example is the Cornell box, which was used to validate radiosity solutions [47]. Rushmeier et al. [62] developed a model of a conference room with four different lighting systems and corresponding photographs of its real world counterpart. Recently, McNamara et al. [46] built a low scale environment, and requested the human observers to compare its appearance against images generated by the Radiance rendering system [80]. These scenes are of relatively low complexity in terms of geometry and lighting, and only the Lambertian and specular reflectance data are posted on the corresponding Web pages. Myszkowski and his students developed more complete data of a complex environment (an atrium at the University of Aizu), which are disseminated through the Internet for public use [51].

In the following section we provide more details on the atrium test scene. Also, we present the obtained validation results for the SHR and hybrid algorithms (refer to [53, 54]) using the

atrium data. Finally, we describe a multi-stage validation procedure designed specifically for global illumination and rendering solutions developed by Myszkowski and Kunii [54].

### 3.1 Experiments with the Atrium Scene

The main goal of our experiments with the atrium test scene was to predict the appearance of a full-scale, real world architectural environment of significant complexity given all required input data such as geometry, description of light sources, surface reflectance characteristics, and textures. The atrium at the University of Aizu was chosen for its challenging lighting model, the richness and diversity of the materials employed in the construction, and the geometrical purity of its architectural structure. Two versions of the atrium data have been developed by the author and his students within the last six years.

In the earlier version of the atrium data, the geometrical model was significantly simplified (it was built of about 28,000 polygons) partially to account for the performance of graphics workstations available in the mid-nineties. Since the reflectance and color characteristics of construction materials used in the atrium were not available at that time as well, their manual tuning was performed to reduce the differences between the synthesized image and the corresponding photograph. In Section 3.1.1 we describe the overall procedure which we took to obtain a good match between these images.

In the current version of the atrium data, special care was taken to reconstruct a vast majority of the atrium details. The model is built of almost 700,000 polygons. The reflectance characteristics of the most important construction materials were experimentally measured, and BRDFs of high sampling density were obtained. Also, the lighting distribution on the floor was measured at selected sample points. The goal of the experiments with these data was twofold:

- Estimation of the lighting simulation accuracy through comparison with measurement data;
- Evaluation of computer images fidelity in respect to the real world environment.

In Section 3.1.2 we summarize the obtained results.

#### 3.1.1 Simplified Model

The goal of the experiment performed with the earlier version of data was to match the synthesized image to the photograph. The atrium geometry, although simplified, was properly reconstructed for all major surfaces taking part in lighting interactions. Furthermore, the spatial candle-power characteristics (so called goniometric diagrams) of two types of light sources installed in the atrium were obtained from their manufacturer (Matsushita Electric, Inc.). In total 108 light sources were modeled. However, the reflectance characteristics of the materials used in the atrium were not available. As the first approximation, the reflectance coefficients were assigned based on the data which are available in literature (e.g., [59]) for similar materials. Only specular and diffuse reflectance coefficients were used, and their values were manually adjusted to reduce the visible differences between the image and the photograph.

The question of media used for judging such differences immediately arises. If the media are the same, and viewed under similar conditions (e.g., photographs compared side-by-side), then the observer's state of adaptation is also very similar, and his judgment becomes quite reliable.

Obviously, the fidelity problem of image reproduction on a given medium must be properly addressed, which in practice proves to be a very difficult problem [15]. In our experiments, we used the CRT display device to compare images, because this method offers more flexibility in contrast and brightness manipulations than the hard-copy photographs.

We took the following procedure for the display of synthetic and photographic images. First, we confirmed that we were able to present the photographic image under display conditions that gave the best match between this image and the real-world environment. For this purpose, we adjusted contrast, gamma correction and brightness of the photographic image. Next, we proceeded with displaying the synthetic image using the TMO which was originally developed by Tumblin and Rushmeier [71], but was adapted to our needs. The synthetic image was generated using the ray tracing technique based on the view-independent illumination maps, which were produced using the SHR approach. Luminance values were computed for every pixel of the image, and transformed to perceived brightness values which were predicted using Stevens' power law [71] for given observation conditions. Assuming that the lighting simulation results are correct, these brightness values should correspond to those of the real-world scene. The brightness transformation was also performed for the range of luminances produced by the display device. Then the image brightness values were mapped to those of the display device while preserving the contrast relation in respect to the maximum brightness value in the image and the maximum brightness value which can be reproduced by the CRT display device. The goal of this mapping was to overcome limitations of the CRT device in the dynamic luminance range reproduction, while obtaining a believable appearance of the displayed image in respect to the real-world scene appearance. Then, we adjusted the parameters of the brightness transformation until the appearance of the synthetic image (Figure 3.1b) best matched that of the photographic image (Figure 3.1a). Since we were unable to determine directly a given viewer's light adaptation level, this method allowed the brightness function to be chosen that worked best for that viewer under those viewing conditions. Note that the chroma was maintained separately to avoid color shifts, which could arise from applying the non-linear brightness transformation to every RGB channel.

While our atrium rendering was far from being perfect, first impressions when observing the rendered image and comparing its appearance with the photographic image were quite favorable. In fact, many viewers who were quite familiar with the real atrium thought that they were viewing the actual photographic images, when they first viewed the synthetic images. This means that, in terms of the absolute evaluation, the quality of our images was acceptable. However, when the same viewers compared the synthetic images to the photographic images, they were able to find many differences and therefore were able to provide us very useful feedback. Apart from clearly visible geometry simplifications in the model, the viewers detected our unrealistic reflections on the side atrium wall, which were originally built of glossy metallic panels. It became clear that simple Lambertian and specular reflectance functions cannot properly reproduce the appearance of such surfaces, and an exact BRDF must be considered instead. Also, some viewers noticed that while the image and photographs have similar appearance, they poorly reconstruct the real-world atrium appearance.

The atrium is illuminated by mixed fluorescent and incandescent lighting, and the standard photographic techniques failed to produce satisfactory results. The daylight as well as tungsten films resulted in unnaturally yellowish and bluish images (the question arises whether the commonly used term "photorealistic rendering" is adequate, or rather just "realistic rendering" should be

used instead). We chose the daylight film, which subjectively produced more believable results. Since we adjusted manually surface reflectance coefficients and color to reduce the differences between the resulting rendered image and the photograph, the image became yellowish as well. In other words, a good match to the photograph was obtained, which in this case was not equivalent to a good reproduction of the real-world atrium appearance.

In the following section we describe the next generation of the atrium model in which a vast majority of the drawbacks discussed were removed.

### 3.1.2 Complete Model

The main motivation of further work on the atrium model was to overcome its most annoying drawbacks that were pointed out by the viewers judging the atrium rendering quality. The following measures were taken to improve the fidelity of rendered atrium images in respect to the real-world scene:

- **Geometry.** The geometrical model was very carefully prepared based on the blueprints and direct measurements within the atrium. Every detail of the model was described with a very high accuracy.
- **Surface reflectance characteristics.** The BRDF measurement for six major construction materials was performed in cooperation with the Integra, Inc. company. The materials selected for measurement represent more than 80% of the total surface area in the atrium. The reflectance attributes for the remaining surfaces were estimated based on available literature, e.g., [59]. To derive the color of painted surfaces, which were not selected for the BRDF measurement, the Standard Paint Color charts edited by the Japan Paint Manufacturers Association were used, and the corresponding RGB values were derived from a Toyo 88 RGB color finder.
- **Textures.** Texture acquisition was performed using a digital camera. Standalone samples of the major construction materials, allowed for much better control of the illumination of the sample than would have been possible directly in the atrium environment.
- **Light sources.** The same goniometrical diagrams as in the earlier version of the atrium model were used. While these diagrams are highly accurate, a practical problem arises with estimation of the maintenance factor for every light source to account for its utilization level, accumulated dirt and so on. Some atrium light sources are used on a daily basis, while others are usually switched off to cut down electricity costs. To overcome this problem, the illumination due to every light source was measured trying to minimize the influence of indirect lighting. A luxmeter probe was placed in a deep box in such a way that the probe was directly illuminated, while indirect lighting influence was suppressed by the box. To reduce interreflections within the box all its sides were covered with a black fabric. Measurements to determine a reference were taken from newly installed bulbs in a dirt free fixture. The maintenance factor for this reference light source was assumed to be 100%. The maintenance factors of the other light sources were scaled in respect to the reference, based on measurements of illumination at similarly selected sample points. The lowest maintenance factor was 62%, which clearly shows that the catalogue data should be properly interpreted to account for characteristics of the real world environment.



a)



b)

Figure 3.1: An atrium of the Research Quadrangle at the University of Aizu: a) photograph, b) rendering based on the simplified atrium model.

More details on the atrium modeling issues are available on the atrium project Web pages [51].

We used the atrium data to validate the lighting simulation accuracy of our hybrid algorithm described in Section 1.4.2. Also, we investigated to what extent the image appearance is influenced by the replacement of the measured BRDF by a less rigorous approach to light scattering, in which simple approximations of the surface reflectance by Lambertian and mirror reflection models is considered. To estimate these coefficients hemispherical integration of the measured BRDF for various incident angles was performed, and the obtained values were averaged. This corresponds to the situation when the approximated reflectance can be somehow acquired, e.g., by using a simple measurement device, or by taking standard values from textbooks for the most common materials. This scenario is also useful when the global illumination and rendering software does not support complex reflectance functions described using BRDFs.

We measured the lighting distribution at the atrium floor for 84 sample points. Although all measurement points were directly illuminated, indirect lighting was also significant accounting for about 30–55% of the total illumination. The graphs in Figure 3.2 show the distribution of measured and simulated illumination at selected sample points, which are located at the atrium floor along three different lines (refer to the atrium project Web pages [51] for a complete documentation of experiments performed with tabulated illumination and error values). The best results were obtained from the measured BRDF, in which case the average simulation error in respect to the measured illumination was 10.5%. For the approach based on the averaged diffuse and specular reflectance, the error increased to 18.2%. Errors below 5%, 10%, and 20% were obtained for about 40%, 75%, and 93% of the sample points, respectively, when the measured BRDFs were used. When the averaged reflectance coefficients were used errors below 5%, 10%, and 20% were obtained for 8%, 31%, and 68% of the sample points, respectively.

The obtained results may appear pessimistic in terms of the simulation accuracy achieved. However, the following tolerances for lighting design applications are proposed in the guidelines issued by Commission Internationale de l'Éclairage [17]: 10% for average illuminance calculations and 20% for measured point values. Such a high tolerance is a result of the realistic evaluation of the accuracy of input data for lighting simulation such as the BRDF of materials, description of light sources, and simplifications of geometrical models. Taking into account some possible inaccuracies in our atrium model, it can be stated that our hybrid technique mostly meets the requirements imposed on the simulation accuracy in lighting engineering applications. However, this is the case only when the measured BRDFs were used, while for the averaged diffuse and specular reflections the errors are too big. On the other hand, such inaccuracies usually do not affect significantly the image quality. To our experience it is important to model properly the mirror-like reflections for glossy surfaces. For the BRDF approach this is an easy task, but for the averaged diffuse and specular reflections some tricks are required, e.g., activating the distributed ray tracing [23] with many reflected rays spread within a solid angle centered around the mirror reflection direction.

A robust prediction of real-world scene appearance based on valid input data opens many important applications for realistic rendering. We performed psychophysical experiments involving 25 subjects (12 of them got basic training in realistic image synthesis and could be considered as experts) to check how different are rendered images in respect to their real-world counterpart. The settings used in our experiments are shown in Figure 3.3. We did not use photographs as reference images in this comparison because of the fidelity problems discussed in

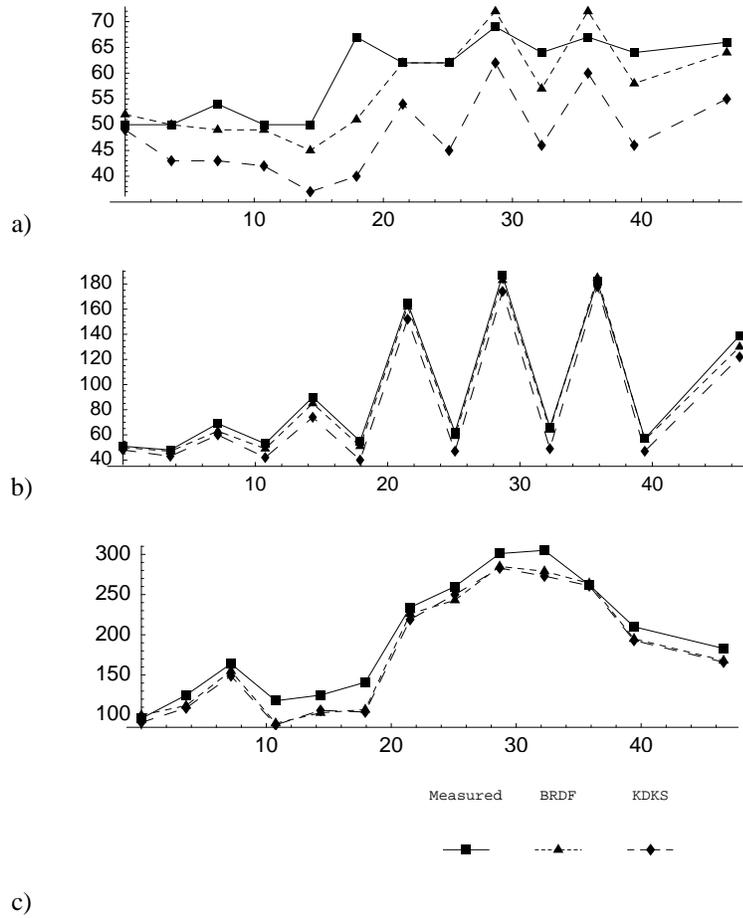


Figure 3.2: Illumination values [lx] distribution on the atrium floor along a) the left-hand side corridor (near the wall), b) the left-hand side corridor (near columns), and c) the central part of the atrium. Measured illumination is compared against the results of simulation in which the BRDF, and averaged diffuse and specular reflectance coefficients (in the graph we mark this approach by “KDKS”)

Section 3.1.1. Instead, we chose the cross-media comparison [15] in which images displayed on the CRT screen were compared directly against the real-world atrium. For the rendered images we used two methods of acquiring surface reflectance characteristics: (1) through the BRDF measurement, and (2) through the manual adjustment driven by the image appearance (as discussed in Section 3.1.1). The first method we dub the predictive rendering, and the second one the artistic rendering, since the latter approach is commonly used in the film and entertainment industry [2]. Figures 3.4a and b show example images obtained using the predictive and artistic approaches (refer to the atrium project Web pages [51] for all images used in our experiments). We conducted two sets of experiments the goal of which was to study the differences that can be perceived during a very limited and an unlimited observation time.

In the first group of experiments, the subjects were allowed to watch the atrium scene for 10 seconds, and then they were exposed randomly for the same period of time to the predictive image, the artistic image, or just the digital photograph. The subjects were asked how well the real-world atrium is reproduced by every presented image. Also, more specific questions concerning lighting reconstruction and tone reproduction on the CRT device were posed. In all cases the photographs got the highest scores with the mean ranking falling into the range 7.36–8.48 (for the scale spanning the range 0–10). The artistic and predictive approaches got lower scores falling into the ranges 6.88–7.24 and 6.56–6.96, respectively. What is remarkable is that the differences in the mean ranking of photographs and computer images were rather small, below the value of standard deviation error. The artistic approach which is much easier and cheaper (no expensive BRDF measurement is required) was slightly higher ranked than the predictive rendering, which to certain extent explains why pragmatic industry always chooses tweaking rendering parameters instead of performing full-fledged physically-based computations. The artist can compensate for simplifications in the real-world modeling as well as for the errors accumulated during input data acquisition (e.g., the measurement of BRDFs and goniometrical diagrams). However, such an artistic approach is likely to fail because of the lack of reference views for the tasks involving rendering of scenes whose appearance is unknown but must be properly predicted. This is a common problem in engineering applications such as urban planning, architecture and interior design, and product appearance design.

In the second group of experiments the observation time was not limited, and the photographs were not considered anymore. Detailed questions were posed to the subjects concerning the quality of lighting and shadow reconstruction, texture and light source rendering, appearance of highlights and reflections, and contrast reproduction. In the majority of cases the artistic approach obtained slightly higher mean ranking scores, while the fidelity of the appearance of highlights and reflections were ranked higher for the predictive approach. This suggests that by measuring the BRDF better image fidelity can be expected, especially for materials with complex reflectance characteristics. The detailed comments provided by the subjects revealed that lack of contrast in highlight and shadow areas is particularly annoying. This could be improved by more careful selection of the TMO. The architectural perfection of the model does not account for variations in construction and disparities of materials. This becomes highly visible near highlights and reflection regions, which are distorted because of inaccuracies in the positioning of tiles. Also, in the real-world scene subjectively stronger specularities can be observed for pink tiled columns and brown painted panels in the central part of the atrium. This disagreement in appearance can be explained by inaccuracies in the capturing of rapidly changing specular reflectance during the BRDF measurement. Also, the limited dynamic range of the



Figure 3.3: Settings used for experimental comparison of the atrium images displayed at the CRT device against the real-world view. The subject observed the atrium through an aperture which limited the field of view to the one similar to the displayed images.

CRT device may contribute to this disagreement. A more detailed report on the psychophysical experiments is presented on the atrium project Web pages [51].

The atrium test is a good tool to evaluate the overall rendering performance and quality. However, it is too complex to validate isolated components of the global illumination algorithms. This requires some step-by-step validation procedure. In the following section we present an attempt to establish such a procedure.

### 3.2 Multi-Stage Validation Procedure

Myszkowski and Kunii [54] proposed a systematic approach toward validation of global illumination and rendering solutions. As the result, a multi-stage validation procedure was designed, which is an attempt of testing lighting simulation and image display solutions proposed in the realistic image synthesis literature. The SHR approach (refer to [53, 54]) was chosen as a case study. The following validation stages were considered:

1. **Comparison with analytically derived data.** The advantages of such an approach are easy access to the reference data and very good accuracy of such data. For example, this makes it possible to check whether the simulation results converge to the correct solution, or some bias is introduced by an examined algorithm. However, the analytical solutions



Figure 3.4: Rendered images using the more complete atrium model a) the predictive rendering and b) the artistic rendering.

of the rendering equation exist only for extremely simple scenes.

In [54] two simple scenes were considered: an empty cube and an empty sphere tessellated by planar mirrors. For the cube test the Lambertian surface reflectance characteristics were assumed. This test was used to investigate the influence of the discretization error caused by meshing solutions of various complexity which are used in radiosity computation. In particular, it turned out that a lighting storage scheme based on a non-adaptive mesh can lead to significant discretization errors, even if the lighting simulation is of a good accuracy.

The test with an empty sphere tessellated by one or two mirror planes was designed specifically to check the modeling accuracy of specular light reflection. Since modeling of such reflections for indirect lighting is costly within the radiosity algorithm framework, the goal of this test was to investigate how clustering of secondary emitters trades the simulation accuracy for the computation efficiency. The results presented in [54] showed that clustering dramatically reduces the computation costs while it affects the solution accuracy only slightly.

- 2. Comparison with experimentally measured data.** More advanced tests involving visibility computations must be performed for the completeness of the validation procedure. The only way to obtain reliable lighting distribution data for complex environments is a direct measurement of illumination for real-world scenes.

In [54] two sets of measurement data were compared against simulation results. The data were obtained in the specifically designed measurement room owned by Toshiba Lighting Corporation. Real-world light sources (not idealized point light sources used in the analytical tests, which distribute energy uniformly in all directions) were used to illuminate the room, and their spatial energy distribution was specified by their goniometrical diagrams. This made it possible to test whether such diagrams are handled properly by the SHR software. While originally not considered in [54], the most recent version of the atrium test (refer to Section 3.1.2) with measured distribution of illumination on the floor, could also be used. However, this atrium test is especially suitable for validation of algorithms supporting complex BRDFs, which is not the case for the SHR technique.

- 3. Comparison of rendered images and photographs.** The validation procedure discussed so far was used to test all major components of the SHR algorithm in terms of the accuracy of lighting simulation. However, in many applications what really matters is the quality of rendered images as perceived by the human observer.

In [54] the atrium scene discussed in Section 3.1.1 was chosen to compare simultaneously viewed synthetic images and photographic images representing the same scene. Obviously, the latter version of the atrium test (refer to Section 3.1.2) could be considered, but at the time of preparing [54] it was not available. As discussed in Section 3.1.1, the comparison against the photographic image might involve the problem of fidelity to real-world. However, such a comparison proved to be useful in evaluating some of aspects of rendering such as the fidelity of geometry and light distribution modeling. Also, it is much easier to use photographs, especially by the third party users, who were not involved in the development of test data, and who could not see the real-world scene used for the test.

---

As concluded in [54] the SHR algorithm positively passed all stages of validation, and can be considered as an efficient and reliable tool for lighting simulation and predictive rendering of high quality images. The hybrid algorithm (refer to in Section 1.4.2) which was developed more recently [75] underwent a similar validation procedure. The obtained lighting simulation accuracy and quality of images (refer to Section 3.1.2) were also very good. The hybrid algorithm is our default choice for rendering of complex scenes, which feature surfaces with the complex BRDFs. We prefer to use the SHR algorithm for simpler scenes with mostly Lambertian surface reflectance characteristics.

---

## References

- [1] <http://www.mpi-sb.mpg.de/resources/vdp> - the Web page with documentation of the VDP validation experiments.
- [2] A.A. Apodaca. Photosurrealism. In G. Drettakis and N. Max, editors, *Rendering Techniques '98*, pages 315–322. Springer, 1998.
- [3] M.R. Bolin and G.W. Meyer. A perceptually based adaptive sampling algorithm. In *Proc. of Siggraph'98*, pages 299–310, 1998.
- [4] P.J. Burt. Fast filter transforms for image processing. *Computer Vision, Graphics and Image Processing*, 21:368–382, 1981.
- [5] M. F. Cohen, D. P. Greenberg, D. S. Immel, and P. J. Brock. An efficient radiosity approach for realistic image synthesis. *IEEE Computer Graphics and Applications*, 6(3):26–35, March 1986.
- [6] M.F. Cohen, S.E. Chen, J.R. Wallace, and D.P. Greenberg. A progressive refinement approach to fast radiosity image generation. In *Proc. of Siggraph '88*, pages 75–84, 1988.
- [7] S. Comes, O. Bruyndonckx, and B. Macq. Image quality criterion based on the cancellation of the masked noise. In *Proc. of IEEE Int'l Conference on Acoustics, Speech and Signal Processing*, pages 2635–2638, 1995.
- [8] S. Daly. The Visible Differences Predictor: An algorithm for the assessment of image fidelity. In A.B. Watson, editor, *Digital Image and Human Vision*, pages 179–206. Cambridge, MA: MIT Press, 1993.
- [9] S. Daly. Engineering observations from spatiovelocity and spatiotemporal visual models. In *Human Vision and Electronic Imaging III*, pages 180–191. SPIE Vol. 3299, 1998.
- [10] S. Daly. Important issues for automating image quality estimation. *Siggraph '2000 Course Notes: Image quality metrics*, 44, July 2000.
- [11] R.L. De Valois and DeValois K.K. *Spatial vision*. Oxford University Press, Oxford, 1990.
- [12] G. Drettakis and F.X. Sillion. Accurate visibility and meshing calculations for hierarchical radiosity. In X. Pueyo and P. Schroder, editors, *Rendering Techniques '96*, pages 269–278. Springer, June 1996.
- [13] M.P. Eckert and G. Buchsbaum. The significance of eye movements and image acceleration for coding television image sequences. In A.B. Watson, editor, *Digital Image and Human Vision*, pages 89–98. Cambridge, MA: MIT Press, 1993.
- [14] R. Eriksson, B. Andren, and K. Brunnstrom. Modelling of perception of digital images: a performance study. In *Human Vision and Electronic Imaging III*, pages 88–97. Proc. of SPIE Vol. 3299, 1998.
- [15] M. D. Fairchild. *Color appearance models*. Addison-Wesley, Reading, Massachusetts, 1998.

- 
- [16] J.A. Ferwerda, S. Pattanaik, P. Shirley, and D.P. Greenberg. A model of visual masking for computer graphics. In *Proc. of Siggraph'97*, pages 143–152, 1997.
- [17] A. Fisher. Tolerances in lighting design. In CIE Central Bureau, editor, *Proc. of the CIE Seminar on Computer Programs for light and lighting*, 1992.
- [18] A. Gaddipatti, R. Machiraju, and R. Yagel. Steering image generation with wavelet based perceptual metric. *Computer Graphics Forum (Eurographics '97)*, 16(3):241–251, September 1997.
- [19] S. Gibson and R. J. Hubbard. Perceptually-driven radiosity. *Computer Graphics Forum*, 16(2):129–141, 1997.
- [20] S.J. Gortler, R. Grzeszczuk, R. Szeliski, and M.F. Cohen. The lumigraph. In *Proc. of Siggraph'96*, pages 43–54, 1996.
- [21] D.P. Greenberg. A framework for realistic image synthesis. *Communications of the ACM*, 42(8):43–53, August 1999.
- [22] D.P. Greenberg, K.E. Torrance, P. Shirley, J. Arvo, J.A. Ferwerda, S.N. Pattanaik, E.P.F. Lafortune, B. Walter, S.C. Foo, and B. Trumbore. A framework for realistic image synthesis. In *Proc. of Siggraph'97*, pages 477–494, 1997.
- [23] P. Heckbert. Adaptive radiosity textures for bidirectional ray tracing. In *Proc. of Siggraph'90*, pages 145–154, August 1990.
- [24] D. Hedley, A. Worrall, and D. Paddon. Selective culling of discontinuity lines. In J. Dorsey and P. Slusallek, editors, *Rendering Techniques '97*, pages 69–80. Springer, 1997.
- [25] R.F. Hess and R.J. Snowden. Temporal properties of human visual filters: number shapes and spatial covariation. *Vision Research*, 32:47–59, 1992.
- [26] W.G. Hunt. *The reproduction of colour*. Fountain Press, England, 1995.
- [27] H. W. Jensen. Global Illumination Using Photon Maps. In X. Pueyo and P. Schroder, editors, *Rendering Techniques '96*, pages 21–30. Springer, 1996.
- [28] J.T. Kajiya. The rendering equation. In *Proc. of Siggraph'86*, pages 143–150, 1986.
- [29] D.H. Kelly. Motion and Vision 2. Stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America*, 69(10):1340–1349, 1979.
- [30] R. Klette, K. Schluns, and A. Koschan. *Computer Vision: Three-Dimensional Data from Images*. Springer-Verlag, 1998.
- [31] M. Levoy and P. Hanrahan. Light field rendering. In *Proc. of Siggraph'96*, pages 31–42, 1996.
- [32] B. Li, G.W. Meyer, and R.V. Klassen. A comparison of two image quality models. In *Human Vision and Electronic Imaging III*, pages 98–109. SPIE Vol. 3299, 1998.

- 
- [33] D. Lischinski and A. Rappoport. Image-based rendering for non-diffuse synthetic scenes. In G. Drettakis and N. Max, editors, *Rendering Techniques '98*, pages 301–314. Springer, 1998.
- [34] D. Lischinski, B. Smits, and D.P. Greenberg. Bounds and error estimates for radiosity. In *Proc. of Siggraph'94*, pages 67–74, 1994.
- [35] D. Lischinski, F. Tampieri, and D. P. Greenberg. Discontinuity meshing for accurate radiosity. *IEEE Computer Graphics and Applications*, 12(6):25–39, November 1992.
- [36] D. Lischinski, F. Tampieri, and D. P. Greenberg. Combining hierarchical radiosity and discontinuity meshing. In *Proc. of Siggraph'93*, pages 199–208, 1993.
- [37] C. Lloyd and R.J. Beaton. Design of spatial-chromatic human vision model for evaluating full-color display systems. In *Human Vision and Electronic Imaging: Models, Methods, and Appl.*, pages 23–37. SPIE Vol. 1249, 1990.
- [38] J. Lubin. A visual discrimination model for imaging system design and development. In Peli E., editor, *Vision models for target detection and recognition*, pages 245–283. World Scientific, 1995.
- [39] J. Lubin. A human vision model for objective picture quality measurements. In *Conference Publication No. 447*, pages 498–503. IEE International Broadcasting Convention, 1997.
- [40] S. Marcelja. Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America*, 70:1297–1300, 1980.
- [41] W.R. Mark, L. McMillan, and G. Bishop. Post-rendering 3D warping. In *1997 Symposium on Interactive 3D Graphics*, pages 7–16. ACM Siggraph, 1997.
- [42] W.L. Martens and K. Myszkowski. Psychophysical validation of the Visible Differences Predictor for global illumination applications. In *IEEE Visualization '98 (Late Breaking Hot Topics)*, pages 49–52, 1998.
- [43] I. Martin, X. Pueyo, and D. Tost. An image-space refinement criterion for linear hierarchical radiosity. In *Graphics Interface '97*, pages 26–36, 1997.
- [44] R.A. Martin, A.J. Ahumada, and J.O. Larimer. Color matrix display simulation based upon luminance and chrominance contrast sensitivity of early vision. In *Human Vision, Visual Processing, and Digital Display III*, pages 336–342. SPIE Vol. 1666, 1992.
- [45] L. McMillan. *An Image-Based Approach to 3D Computer Graphics*. Ph.D. thesis, North Carolina University, Chapel Hill, 1997.
- [46] A. McNamara, A. Chalmers, T. Troscianko, and I. Gilchrist. Comparing real and synthetic scenes using human judgement of lightness. In B. Peroche and H. Rushmeier, editors, *Rendering Techniques '00*, pages 207–218. Springer, 2000.
- [47] G.W. Meyer, H.E. Rushmeier, M.F. Cohen, D.P. Greenberg, and K.E. Torrance. An experimental evaluation of computer graphics imagery. *ACM Transactions on Graphics*, 5(1):30–50, January 1986.

- 
- [48] G. Miller, S. Rubin, and D. Poncelen. Lazy decomposition of surface light fields for pre-computed global illumination. In G. Drettakis and N. Max, editors, *Rendering Techniques '98*, pages 281–292. Springer, 1998.
- [49] J.A. Movshon, E.H. Adelson, M.S. Gizzi, and W.T. Newsome. The analysis of moving visual patterns. In C. Chagas, R. Gattas, and C. Gross, editors, *Pattern recognition mechanisms*, pages 117–151. Rome: Vatican Press, 1985.
- [50] K. Myszkowski. The Visible Differences Predictor: Applications to global illumination problems. In G. Drettakis and N. Max, editors, *Rendering Techniques '98*, pages 223–236. Springer, 1998.
- [51] K. Myszkowski and F. Drago. Validation proposal for global illumination and rendering techniques. In *the atrium project Web pages available under the URL <http://www.mpi-sb.mpg.de/resources/atrium>*, 2000.
- [52] K. Myszkowski, A. B. Khodulev, and E. A. Kopylov. Validating global illumination algorithms and software. In *Visual Proceedings, Technical Sketch at ACM Siggraph'97*, page 156, 1997.
- [53] K. Myszkowski and T.L. Kunii. An efficient cluster-based hierarchical progressive radiosity algorithm. In *ICSC '95*, volume 1024 of *Lecture Notes in Computer Science*, pages 292–303. Springer-Verlag, 1995.
- [54] K. Myszkowski and T.L. Kunii. A case study towards validation of global illumination algorithms: progressive hierarchical radiosity with clustering. *The Visual Computer*, 16(5):271–288, 2000.
- [55] K. Myszkowski, P. Rokita, and T. Tawara. Perceptually-informed accelerated rendering of high quality walkthrough sequences. In D. Lischinski and G.W. Larson, editors, *Rendering Techniques '99*, pages 5–18. Springer, 1999.
- [56] K. Myszkowski, P. Rokita, and T. Tawara. Perception-based fast rendering and antialiasing of walkthrough sequences. *IEEE Transactions on Visualization and Computer Graphics*, 6(4):360–379, 2000.
- [57] K. Myszkowski, A. Wojdala, and K. Wicynski. Non-uniform adaptive meshing for global illumination. *Machine Graphics and Vision*, 3(4):601–610, 1994.
- [58] W. Osberger, A.J. Maeder, and N. Bergmann. A perceptually based quantization technique for MPEG encoding. In *Human Vision and Electronic Imaging III*, pages 148–159. Proc. of SPIE Vol. 3299, 1998.
- [59] E.D. Palik. *Handbook of Optical Constants of Solids*. Academic Press, New York, 1985.
- [60] J. Prikryl and W. Purgathofer. State of the art in perceptually driven radiosity. In *State of the Art Reports*. Eurographics, 1998.
- [61] M. Ramasubramanian, S.N. Pattanaik, and D.P. Greenberg. A perceptually based physical error metric for realistic image synthesis. In *Proc. of Siggraph'99*, pages 73–82, 1999.

- 
- [62] H. Rushmeier, G. Ward, C. Piatko, P. Sanders, and B. Rust. Comparing real and synthetic images: some ideas about metrics. In P. Hanrahan and W. Purgathofer, editors, *Rendering Techniques '95*, pages 82–91. Springer, 1995.
- [63] J.W. Shade, S.J. Gortler, L. He, and R Szeliski. Layered depth images. In *Proc. of Siggraph'98*, pages 231–242, 1998.
- [64] M. Shinya. Spatial anti-aliasing for animation sequences with spatio-temporal filtering. In *Proc. of Siggraph'93*, pages 289–296, 1993.
- [65] F.X. Sillion and C. Puech. *Radiosity and Global Illumination*. Morgan Kaufmann, San Francisco, 1994.
- [66] C.C. Taylor, Z. Pizlo, J. P. Allebach, and C.A. Bouman. Image quality assessment with a Gabor pyramid model of the Human Visual System. In *Human Vision and Electronic Imaging*, pages 58–69. SPIE Vol. 3016, 1997.
- [67] A.M. Tekalp. *Digital video Processing*. Prentice Hall, 1995.
- [68] P.C. Teo and D.J. Heeger. Perceptual image distortion. pages 127–141. SPIE Vol. 2179, 1994.
- [69] X. Tong, D. Heeger, and C. van den Branden Lambrecht. Video quality evaluation using ST-CIELAB. In *Human Vision and Electronic Imaging IV*, pages 185–196. Proc. of SPIE Vol. 3644, 1999.
- [70] J. Tumblin, J. Hodgins, and B. Guenter. Two methods for display of high contrast images. *ACM Transactions On Graphics*, 18(1):56–94, January 1999.
- [71] J. Tumblin and H.E. Rushmeier. Tone reproduction for realistic images. *IEEE Computer Graphics and Applications*, 13(6):42–48, November 1993.
- [72] C.J. van den Branden Lambrecht. *Perceptual models and architectures for video coding applications*. Ph.D. thesis, Ecole Polytechnique Federal de Lausanne, 1996.
- [73] E. Veach. *Robust Monte Carlo methods for lighting simulation*. Ph.D. thesis, Stanford University, 1997.
- [74] C. Vedel and C. Puech. A testbed for adaptive subdivision in progressive radiosity. In P. Brunet and F.W. Jansen, editors, *Photorealistic Rendering in Computer Graphics*, 1991.
- [75] V. Volevich, K. Myszkowski, A. Khodulev, and Kopylov E.A. Using the Visible Differences Predictor to improve performance of progressive global illumination computations. *ACM Transactions on Graphics*, 19(2):122–161, 2000.
- [76] V. Volevich, K. Myszkowski, A.B. Khodulev, and E.A. Kopylov. Perceptually-informed progressive global illumination solution. Technical Report TR-99-1-002, Department of Computer Science, Aizu University, February 1999.
- [77] B.J. Walter. *Density estimation techniques for global illumination*. Ph.D. thesis, Cornell University, 1998.

- 
- [78] B.J. Walter, P.M. Hubbard, P. Shirley, and D.P. Greenberg. Global illumination using local linear density estimation. *ACM Transactions on Graphics*, 16(3):217–259, 1997.
- [79] B.A. Wandell. *Foundations of vision*. Sinauer Associates, Inc., Sunderland, Massachusetts, 1995.
- [80] G.J. Ward. The RADIANCE lighting simulation and rendering system. In *Proc. of Siggraph'94*, pages 459–472, 1994.
- [81] A.B. Watson. Temporal sensitivity. In *Handbook of Perception and Human Performance, Chapter 6*. John Wiley, New York, 1986.
- [82] A.B. Watson. The Cortex transform: rapid computation of simulated neural images. *Comp. Vision Graphics and Image Processing*, 39:311–327, 1987.
- [83] A.B. Watson. Toward a perceptual video quality metric. In *Human Vision and Electronic Imaging III*, pages 139–147. Proc. of SPIE Vol. 3299, 1998.
- [84] A.B. Watson, J. Hu, J.F. McGowan III, and J.B. Mulligan. Design and performance of a digital video quality metric. In *Human Vision and Electronic Imaging IV*, pages 168–174. Proc. of SPIE Vol. 3644, 1999.
- [85] H. Weghorst, G. Hooper, and D.P. Greenberg. Improved computational methods for ray tracing. *ACM Transactions on Graphics*, 3(1):52–69, January 1984.
- [86] S.J.P. Westen, R.L. Lagendijk, and J. Biemond. Perceptual image quality based on a multiple channel HVS model. In *Proc. of IEEE Int'l Conference on Acoustics, Speech and Signal Processing*, pages 2351–2354, 1995.
- [87] J.H.D.M. Westerink and C. Teunissen. Perceived sharpness in moving images. pages 78–87. Proc. of SPIE Vol. 1249, 1990.
- [88] H.R. Wilson. Psychophysical models of spatial vision and hyperacuity. In D. Regan, editor, *Spatial vision, Vol. 10, Vision and Visual Dysfunction*, pages 179–206. Cambridge, MA: MIT Press, 1991.
- [89] H.R. Wilson and D.J. Gelb. Modified line-element theory for spatial-frequency and width discrimination. *Journal of the Optical Society of America*, 1(1):124–131, 1984.
- [90] S. Winkler. A perceptual distortion metric for digital color video. In *Human Vision and Electronic Imaging IV*, pages 175–184. Proc. of SPIE Vol. 3644, 1999.
- [91] E.M. Yeh, A.C. Kokaram, and N.G. Kingsbury. A perceptual distortion measure for edge-like artifacts in image sequences. In *Human Vision and Electronic Imaging III*, pages 160–172. Proc. of SPIE Vol. 3299, 1998.
- [92] C. Zetsche and Hauske G. Multiple channel model for the prediction of subjective image quality. In *Human Vision, Visual Processing, and Digital Display*, pages 209–216. SPIE Vol. 1077, 1989.