# Image Quality Metrics

Alan Chalmers    Scott Daly    Ann McNamara
Karol Myszkowski    Tom Troscianko

Course #44      SIGGRAPH 2000

New Orleans, USA
23-28 July 2000

# About The Authors

Dr 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 photo-realistic graphics. He is currently vice chair of ACM SIGGRAPH. His research is investigating the use of photo-realistic graphics in the 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 research member of the Vision and Graphics Laboratory at the University of Bristol, where she works on image evaluation and comparison using data collected from psychophysical experiments. Her area of interest mainly focuses on evaluating the fidelity of computer imagery to real scenes, with an emphasis on human observer evaluation. She has published a number of papers outlining these results

Dr Karol Myszkowski is currently a visiting senior researcher at the Max-Planck-Institute for Computer Science, Germany. From 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.

Dr Tom Troscianko heads the Perceptual Systems Research Centre in the University of Bristol. He has a background in vision science, and his current research is investigating the way in which human vision samples the natural visual environment, especially on the role of colour information in such environments.

# Contents

# Chapter 1

# Introduction

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 includ-

ing Visual Difference Predictors, the Sarnoff model and Animation Quality Metrics; and psychophysics.

## 1.1 Course Syllabus

Introduction to Image Quality (Chalmers 15 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

Subjective Image Quality Metrics (McNamara & Troscianko 45 mins)

- 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 45 mins)

- visual perception
- computer models of visual system

Objective Image Quality Metrics (Daly 45 mins)

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

Applications in Rendering and Animation (Chalmers & Myszkowski 40 mins)

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

Summary, discussion and questions (All 20 mins)

## 1.2   Structure of Notes

The notes contain important background information as well as detailed descriptions of the Image Quality Metrics described. The notes are arranged as follows:

Chapter 2 introduces the nature and behaviour of light, along with an outline of the radiometric concepts, photometric concepts and terms relevant to illumination engineering. This is followed by a brief outline of some image synthesis techniques (illumination models). Aspects of the HVS are outlined detailing the physiological characteristics that a global illumination framework must observe to effectively reproduce an authentic simulated response to a (real or virtual) scene.

Chapter 3 describes how knowledge about human visual perception can be employed to the advantage of realistic image synthesis. In particular, we focus on perception driven techniques, perception based metrics, and effective display methods.

Chapter 4 concentrates on perception-driven global illumination and rendering techniques. Some of the methods described in chapter 3 are briefly reviewed before the Visual Differences Predictor is described in detail.

Chapter 5 introduces a new methodology to enable the comparison of synthetic imagery to the **real** environment, using human observers. We begin by presenting the conception of a small pilot study, demonstrating the feasibility of the approach. Building on this study extends image comparison to real world using a more complex test scene containing three dimensional objects in varying shades of grey, enabling examination of effects such as partial occlusion and shadowing.

Chapter 6 describes some issues relating to preception rendering of walkthrough animations.

Finally, some detailed slides are presented by Scott Daly on important issues for Image Quality Metrics.

# Chapter 2

# Illumination: Simulation & Perception

Since its conception, the pursuit of realistic computer graphics has been the creation of representative, high quality imagery [45, 44, 42, 103]. The production (rendering) of realistic images in particular requires a precise treatment of lighting effects. This can be achieved by simulating the underlying physical phenomena of light emission, propagation, and reflection. The environment under consideration is first modelled as a collection of virtual light, objects and a camera (or eye) point. A *rendering algorithm* then takes this model as input and generates the image by simulating the light and its interaction with the environment [38, 29, 117]. *Physically-based* rendering algorithms [40, 5, 56, 95] focus on producing realistic images by simulating the light energy, or *radiance*, that is visible at every pixel of the image. Finally the radiance values computed in the rendering stage must be mapped to values suitable for display on some display device [104]. This so called rendering pipeline [44] is illustrated in figure 2.1.
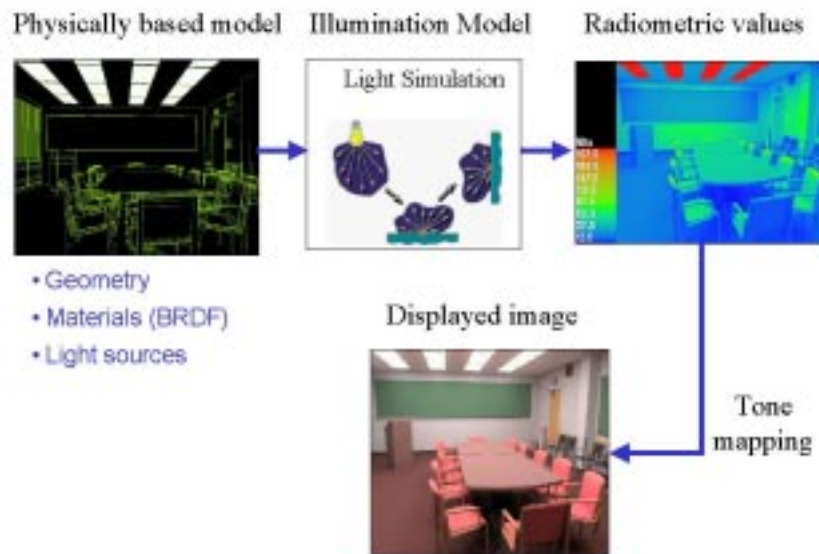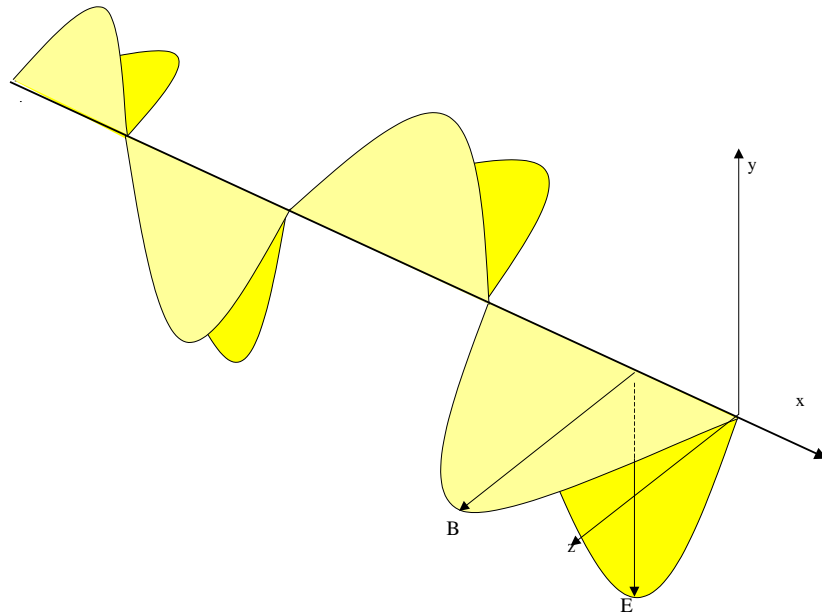


Figure 2.1: The rendering pipeline

Figure 2.2: Mutually orthogonal E and B fields of an electromagnetic wave propagating in the x axis

## 2.1 Light and Materials

Understanding the natural illumination process, and how to quantify of illumination provides the foundations for designing and controlling physically based image synthesis algorithms. A precise terminology exists to quantify illumination [13], from this terminology the underlying equations used to build the mathematical models for illumination simulation algorithms are derived. Also, certain aspects of the human visual system 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.

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 [52, 6]

- **Radiometry** is the science of measuring radiometry 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. The terms and units have been standardised in the ANSI publication, [49]

- **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 [125]. Photometry deals with perceptual issues; if a surface radiates a given amount of energy, then how bright does that surface *appear* to an average viewer? By standardising the *luminous efficiency* of the human visual system, the subjective nature of photometric measurement may be eliminated. The was done in 1924 by the *Commission Internationale d'Eclairages* or CIE, by performing empirical tests with over one hundred observers.[13]

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
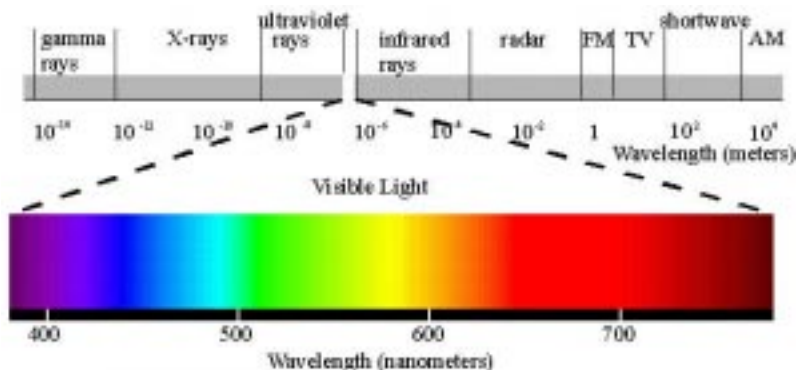
Figure 2.3: The visible portion of the electromagnetic spectrum

wave,(figure 2.2). The portion of light 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-rays, gamma rays. The range of visible light, which lies approximately between 380nm and 780nm, is placed in the context of the whole electromagnetic spectrum is depicted in figure 2.3. The scenes human perceive are based on an integration over the visible spectrum of incoming radiation. Most of the following definitions are taken from the Illumination Engineering Society Lighting Handbook, given by the IES [54].

**Illuminating hemisphere**($\Omega$): The illuminating hemisphere is a convenient notation to
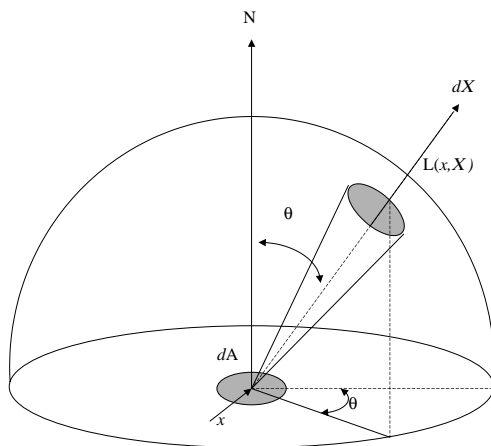


Figure 2.4: The Illumination Hemisphere

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. The illuminating hemisphere is depicted in Figure 2.4. Using this form, the illumination at a given point can be computed by considering **all** illumination events captured on the illumination hemisphere.
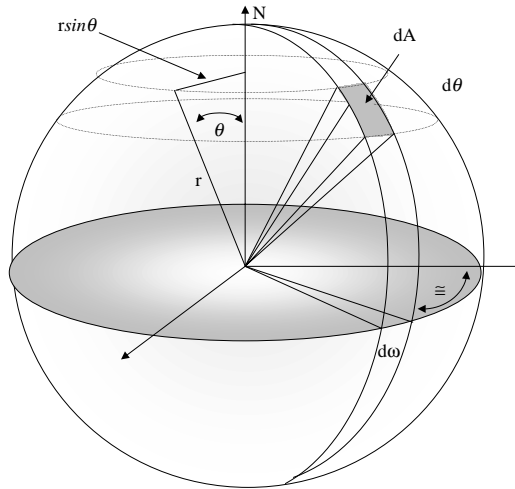
Figure 2.5: Calculating the Solid Angle.

**Solid Angle**: Solid angles are the solid geometry equivalent of angles in plane geometry, figure 2.5. Solid angles are measured in *steradians* or *sr*. The solid angle is defined as the area of the cone, with apex at the centre of the sphere, cut out from a sphere of radius 1. The solid angle of the entire sphere is *4πsr*, thus that of a hemisphere is *2πr*. A small circular area on the surface of a sphere may be approximated by a flat section. Thus the solid angle subtended at the centre of the sphere by this small area may be expressed as:

$$d\omega = \frac{\pi (R \sin \alpha)^2}{R^2} = \pi \sin^2 \alpha$$

**Projected Area** $dA_i$: 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 2.6.

$$dA_i = dA \cos \theta$$

Clearly projected area varies according to viewing direction.

### 2.1.1 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)*, photons of a certain frequency have a specific quantum of energy, defined by $E = hf$, where $h$ is Plank's Constant [1] and $f$ is the

---

[1]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 $6.626x10^{-20} J - s$
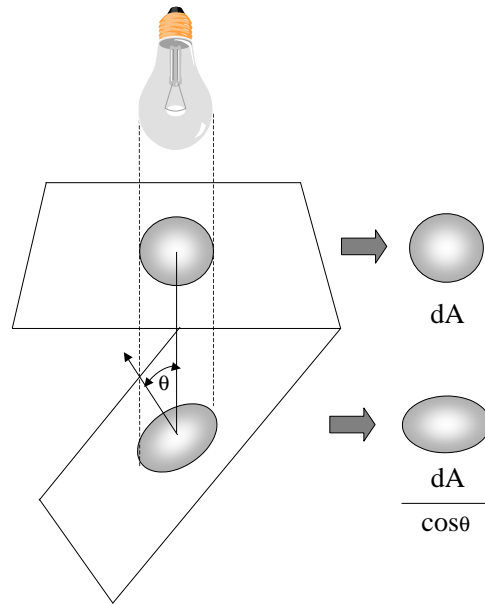
Figure 2.6: The greater the angle the greater the area over which light is distributed, so energy at a given point will be proportionally less, after Collins

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 meter ($Wm^2$).* The quotient of the radiant flux incident on or emitted by 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 leaving a surface.

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

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

**Radiant Intensity(I)**:*Measured in watts per steradian ($Wsr^1$).* Radiant Intensity represents the radiant flow 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 meter squared ($Wsr^1m^2$).* 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.

### 2.1.2 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 equivalent to radiometric terms weighted by $V(\lambda)$, the *photopic*
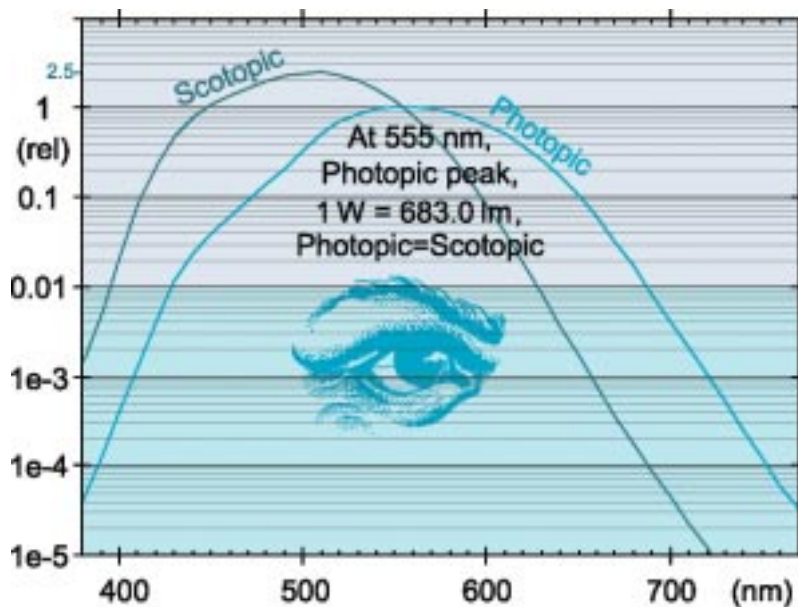


Fig. 2.3 CIE Photopic and Scotopic sensitivity curves.

Figure 2.7: Luminous Efficiency Curve, after Ryer

*spectral luminous efficiency* curve, figure 2.7. Radiations outside the visible spectrum do not play a role in photometry. The photopic metrics 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 *555nm* whose radiant flux is (1/663)*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 (1/663)W where $f$ is

the reciprocal of the sensitivity of the corresponding wavelength relative 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)**: 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 Candela per square meter*. 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 2.1: Radiometric and Photometric Quantities

Each of the radiometric quantities listed earlier has their photometric counterpart. Both radiometric and photometric quantities are shown in Table 2.1, along with their units.

### 2.1.3  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 very different ways, for example a matt house paint reflects light very differently than the often highly specular paint used on a sports car. *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"*, [82].

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 [17, 38]. Hence, the BRDF is the ratio of outgoing intensity to incoming energy, figure 2.8. Generally we define BRDF as:

$$R_{bd}(\lambda_i, \phi_i, \theta_i, \phi_\nu, \theta_\nu)$$

This relates the light in the direction $(\lambda_i, \theta_i)$ to outgoing light in the direction $(\phi_\nu, \theta_\nu)$. BRDF is a function of wavelength.

$$R_{bd}(\lambda_i, \phi_i, \theta_i, \phi_\nu, \theta_\nu) = \frac{I_\nu(\phi_i, \theta_i, \phi_\nu, \theta_\nu)}{E_i(\phi_i, \theta_i)}$$

Incoming energy is related to outgoing intensity by

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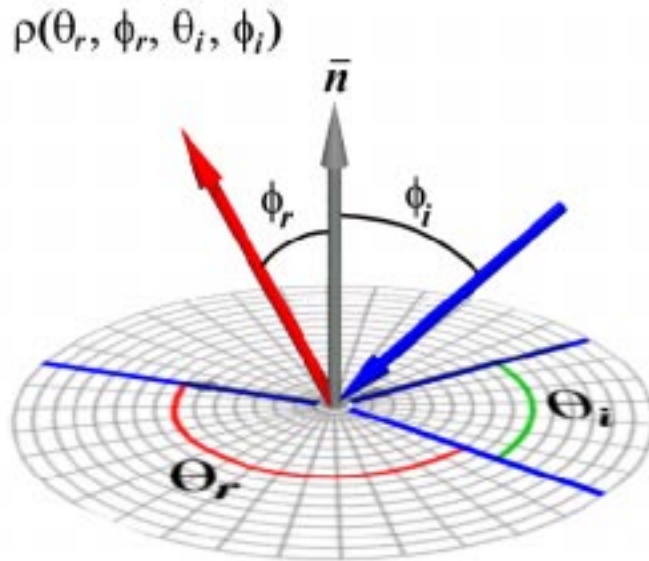


Figure 2.8: Geometry of the BRDF

Figure 2.9 shows different types of material behaviour, which are defined as follows, [38]:

**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 night time 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 don't fall exactly into one 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.
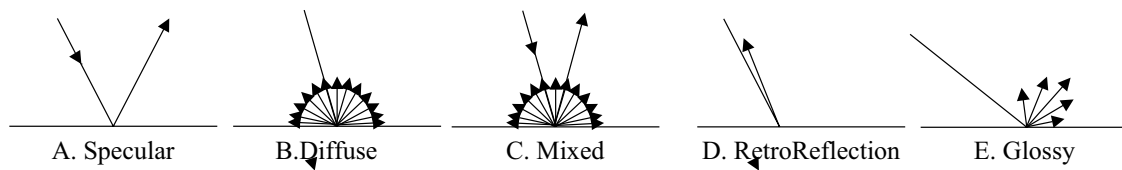


Figure 2.9: Types of Reflection. a. Specular, b. Diffuse, c. Mixed, d. Retro-Reflection, e. Gloss

## 2.2 Illumination Models
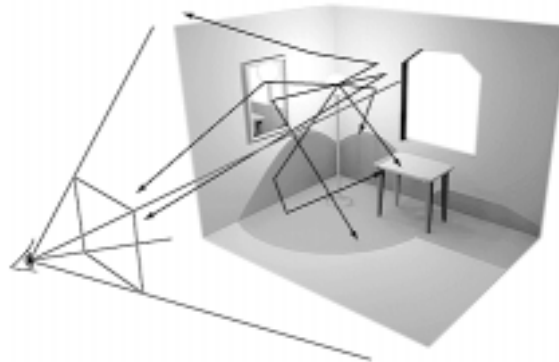


Figure 2.10: Light behaviour in an environment, after [7]

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 by, and only by, the characteristics of the surface itself and those of the light source. The Phong illumination model [85] is 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 surfaces 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*, $I_a k_a$.

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 2.10. 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. This transfer of radiant energy is governed by the laws of physics. The complexities of the interaction of light and surfaces in an environment can be neatly described in a compact form by *rendering equation* [51]:

$$I(x, x') = g(x, x')[\epsilon(x, x') + \int_{x' \epsilon S} \rho(x, x', x'')I(x', x'')dx'']$$

where

$f(x, x')$ relates to the intensity of light passing from point $x$ to point $x'$
$g(x, x')$ is a "geometry" term
$\epsilon(x, x')$ is related to the intensity of emitted light from $x'$ to $x$
$\rho(x, x', x'')$ is related to the intensity
of light scattered from $x''$ to $x$ by a surface element a $x'$
$S$ is the union of all the surfaces in the environment.

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

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 [2]. 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.

---

[2]The rendering equation is a linear inhomogeneous Fredholm integral equation of the second kind, which exhibits a recursive nature making it difficult to evaluate.
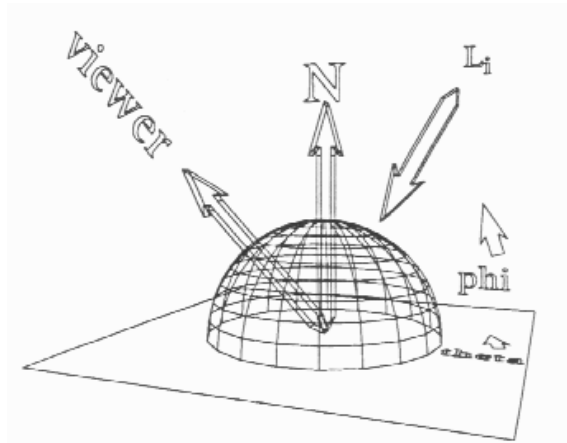
Figure 2.11: Illustration of the Rendering Equation which determines radiance by summing self emitting radiance and reflected radiance

## 2.2.1 Raytracing
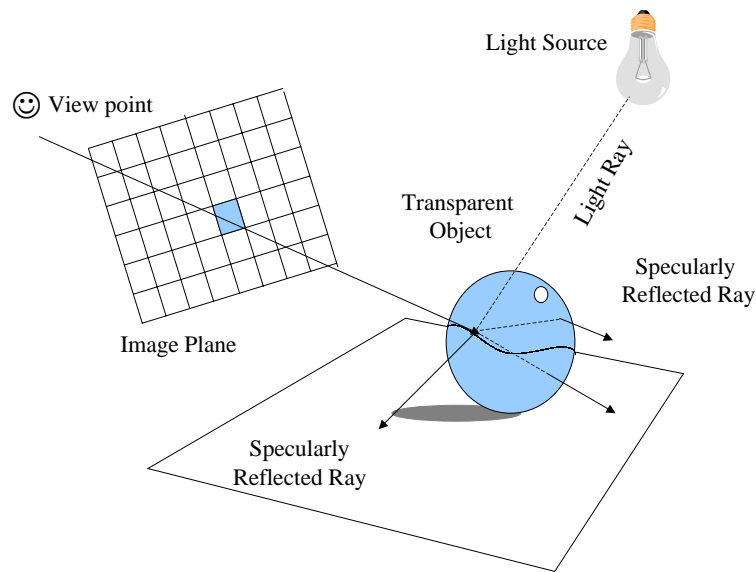


Figure 2.12: Raytracing : Rays are traced from the eye into the scene in an attempt to capture specular reflection, transparency effects and shadowing

Raytracing is a versatile technique for computing images by tracing individual paths of light through an scene. Raytracing algorithms attempt to capture *view-dependent* specular effects as well as reflections and transmissions [4, 121]. 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 photon striking the eye are needed to construct 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.
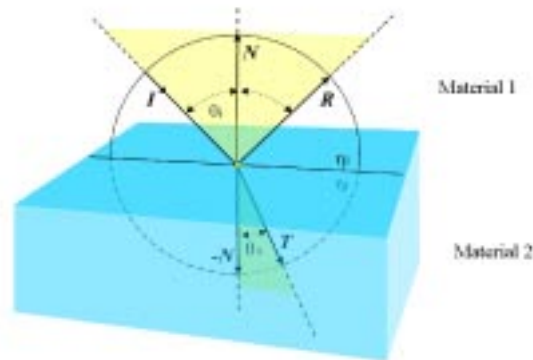


Figure 2.13: Behaviour of light ray incident on a surface

In the case of reflective or transparent surfaces, the direction in which light arrives by reflection or transmission in also needed. Reflected rays are easily computed since the angle of reflection is equal to the angle of incidence, figure 2.13. 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 product of the refractive index is the sine angle of incidence of a ray in one medium is equal to the product of the refractive index and the sine angle of refraction in a successive medium, figure 2.13.
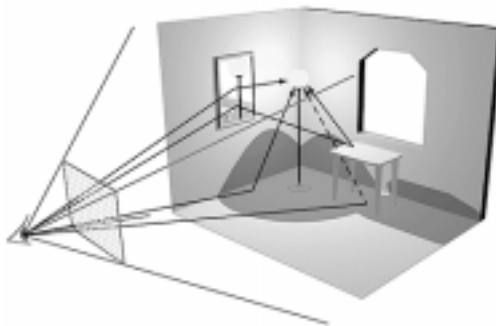


Figure 2.14: Raytracing, after [7]

19

A *recursive* evaluation must be employed, at each surface, figure 2.14. 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, traditionally this indirect illumination is referred to as *ambient light* and is accounted for by a constant *ambient term*, which is usually assigned an arbitrary value. The following pseudo code illustrates the recursive raytracing procedure:

```
For each pixel, p, in an image
Set I = ray starting at eye though pixel p
rad = Trace(I);
DrawPixel(p, I);
Radiance Trace(Ray I){
Radiance 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); } }
```

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* [30, 37], 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-regions (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 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 3DDDA, speed can be further augmented by implementing this in hardware.
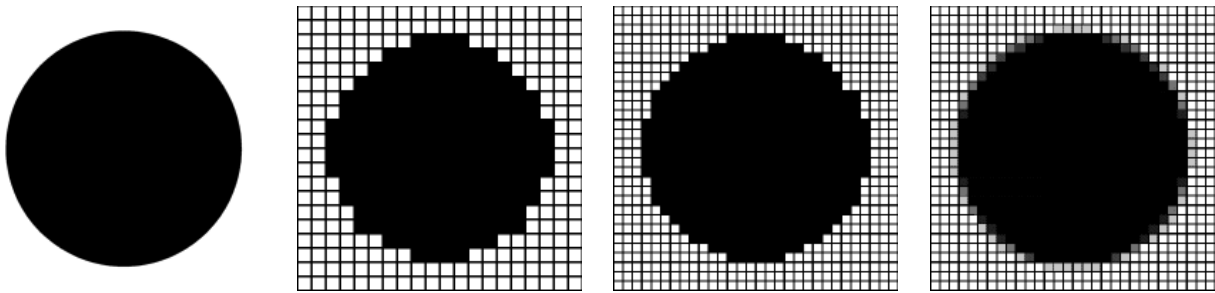


Figure 2.15: Antialiasing: a) A circle b) Strongly aliased circle c) Aliased Circle at Higher Resolution d) Antialiased Circle

*Aliasing* effects, (figure 2.15), 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 2.15 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 as shown in figure. *Distribution Raytracing* [19, 20] 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 was achieved by distributing rays over several domains (pixel positions, lens position, area sampling position etc). In distribution 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 raytracing may be found in, for example [39]

### 2.2.2 Radiosity

Radiosity methods [40, 83, 15] attempt to capture *view-independent* diffuse interreflections in a scene, figure 2.16. Techniques originally developed to compute the radiant interchange between surfaces, were first applied to the global illumination problem in the mid 1980s. *Radiosity*[3] 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

---

[3]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.
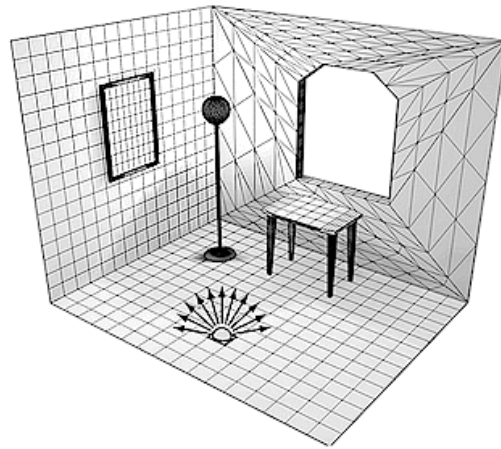
Figure 2.16: Radiosity, after [7]

effects of multiple reflections. To accomplish this the surfaces of a scene are first divided into a *mesh* of *patches*, and the radiance of these patches is computed by solving a system of equations, figure 2.17. 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.



Figure 2.17: Radiosity: An image on the left, meshed representation the right, after [2]

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

$$B_i = E_i + \rho_i \sum 1 - njF_{ij}B_j$$

$B_i$ is the radiosity of patch $i$. (energy per unit area per unit time)
$E_i$ 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 reflectivity of patch $i$.
$n$ is the number of patches in the environment.

The *form-factor*, figure 2.18, $F_{ij}$ is the fraction of energy transferred from patch $i$ to patch $j$. The *reciprocity relationship* [94] states:

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

Figure 2.18: The relationship between two patches

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

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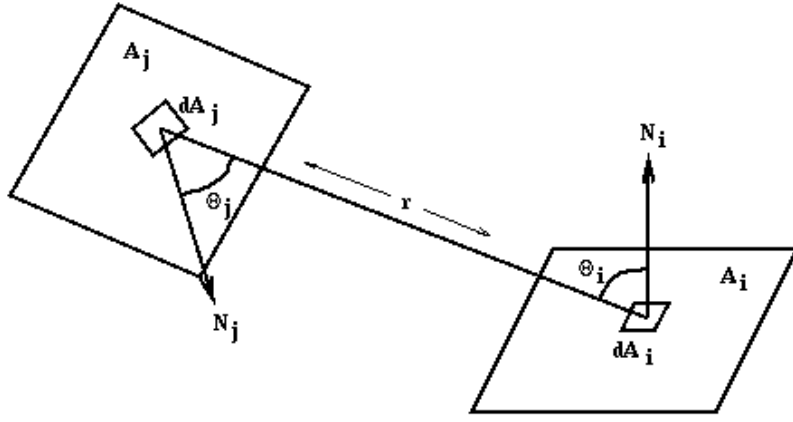. \\
E_n
\end{pmatrix}
$$

A patch can contribute to its own reflected energy (in the case of convex 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. 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.

*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 2.18, 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, $H_{ij}$ take the value of 1 or 0 depending on whether or not $dA_i$ is visible from $dA_j$. To calculate the form factor, $F_{i-j}$. from differential area $dA_i$ to finite area $A_j$ integrate over the area of patch $j$:

$$
F_{di-j} = \int \frac{\cos \theta_i \cos \theta j}{\pi r^2} H_{ij} dA_j dA_i
$$

23

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

$$F_{i-j} = \frac{1}{A_i} \int_{Ai} \int_{Aj} \frac{\cos\theta_i \cos\theta j}{\pi r^2} H_{ij} dA_j dA_i$$

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

An equivalent to computing form-factors, Nusselt 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 2.19. 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 $\pi$.
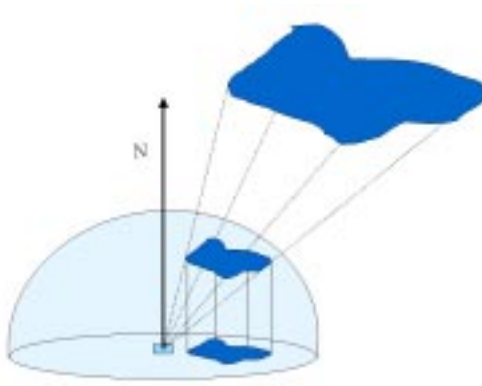


Figure 2.19: The Nusselt Analog

An alternative algorithm, proposed by Cohen and Greenberg, projects onto the upper half of a cube, *hemicube*, centred about $dA_i$, with the cube's top parallel to the surface, figure 2.20. The hemicube 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 hemicube.

Each cell, $p$, of the hemicube has a precomputed *delta form factor* associated with it:

$$\triangle F_p = \frac{\cos\theta_i \cos\theta j}{\pi r^2} \triangle A$$

$\theta_p$ 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, figure. Assigning the hemicube a $(x, y, z)$coordinate system, with the origin at the centre of the bottom face, then for the top face:

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

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

24

Figure 2.20: The Hemicube

$x_p$ and $y_p$ are the coordinates of the hemicube cell.

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

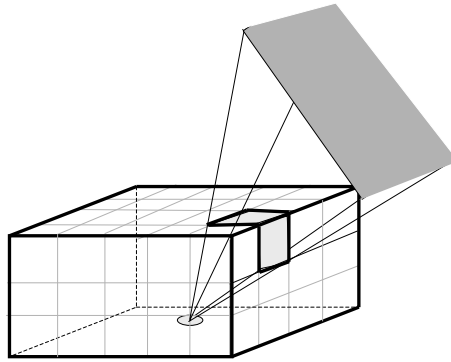The full matrix algorithm solves each $B_i$ value one at a time by "gathering" light contributions from all other patches in the scene. One of the disadvantages of this method is only *after* all radiosities have been computed can the resultant image be displayed. For complex environments the time taken 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 proposed the *progressive refinement radiosity* which uses the notion of *adaptive refinement* of images, to provide the user as soon as possible 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 meet to be stored *progressive refinement* radiosity allows partial solutions to be displayed. The progressive refinement approach simultaneously solves all patch radiosities by repeatedly choosing a patch to "shoot" and distributing that patches energy to all other patches. This is attractive as it provides a very good approximation to the final solution after only a few iterations.

More details of the radiosity method may be found in, for example [6, 17]

*Hierarchical radiosity* attempts to minimise the number of form-factor computations by approximating blocks of the form-factor matrix with a single value. *Form Factor Computation*

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

## 2.3 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.
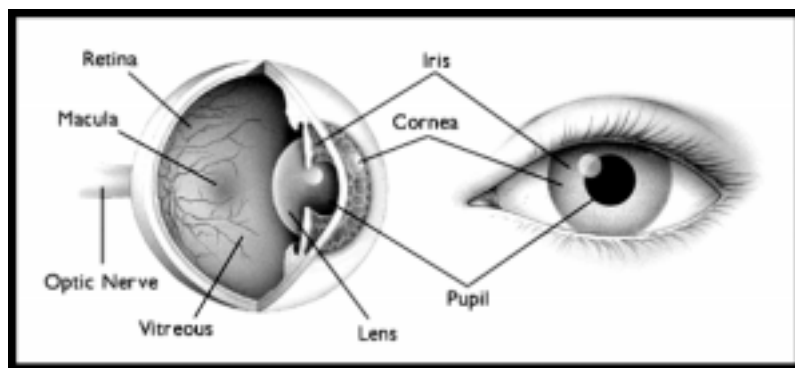
### 2.3.1 The Human Visual System



Figure 2.21: Cross Section of the human Eye

Vision is a complicated 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, 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 2.22. The main structures are the iris, lens, pupil, cornea, retina, vitreous humor, optic disk and optic nerve.

The path of light through the visual system begins at the *pupil*, is focused by the *lens*, then passes onto the *retina*, figure 2.24, 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 2.21 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 schlera 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 though the lens and proceeds through the *vitreous humor*, a transparent substance, to the rear wall of the eye, the *retina*. The retina has photo-receptors ouplked to nerve cells,
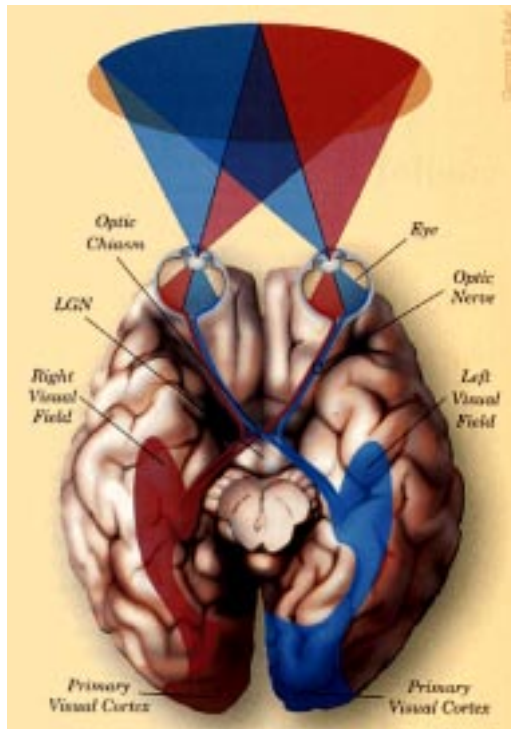
Figure 2.22: The components of the HVS (http://www.hhmi.org/senses/front/110.htm))

which intercept incoming photons and outputting neural signals, which 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 levels*(photopic)* levels of illumination. A schematic drawing of rod and cone cells are shown in Figure 2.24. 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.

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. Cones contain three different pigments, which respond either to blue, red, or green wavelengths of light. A person 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

27

mainly horizontal links, indicating a preferential processing of signals in the horizontal plane.
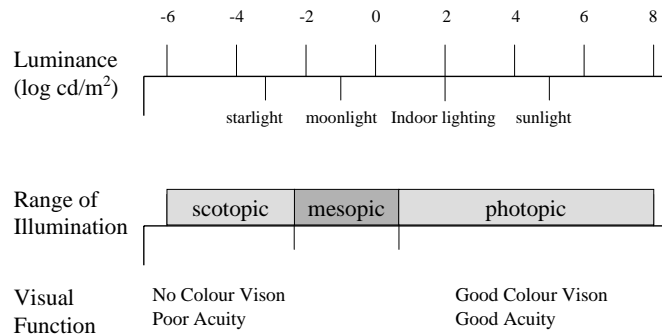


Figure 2.23: The range of luminances in the natural environment and associated visual parameters, after Ferwerda et al.

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 2.23 shows the range of luminances 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 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 $180^o$.

Contrast is defined as by $(l_{max} - l_{min})/(l_{max} + l_{min})$, where lmax and lmin are maximal and minimal luminances. 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 conditional contrast is illustrated in 2.25. Despite the fact that *all* centre squares are the same brightness, they are perceived as different due to the different background brightness.

*Depth Perception* is the ability to see the world in three dimensions and to perceive distance. Images projected onto the retina are two dimensional, 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 which are effective when viewed with only one eye, these include 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
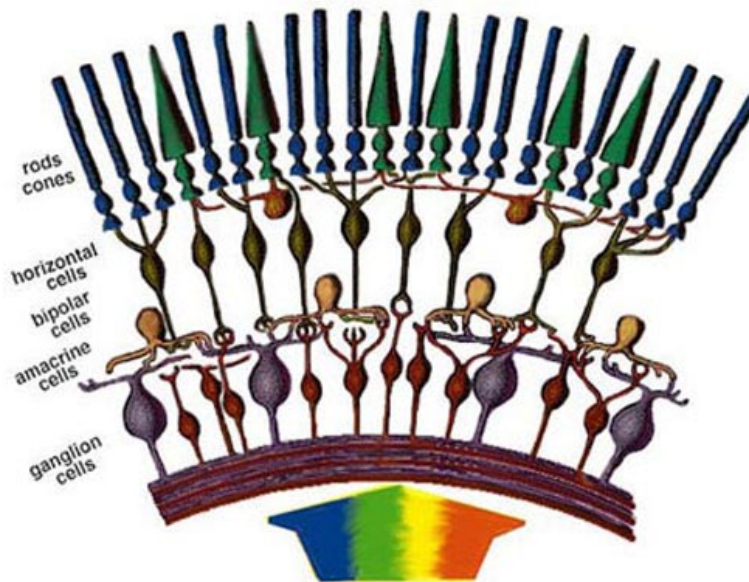
Figure 2.24: Retinal structure, after [47].



Figure 2.25: 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

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 colour 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.

### 2.3.2 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, so it makes sense to examine the HVS as a whole. Instead of reusing information from previous psychophysical experiments, new experiments are needed which examine the HVS as a whole rather than trying to probe individual components. Some examples will support this.

A Benham's disk is a flat disc half of which is black and the other half has three sets of lines like the groves on a record but more spaced out, figure. 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, in fact human percepts are in fact the product of complex multidimensional response.

A second example, figure 2.26, shows the panels in checkerboard block on the left and a flat pattern on the right, which have the same *reflectances* but differences in their three-dimensional organisation means they are perceived differently. The two panel marked with X's have the same reflectance but on the block they appear to have different reflectances 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.
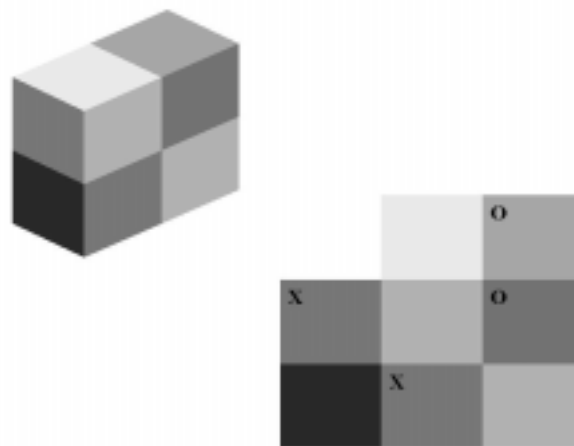


Figure 2.26: Interaction between apparent reflection, apparent illumination and apparent three-dimensional shape. Corresponding panels in the two patterns have the same physical reflectance. Differences in the perceived spatial organisation of the patterns produces differing interpretations in terms of lightness(apparent reflectance) and brightness(apparent illumination), after Adelson

### 2.3.3   Lightness Perception

Gilchrist is a firm believer in the systematic study of lightness error as and understanding of the HVS as, firstly there are always errors, and secondly 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 these notes involved asking human observers to match the reflectance of real world objects to a Munsell chart, this 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.

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. 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 this experimental framework, apart from being simple to perform it also allows excellent control over experimental stimuli.

# Chapter 3

# Perception and Graphics

Realistic image synthesis is defined as the computation of images that are faithful representations of a real scene. For computer generated imagery to be predictive and have use beyond illustration and entertainment then *realism* is the key. Generally, the overall level of accuracy required is determined by the application. For certain applications where viewers simply need to be convinced the scene could be real (children's education, entertainment, games). In such a case, *empirical models* of light simulation may be sufficient. However, for a predictive applications (industrial simulation, archaeological reconstruction) where the aim is to present the user with the same visual experience as if they were actually in the scene, *physically based models* are normally employed. In many cases, but not always, the images are intended for viewing by human observers. It follows then that the required level of accuracy in the image synthesis process is dictated by capability of the human visual system. Recall the image synthesis pipeline illustrated in figure 2.1, which can be loosely categorised into three stages: model, render and display. What is known about human vision can be applied in various ways at each of these stages. The level of detail at the *modeling* stage is prescribed by the level of detail visible to a human observer. During the *rendering* stage the limitations of the HVS can be exploited to speed up rendering without sacrificing the visual quality of the result. Perception based metrics can be used to evaluate rendered images, demonstrating that the image is comparable to what an observer standing in the physical scene would see. Such metrics can be used to determine the visible differences between two images (or successive images in a progressive computation), enabling stopping conditions to be enforced when the differences between computed image and target image are not perceptible. Having spent considerable time and effort computing the image, care must be taken when presenting the result through some display device, so as not to waste this effort. The display, and hence correct perception, of an image involves converting computed values to screen values and finally to values perceptible by the HVS.

## 3.1   Using Perception to increase efficiency

Knowledge of the behaviour of the HVS can be used to speed up rendering by focusing computational effort into areas of an image with perceivable errors. Accounting for such HVS limitations enables computational effort to be shifted from areas of a scene deemed to have a visually insignificant effect on the solutions appearance, and shifted into those areas that are most noticeable. Several attempts have been made to incorporate what is known about the HVS into the image synthesis process.

Mitchell [75] introduced a ray tracing algorithm that took advantage of the HVS's poor sensitivity to high spatial frequency, to absolute physical errors (*threshold sensitivity*) and to the high and low wavelength content of a scene. Using a simple model of vision he managed to generate antialiased images using a low sampling density. Initially the image is sampled at a low density using a Poisson disk sampling strategy. Then an adaptive sampling rate is defined according to the frequency content, thus enabling aliasing noise to be concentrated into high frequencies where artifacts are less conspicuous. To determine which areas of the image needed further refinement, a contrast metric which operates in RGB space was used, this was an attempt to obtain a perceptually based measure of the variation. A differential weighting was then applied to each of the RGB channels to account for colour variation in the spatial sensitivity of the HVS. Finally using multi-stage filters to interpolate the non-uniform samples into the completed image. While this approach has the beginnings of a perceptual method it is at best a crude model for the visual system. This approach fails to consider the phenomenon of masking, and only uses two levels of adaptivity in sampling.

Noting the fact that the HVS has *poor spatial acuity*, Meyer and Li developed an adaptive image synthesis algorithm. An opponents processing model of colour vision, comprising chromatic and achromatic colour channels, forms the core of this algorithm. By altering a screen subdivision raytracer to control the amount of chromatic and achromatic detail present at the edges in a scene. To take advantage of colour acuity deficiencies, an image synthesis procedure which computes low spatial frequency first, and high spatial information last is required. This enables control of the amount of refinement used to generate colour spatial detail. To meet this requirement, Painter and Sloan [84]adaptive subdivision was used to generate a K-D tree representation of an images. Areas of the image containing high frequency information are stored in the lower levels of the tree. Traversing the tree to a lesser depth determined the chromatic channels of the final images. The technique has been validated using a simple psychophysical test, demonstrating that the quality of the result is not degraded by employing the method.

Based on *the visibility of sampling artifacts*, Bolin and Meyer [8] invented a frequency based raytracer which controls the distribution of rays cast into a scene. Their approach synthesised images directly into the frequency domain, enabling them to employ a simple vision model to decide where rays are cast into a scene and determine how to spawn rays that intersect objects in the scene. Following this procedure the artifacts that are most visible in the scene can be eliminated from the image first then noise can be channeled into areas of the image where it is not significant in terms of noticeability. This technique is an improvement on Micthell's because the vision model employed exploits *Contrast Sensitivity:* The response of the eye is non-linear. *Spatial Frequency* The response of the eye is less for patterns of pure colour than for patterns including luminance differences. *Masking* High spatial frequency in the field of view can *mask* the presence of other high frequency information.

Gibson and Hubbold 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.

Myskowski [77] 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, this 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 perceivable threshold and thus not visible to a human observer). This perceptual error metric takes the form of Daly's [21] **V**isible **D**ifference **P**redictor. The VDP takes as input a pair of images. A model of human vision is then applied to these images, transforming them into a visual representation. The "distance" between the images is then computed to form a local visual difference map. This map is then compared against a perceptual threshold value to ascertain the whether or not the difference is perceptible. The VDP is used by Myszkowski by applying it to two intermediate images computed at consecutive time steps of the solution to give a functional error estimate.

Bolin and Meyer [9] devised a similar scheme, also using a sophisticate vision model in an attempt to make use of all HVS failings. They used Sarnoff VDM during image generation to direct subsequent computational effort. They used upper and lower bound images from the computation results at intermediate stages and used the predictor to get an error estimate for that stage. Hence this approach estimates the error bounds.

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 multiscale spatial processing involved in some of its components. This means that the cost of recomputing the vision model may offset the savings gained by employing the perceptual error metric to speed up the rendering algorithm. To combat this Ramasubramanian [88] introduced a metric which handles luminance-dependent processing and spatially-dependent processing independently, allowing the expensive spatially-dependent component to be *precomputed*.

## 3.2   Perceptually Based Image Quality Metrics

Reliable image quality assessments are necessary for the evaluation of realistic images 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 3.1.

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 refjan so 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) assesement of images, afterall most final uses of computer generated images will be viewed by human observers.

A number of experimental studies have demonstrated many features of how the HVS
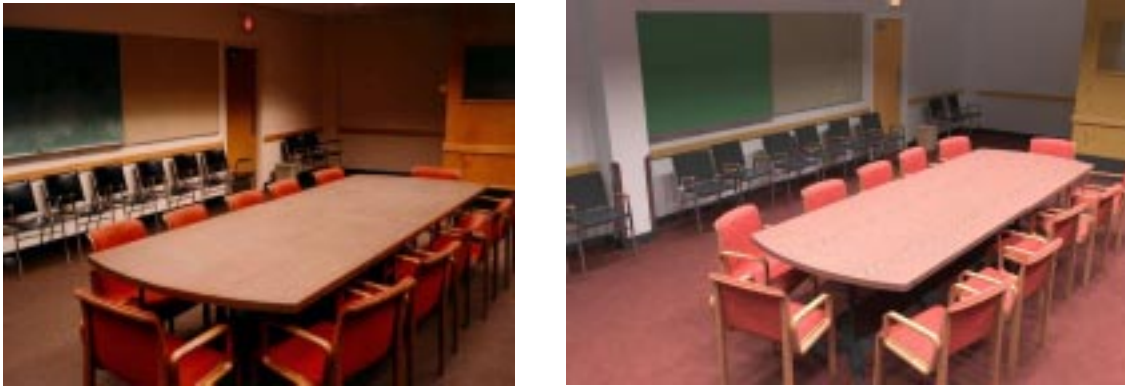
Figure 3.1: Photograph of a Conference Room(left) & Photo-Realistic Rendering(right)

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 in chapter 2, the HVS comprises complex mechanisms, which rather than working independently, often features work together, so it makes sense to examine the HVS as a whole. Instead of reusing information from previous psychophysical experiments, new experiments are needed which examine the HVS as a whole rather than trying to probe individual components.

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 corners could be cut with least visual impact.

Using a simple five sided cube as their test environment Meyer et al [74] presented an approach to image synthesis comprising separate physical and perceptual modules. They chose diffusely reflecting materials to built 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. The calculated values are much higher than the measurements. 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. A group of twenty experiment participants were asked to differentiate between 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 cited
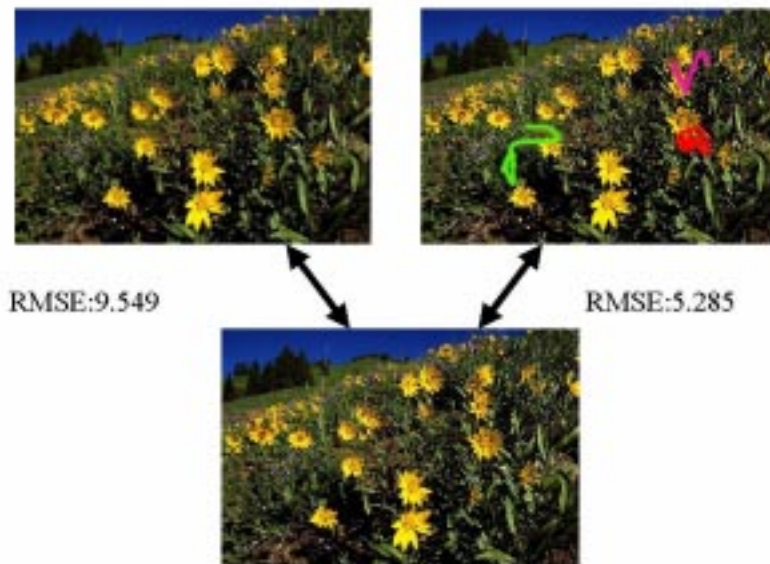
Figure 3.2: **Comparing top images to the image on the bottom use RMSE:** The image on the left has been slightly blurred, the image on the right has deliberate scribbles. The rmse value for blurred image is markedly higher than the rmse for the image on the right. However, a human observer might indicate a higher correlation between the image on the left. This illustrates the use of rmse is not sufficient, after [86]

in the sharpness of the shadows (a consequence of the discretisation in the simulation) and in 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.

Rushmeier et al. [89] explored using perceptually based metrics, based on image appearance, to compare image quality to a captured image of the scene being represented. The following image comparison metrics were derived from [21],[32], [65] in a study which compared real and synthetic images by Rushmeier *et al* [89]. Each is based on ideas taken from image compression techniques. The goal of this work 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* [65]: 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. 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* [32]: This model 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. An 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 [32].

*Model 3 After Daly:adapted from* [21]: In this model the effects of adaptation and non-linearity are combined in one 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. An FFT is applied to each transformed luminance and then they are filtered by a CSF (computed for a level of 50 cd/m$^2$). The distance between the two images is then computed using MSE as in model 1.

The **V**isible **D**ifference **P**redictor (VDP) is a perceptually based image quality metric proposed by Daly [21]. Myskowski [77] 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. Then using 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. 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 with VDP results for shadow and lighting pattern masking by masking and in comparison of the perceived quality of images generated as subsequent stages of indirect lighting solutions.

The VDP is one of the key Image Quality Metrics presented in this course and as such it will be described in detail in the next chapter.

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 enables calibration and validation of the models to assess the level of fidelity of the images produced. This will be described in detail in chapter 5.

## 3.3   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. Both the absolute and dynamic ranges of light we encounter in natural environments are 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, on 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 by the device in question, be is 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 the fact that to 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 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. The human eye is sensitive to *relative* luminances rather than *absolute* luminances. Taking advantage of this 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-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.

### 3.3.1 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 [105]. Using a psychophysical model of brightness perception first developed by Stevens and Stevens, 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. 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 $cdm^{-2}$. 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:

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

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:

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

giving

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

Taking a slightly approach, Ward [112] searched for a linear transform 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, Ward defines a relationship which states that if the eye is adapted to luminance level $L_a$, the smallest alteration in luminance that can be seen satisfies:

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

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

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

Where $L_{aw}$ and $L_{a(d)}$ 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 :

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

To estimate the adaptation levels, $L_{aw} to L_{ad}$, Ward assumed 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 $L_{dmax}$, dividing by $L_{dmax}$ then gives values in the required range from [0..1]. The scaling factor is then given by:

$$m = \frac{1}{L_{dmax}}[\frac{1.219 + (L_{dmax/2})^0.4}{1.219 + (L_{a(w)}^0.4)}]^{2.5}$$

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

In 1996, Ferwerda et al 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 meopic range.

To account for loss in 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 consider by Ferwerda, a parametric constant, $b$ is added to the display luminance, the value of which changes over time. The value of b is set so that...remains constant over time. This means the overall luminance of the displayed image remains the same during the time dependent adaptation process.

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 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, we need to develop a more complete visual model of adaptation. This model will be especially important for immersive display systems that occupy the whole visual field and therefore determine the viewer's visual state.

### 3.3.2  Multi Scale Tone Reproduction Operators

After careful investigation of the effects of tone mapping of a small test scene illuminated only by a single incandescent bulb, Chiu et al [12] 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 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, [91] 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, no attempt is made to employ psychophysical models of the HVS.

Larson et al, [57] developed a histogram equalization 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. Fist 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 $1^o$ of visual field. A histogram a cumulative distribution function were the 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 luminance at a given level of adaptation. This function can then be used to guide algorithms, such a s discontinuity meshing where there is a need to determine whether some process would be noticeable or not to the end user.

## 3.4   Summary

Recent years have seen an increase in the application of visual perception to computer graphics. 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. But realism implies computational expense, 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. Visual perception is used in many guises in graphics to achieve this required level of realism. As the application of graphics continues to diversify to increasingly safety critical applications such as industrial and military applications the correct perception of imagery becomes a priority. In principle, physically based image synthesis can potentially generate images that are faithful visual representation of real or imaginary environments, thus enabling image synthesis to be usefully applied in a wide variety of disciplines.

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 sufficient 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.

There is a strong need for the models of human vision currently used in image synthesis computations to be *validated* to demonstrate their performance is comparable to the actual performance of the HVS.

# Chapter 4

# Perception-driven global illumination and rendering computation

[1]

As we stated in the Introduction the basic goal 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 minimised. Thus, perception issues are clearly involved in realistic rendering, and should be considered at various stages of computation such as image display, global illumination computation and rendering.

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 of the computer graphics community [41, 43], motivated by the progress in physiology, psychophysics, and psychology in providing computational models of the HVS. Since global illumination solutions are costly in terms of computation, there are high prospects for their efficiency improvements 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 [59]. 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 4.1 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, 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

---

[1]written by Karol Myszkowski

derive a relative metric which predicts the perceived differences between a pair of images [90]. (It is well-known that a common mean-squared error metric usually fails in such a task [21, 99, 90, 31].) 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 4.2, we briefly overview applications of such local metrics to guide the global illumination and rendering solutions. Such metrics usually involve advanced HVS models, and might incur non-negligible computation costs. An important issue becomes whether these costs can be compensated by the savings in computation that are obtained through the usage of such metrics.

A representative example of such an advanced image fidelity metric is the Visible Differences Predictor (VDP) developed by Daly [21]. In Section 4.2.1, 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 work 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 4.2.2, 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 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 4.2.3). 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 were able to apply the metric in 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 4.3.1). Based on the obtained results, we propose a novel global illumination algorithm which is a mixture of stochastic (density estimation) and deterministic (adaptive mesh refinement) algorithms that are used in a sequence optimised to reduce the differences between the intermediate and final images as perceived by the human observer in the course of lighting computation (Section 4.3.2). 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 switch-over 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 in perceivable changes to the quality of the resulting images (Section 4.3.3).

## 4.1   Low-level perception-based error metrics

One of the research directions towards the perception-driven improvement of global illumination computation performance relied on direct embedding of some simple error metrics in decision making on the level of light interactions between surfaces. Gibson and Hubbold [33] proposed a perception-driven hierarchical algorithm in which TMO and perceptually uniform colour space CIE $L^*u^*v^*$ are used to decide when to stop the hierarchy refinement. Links between patches are not refined anymore 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 feelers that should be used in visibility test for these patches. A similar strategy was assumed by Martin et al. [70], their 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 [87].

Perceptually-informed error metrics were also successfully introduced to control the adaptive mesh subdivision [81, 33, 46] and mesh simplification [111] in order to minimise the number of mesh elements used to reconstruct lighting function without 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 [61, 25] and adaptive mesh subdivision [16, 107, 60] 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 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. [81] noticed that mesh refinement can be driven by some metrics measuring 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 [105], 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. Ideally, the local illumination should correspond to the global illumination. However, in the radiosity technique [78] only direct illumination is known at the stage of mesh refinement, which might result in a too 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 [81]. Gibson and Hubbold [33] showed that the meshing performance can be improved even if some crude approximation of global illumination such as the ambient correction term [14] is used. Also, Gibson and Hubbold improved [81] further by introducing colour considerations into their mesh subdivision criteria. Further improvement of meshing performance was achieved by Volevich et al. [108], whose lighting simulation algorithm (discussed in more detail in Section 4.3.2) provides local estimates of global illumination quickly. Those estimates are available at the mesh refine-

ment stage, which make it a possibly more reliable evaluation of the contrast at lighting discontinuities. Thus, the prediction of discontinuities perceivability becomes also more robust, and excessive mesh subdivision can be avoided. In the example given in [108], the uniform mesh built of 30,200 triangles was subdivided into 121,000, 97,000, and 86,000 elements using techniques proposed in [81], [33], and [108], respectively, without any noticeable difference in the resulting images quality.

The perception-based criteria have also been used to remove superfluous mesh elements in the discontinuity meshing approach [46]. Also, a similar perception-driven mesh simplification was performed as post-processing to a density estimation solution applying a dense, uniform mesh [111].

All techniques discussed so far used perceptual error metrics on the atomic level (e.g., every light interaction between patches, every mesh element subdivision), which put a significant amount of 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 overview briefly applications of such metric to global illumination and rendering solutions.

## 4.2 Advanced perception-based error metrics

The scenario of embedding advanced HVS models into global illumination and rendering algorithms is very attractive, because computation can be perception-driven specifically for a given scene. Bolin and Meyer [9] have developed an efficient approximation of the Sarnoff Visual Discrimination Model (VDM) [63], which made it possible to use this model to guide 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 [77] 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.* [88] 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 possible more efficient distribution of indirect lighting samples 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 [53], and a significant speedup of computations was reported compared to the sample distribution based on purely stochastic error measures.

A practical problem arises that the computational costs incurred by the HVS models introduce an overhead to the actual lighting computation, which may become the more significant the more rapid is the lighting computation. This means that the potential gains of such perception-driven computation can be easily cancelled by this overhead depending on many factors such as the 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.* [88] 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 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 [21] takes about 20 seconds on a R10000, 195 MHz processor, and such processing should be repeated a number of times to get a reasonable monitoring of the image quality progression. In this work, we explore approaches in which the advanced HVS models are used both off-line and 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 the computation time for lighting simulation, and the costs of HVS processing (which is performed off-line) are of secondary importance. In the latter case, the VDP processing is performed along with 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.

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

### 4.2.1   Visible Differences Predictor

Although, substantial progress in physiology and psychophysics studies has been achieved in recent years, the Human Visual System (HVS) as the whole, and in particular, the higher order cognitive mechanisms, are not 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 [23]. It is believed that the internal representation of an image by cells in the visual cortex is based on spatial frequency and orientation channels [66, 114, 122]. The channel model explains such visual characteristics well 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 aftereffects - as result of the eye 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
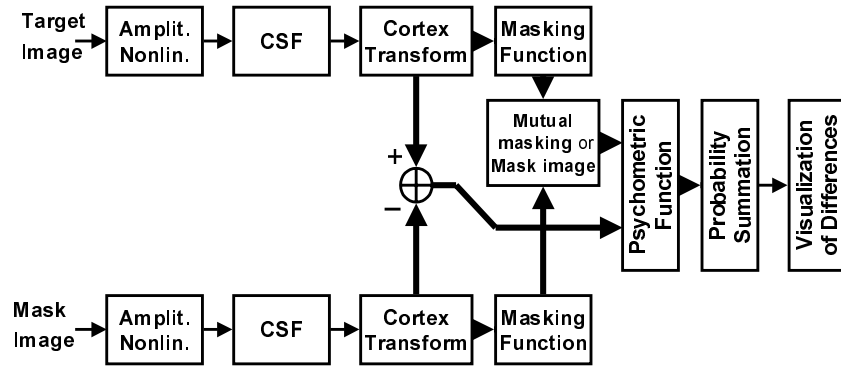
Figure 4.1: Block diagram of the Visible Differences Predictor (heavy arrows indicate parallel processing of the spatial frequency and orientation channels).

observed between two images. Therefore, we were most interested in the HVS models developed for similar tasks [127, 62, 71, 21, 99, 18, 119, 28, 31, 97], 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 [21] 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 to predicting the differences between images that can be perceived by the human observer [58]. 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 an identical initial processing (Figure 4.1). At first, the original pixel intensities are compressed by the amplitude non-linearity based on the local luminance adaptation, simulating Weber's law-like behavior. Then the resulting image is converted into the frequency domain and processing of 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. Then the individual channels are transformed back to the spatial domain, in which visual masking is processed. For every

channel and for every pixel, the elevation of detection threshold is calculated based on the mask contrast for that channel and that pixel. The resulting threshold elevation maps can be computed for the mask image, or mutual masking can be considered by taking the minimal threshold elevation value for the corresponding channels and pixels of the two input images. These 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 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 standard threshold value for discrimination tasks [122]. 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 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 current tasks 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 WWW page [1]).

### 4.2.2 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. [97] this is a potential threat for the VDP integrity. The approach promoted in [97, 123] 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 Daly model integrity 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: [21], [71, 28], and [31]. 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 [31] 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 [21] with 6 spatial and 6 orientation channels (on the WWW page [1] we show a typical output of every channel for our standard test image), and the band-pass (Laplacian) pyramid proposed by Burt [10] 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

low-pass 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 that the differences can be perceived which are reported for every pixel), the distribution of predicted perceivable differences over the image surface usually matched quite well. On the WWW page [1], we provide a comparison of the VDP output for all experiments discussed in this section.

In [77] we report representative results of more specialised VDP experiments, which were focused on prediction of the perceived shadow quality as a function of the visual masking by texture, the CRT device observation distance, and the global illumination solution convergence. In all cases tested we obtained predictions that matched well with our subjective judgements. On the WWW page [1], we provide input images along with the VDP predictions for the full set of experiments that we performed. We disseminate this material so that it can be used for testing other metrics of differences between images.

### 4.2.3   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 development of our novel global illumination algorithm described in Section 4.3.2. 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 [69, 68]. As postulated in [41] the experiments were performed in cooperation with an experimental psychologist.

In the experiment reported here, subjective judgements 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 WWW page [1]). The prediction of differences for the same pairs of images was computed using the VDP, and compared against the subjects' judgements. Figure 4.2 summarises the obtained results. A good agreement was observed between VDP results and subjective ratings, as indicated by the high coefficient of determination, $R^2 = 0.93$. The standardised VDP values (circular symbols) almost always lay within one standard error of the mean standardised rating. This means that as the progressive radiosity solution converged, close agreement between the VDP predictions and the subjective judgements was maintained.
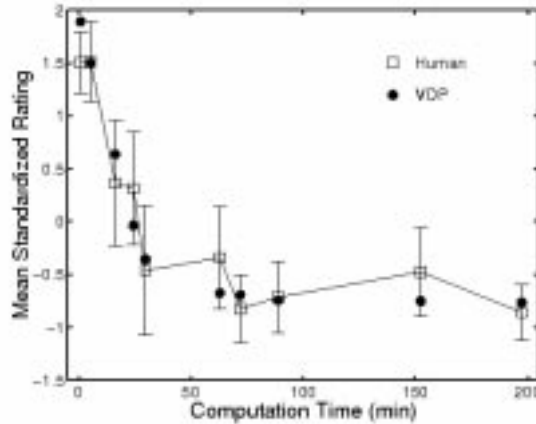
Figure 4.2: The standardised mean ratings (squares) at each of 10 cumulative computation times are shown along with corresponding VDP predictions (filled circles).

Encouraged by the positive results of VDP validation in psychophysical experiments and integrity tests, we used the VDP in actual applications whose 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.

## 4.3 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. Myszkowski [77] has shown that the RMS error it is not suitable to monitor the progress of computation because it poorly predicts the differences as perceived by the human observer (a similar conclusion on this metric, although reached for different applications was reported in [21, 99, 90, 31]). 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. Myszkowski [77] used this observation to measure and compare the performance of various global illumination algorithms using the VDP responses. We discuss this VDP application more in detail in Section 4.3.1. As the result of such a comparison, a hybrid global illumination solution has been proposed in [108], in which the technique that performs best in terms of the perceptual convergence is selected at every stage of computation. We discuss this hybrid technique in Section 4.3.2.

As can be seen in Figure 4.2 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 4.3.3.

51

### 4.3.1 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. In Section 4.2.3 we discussed a new measure of the solution progression, which we called the perceptual convergence in image quality. We used the VDP to provide the quantitative measures of the perceptual convergence by predicting the perceivable differences between the intermediate and final images [77].

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

- Deterministic Direct Lighting (DDL) computation with perceptually-based adaptive mesh subdivision [81].

- Shooting iteration Hierarchical (link-less and cluster-based) Radiosity (SHR) [78, 79] 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 [108]. 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.

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 4.2.1, 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 the difference detection is over 0.75, which is the standard threshold value for discrimination tasks [122]. In all tests performed we used images of resolution $512 \times 512$. The diagonal of images displayed on our CRT device was 0.2 meter, and we assumed that images were observed from the distance of 0.5 meter.

We assumed that the final images used for the VDP computation are based on the DDL+SHR and DDL+IDEPT global illumination solutions which are converged within some negligible error tolerance. The final images obtained using these methods are usually only slightly different (these 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 every 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 [108] 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.
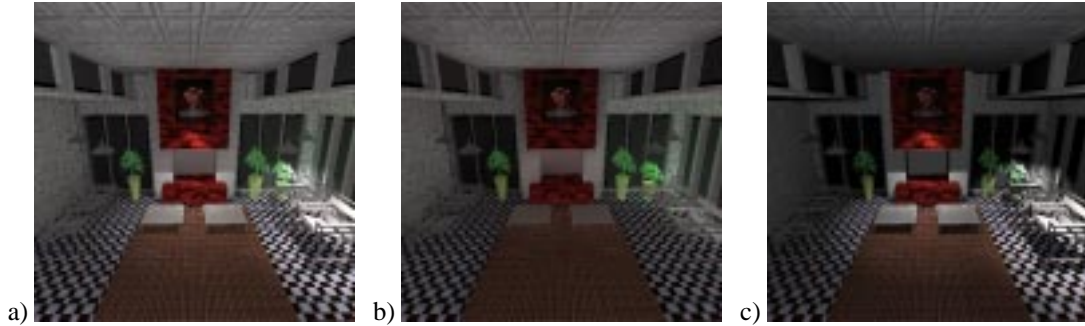


Figure 4.3: Test scene POINT: a) full global illumination solution, b) indirect lighting only, c) direct lighting only.
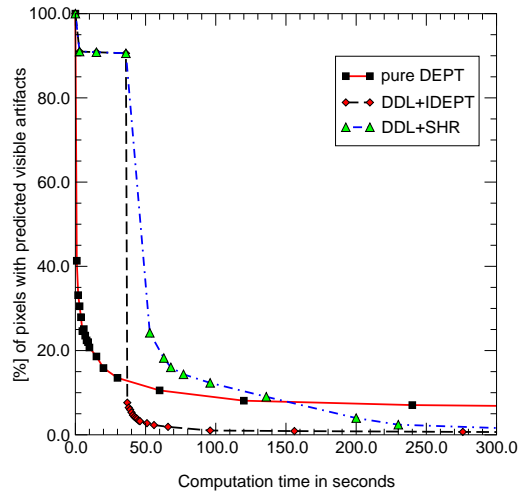


Figure 4.4: Plots of the VDP results (predicted differences between the intermediate and final images) measuring the performance of the DEPT, DDL+IDEPT, and DDL+SHR algorithms for the POINT scene.

The graph in Figure 4.4 shows that the perceptual convergence of the indirect lighting solution for the SHR technique is slower than for the IDEPT approach (direct lighting is computed using the same DDL method). We did not use the ambient light approximation or overshooting techniques [96], because we are interested in physically-sound intermediate results. 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 only for simple scenes. Based on these results, we use the DDL+SHR technique for scenes built of less than 500 polygons. For scenes of more practical complexity we consider the DDL, DDEPT and IDEPT techniques to optimise the progressive refinement of image quality.

The graph in Figure 4.4 shows that at initial stages of computation the combination of DDEPT+IDEPT provides the best performance and rapidly gives meaningful feedback to

the user. At later stages of computation the DDL+IDEPT hybrid shows faster perceptual convergence to the final image. In both cases, we used the same fixed mesh to bucket photons. Because of 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 graph in Figure 4.4 shows that the algorithms examined have different performance at different stages of computation. This makes possible development of a hybrid (composite) algorithm which uses the best candidate algorithm at a given stage of computation. This idea is investigated in the following section.

### 4.3.2 Optimising 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 [108]), a new hybrid technique which uses DDEPT, IDEPT and DDL can be proposed:

1. At first stochastic computations of direct and indirect lighting should be performed.

2. Then the stochastically computed direct component should be gradually replaced by its deterministically computed counterpart, to reconstruct fine details of the lighting function.

3. Finally, the 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 [77].

All discussed algorithms use mesh vertices to separately store the results of direct and indirect lighting computations, so switching between them can easily be performed. Only in the case of the DDL technique, the mesh is adaptively refined to fit the lighting distribution better, 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 [108]), the optimal selection of switch-over points between the sequentially executed algorithms depended on a given scene characteristics. Ideally, the switch-over 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 4.2). To overcome this problem we decided to elaborate a robust heuristic of the switch-over points selection which provides good progression of the image quality for a wide range of indoor scenes. For this purpose, we designed another experiment involving the VDP off-line, and our experimental setting is shown in Figure 4.5.[2] Within this frame-

---

[2]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.
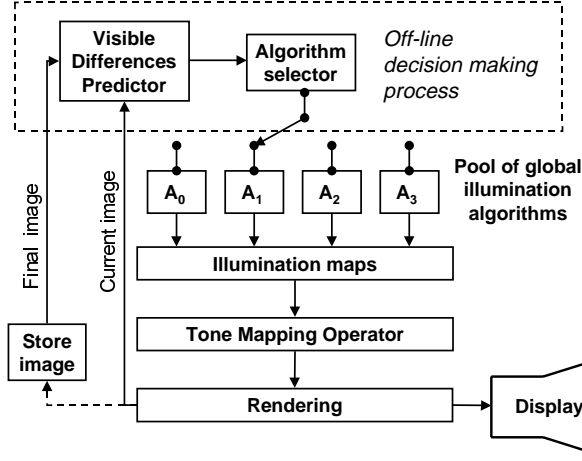
Figure 4.5: Experimental setting for evaluation of the image quality progression and selection of the switch-over points between global illumination algorithms (the human-assisted selection is based on minimising the perceptual distance between the intermediate and final images).

work, 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 are summarised in Figure 4.6a. The thick line between two switching 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, which involves $N$ switching points $T_1, \ldots, T_N$. We investigated various choices of $T_i$, 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 4.6a shows an envelope of all graphs depicting our composite algorithm per-formance for all combinations of switching points investigated by us. This envelope ap-proximates the best expected performance of our composite technique assuming an "opti-mal" switching strategy between the DDL and IDEPT algorithms with multiple switching points $T_1, \ldots, T_N$. As can be seen, gains in performance achieved using the $T_1, \ldots, T_N$ strategy were negligible compared to the strategy based on well chosen switching points $T_1$ and $T_2$.

For the sake of simplicity of our composite algorithm, we decided to use just two switching points $T_1$ and $T_2$. We investigated various choices of $T_1$, which measures dura-tion 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 4.6b. As can be seen our composite algorithm performs much bet-ter than the standalone DDL+IDEPT or DEPT methods for all choices of $T_1$ which are considered in Figure 4.6b. In [108] 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 poor quality 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
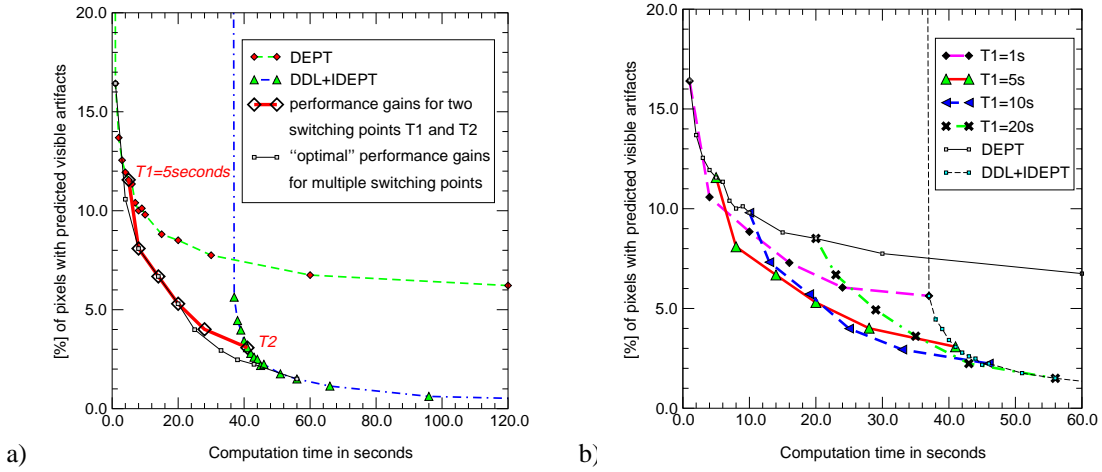
Figure 4.6: Plots of the VDP results (magnified from Figure 4.4) measuring the performance of DEPT and DDL+DEPT algorithms for the POINT test. a) The thick line between two switching points $T_1$ and $T_2$ depicts possible performance gains if the DEPT is replaced by the DDL at $T_1$, and then the IDEPT is activated at $T_2$. The thin line depicts an "optimal" switching strategy between the DDL and IDEPT algorithms with multiple switching points $T_1, \ldots, T_N$. b) Performance gains for various choices of switching time $T_1$.

shadows and other shading details. Because of this, in our heuristic of $T_1$ selection, we assumed that the upper bound for $T_1$ should be comparable with the computation time of the first iteration $T_{i0}$ of the DDL processing, after which the first rendering of a complete direct lighting distribution becomes possible. We can estimate $T_{i0}$ quite well by measuring timings of tracing pilot photons, and knowing the number of the 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 because in our implementation we assumed that 0.5 seconds is the minimal interval for the DEPT solution error sampling). Then, we 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 [109]). 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 [96], 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 switching point $T_1$ selection. However, we found that this heuristic provided quite stable progressive enhancement of the rendered images quality for all performed tests with multiple scenes. This robust behaviour of our heuristic can be explained by the relative

56

insensivity of our composite algorithm to the $T_1$ selection [108], and strong low-pass filtering properties of our lighting reconstruction method at the initial stages of computation.

Figure 4.7 shows an example of fast perceptual convergence of the intermediate solutions in terms of the perceived quality of the corresponding images. The THEATRE scene is built of 17,300 polygons (tessellated into 22,300 mesh elements) and is illuminated by 581 light sources. Figures 4.7a and b depict non-filtered and filtered illumination maps, which were obtained after 30 seconds of the DEPT computation. Figure 4.7b closely resembles the corresponding image in Figure 4.7c, which took 20 and 68 minutes of the DEPT and DDL computations, respectively. The final antialiased image (Figure 4.7d) was rendered using ray tracing, which took 234 minutes (the image resolution was 960 × 740 pixels). In ray tracing computation, the direct lighting was recomputed for every image sample. This solution is typical for multi-pass approaches, e.g., [50]. 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 4.7 exhibit the Lambertian properties of light reflection, the illumination maps (Figures 4.7b and c) are of similar quality to that obtained using the ray tracing computation (Figure 4.7d). 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.
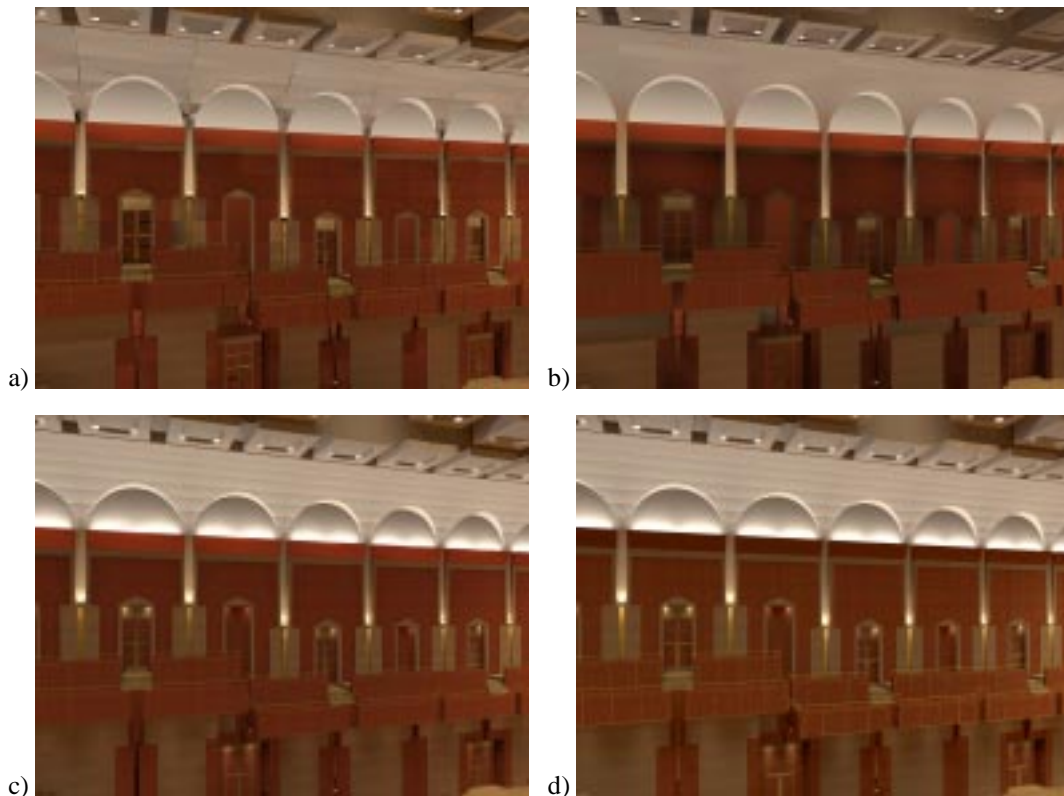


Figure 4.7: Comparison of various renderings for the THEATRE scene: a) photon tracing without illumination map filtering (30 seconds), and b) photon tracing with filtering (30 seconds), c) enhanced accuracy of direct illumination (88 minutes), d) ray traced image (234 minutes).

The composite algorithm discussed in this section was implemented in the commercial package Specter (Integra, Inc., Japan), and was selected as a default global illumination solution because of rapid and meaningful responses for interactive scene changes performed by the user.

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 the 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.

### 4.3.3 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 [77]. 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, and 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 a 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 [106, 110, 109].

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 are not reliable as well, and visible image artifacts may appear although 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 4.2.3. 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 discussed in Section 4.3.1. In this work, we discuss the results obtained for a scene shown in Figure 4.3. However, we achieved similar results for other scenes we tested.

Let us assume that the current image $I_T$ is obtained after the computation time $T$, and let us denote by $VDP(I_T, I_{\alpha T})$ the VDP response for a pair of images $I_T$ and $I_{\alpha T}$ where $0 < \alpha < 1$. We should find an $\alpha$ to get a reasonable match between $VDP(I_T, I_{\alpha T})$ and
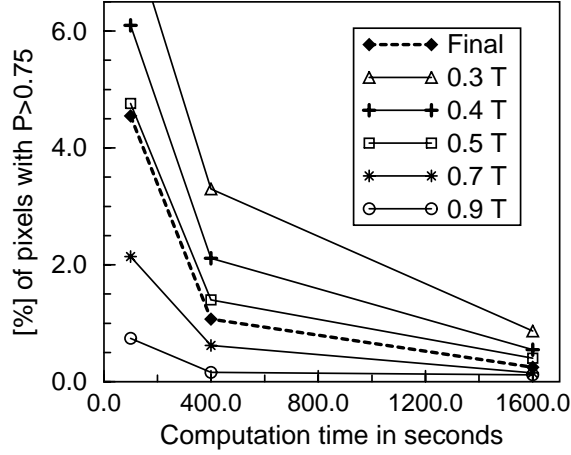
Figure 4.8: The VDP predicted differences between $I_C$ and $I_T$, and $I_T$ and $I_{\alpha T}$ images.

$VDP(I_C, I_T)$, where $I_C$ is an image for the fully converged solution. Figure 4.8 shows the numerical values of $VDP(I_C, I_T)$ and $VDP(I_T, I_{\alpha T})$ for $T = \{100, 400, 1600\}$ seconds and various $\alpha$, for scene shown in Figure 4.3. While the numerical values of $VDP(I_T, I_{0.5T})$ provide the upper bound for $VDP(I_C, I_T)$ 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 WWW page [1] for colour images with $VDP(I_C, I_T)$ and $VDP(I_T, I_{0.5T})$). This means that for certain regions of $I_{0.5T}$ and $I_T$ the variance of the luminance estimate is very small (below the perceived level), and it is likely that it will be so for $I_C$. For other regions such variance is high, and it is likely that luminance estimates for $I_{0.5T}$ and $I_T$ which fluctuate around the converged values for $I_C$ will be different, and can be captured by the VDP. Thus, the choice of $\alpha$ is a trade-off. The $\alpha$ should be small enough to capture such perceivable fluctuations. However, it cannot be too small because $I_{\alpha T}$ may exhibit high variance in the regions in which the solution for $I_T$ converged to that of $I_C$, 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 $I_T$ and $I_{0.5T}$) is such a reasonable trade-off.

# Chapter 5

# A Psychophysical Investigation

This Chapter outlines the steps involved in building a psychophysical experiment to facilitate easy comparison of real scenes and synthetic images by a human observer [73, 102]. This comprises, a calibrated light source, a well articulated scene containing three dimensional objects placed within a custom built environment to evoke certain perceptual cues such as lightness constancy, depth perception and the perception of shadows (the real scene), paired with various synthetic images of that scene. Measurements of virtual environments are often inaccurate. For some applications [1] such estimation of input may be appropriate. However, for the application we are considering an accurate description of the environment is essential to avoid introducing errors at such an early stage. Also, once the global illumination calculations have been computed, it is important to display the resulting image in the correct manner while taking into account the limitations of the display device. The measurements required for this study, the equipment used to record them are described herein, along with the rendering process and the final, important, but often neglected, stage of image display. In this study we conduct a series of psychophysical experiments to assess the fidelity of graphical reconstruction of real scenes. A 3D environment of known composition and methods developed for the study of human visual perception (psychophysics) 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. This investigation is a step toward providing a quantitative answer to the question of just how "real" photo-realism actually is.

## 5.1   Psychophysics

Psychophysics is used to determine the sensitivity of perceptual systems to environmental stimuli. This study seeks to study to establish an empirical relationship between **stimuli** (aspects of the **environment**), and **responses** (aspects of behaviour). This involves the study of the response (psycho) to a known stimulus (physics). 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, this thesis is concerned with the levels of visual responses. More precisely, psycho-physicist are interested in the exploration of thresholds, i.e. what is the minimum

---

[1]The level of realism required is generally application dependent. In some situations a high level of realism is not required, for example games, educational techniques and graphics for web design

change in brightness sensible by the human eye. The goal of this thesis is to discover facts about the visual perception of computer imagery though psychophysics.

Our use of the psychophysical lightness-matching procedure is chosen because it is sensitive to errors in perceived depth. Lightness constancy depends on a correct representation of the three dimensional structure of the scene [35], [34]. Any errors in depth perception, when viewing the computer model, will result in errors of constancy, and thus a poor psychophysical matching performance.

The lens of the human eye projects onto the retina a two dimensional images of a three dimensional physical world. The visual system has a remarkable ability to correctly perceive the physical colour of a surface in spite of wide variations in the mixture of wavelengths in the illumination. This is the phenomenon of colour constancy. Again the visual system has an exceptional capability to determine the lightness of a neutral surface in spite of wide variations in the intensity of the illumination. Such it the phenomenon of lightness constancy, which is analogous to the phenomenon of colour constancy for coloured surfaces.

## 5.2   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. For these reasons a small preliminary study preceded the main experiments. For the pilot study a simple test scene was constructed which allows implementation and testing of various conditions. The main function of this section is to describe precisely how the pilot study was conducted, discuss the results obtained and the modifications necessary to eliminate some unwanted influences present in the pilot method. Following the pilot experiments were more intensive studies, involving more complex conditions. In this work we inspect perceptual, as opposed to physical, correspondence between a real and graphical scene by performing tests of visual function in both situations and establishing that both sets of results are identical (allowing for measurement noise). Such an identity will provide strong evidence that no significantly different set of perceptual parameters exist in each situation.

### 5.2.1   Participants in the Pilot Study

Fifteen 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 this chapter.

### 5.2.2  Apparatus

This study required an experimental set-up comprised of a real environment and a computer representation of that environment. 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 representation.

### 5.2.3  The Real Scene

The test environment was a small box of 557 mm high, 408 mm wide and 507 mm deep, with an opening on one side, figure 5.1. All interior surfaces of the box were painted with white matt house paint. To the right of this enclosure a chart showing thirty gray level
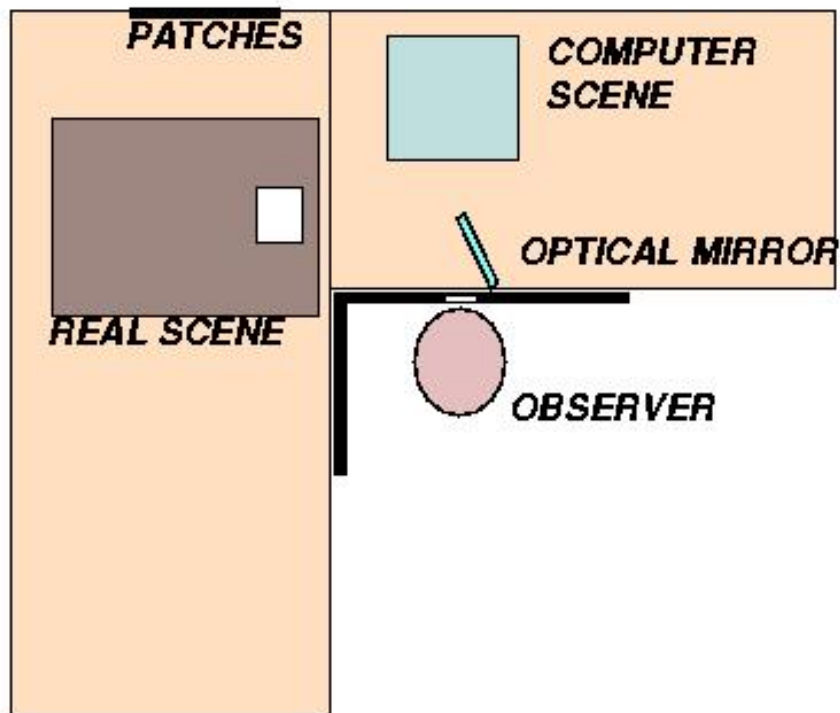


Figure 5.1: Experimental Set up

patches, labelled as in figure 5.2, 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 [125]. A series of fifteen of these gray level patches were chosen at random, reshaped, and placed in no particular order within the physical environment.

A small front-silvered, high quality mirror was incorporated into the set up to allow the viewing conditions to facilitate alternation between the two settings, viewing of the original scene or viewing of the modelled scene on the computer monitor. When the optical mirror was in position, subjects 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 subject 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.
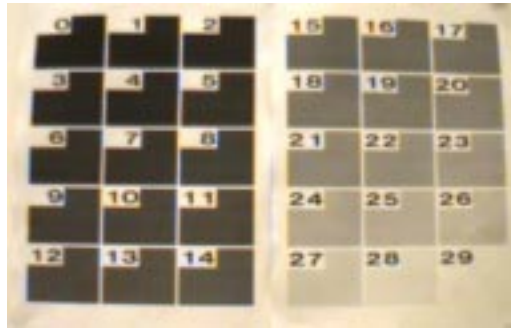
Figure 5.2: Reference patches

### 5.2.4 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 sheets, were placed framing the screen and the open wall of the 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.

### 5.2.5 The Graphical Representation

The geometric model of the real environment was created using Alias Wavefront [118]. The photometric instrument used throughout the course of the 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. 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. For input into the graphical modelling process the following measurements were taken.

*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.

*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) x, y chromaticity coordinates of each primary were obtained and the relative luminances for each phosphor were recorded using the chroma meter. Following Travis [101], these values were then transformed to screen RGB tristimulus values as input to the renderer by applying a matrix based on the chromaticity coordinates of the monitor phosphors and a monitor white point.

*Illumination:* 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.

The rendered image was created using the Radiance Lighting simulation package [112]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.

The entire experimental set-up resides in an enclosed dark laboratory in which the only light sources are the DC bulb (shielded from direct view) or illumination from the monitor. Gilchrist [11],[35],[36] has shown that such an experimental environment is sufficient for the purposes of this experiment.

### 5.2.6  Procedure

The subjects' task was to match gray level patches within a physical environment to a set of control patches. Then subjects 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 5.3. In the *Original Scene*, physical stimuli were presented in the test environ-



Figure 5.3: Rendered Image (left) with blurring (right)

ment, described in the following section. Subjects viewed the screen monocularly through a fixed viewing position. The experiment was undertaken under constant and 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, subjects viewed the screen monocularly through a fixed viewing position.

### 5.2.7  Results and Discussion of Pilot Study

For the pilot study data were obtained for fifteen subjects; fourteen of these were naive as to the purpose of the experiment. Subjects had either normal or corrected-to-normal vision. Each subject performed a number of conditions, in random order, and within each condition the subject's task was to match the fifteen gray test patches to a reference chart on the wall. Each patch was matched once only and the order in which each subject performed the matches was varied between subjects and conditions.

Figure 5.4 shows the results obtained for comparison of a rendered and a blurred scene to the real environment. The x-axis gives the *actual* Munsell value of each patch, the y-axis gives the *matched* Munsell value, averaged across the subjects. A perfect set of data would lie along a $45^0$ diagonal line. 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 constance is not perfect for the original scene. What
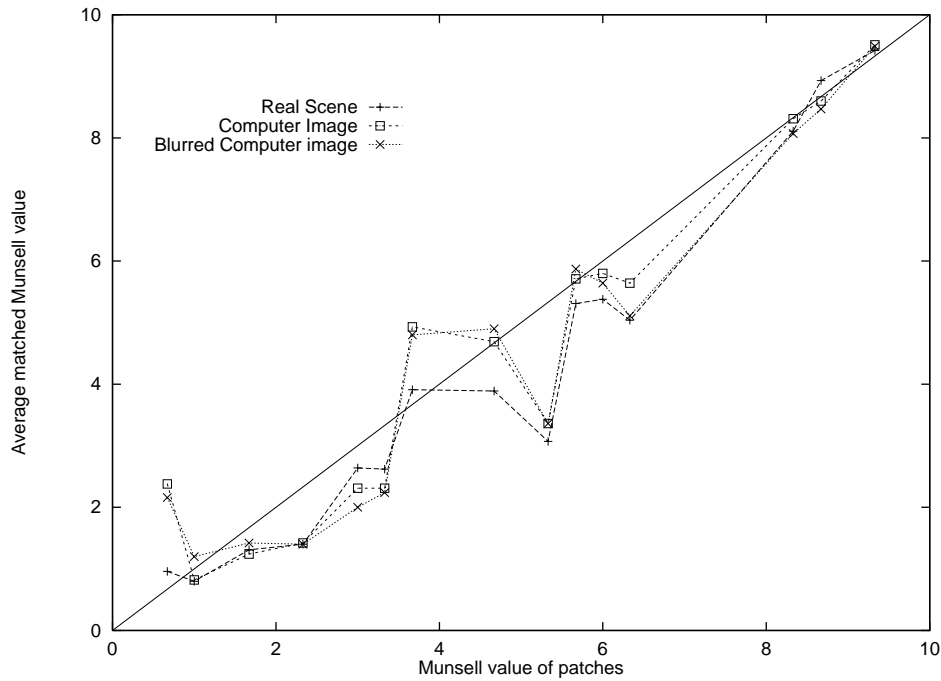
Figure 5.4: Comparison of average matchings of Munsell values

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. The two other graphs relating to the rendered and the blurred rendered images are plotted 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 rendered and the blurred rendered images. This, in turn suggests that there is no significant perceptual difference between the original scene and both the rendered version and the blurred 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; also, under our conditions the blurred version looked very different subjectively, but again similar data were obtained. 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 and blurred rendered functions.

- Compute the mean and standard deviation of these

Table 5.1 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. How do these values compare to other methods? Using the algorithm of Daly [21] we found a 5.04% difference between the rendered and blurred rendered images. 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

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

Table 5.1: Comparison of Rendered and Blurred Scene to Real Environment

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. Further work is planned to attempt to capture a scene without errors of reproduction. Figure 5.5 shows the results obtained for comparison of the real scene and rendered images of the environment after the depth has been altered by 50% and the patches specularity increased from 0% to 50%. As can be seen from the graph, for simple scenes lightness constancy is ex-
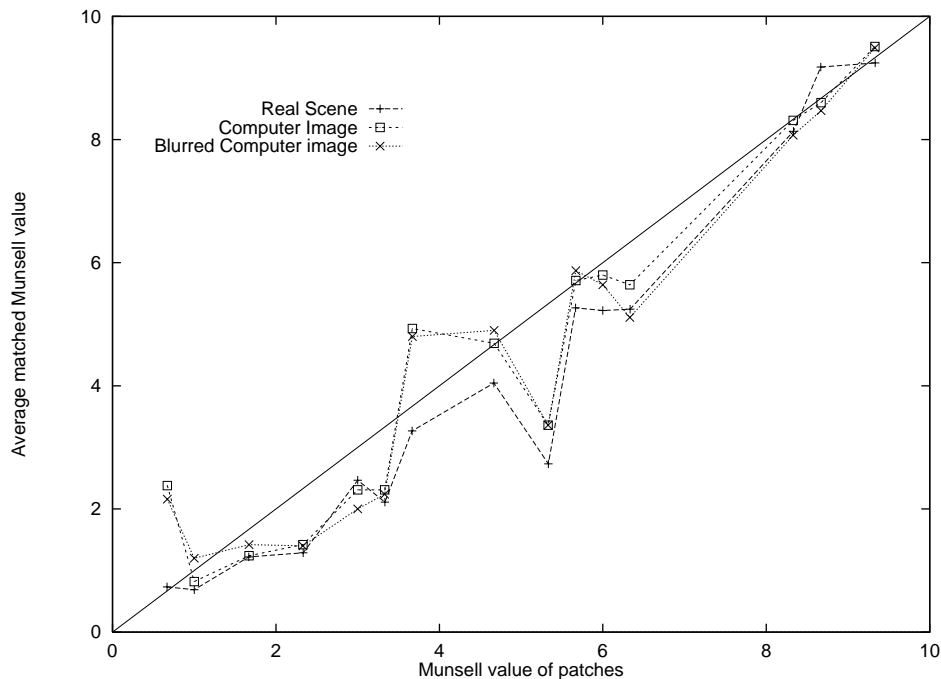


Figure 5.5: Comparison of average matchings of Munsell values

tremely robust against changes in depth, specularity and blurring. In summary: the results 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.

## 5.3 Necessary Modifications to Pilot Study

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 algorithm. To introduce more complexity into the environment, two dimensional patches were extended to three dimensional objects allowing the exploration of effects such as shadowing and depth perception

### 5.3.1 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. In 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. To avoid this scenario images are presented randomly, including presentation of the real environment.

### 5.3.2 Custom Paints

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 achieve, 30ml syringes were used to mix paint in parts as shown in Table 5.3.2

### 5.3.3 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 wooden pillars. The objects and their dimensions are given in table 5.3.3

| Appearance | Parts Black Paint | Parts White Paint | Percentage White |
|---|---|---|---|
| Black | 1 | 0 | 0% |
| Dark Gray | 9 | 1 | 10% |
| Dark Gray | 8 | 0 | 20% |
| Dark Gray | 7 | 0 | 30% |
| Dark Gray | 6 | 0 | 40% |
| Gray | 5 | 0 | 50% |
| Gray | 4 | 0 | 60% |
| Light Gray | 3 | 0 | 70% |
| Light Gray | 2 | 0 | 80% |
| Light Gray | 1 | 0 | 90% |
| Almost White | .5 | 0 | 95% |
| Almost White | .25 | 0 | 97.5% |
| Almost White | .625 | 0 | 98.25% |
| White | 0 | 0 | 100% |

Table 5.2: Quantities of Black and White Paint used to Achieve Grey Levels

### 5.3.4  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 accommodation effects. A new method of matching to patches was devised for the main experiment. Subjects were asked to match numbered patches to a random selection of patches as shown in figure 5.6, then asked to repeat the process with the numbered chart removed. The difference between the two can be accounted for and then scaled back to match. For example, consider a patch number 6. If in the presence of the chart participants matched patch 6 to patch 7, and in the absence of the chart it is matched to patch 8, then this is corrected for after all results are collected.



Figure 5.6: Patches

## 5.4  Conclusions Drawn from 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. The results are summarised in figure. Shadows are important for the correct perception of a scene. Although the pilot study gave..

| Object | Dimensions | Paint |
|---|---|---|
| Pyramid | 1 | 1:9 |
| Small Cylinder | 9 | 0:1 |
| Ledge on Small Cylinder | 8 | 3:2 |
| Small Sphere at Front | 7 | 4:1 |
| Small Cube at Front | 6 | 9:1 |
| Tall Rectangle on Right | 5 | 2:3 |
| Large Sphere | 4 | 2:3 |
| Tall Cylinder on Right | 3 | 1:4 |
| Ledge on Tall Cylinders | 2 | 1:1 |
| Small Cylinder | 1 | 0:1 |
| Tall Cylinder on Left | .5 | 0:1 |
| Tilted Box | .25 | 2:1 |
| Box Under Tilted Box | .625 | 1/4:93/4 |
| Ledge on Rectangle on Right | .625 | 3:7 |
| Tall Rectangle of Right | 0 | 1/8:97/8 |
| Right Wall | 0 | 0:1 |
| Left Wall | 0 | 0:1 |
| Back Wall | 0 | 0:1 |
| Floor | 0 | 0:1 |
| Ceiling | 0 | 0:1 |

Table 5.3: Objects, their placement and assigned paint

This suite of pilot experiments was conducted using fifteen participants. This instilled confidence in the methodology, while establishing some common methods and conditions.

## 5.5 Modifications to the Original Apparatus

To extend the environment to a more complex scene, some additional measurements were needed. In the pilot study the patches were generated using known reflectance, then verified using the Minolta CS-100 Chromameter. For the main experiment, although the *ratios* of the paint were known, their *reflectance* needed to be measured.

The test environment was a five sided box. The dimensions of the box are shown in figure as are the dimensions of the objects that were placed within the box for examination. The spectral reflectance of the paints were measured using a TOPCON-100 spectroradiometer.

### 5.5.1 Spectral Data to RGB

Resulting chromaticity values were converted to RGB triplets by applying a matrix based on the chromaticity coordinates of the monitor phosphors and a monitor white point, described as follows [101].

The CIE (1931) x, y chromaticity co-ordinates of each primary were obtained using the Minolta chroma meter. Then these values were transformed to screen RGB tristimulus values as input to Radiance using the following method. Using these x, y, values the z for each primary can be computed 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 normalisation being achieved by dividing by the sum of the three luminances. 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

$$X_r = Y_r * \frac{x_r}{y_r}$$
$$Z_r = Y_r * \frac{z_r}{y_r}$$

And the same for blue and green primaries.

By this method we can construct matrix

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

(Note Y tristimulus values total to give 1)

Now in order to compute the tristimulus (RGB) values given chromaticity co-ordinates x and y and luminance Y. First, the tristimulus values need to be calculated:

$$(X = x * fracYy) \, (Y = Y) \, (Z = (1 - x - y) * fracYy)$$

Then

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

### 5.5.2 Spectroradiometer

Radiometry is the measurement of radiation in the electromagnetic spectrum. This includes ultraviolet (UV), visible and infrared (IR) light. Electromagnetic radiation is characterised by its frequency of oscillation. This frequency determines the "colour" of the radiation.

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) [2]. The sr-1 outputs the spectral radiance of the sample under examination, in 5nm increments.

## 5.6 Display Devices

Display processes are limited in dynamic range, linearity repeatability and colour range that can be produced. The problem of verification of correctness is amplified by these

---

[2] kindly on loan from DERA

limitations of colour display devices. While it is not yet possible to eliminate these problems, understanding them can provide and insight into compensations that minimise the adverse affects of these limitations.

A minimum warm up time of thirty minutes in a stable environment is required for all monitors. It is standard practice to leave monitors running at all times, as was the case for this set of experiments.

### 5.6.1 Gamma Correction

Regardless of the care taken in modelling the behaviour of light and maintaining the spectral information, an image will not look right unless the monitor on which it is displayed is correctly calibrated and placed in a reasonable viewing environment. In an ideal display, the function relating display luminance and voltage would be linear. In practice this function is non-linear, the actual luminance is typically modelled by a function of the form

$L = D^\gamma$

where $\gamma$ is constant. Gamma correction is the process where voltage values are computed to generate required luminances.

## 5.7 Experimental Design

### 5.7.1 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.

### 5.7.2 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, last say. 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 it is possible to draw four aces from a deck of cards without cheating.

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 treatments 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 subjects, 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.

## 5.8 Procedure

The experimental conditions were kept constant over each subject, 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.

## 5.9 Experiment

### 5.9.1 Training on Munsell Chips

Observers were asked to select, from a grid of 30 achromatic Munsell chips presented on a white background, a sample to match a second unnumbered grid simultaneously displayed on the same background, under constant illumination. 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, there was constant illumination across the wall. 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 [3] 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. Figure ? shows the

---

[3]non-invasive medium

results of this training exercise, data from a student t-test, which gives the probability that two sets of data are from the same source, give confidence that this method is sufficient to allow the numbered chart used in the pilot study to be eliminated from the set up and training to be used in its place. 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.

### 5.9.2 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. The images presented were : Three conditions were run, each
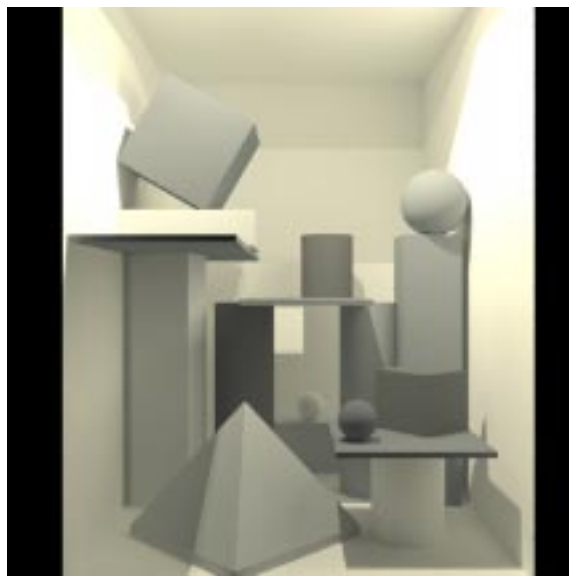


Figure 5.7: Mixed Environment

condition having a number of variations. The environments are shown in figure 5.7

### 5.9.3 Instructions

Each observer was asked to match the small squares on the left to one of the squares 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. In general such an explanations were given when the question was raised by an observer, the task being clear to most observers.

## 5.10 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

| | Mixed Environment | Dark Environment | Light Environment |
|---|---|---|---|
| Real Environment | √ | √ | √ |
| 2 ambient bounces | √ | √ | √ |
| 8 ambient bounces | √ | √ | √ |
| Photograph | √ | √ | √ |
| Tone Mapped | √ | √ | √ |
| Raytracing | √ | | √ |
| Radiosity | √ | | √ |
| Guessed Illumination | √ | √ | √ |
| Guessed Materials | √ | | √ |
| Default | √ | √ | √ |
| Brighter | √ | √ | √ |

Table 5.4: Experimental Conditions: Notes some of the Dark Environments were too dark to use so were not considered

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 [4]. 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.

---

[4] *albedo* is the proportion of light or radiation reflected by a surface

# Chapter 6

# Perception-driven rendering of high-quality walkthrough animations

[1]

Rendering of animated sequences proves to be a very computation intensive task, which in professional production involves specialised 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 [3], 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 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 [24]. 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 [55, 113]. 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 [120]. 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. In our approach, computation is focused on those selected frames (keyframes) and frame fragments (inbetween frames), which strongly affect the whole animation appearance 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 global illumination solution), which provides images of high quality. The remaining pixels are derived using inexpensive Image-Based Rendering (IBR) techniques [72, 67, 92]. Ideally, the

---

[1]written by Karol Myszkowski

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

## 6.1   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 the high level of realism of the derived frames can be obtained while tedious geometric modelling 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, 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 [92]). Figure 6.1 depicts the process of acquiring an image for the desired view based on two reference images, and corresponding depth maps (the distance to the object is encoded in grey scale). At first 3D warping [72] and reprojection of every pixel in the reference image to its new location in the desired image 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 blending step (in the example shown in Figure 6.1 just two images are blended), 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 walkthrough animation along a predefined path a proper selection of reference (keyframe) images is usually easier because of 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 6.3.

3D warping [72] 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 6.2). Based on the displacement vector, and knowing the time span between the subsequent animation frames (e.g., in the PAL composite video standard 25 frames per second are displayed)
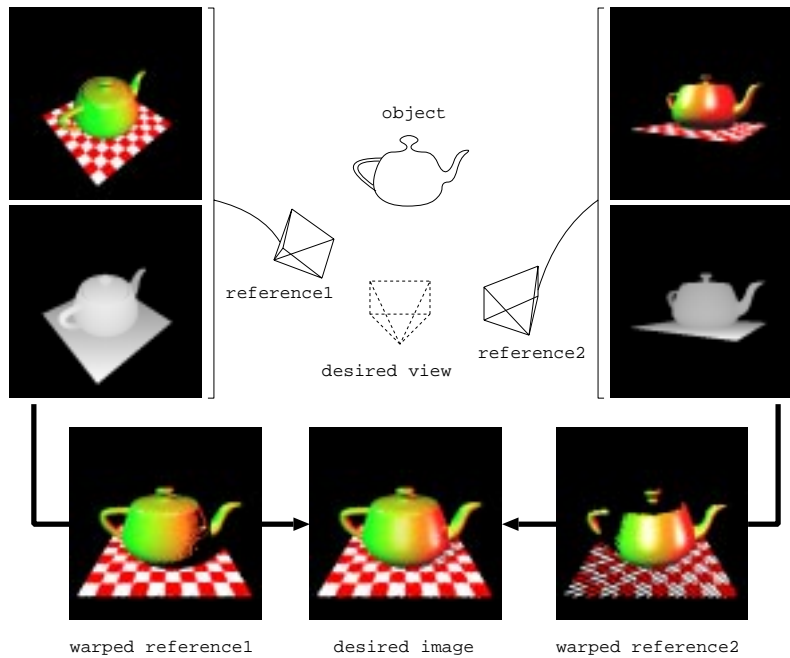
Figure 6.1: IBR: derivation of an image for the desired view based on two reference images.

it is easy to compute the corresponding velocity 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 [98]. In this work, we focus on walkthrough animation sequences that deal exclusively with changes of camera parameters,[2] in which case the PF of good accuracy can be derived using the computationally efficient 3D warping technique. In Section 6.2.1, we show an application of the PF to estimate the human eye sensitivity to spatial patterns moving across the image plane.
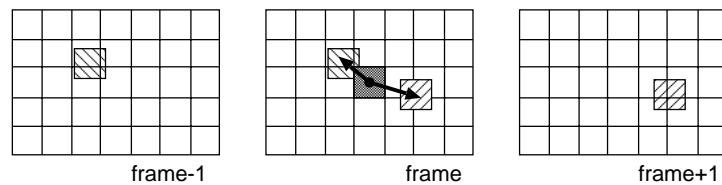


Figure 6.2: Displacement vectors for a pixel of the current frame in respect to the previous (frame-1) and following (frame+1) frames in an animation sequence. All marked pixels depict the same scene detail.

---

[2]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 [93]. For the natural image sequences sufficient spatial image gradients must exist to detect pixel displacements, in which case so called the optical flow can be computed [98]. The optical flow computation is usually far less accurate and more costly than the PF computation for synthetic sequences.

## 6.2   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 [100]. In recent years, a number of automatic video quality metrics, based on the computational models of human vision, have been proposed. Some of these metrics were designed specifically for video [24, 124], and are often specifically tuned (refer to [126]) for the assessment of perceivability of typical distortions arising in lossy video compression such as blocking artifacts, blurring, colour shifts, and fragmentation. Also, some well-established still image quality metrics were extended into the time domain [64, 115, 100].

Lack of comparative studies makes it difficult to evaluate the actual performance of the discussed metrics. It seems that the Sarnoff's Just-Noticeable Difference (JND) Model [64] is the most developed, while a model based on the Discrete Cosine Transform which has been proposed by Watson [115] is computationally efficient and retains many basic characteristics of the Sarnoff model [116]. In this work, we decided to use our own metric of animated sequence quality, which is specifically tuned for synthetic animation sequences.

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.

### 6.2.1   Spatiovelocity CSF model

Spatio-temporal sensitivity to contrast, which varies with the spatial and temporal frequencies is an important characteristics of the HVS. The sensitivity is characterised by so called spatiotemporal CSF, which defines the detection threshold for a stimulus as a function of its spatial and temporal frequencies. The spatiotemporal CSF is widely used in multiple applications dealing with motion imagery. One of the most commonly used analytical approximations of the spatiotemporal CSF are the formulas derived experimentally by Kelly [55]. Instead of experimenting with flickering spatial patterns, Kelly measured contrast sensitivity at several fixed velocities for travelling waves of various spatial frequencies. Kelly used the well-known relationship of equivalence between the visual patterns flickering with temporal frequency $\omega$, and the corresponding steady patterns moving along the image plane with velocity $\vec{v}$ such that [113]:

$$\omega = v_x \rho_x + v_y \rho_y = \vec{v} \cdot \vec{\rho} \qquad (6.1)$$

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 $\rho_x$ and $\rho_y$ are the corresponding components of the spatial frequency $\vec{\rho}$. 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 (6.1).

We use the spatiovelocity CSF model provided by Daly [22], who extended Kelly's model to accommodate for the requirements of current CRT display devices (characterised by the maximum luminance levels of about $100\,cd/m^2$), and obtained the following formula:

$$CSF(\rho, v) = c_0(6.1 + 7.3|\log(\frac{c_2 v}{3})|^3)c_2 v(2\pi c_1\rho)^2\exp(-\frac{4\pi c_1\rho(c_2 v + 2)}{45.9}) \quad (6.2)$$

where $\rho = \|\vec{\rho}\|$ 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 [22, 80] more extended discussion on estimates of the retinal velocity is available, which takes into account the eye natural drift, smooth pursuit, and saccadic movements.

Although, the spatiotemporal CSF is used by widely known video quality metrics, we chose to include the spatiovelocity CSF to our animation quality metric. In this design decision, we were encouraged by an observation that it is not clear whether the vision channels are better described as spatiotemporal (e.g., Hess and Snowden [48] and many other results in psychophysics) or spatiovelocity (e.g., Movshon et al. [76] 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 [22]. Finally, the widely used spatiotemporal CSF was in fact derived from Kelly's spatiovelocity CSF, which was measured for moving stimuli (travelling 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 shown in Section 6.1). 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.

### 6.2.2 AQM algorithm

As the framework of our Animation Quality Metric (AQM) we decided to expand the perception-based visible differences predictor for static images proposed by Eriksson et al. [27]. 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 [27]. Also, we found that the responses of this predictor are very robust, and its architecture was suitable for incorporation into the spatiovelocity CSF.

Figure 6.3 illustrates the processing flow of the AQM. Two comparison animation sequences are provided as input. As for the VDP (refer to Section 4.2.1), for every pair of input frames a map of probability values is generated as output, which characterises the difference in perceivability. Also, the percentage of pixels with the predicted differences over the Just Noticeable Differences (JND) unit [64, 21] is reported. Each of the compared animation frames undergoes the identical initial processing. At first, the original pixel intensities are compressed by amplitude non-linearity and normalised 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 [21] 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 [27] 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 Kelly model. The contrast sensitivity values are calculated using equation (6.2) for the centre frequency $\rho$ of

each Cortex frequency band. 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 6.2). As we discussed in Section 6.1, the PF can be estimated rapidly using the 3D warping technique, which requires access to the range data of current frame, and the camera parameters for all three involved frames. This means that the access to well localised 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 [26], we project the retinal velocity vector onto the direction of the filter band orientation. The contrast sensitivity values resulting from such processing are used to normalise the contrasts in every frequency-orientation channel into the JND units. Next the visual masking is modelled using the threshold elevation approach [27]. The final stage is error pooling across all channels.
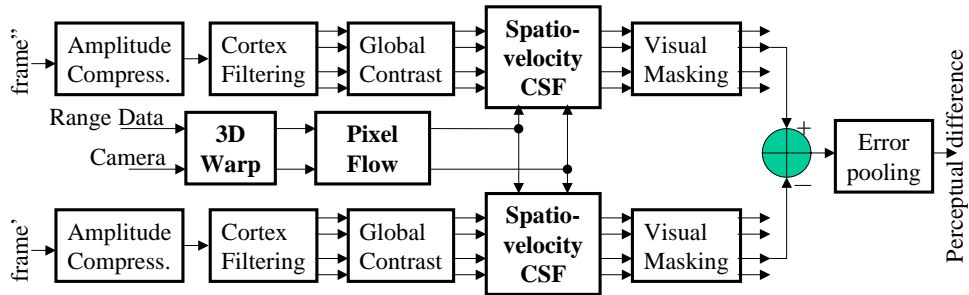


Figure 6.3: Animation Quality Metric. The spatiovelocity CSF requires the velocity value for every pixel, which is acquired from the PF. The PF is computed for the previous and following frames along the animation path in respect to the input frame' (or frame'' which should closely correspond to frame').

In this work, we apply the AQM to guide inbetween frame computation, which we discuss in the following section.

## 6.3   Rendering of the animation

For animation techniques relying on keyframing the rendering cost depend heavily upon the efficiency of inbetween frame computation because the inbetween frames usually significantly outnumber the keyframes. We use IBR techniques [72, 67] described in Section 6.1 to derive the inbetween frames. Our goal is to maximise the number of pixels computed using the IBR approach without deteriorating the animation quality. An appropriate selection of keyframes is an important factor in achieving this goal. We assume that initially the keyframes are placed sparsely and uniformly, and then adaptive keyframe selection is performed, which is guided by the AQM predictions. At first the initial keyframe placement is decided, and then every resulting segment $S$ is processed separately applying the following recursive procedure.
The first frame $k_0$ and the last frame $k_{2N}$ in $S$ are generated using ray tracing (keyframes are shared by two neighbouring segments and are computed only once). Then 3D warping [72] is performed, and we generate two frames corresponding to $k_N$ as follows: $k'_N = Warp(k_0)$ and $k''_N = Warp(k_{2N})$. Using the AQM we compute a map of perceivable differences between $k'_N$ and $k''_N$. As explained in Section 6.2.2, this quality metric

incorporates the PF between frames $k_{N-1}$ and $k_N$, and frames $k_N$ and $k_{N+1}$ to account for the temporal sensitivity of the human observer.

In an analysis process, we first search for perceivable differences in the images of objects with strong specular, transparent and glossy properties, which we identify using the item buffer of frame $k_N$. Such surfaces cannot be properly reconstructed using basic IBR techniques described in Section 6.1. Because of that all pixels depicting such objects for which significant differences are reported in the perceivable differences map will be recalculated using ray tracing. We mask out all ray-traced pixels from the map. In the same manner, we mask out holes composed of pixels which could not be derived from the reference images using 3D warping. If the masked-out difference map still shows significant discrepancies between $k_N'$ and $k_N''$ then we split the segment $S$ in the middle and we process recursively two resulting sub-segments using the procedure described in the previous paragraph. Otherwise, we blend $k_N'$ and $k_N''$ (with correct processing of depth [92]), and ray trace pixels for the remaining holes, and masked out specular objects to derive the final frame $k_N$.

In the same way, we generate all the remaining frames in $S$. To avoid image quality degradation resulting from multiple resamplings, we always warp the fully ray-traced reference frames $k_0$ and $k_{2N}$ to derive all inbetween frames in $S$.

We evaluate the animation quality metric only for frame $k_N$. We assume that derivation of $k_N$ applying the IBR techniques is the most error-prone in the whole segment $S$ because its minimal distance along the animation path to either the $k_0$ or $k_{2N}$ frames is the longest one. This assumption is a trade off between the time spent for rendering and for the control of its quality but in practice, it holds well for typical animation paths.

Figure 6.4 summaries 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 utilises the PF to perform motion-compensated filtering (refer to [80] for more details).

On the Web page located under the URL:

`http://www.u-aizu.ac.jp/labs/csel/aqm/,`

we provide the walkthrough animation sequences which result from our techniques of adaptive keyframe selection guided by the AQM predictions.
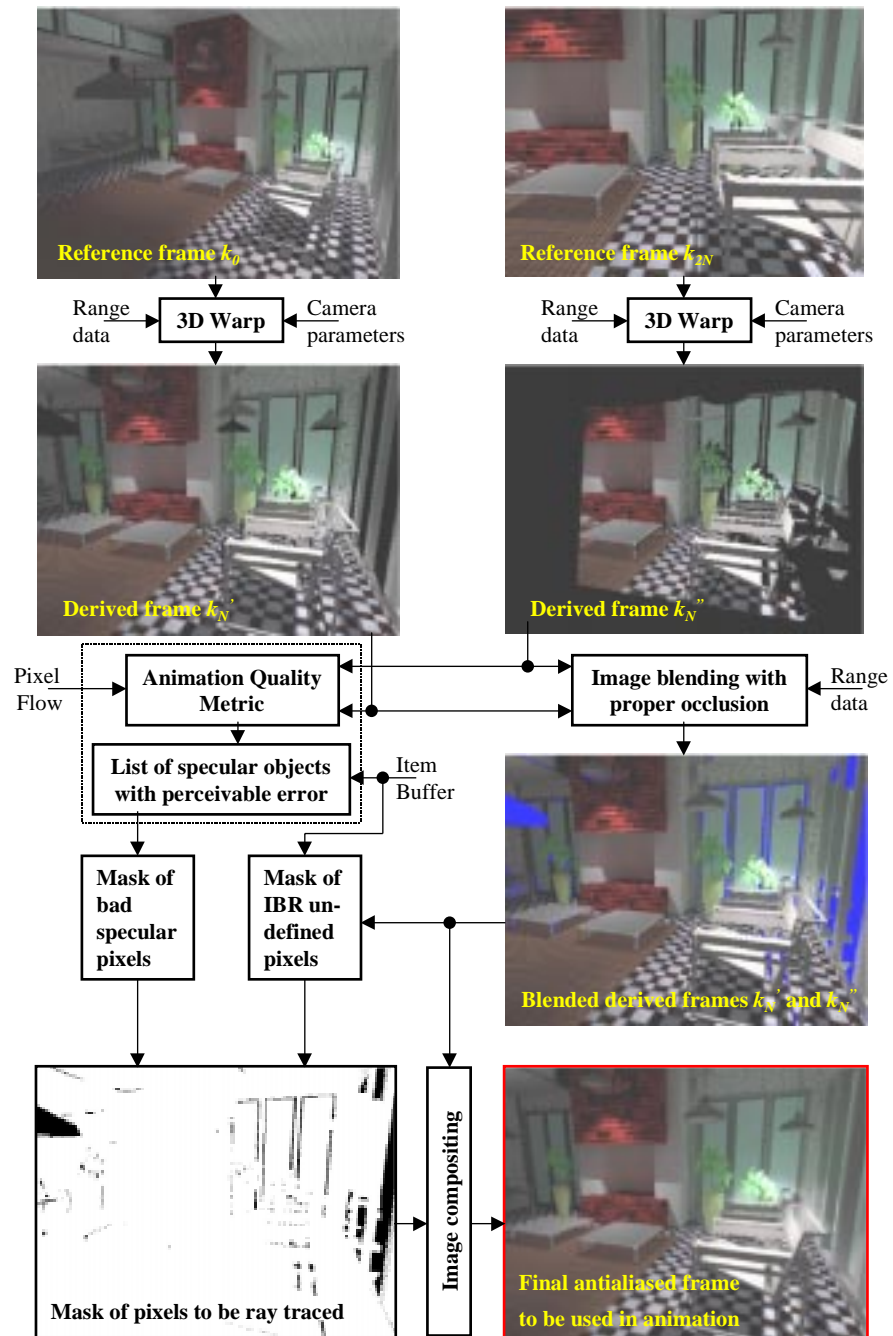
Figure 6.4: The processing flow for inbetween frames computation.

# Bibliography

[1] `http://www.u-aizu.ac.jp/labs/csel/vdp/` - the Web page with documentation of the VDP validation experiments.

[2] M. M. 1994 by Charles Ehrlich, 1994. http://radsite.lbl.gov.

[3] A. Apodaca. Photosurrealism. In *Rendering Techniques '98 (Proceedings of the Nineth Eurographics Workshop on Rendering)*, pages 315–322. Springer Wien, 1998.

[4] A. Appel. Some techniques for shading machine renderings of solids. In *AFIPS 1968 Spring Joint Computer Conf.*, volume 32, pages 37–45, 1968.

[5] J. Arvo and D. Kirk. Particle transport and image synthesis. volume 24, pages 63–66, Aug. 1990.

[6] I. Ashdown. *Radiosity A Programmers Perspective*. John Wiley and Sons, New York, NY, 1994.

[7] I. Ashdown, 1999. http://www.ledalite.co/library-/rad.htmm.

[8] M. Bolin and G. Meyer. A frequency based ray tracer. In *ACM SIGGRAPH '95 Conference Proceedings*, pages 409–418, 1995.

[9] M. Bolin and G. Meyer. A perceptually based adaptive sampling algorithm. In *ACM SIGGRAPH '98 Conference Proceedings*, pages 299–310, 1998.

[10] P. Burt. Fast filter transforms for image processing. *Computer Vision, Graphics and Image Processing*, 21:368–382, 1983.

[11] J. Cataliotti and A. Gilchrist. Local and global processes in lightness perception. In *Perception and Psychophysics*, volume 57(2), pages 125–135, 1995.

[12] 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.

[13] CIE. *CIE Proceedings*. Cambridge University Press, Cambridge, 1924.

[14] M. Cohen, S. E. Chen, J. R. Wallace, and D. P. Greenberg. A progressive refinement approach to fast radiosity image generation. In *Computer Graphics (SIGGRAPH '88 Proceedings)*, volume 22, pages 75–84, August 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, D. P. Greenberg, D. S. Immel, and P. J. Brock. An efficient radiosity approach for realistic image synthesis. *IEEE CG&A*, 6(3):26–35, March 1986.

[17] M. F. Cohen and J. R. Wallace. *Radiosity and Realistic Image Synthesis*. Academic Press Professional, Boston, MA, 1993.

[18] 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.

[19] R. L. Cook. Stochastic sampling in computer graphics. *ACM Transactions on Graphics*, 5(1):51–72, January 1986.

[20] R. L. Cook, T. Porter, and L. Carpenter. Distributed ray tracing. volume 18, pages 137–145. ACM SIGGRAPH, July 1984.

[21] 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.

[22] S. Daly. Engineering observations from spatiovelocity and spatiotemporal visual models. In *Human Vision and Electronic Imaging III*, pages 180–191. SPIE Vol. 3299, 1998.

[23] R. De Valois and D. K.K. *Spatial vision*. Oxford University Press, Oxford, 1990.

[24] C. den Branden Lambrecht. *Perceptual models and architectures for video coding applications*. Ph.D. thesis, 1996.

[25] G. Drettakis and F. X. Sillion. Accurate visibility and meshing calculations for hierarchical radiosity. *Eurographics Rendering Workshop 1996*, pages 269–278, June 1996. Held in Porto, Portugal.

[26] M. Eckert and G. Buchsbaum. The significance of eye movements and image acceleration for coding television image sequences. In A. Watson, editor, *Digital Image and Human Vision*, pages 89–98. Cambridge, MA: MIT Press, 1993.

[27] R. Eriksson, B. Andren, and K. Brunnstrom. Modelling of perception of digital images: a performance study. pages 88–97. Proceedings of SPIE Vol. 3299.

[28] J. Ferwerda, S. Pattanaik, P. Shirley, and D. Greenberg. A model of visual masking for computer graphics. In *ACM SIGGRAPH '97 Conference Proceedings*, pages 143–152, 1997.

[29] 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.

[30] A. Fujimoto, T. Tanaka, and K. Iwata. ARTS: Accelerated ray tracing system. *IEEE Computer Graphics and Applications*, 6(4):16–26, 1986.

[31] A. Gaddipatti, R. Machiraju, and R. Yagel. Steering image generation with wavelet based perceptual metric. *Computer Graphics Forum (Eurographics '97)*, 16(3):241–251, Sept. 1997.

[32] J. Gervais, J. L.O. Harvey, and J. Roberts. Identification confusions among letters of the alphabet. In *Journal of Experimental Psychology: Human Perception and Perfor mance*, volume 10(5), pages 655–666, 1984.

[33] S. Gibson and R. J. Hubbold. Perceptually-driven radiosity. *Computer Graphics Forum*, 16(2):129–141, 1997.

[34] A. Gilchrist. Lightness contrast and filures of lightness constancy: a common explanation. In *Perception and Psychophysics*, volume 43(5), pages 125–135, 1988.

[35] A. Gilchrist. *Lightness, Brightness and Transparency*. Hillsdale: Lawerence Erlibaum Associates, 1996.

[36] A. Gilchrist and A. Jacobsen. Perception of lightness and illumination in a world of one reflectance. *Perception*, 13:5–19, 1984.

[37] A. S. Glassner. Space sub-division for fast ray tracing. *IEEE Computer Graphics and Applications, October 1984*, 4:15–22, 1984.

[38] A. S. Glassner. *Principles of Digital Image Synthesis*. Morgan Kaufmann Publishers, San Francisco, California, 1995.

[39] 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*. July 1987.

[40] C. M. Goral, K. K. Torrance, D. P. Greenberg, and B. Battaile. Modelling the interaction of light between diffuse surfaces. *Computer Graphics*, 18(3):213–222, July 1984.

[41] D. Greenberg, K. Torrance, P. Shirley, J. Arvo, J. Ferwerda, S. Pattanaik, E. Lafortune, B. Walter, S. Foo, and B. Trumbore. A framework for realistic image synthesis. In *ACM SIGGRAPH '97 Conference Proceedings*, pages 477–494, 1997.

[42] D. P. Greenberg. A framework for realistic image synthesis. *Communications of the ACM*, 42(8):44–53, Aug. 1999.

[43] D. P. Greenberg. A framework for realistic image synthesis. *Communications of the ACM*, 42(8):43–53, August 1999.

[44] D. P. Greenberg, K. E. Torrance, P. Shirley, J. Arvo, J. A. Ferwerda, S. Pattanaik, E. P. F. Lafortune, B. Walter, S.-C. 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.

[45] R. A. Hall and D. P. Greenberg. A testbed for realistic image synthesis. *IEEE Computer Graphics and Applications*, 3(8):10–20, Nov. 1983.

[46] D. Hedley, A. Worrall, and D. Paddon. Selective culling of discontinuity lines. In *8th Eurographics Workshop on Rendering*, pages 69–80, 1997.

[47] R. N. Helga Kolb, Eduardo Fernandez, 2000. http://www.insight.med.utah.edu/Webvision/.

[48] R. Hess and R. Snowden. Temporal properties of human visual filters: number shapes and spatial covariation. *Vision Research*, 32:47–59, 1992.

[49] 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.

[50] H. W. Jensen. Global Illumination Using Photon Maps. *Seventh Eurographics Workshop on Rendering*, pages 21–30, 1996.

[51] 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.

[52] J. T. Kajiya. Radiometry and photometry for computer graphics. In *SIGGRAPH '90 Advanced Topics in Ray Tracing course notes*. ACM SIGGRAPH, Aug. 1990.

[53] J. T. Kajiya. The rendering equation. *Computer Graphics (SIGGRAPH '86 Proceedings)*, 20(4):143–150, August 1986. Held in Dallas, Texas.

[54] J. E. Kaufman and H. Haynes. *IES Lighting Handbook*. Illuminating Engineering Society of North America, 1981.

[55] D. Kelly. Motion and Vision 2. Stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America*, 69(10):1340–1349, 1979.

[56] E. Lafortune. *Mathematical Models and Monte Carlo Algorithms for Physically Based Rendering*. Ph.D. thesis, Department of Computer Science, Katholieke Universitiet Leuven, Leuven, Belgium, Feb. 1996.

[57] 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.

[58] B. Li, G. Meyer, and R. Klassen. A comparison of two image quality models. In *Human Vision and Electronic Imaging III*, pages 98–109. SPIE Vol. 3299, 1998.

[59] D. Lischinski, B. Smits, and D. P. Greenberg. Bounds and error estimates for radiosity. In A. Glassner, editor, *Proceedings of SIGGRAPH '94 (Orlando, Florida)*, Computer Graphics Proceedings, Annual Conference Series, pages 67–74, July 1994.

[60] D. Lischinski, F. Tampieri, and D. P. Greenberg. Discontinuity meshing for accurate radiosity. *IEEE CG&A*, 12(6):25–39, November 1992.

[61] D. Lischinski, F. Tampieri, and D. P. Greenberg. Combining hierarchical radiosity and discontinuity meshing. *Computer Graphics (SIGGRAPH'93 Proceedings)*, 27:199–208, August 1993.

[62] C. Lloyd and R. 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.

[63] 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.

[64] J. Lubin. A human vision model for objective picture quality measurements. In *Conference Publication No. 447*, pages 498–503. IEE International Broadcasting Convention, 1997.

[65] 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.

[66] S. Marcelja. Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Am.*, 70:1297–1300, 1980.

[67] W. Mark, L. McMillan, and G. Bishop. Post-rendering 3D warping. In *1997 Symposium on Interactive 3D Graphics*, pages 7–16. ACM SIGGRAPH, 1997.

[68] W. Martens and K. Myszkowski. Appearance preservation in image processing and synthesis. In *Fifth International Workshop on Human Interface Technology*, pages 25–32, 1998.

[69] W. 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.

[70] I. Martin, X. Pueyo, and D. Tost. An image-space refinement criterion for linear hierarchical radiosity. In *Graphics Interface '97*, pages 26–36, 1997.

[71] R. Martin, A. Ahumada, and J. 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.

[72] L. McMillan. *An Image-Based Approach to 3D Computer Graphics*. Ph.D. thesis, 1997.

[73] A. McNamara, A. Chalmers, T. Troscianko, 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.

[74] 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.

[75] D. P. Mitchell. Generating antialiased images at low sampling densities. *Computer Graphics*, 21(4):65–72, July 1987.

[76] J. Movshon, E. Adelson, M. Gizzi, and W. 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.

[77] K. Myszkowski. The Visible Differences Predictor: Applications to global illumination problems. *Eurographics Rendering Workshop 1998*, pages 223–236, June 1998.

[78] K. Myszkowski and T. 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.

[79] K. Myszkowski and T. Kunii. A case study towards validation of global illumination algorithms: progressive hierarchical radiosity with clustering. *The Visual Computer*, 16(?), 2000.

[80] K. Myszkowski, P. Rokita, and T. Tawara. Perceptually-informed accelerated rendering of high quality walkthrough sequences. *Eurographics Rendering Workshop 1999*, June 1999. Held in Granada, Spain.

[81] K. Myszkowski, A. Wojdala, and K. Wicynski. Non-uniform adaptive meshing for global illumination. *Machine Graphics and Vision*, 3(4):601–610, 1994.

[82] 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.

[83] 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.

[84] 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.

[85] B.-T. Phong. Illumination for computer generated pictures. *CACM June 1975*, 18(6):311–317, 1975.

[86] J. Prikryl, 1999. http://www.cs.kuleuven.ac.be/ graphics/SEMINAR/program.html#Wien.

[87] J. Prikryl and W. Purgathofer. State of the art in perceptually driven radiosity. In *State of the Art Reports. Eurographics '98*. 1998.

[88] 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.

[89] 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.

[90] H. Rushmeier, G. Ward, C. Piatko, P. Sanders, and B. Rust. Comparing real and synthetic images: some ideas about metrics. In *Sixth Eurographics Workshop on Rendering*, pages 82–91. Eurographics, June 1995.

[91] C. Schlick. An inexpensive BRDF model for physically-based rendering. *Computer Graphics Forum*, 13(3):C/233–C/246, 1994.

[92] J. Shade, S. Gortler, L. He, and R. Szeliski. Layered depth images. In *SIGGRAPH 98 Conference Proceedings*, pages 231–242, 1998.

[93] M. Shinya. Spatial anti-aliasing for animation sequences with spatio-temporal filtering. In *SIGGRAPH '93 Proceedings*, volume 27, pages 289–296, 1993.

[94] R. Siegel and J. R. Howell. *Thermal Radiation Heat Transfer*. Hemisphere Publishing Corporation, Washington, D.C., 3rd edition, 1992.

[95] 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.

[96] F. Sillion and C. Puech. *Radiosity and Global Illumination*. Morgan Kaufmann, San Francisco, 1994.

[97] C. Taylor, Z. Pizlo, J. P. Allebach, and C. 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.

[98] A. M. Tekalp. *Digital video Processing*. Prentice Hall, 1995.

[99] P. Teo and D. Heeger. Perceptual image distortion. pages 127–141. SPIE Vol. 2179, 1994.

[100] 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. Proceedings of SPIE Vol. 3644, 1999.

[101] D. Travis. *Effective Color Displays*. Academic Press, 1991.

[102] T. Troscianko, A. McNamara, and A. Chalmers. Measures of lightness constancy as an index ot the perceptual fidelity of computer graphics. In *European Conference on Visual Perception 1998, Perception Vol 27 Supplement*, pages 25–25. Pion Ltd, Aug. 1998.

[103] 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.

[104] J. Tumblin and H. E. Rushmeier. Tone Reproduction for Realistic Images. *IEEE Computer Graphics and Applications*, 13(6):42–48, Nov. 1993.

[105] J. Tumblin and H. E. Rushmeier. Tone reproduction for realistic images. *IEEE Computer Graphics and Applications*, 13(6):42–48, Nov. 1993.

[106] E. Veach. *Robust Monte Carlo methods for lighting simulation*. Ph.D. thesis, Stanford University, 1997.

[107] C. Vedel and C. Puech. A testbed for adaptive subdivision in progressive radiosity. *Second Eurographics Workshop on Rendering (Barcelona, Spain)*, May 1991.

[108] V. Volevich, K. Myszkowski, A. Khodulev, and K. E.A. Using the Visible Differences Predictor to improve performance of progressive global illumination computations. *ACM Transactions on Graphics*, 19(3), Sepetmber 2000.

[109] V. Volevich, K. Myszkowski, A. Khodulev, and E. Kopylov. Perceptually-informed progressive global illumination solution. Technical Report TR-99-1-002, Department of Computer Science, Aizu University, Feb. 1999.

[110] B. Walter. *Density estimation techniques for global illumination*. Ph.D. thesis, Cornell University, 1998.

[111] B. Walter, P. Hubbard, P. Shirley, and D. Greenberg. Global illumination using local linear density estimation. *ACM Transactions on Graphics*, 16(3):217–259, 1997.

[112] 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.

[113] A. Watson. Temporal sensitivity. In *Handbook of Perception and Human Performance, Chapter 6*. John Wiley, New York, 1986.

[114] A. Watson. The Cortex transform: rapid computation of simulated neural images. *Comp. Vision Graphics and Image Processing*, 39:311–327, 1987.

[115] A. Watson. Toward a perceptual video quality metric. In *Human Vision and Electronic Imaging III*, pages 139–147. Proceedings of SPIE Vol. 3299, 1998.

[116] A. Watson, J. Hu, J. McGowan III, and J. Mulligan. Design and performance of a digital video quality metric. In *Human Vision and Electronic Imaging IV*, pages 168–174. Proceedings of SPIE Vol. 3644, 1999.

[117] A. Watt. *3D Computer Graphics*. Addison-Wesley, 1993. ISBN 0-201-63186-5.

[118] A. Wavefront. *Alias Wavefront User Manual*. Alias Wavefront, 1997.

[119] S. Westen, R. 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.

[120] J. Westerink and C. Teunissen. Perceived sharpness in moving images. pages 78–87. Proceedings of SPIE Vol. 1249, 1990.

[121] T. Whitted. An improved illumination model for shaded display. *Computer Graphics (Special SIGGRAPH '79 Issue)*, 13(3):1–14, Aug. 1979.

[122] H. Wilson. Psychophysical models of spatial vision and hyperacuity. In D. Regan, editor, *Spatial vision, Vol. 10, Vision and Visual Disfunction*, pages 179–206. Cambridge, MA: MIT Press, 1991.

[123] H. Wilson and D. Gelb. Modified line-element theory for spatial-frequency and width discrimination. *J. Opt. Soc. Am. A*, 1(1):124–131, 1984.

[124] S. Winkler. A perceptual distortion metric for digital color video. In *Human Vision and Electronic Imaging IV*, pages 175–184. Proceedings of SPIE Vol. 3644, 1999.

[125] Wyszecki and Stiles. *Colour Science: concepts and methods, quantitative data and formulae (2nd edition)*. New York: Wiley, 1986.

[126] E. Yeh, A. Kokaram, and N. Kingsbury. A perceptual distortion measure for edge-like artifacts in image sequences. In *Human Vision and Electronic Imaging III*, pages 160–172. Proceedings of SPIE Vol. 3299, 1998.

[127] C. Zetzsche and H. 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.

# Important Issues

Section by:

Scott Daly

Digital Video Department

Sharp Laboratories of America

1

---

## 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
    - Sarnoff (Lubin and Brill)
    - Efficiency Versions
  - Spatiotemporal (Motion)
    - Motion Animation Quality Metric

- Validation of metrics:
  - Modelling published psychophysical data
  - Testing with system-based test targets
  - Testing in actual applications

2

## Visual Model Design and Approaches

Visual Modeling utilizes 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 primarily 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 engineering efforts

3

## Types of Visual Models

- Mathematical, quantitative descriptions of visual response under varying parameters

Historical Examples:
- CIELAB standard lightness response (1976):

$L^* = 116(Y/Y_N)^{1/3} - 16$ ,             Y is luminance,             $Y_N$ is luminance of white point

- Contrast Sensitivity Function (CSF) = spatial frequency response   (Mannos & Sakrison '74)

$CSF(u,v) = 2.6*(0.0192 + 0.144*r^{1/2})*exp(-\{0.144*r^{1/2}\}^{1.1})$     (u, v = horizontal and vertical frequencies)

$r = (u^2 + v^2)^{1/2}$       (radial frequency)

- Image processing models of visual thresholds and appearance (simulations)

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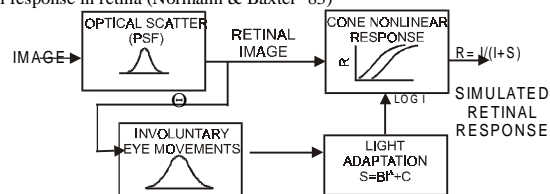


4

## 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 rather than 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

5

## 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

6
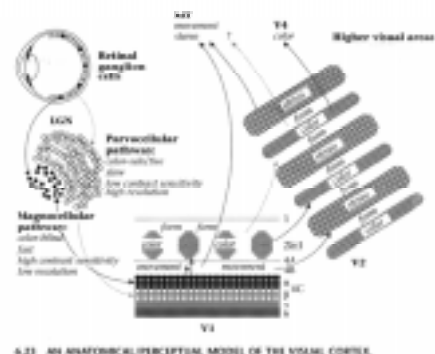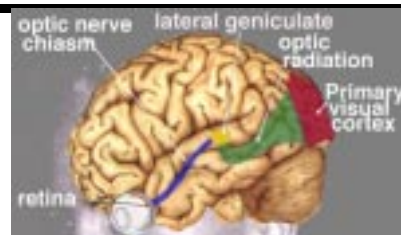
# Properties of Visual System: Luminance Nonlinearity

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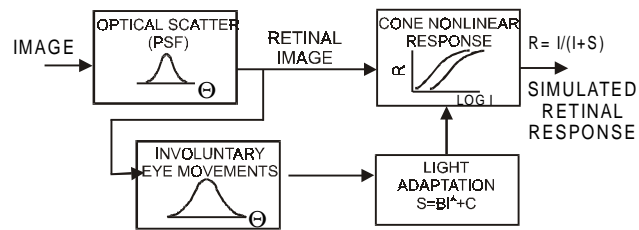


7

# 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 ( $L^{1/3}$ ~= $L^{1/2.4}$ )
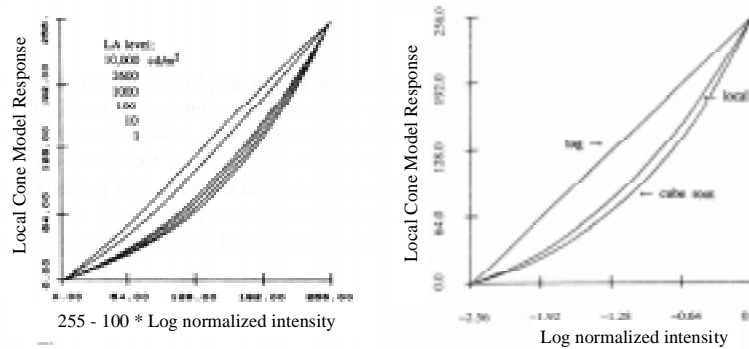


8

## 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



| Linear | Gamma-Corrected | Log |

9

## Properties of Visual System: Luminance Contrast

■ For AC signals contrast is used

■ Linear Amplitude of signal in luminance units does not match perception

■ Contrast of signal is much better match
   • Takes into account signal relative to its mean level
   • Michelson contrast:

$$C = (L_{MAX} - L_{MIN}) / (L_{MAX} + L_{MIN})$$

$$= (L_{MAX} - L_{MEAN}) / L_{MEAN}$$

■ Contrast behaves closer to logarithmic

■ Sensitivity, S, analogous to gain
   • slope of visual response
   $$S = 1 / C_T$$
   $C_T$ = Threshold Contrast



10

5
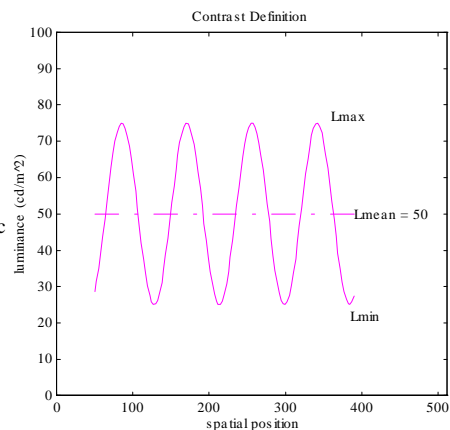
## Properties of Visual System: **Spatial Frequency**

- Spatial behavior constant with visual angle (degrees)
- Spatial frequencies specified in cycles/degree (cpd, cy/deg)
- Spatial frequency behavior described with CSF (contrast sensitivity function)
  - Similar to OTF of optics, MTF of electrical systems, but it is nonlinear and adaptive
  - Measured with psychophysics
- One of the most useful, and widely used properties of visual system
- CSF changes with light adaptation level
- But most practical applications are in range of top curve
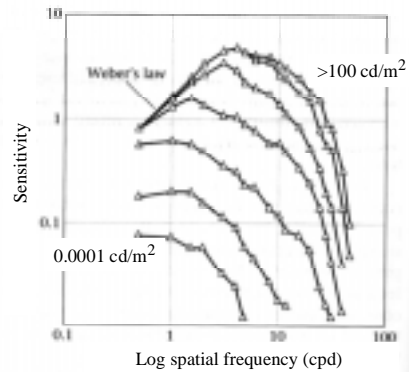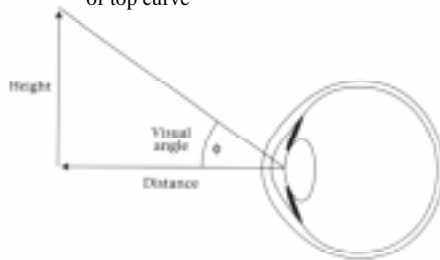


11

---

## Properties of Visual System: **Spatial Frequency**

- **Mapping** visual spatial frequencies to physical or digital frequencies
  - Physical frequencies, examples = cy/mm, dpi, etc.  (when display is known)
  - Digital frequencies = cy/pix, cy/radian, etc
- Since viewing distance used to relate degrees to object size, it is important in applying CSF
  - For physical frequencies, specify distance in physical units
  - For digital frequencies, specify distance in units of pixels (old way used multiples of picture heights)
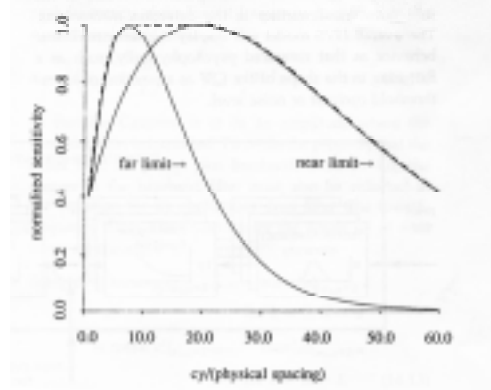


12

# Properties of Visual System: Spatial Frequency

- 2D frequencies important for images
- 2D CSF is <u>not</u> rotationally symmetric (isotropic)
- Lack of sensitivity near 45 degrees, called the oblique effect



2D Spatial CSF
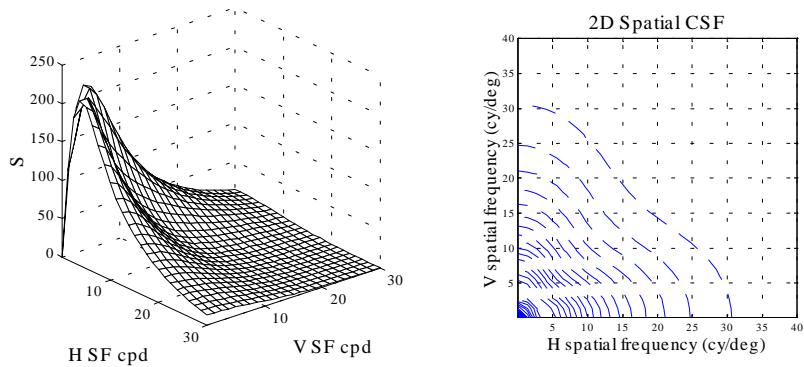
13

# Properties of Visual System: Spatio-Chromatic Frequency

- Color is captured in retina by LMS cones (Long, Middle, Short wavelengths ~= R,G,B cones)
- But converted by ganglion cells and LGN to **opponent color** representation
- L achromatic channel, R-G channel and B-Y channel (difference occurs in nonlinear domain)
  - CIELAB A* ~= R-G channel, B* ~= B-Y channel
- R-G and B-Y channels have no luminance (isoluminant)
- R-G and B-Y spatial frequency bandwidths and sensitivities are lower than L CSF



Luminance and Chromatic CSFs
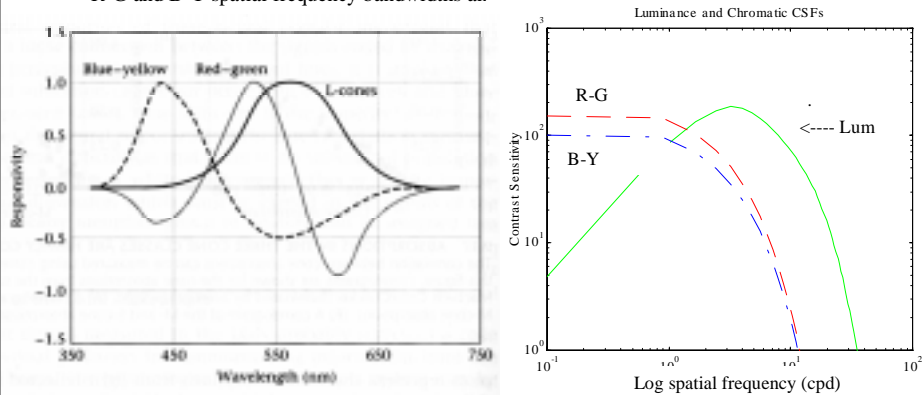
14

## Properties of Visual System: **Local Spatial Content**

- Local spatial content of image affects local visibility within image
- Common engineering knowledge:
    - Busy or complex images hide compression or other distortions
    - Noisy images hide distortions
- But effect is complex:
    - High frequencies don't hide low frequencies
    - Easy to see orthogonal artifacts on edges, such as aliasing artifacts
    - Luminance can sometimes hide color structure, but sometimes can help see it
- Quantitative description referred to as "Masking"

- This property arises from the structure of the visual cortex, shown to right
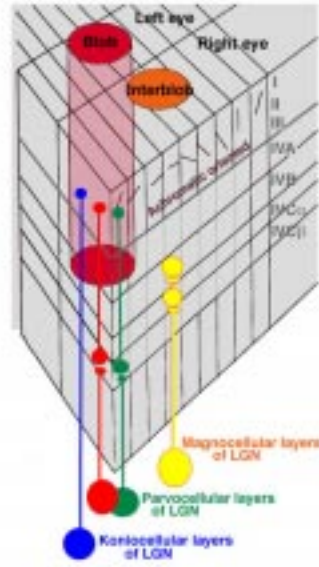


15

## 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



MASK   SIGNAL F = MASK F   SIGNAL + MASK

MASK   SIGNAL F ⊥ MASK F   SIGNAL + MASK

MASK   SIGNAL F = 2 x MASK F   SIGNAL + MASK
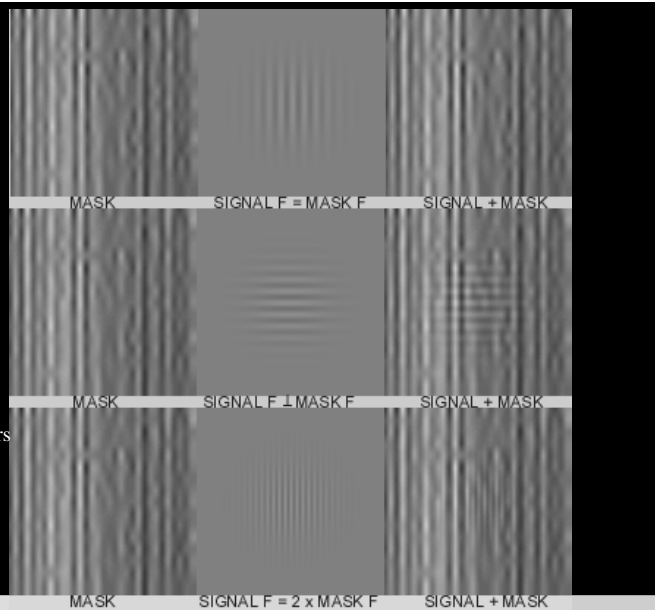
16

## 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:
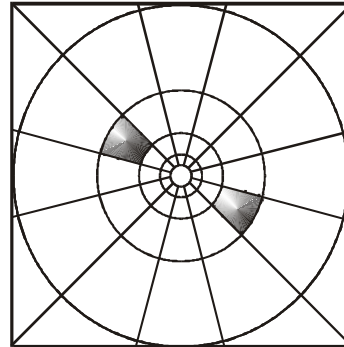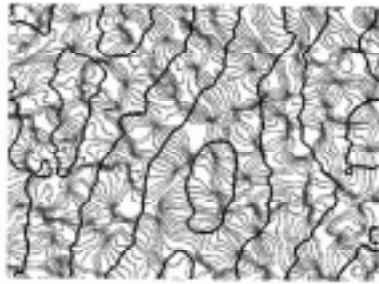


■ Electrophysiology measures of common orientation response in visual cortex
- gray lines = common orientation
- black lines = orientation boundary

Cycles/Pixel:   0   0.125  0.25   0.5
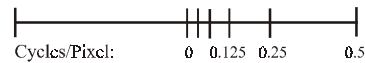
■ Key Features of frequency "channels": Dyadic radial frequencies, orientation selectivity, baseband, space frequency localization

---

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

■ Channels overlap, and have the Fourier symmetry of real signals
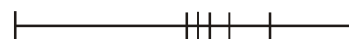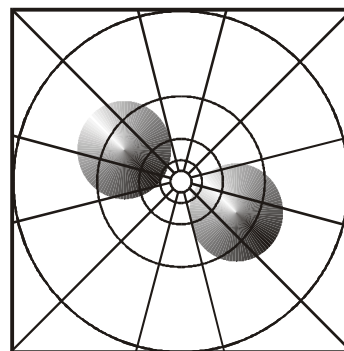
## Properties of Visual System: Frequency Structure

Wavelet Compression (JPEG2000)

- 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
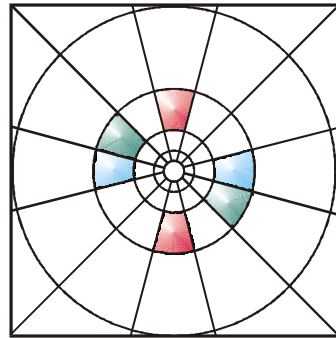


Cycles/Pixel:    0   0.125  0.25    0.5          Cycles/Pixel:    0   0.125  0.25    0.5

- Visual System Frequency Decomposition        Wavelet Algorithm Frequency Decomposition
  (Cortex Transform)                            (JPEG2000)

---

## 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:              $R = f(C)$
  - Threshold:             $T(C) = k / \Delta R = k / f'(C)$
  - Transducer function:   $f(C) = \int 1 / T(C)$
- Transducer functions derived from sine masking and noise masking data are shown to the right
  - Note that plot is now with linear axes
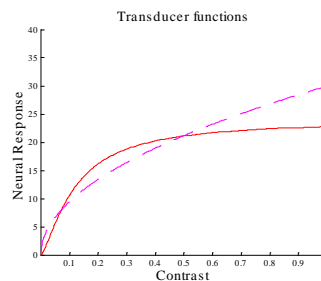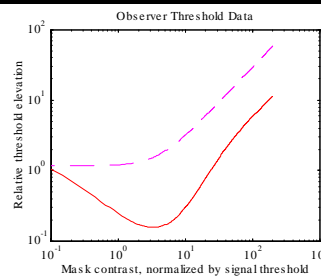
## Properties of Visual System: Temporal Frequency

- 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
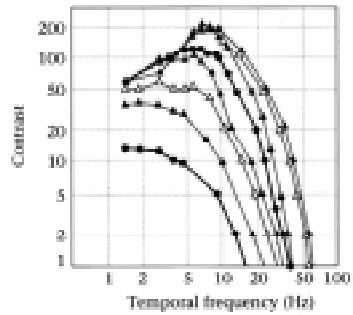
## Properties of Visual System: Spatiotemporal and Motion
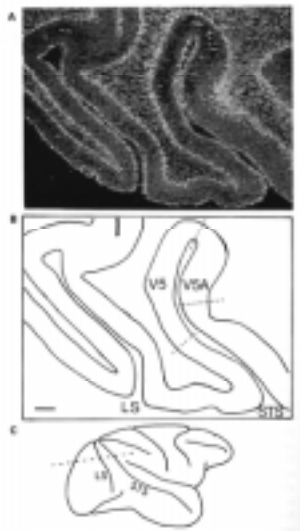


- Motion occurs in area V5 of visual cortex
- Most psychophysical data measures spatio-temporal CSF
- Test signal is product of spatial and temporal frequency
  - Standing Wave
- Data shows max visible temporal frequency near 50 cy/sec
  - Thus 60 fps usually causes no visible flicker
  - Movie film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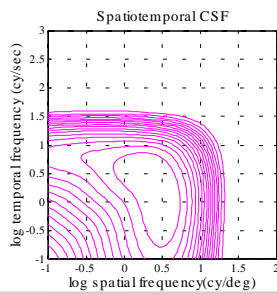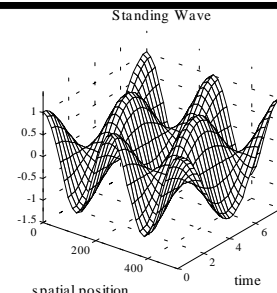
## Properties of Visual System: Motion: Eye Movements

- Velocity as important as flicker
- Smooth tracking eye movements can reduce image velocity on the retina
- Smooth tracking data and model shown below
- $\omega = v\rho$
  - $\omega$ =temporal frequency $v$=velocity $\rho$=spatial frequency
- Rotation into spatiotemporal CSF including effects of eye movements
  - Can be used to assess smoothness of motion



Standing wave = 2 opposing travelling waves



Smooth Pursuit Eye Tracking

Eye Movement Model Spatiotemporal CSF
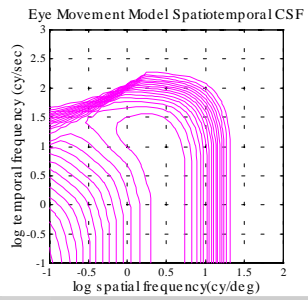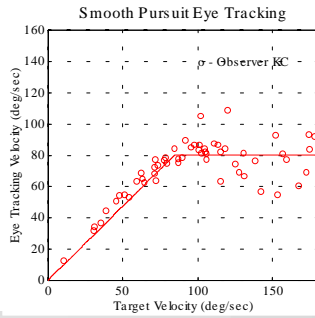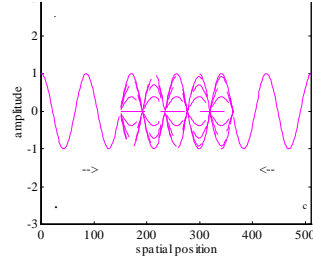
23

---

## Properties of Visual System: Global Color

- Field of Color Reproduction, many standards

24

## 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





CSFs for Eccentricity Range 0-25 deg
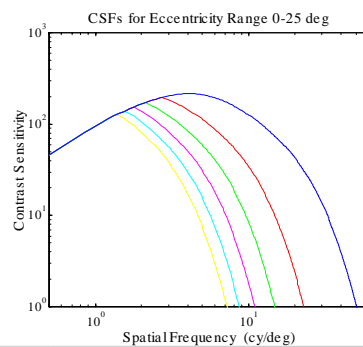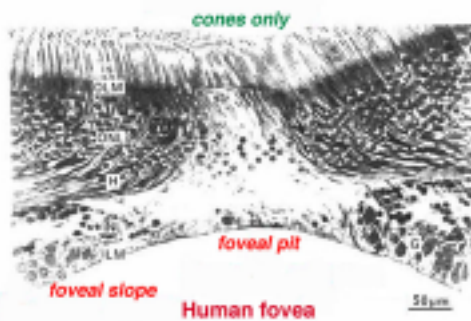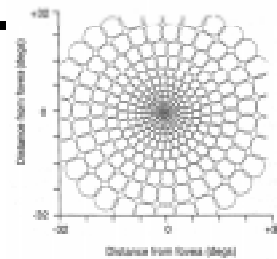
25

---

## 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)
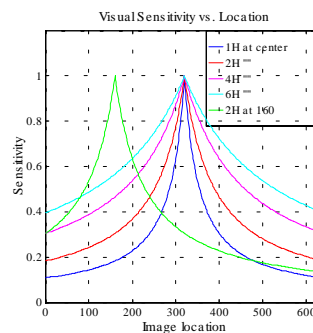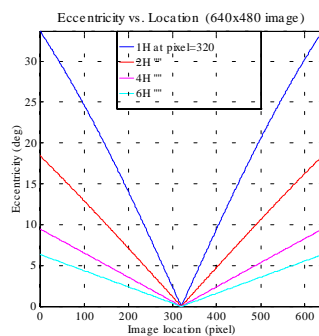




26

## State of the Art Metrics
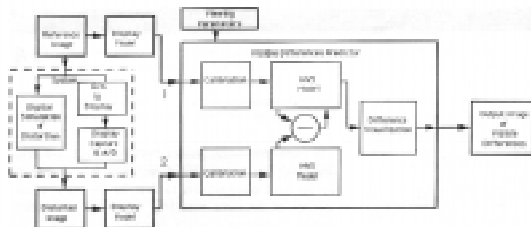


- Spatial and Chromatic:
    - VDP
    - Sarnoff model
    - Efficiency Versions
        - Bradley's wavelet VDP for compression
        - Mark Bolin
        - USC Hadamard model
- Spatiotemporal:
    - Motion Animation Quality Metric

## VDP = Visible Differences Predictor

- Models basic visual effects of amplitude variations, spatial variations, and signal dependent variations s
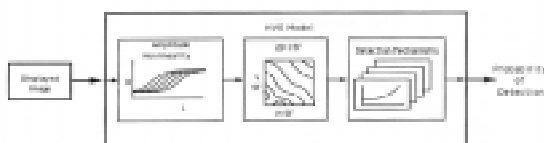


- Amplitude variations: modeled as local cone model described previously, entire image is processed through as point-nonlinearity
- Spatial Variations: the 2D CSF is modeled as a global filter, filter described previously
- Signal Dependent variations (masking) described in block diagram below:
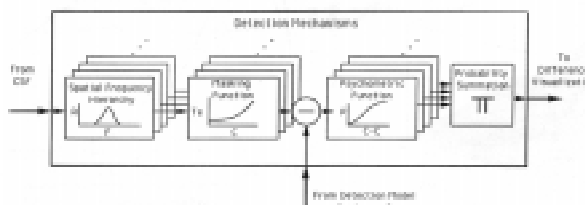
## VDP = Visible Differences Predictor: Channel Design

- Cascade of Isotropic radial filters (DOM filters) with fan filters
- Radial filters shown to right , sum of all filters = 1.0 for reversibility in applications
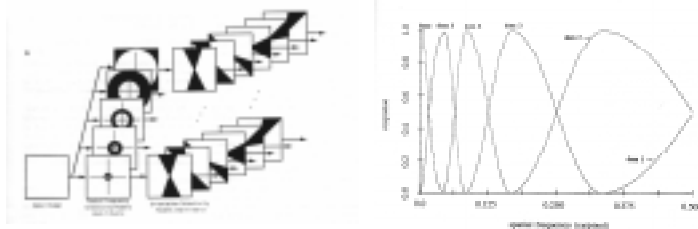


- Resulting dissection of frequency plane:
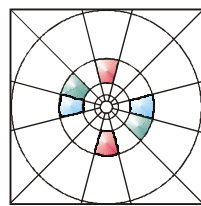


29

---

## VDP = Visible Differences Predictor

- Key effect of masking, while caused by transducer function, described previously, is actually modeled as separate image processing pathway created in addition to signal pathway
- Physiologically unrealistic, but makes easier to test and fine tine to published experiments,
- Masking modeled with possibility to include learning effects, resulting in decreasing masking slope
- 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
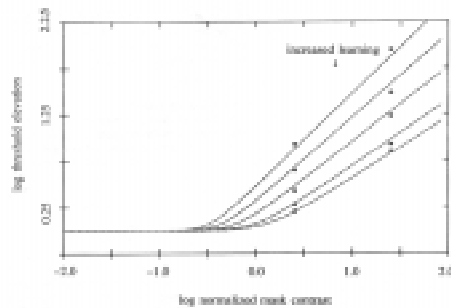


30

## Sarnoff Model

■ Spatial portion originally developed by Jeff Lubin, Color added by Michael Brill
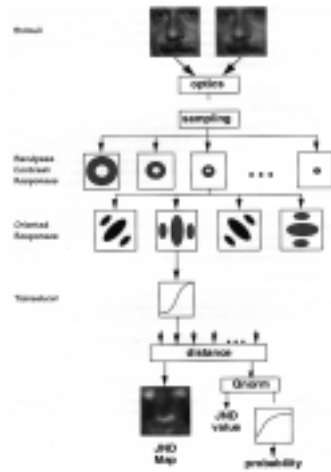
## Sarnoff Model

■ Visualization:
■ JNDs
■ no contrast polarity,
■ therefore no spatial gauge of appearanc
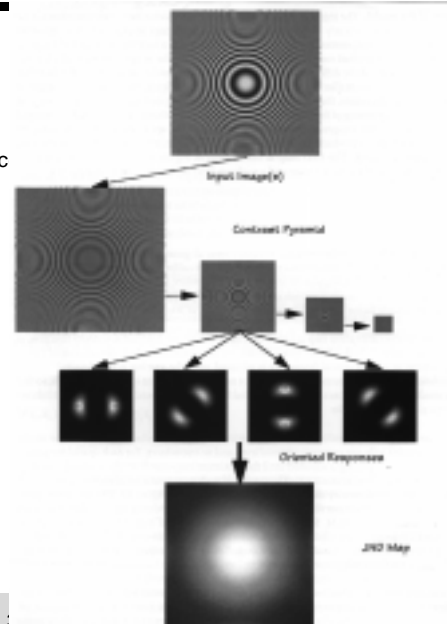
## Key differences between VDP and Sarnoff Models

- See paper by Li, Meyer, Bolin
- 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 supratheshold, 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 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 area)
  - Sarnoff used JND scale, magnitude only, can do suprathreshold but shapes of distortions not simulated

33

## Key Visual Model detail differences

- VDP model cone nonlinearity, Sarnoff models local contrast
- VDP has no spatial pooling other than channel filter support
- Sarnoff has front-end optics to affect LPF before any neural amplitude nonlinearity (correct physiologically), VDP consolidates similar visual effects into basic stages
- Sarnoff has Hilbert pair for each channel (a sine and cosine phase), VDP has all phases combined in each channel (but models phase uncertainty separately)
- more …..

34

## Efficiency Versions

- Bradley's wavelet VDP for compression
- Mark Bolin's Hadamard-based model for computer graphics
- USC Hadamard model for signal processing

## Moving Images, Spatiotemporal channels

- Temporal channels as well as spatial
    - Sustained and Transient
    - Some evidence for 3rd temporal channel at low SFs
    - How are spatio and temporal channels configured:
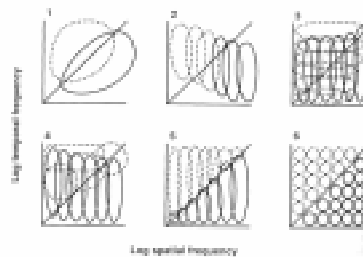


    - Line with slope = 1, 1 deg/sec
    - Solid lines = nondirectional, dashed lines = directional
    - Version 4 is consistent with most amount of data = velocity channels + flicker channel
- Spatiotemporal separability
- Spatiovelocity separability

## Spatiotemporal : Sarnoff Model

■ Spatial model is basically doubled, each one having different temporal prefilter
■ No accounting for eye movements
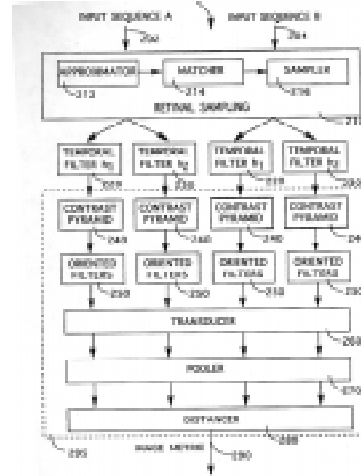
## Motion Animation Quality Metric

■ Spatiovelocity Approach
■ Allows for accounting for eye-movements
■ Details and illustrations by Karol here

## Validation of Metrics

- Testing of the visual model is best to proceed from final detection stage toward front-end processing (reverse visual pathway)
- Example of testing given for VDP (achromatic and still images)
- Fit of model to published psychophysical data

39

## Validation of Metrics

- Each of the following tests will explain what is being tested, and I have slide illustrations of the exact tests and the VDP results in image form showing where detection actually occurs (as indicative of performance of most current visual models )
- Tests psychometric function, and it the ability of all preceding components to not adversely affect its modelling
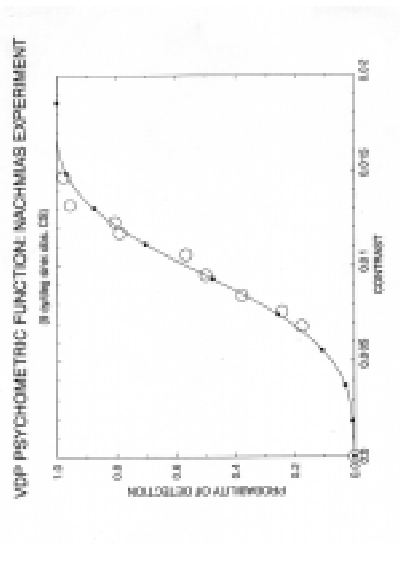


40

## Validation of Metrics

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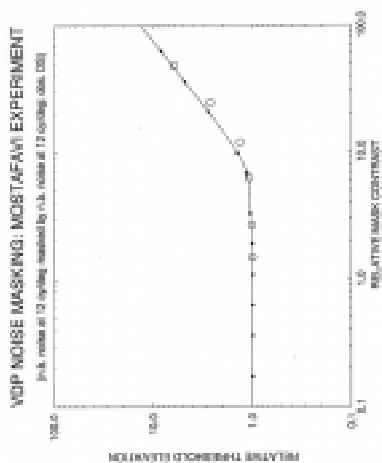


41

## Validation of Metrics

- Tests ability model contrast masking where signal and mask are both narrowband and phase coherent signals
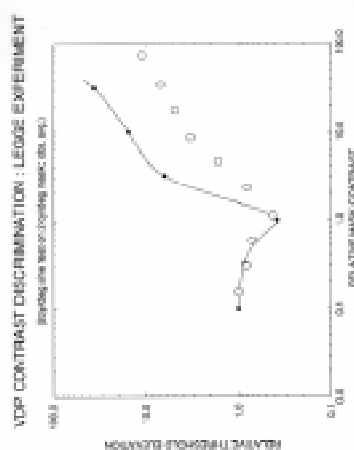


42

## Validation of Metrics

- Test radial frequency channels
- Also tests interaction of psychometric function, probability summation, and CSF effects
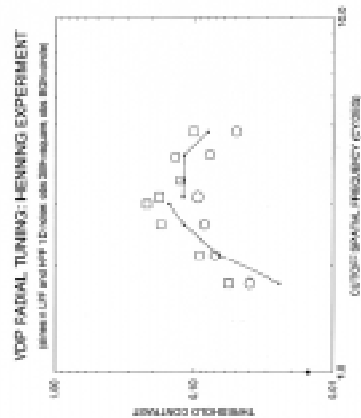
## Validation of Metrics

- Tests shape of fan filters,
- Also interaction of radial fan cascade to create cortex filter
- as well as off-frequency detection, and probability summation
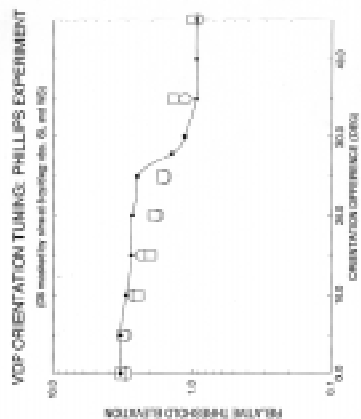
## Validation of Metrics

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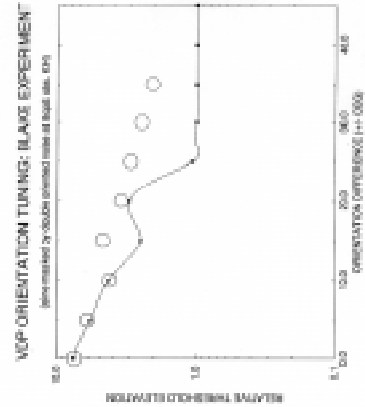


45

## Validation of Metrics

- Tests CSF and its light adaptation capabilities
- Also interaction of CSF with all post-CSF model components
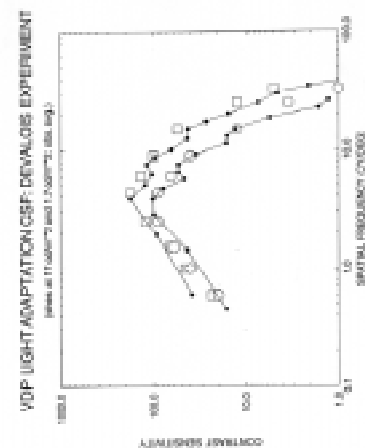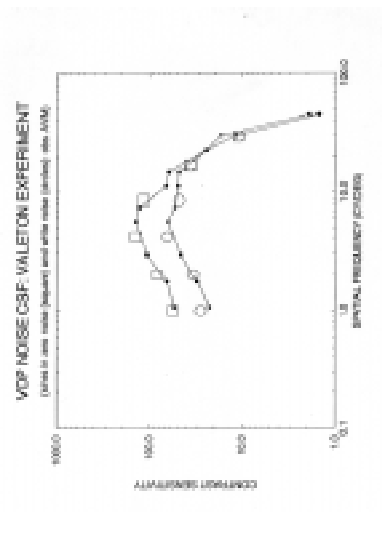


46

## Validation of Metrics

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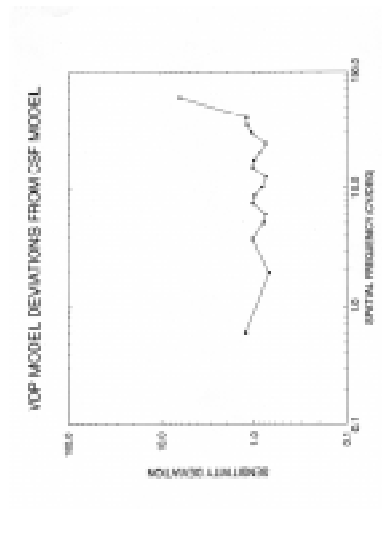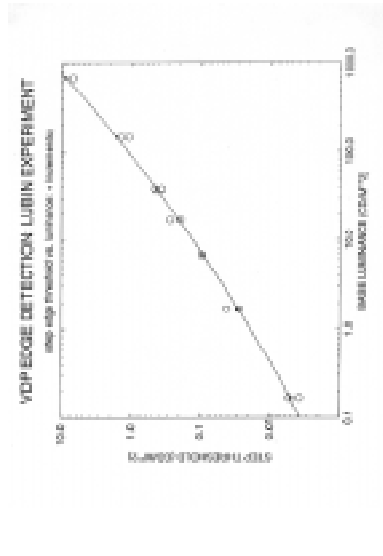


47

## Validation of Metrics

- Tests ability to predict the effects of noise on the CSF
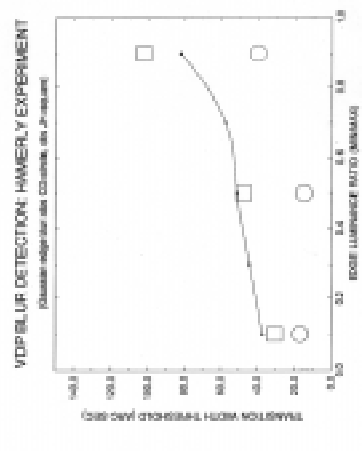- Tests masking in conjunction with CSF



48

## Validation of Metrics

- Tests amplitude nonlinearity in conjunction with rest of model
- Experiment is the detection of sharp luminance edge at different gray levels
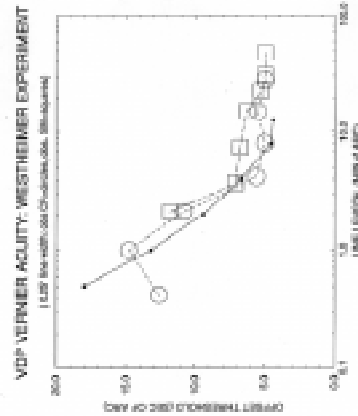


49

## Validation of Metrics

- Tests entire model
- Practical test of ability to see blur as a function of contrast



50

## Validation of Metrics

- Vernier acuity tests entire model
- Practical importance to visibility of jaggies , form various aliasing/ antialiasing tradeoffs.



51

## Validation of Metrics

- Testing of Models in actual image applications
- Demos in 35mm slide and S-VHS formats

52