

High Dynamic Range Imaging & Image-based Lighting

SIGGRAPH 2008
Half-day Class
Los Angeles, CA

Greg Ward
Consultant
Dolby Canada
www.anywhere.com/gward

Erik Reinhard
Lecturer
Department of Computer Science
University of Bristol
University of Central Florida
www.cs.bris.ac.uk/~reinhard

Paul Debevec
Research Associate Professor
Institute for Creative Technologies
University of Southern California
www.debevec.org

High Dynamic Range Imaging & Image-based Lighting

SIGGRAPH 2008 Half-day Class

This class outlines recent advances in high dynamic range imaging (HDRI) - from capture to image-based lighting to display. In a hands-on approach, we show how HDR images and video can be captured, the file formats available to store them, and the algorithms required to prepare them for display on low dynamic range displays. The trade-offs at each step are assessed allowing attendees to make informed choices about data capture techniques, file formats and tone reproduction operators. In addition, the latest developments in image-based lighting will be presented.

Prerequisites

None. This course is intended for students, researchers, and industrial developers in digital photography, computer graphics rendering, real-time photoreal graphics, game design and visual effects production (esp. rendering and compositing).

Syllabus

1. Introduction & Perspective (Ward, 15 minutes)
2. HDR Image Capture & Representation (Ward, 45 minutes)
 - a. Encoding techniques & color
 - b. HDR image & video formats
 - c. Backwards-compatible HDRI compression
 - d. HDR capture refinements
 - e. HDR cameras
3. HDR Tone-mapping & Display (Reinhard, 60 minutes)
 - a. Basic operations
 - b. Global vs. Local TMOs
 - c. TMO comparisons
 - d. Forward + Reverse Sigmoid TMO
4. Image-based Lighting (Debevec, 60 minutes)
 - a. Practical considerations for taking HDR images
 - b. Shooting light probes: acquiring omnidirectional HDRI maps
 - c. Shooting light probes with direct sunlight
 - d. Classical image-based lighting
 - e. Converting light probes to constellations of light sources
 - f. Importance sampling techniques
 - g. Useful IBL approximations
 - h. Image-based relighting
5. Questions (All, 30 minutes)

Instructor Information

Greg Ward is a pioneer in the HDR space, having developed the first widely-used high dynamic range image file format in 1986 as part of the RADIANCE lighting simulation system. Since then, he has developed the LogLuv TIFF HDR image format, the JPEG-HDR format, and authored the application Photosphere, an HDR image builder and browsing program. More recently, he has been involved with Dolby Canada's HDR display developments, which employ dual modulators to show colors 30 times as bright and ten times as dark as conventional monitors. Working in the computer graphics research community for over 20 years, he has developed rendering algorithms, reflectance models and measurement systems, tone reproduction operators, HDR image processing techniques, and photo printer calibration methods. His past employers include the Lawrence Berkeley National Laboratory, EPFL Switzerland, SGI, Shutterfly, Exponent, and BrightSide Technologies. Greg holds a bachelor's in Physics from UC Berkeley and a master's in Computer Science from SF State University. He is currently working as a consultant in Albany, California.

Erik Reinhard is lecturer at the University of Bristol and holds a courtesy appointment at the University of Central Florida. He has a B.S. and a TWAIO diploma in computer science from Delft University of Technology and a Ph.D. in computer science from the University of Bristol. He was a post-doctoral researcher at the University of Utah. He co-authored the first book on High Dynamic Range Imaging (Morgan Kaufmann Publishers, 2005). He is founder and co-editor-in-chief of the journal ACM Transactions on Applied Perception, and guest editor of a special issue on Parallel Graphics and Visualisation for the journal Parallel Computing (March 2003), and a special issue on High Dynamic Range Imaging in the Journal of Visual Communication and Image Representation. He is also co-editor of Practical Parallel Rendering (A K Peters, 2002). His current interests are in visual perception and its application to computer graphics problems such as tone reproduction and color correction.

Paul Debevec is the associate director of graphics research at USC's Institute for Creative Technologies. Debevec's Ph.D. thesis (UC Berkeley, 1996) presented Façade, an image-based modeling and rendering system for creating photoreal architectural models from photographs. Using Facade he led the creation of virtual cinematography of the Berkeley campus for his 1997 film The Campanile Movie whose techniques were used to create virtual backgrounds in the 1999 film The Matrix. Subsequently, Debevec developed techniques for illuminating computer-generated scenes with real-world lighting captured through high dynamic range photography, demonstrating new image-based lighting techniques in his films Rendering with Natural Light (1998), Fiat Lux (1999), and The Parthenon (2004); he also led the design of HDR Shop, the first widely-used high dynamic range image editing program. At USC ICT, Debevec has led the development of a series of Light Stage devices for capturing and simulating how objects and people reflect light, recently used to create realistic digital actors in films such as Spider Man 2 and Superman Returns. He is the recipient of ACM SIGGRAPH's first Significant New Researcher Award and a co-author of the 2005 book *High Dynamic Range Imaging* from Morgan Kaufmann.

HDR Imaging & Image-based Lighting

Greg Ward

Dolby Canada

Erik Reinhard

Bristol University

Paul Debevec

USC Institute for Creative Technologies

HDR Imaging History

- Negative photography always was MDR
- Earliest “exponent” format in Utah Raster Toolkit, and supported by *Rayshade*
 - little-known -e option (Craig Kolb & Rod Bogart)
- *Radiance* incorporated similar RGBE format
 - includes image-processing & conversion tools

Important Milestones in HDRI

- *Radiance* first public release in 1987
- Debevec & Malik's HDR capture in 1997
 - Improvement on 1995 method of Mann
- Debevec's image-based lighting paper 1998
- HDR tone-mapping papers of 2002
 - Durand & Dorsey, Fattal et al., Reinhard et al.
- Seetzen et al. HDR Display Systems 2004

Where HDRI Is Today

- Widespread use in special effects industry
 - Image-based lighting is *the* standard technique
- HDR photography is still in its infancy
 - Flickr and Photomatix -- someone help these guys!
- HDR video -- what's that?
- HDR displays in labs and at trade shows

Where HDRI Is Bound to Go

- Theater and Home Entertainment
 - Like a picture window rather than a picture
- True HDR Photography: “film squared”
- HDR Video -- good-bye exposure problems
- Surround-light with your surround-sound?

Course Overview

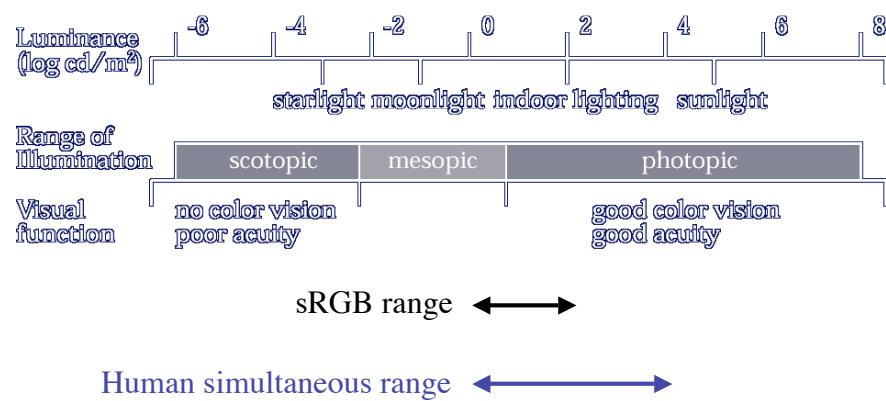
1. Introduction & Perspective
Greg Ward
2. HDR Image Capture & Representation
Greg Ward
3. HDR Tone-mapping & Display
Erik Reinhard
4. Image-based Lighting
Paul Debevec
5. Questions

HDR Image Capture and Representation

Greg Ward
Dolby Canada

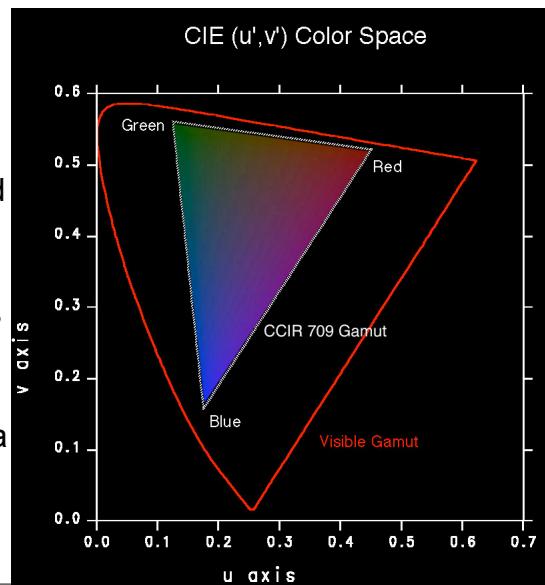
Dynamic Range

From Ferwerda et al, SIGGRAPH '96

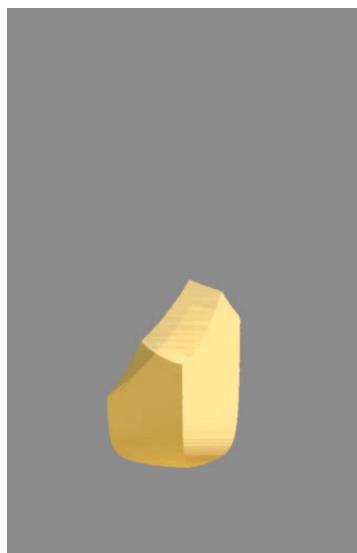


CCIR-709 Color Space

- Human visible gamut is much larger than standard display's
- Saturated blues, greens, and purples are lost in sRGB
- Many HDR image formats also cover a larger color gamut



A Gamut Is a Volume!



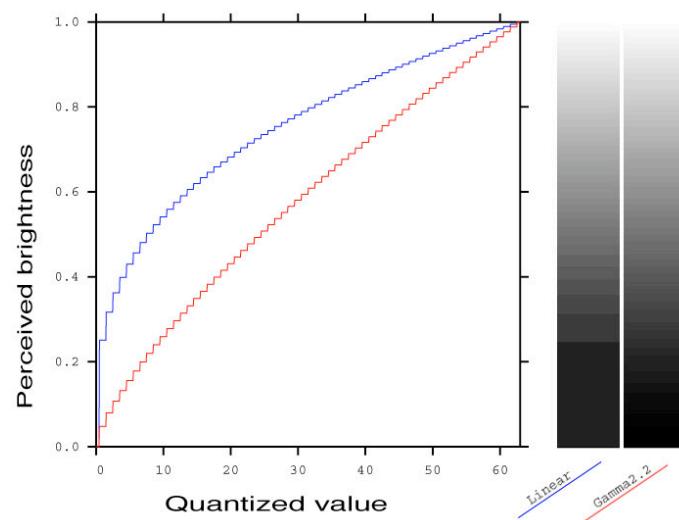
- HDR can represent **brighter** colors
- This *delays* saturation near white
- Result is larger color gamut

← Comparison of standard LCD display to BrightSide HDR display

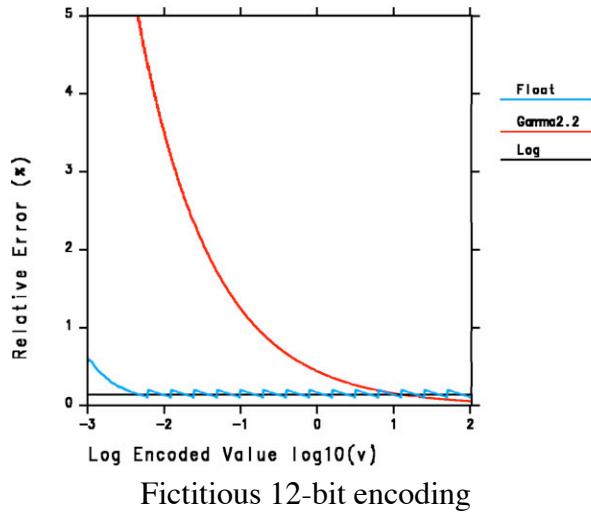
Value Encoding Methods

- Linear quantization
- Gamma function (e.g., CRT curve)
- Logarithmic encoding
- Floating point
- Perceptual

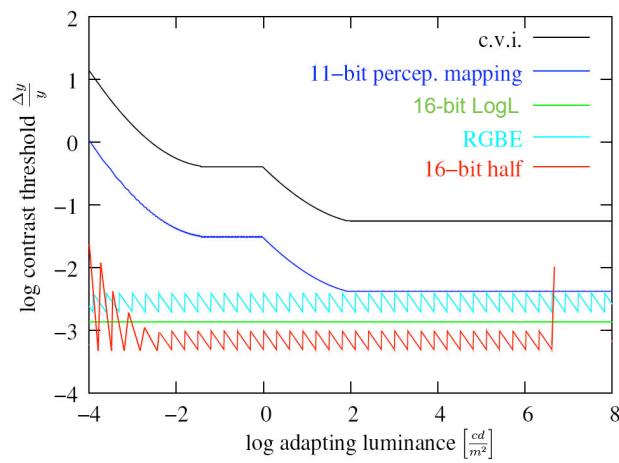
Linear vs. Gamma Encoding



Logarithmic vs. Floating Point



Perceptual Encoding



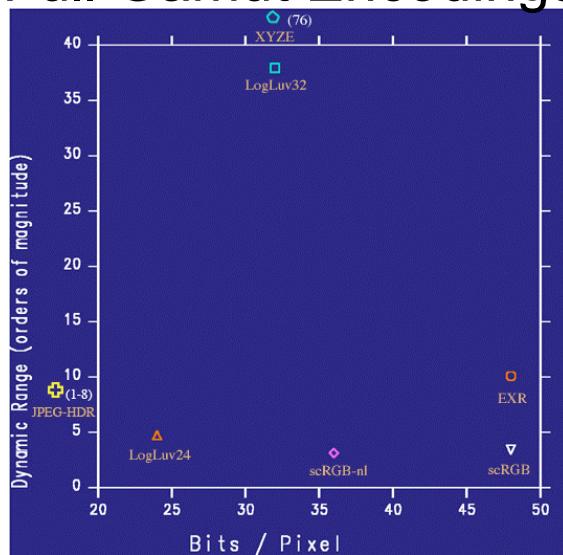
HDR Image & Video Formats

- Available high dynamic-range formats:
 - *Radiance* 32-bit RGBE and XYZE pictures
 - TIFF 48-bit integer and 96-bit float formats
 - SGI 24-bit and 32-bit LogLuv TIFF
 - ILM OpenEXR format
 - BrightSide JPEG-HDR format
- Proposals and extensions:
 - HDR extensions to MPEG from MPI [[Mantiuk et al. '04, '06](#)]
 - HDR extensions to JPEG 2000 from UFL [[Xu et al. 2005](#)]
 - HDR texture compression (Lunds Univ. & Nokia papers)

Encoding Comparison Chart

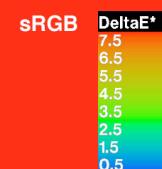
Encoding	Bits / pixel	Dynamic Range	Quant. Step	Covers Gamut
sRGB	24	$1:10^{1.6}$	Variable	No
Radiance RGBE	32	$1:10^{76}$	1%	No
Radiance XYZE	"	"	"	Yes
LogLuv 24	24	$1:10^{4.8}$	1.1%	Yes
LogLuv 32	32	$1:10^{38}$	0.3%	Yes
OpenEXR	48	$1:10^{10.7}$	0.1%	Yes
JPEG-HDR	1-7	$1:10^{9.5}$	Variable	Can

Full Gamut Encodings



HDR Acid Test Image

1:10⁸ dynamic range, covering visible gamut



Visible error for 24-bit/pixel sRGB encoding

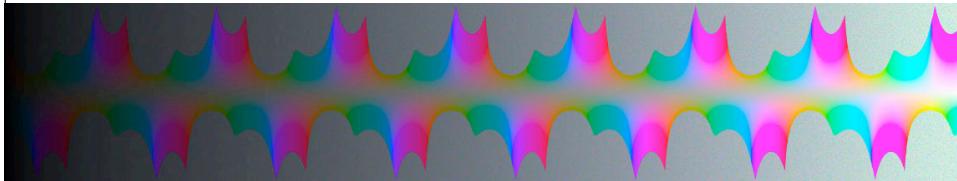
IEEE 96-bit TIFF

- Most accurate representation
- Support (with compression) in Photoshop CS2
- Uncompressed files are enormous
 - 32-bit IEEE floats look like random bits

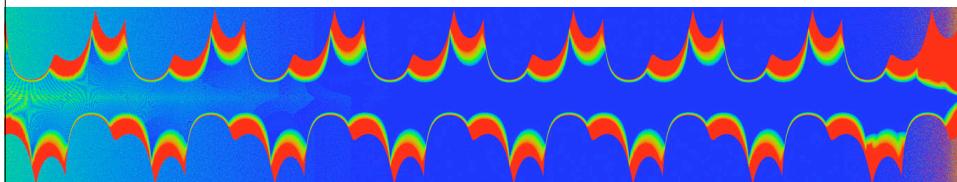
16-bit/sample TIFF (RGB48)

- Supported by Photoshop and TIFF library
- 16 bits each of log red, green, and blue
- 5.4 orders of magnitude in < 1% steps
- LZW lossless compression available
- Does not cover visible gamut
- Most applications think of max. value as “white”

48-bit RGB TIFF Accuracy



1:10⁸ dynamic range, covering visible gamut

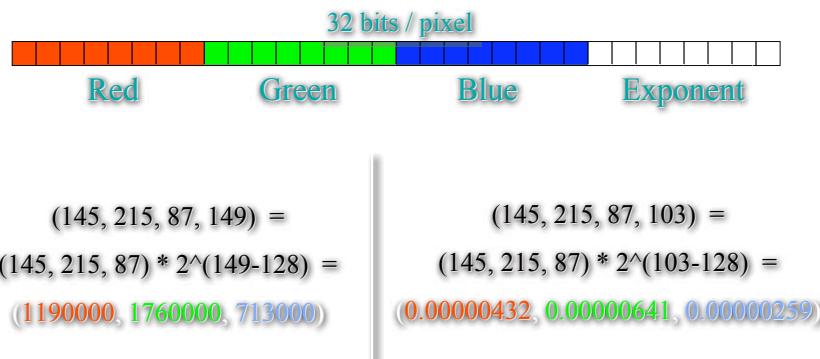


Visible error for 48-bit/pixel RGB encoding ($\gamma = 2.2$)

Radiance RGBE and XYZE

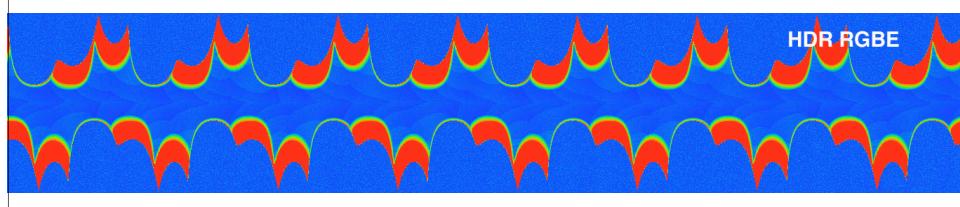
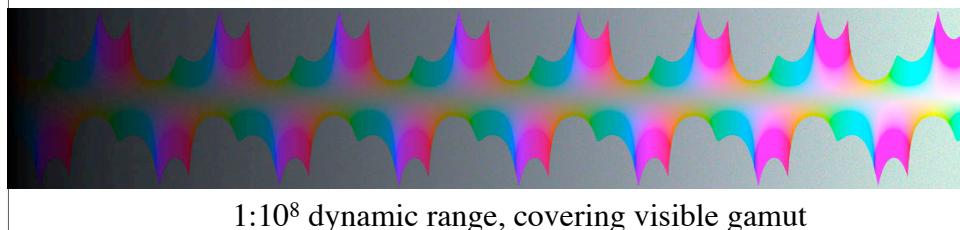
- Simple format with free source code
- 8 bits each for 3 mantissas and 1 exponent
- 76 orders of magnitude in 1% steps
- Run-length encoding (20% avg. compr.)
- **RGBE format does not cover visible gamut**
- Dynamic range at expense of accuracy
- Color quantization not perceptually uniform

Radiance Format (.pic, .hdr)

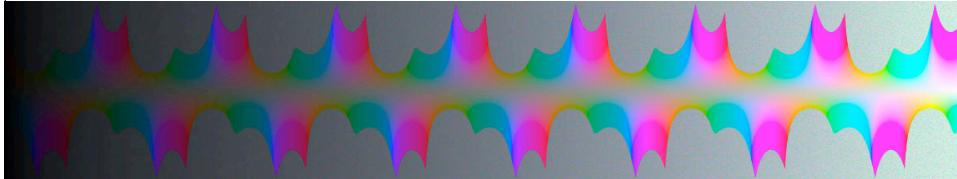


Ward, Greg. "Real Pixels," in Graphics Gems IV, James Arvo ed., Academic Press, 1994

Radiance RGBE Accuracy



Radiance XYZE Accuracy



1:10⁸ dynamic range, covering visible gamut

HDR XYZE

Visible error for 32-bit/pixel XYZE encoding

SGI 24-bit LogLuv TIFF Codec

- Implemented in Leffler's TIFF library
- 10-bit LogL + 14-bit CIE (u',v') lookup
- 4.8 orders of magnitude in 1.1% steps
- Just covers visible gamut and range
- Amenable to tone-mapping as look-up
- Dynamic range is less than we would like
- No compression

24-bit LogLuv Pixel

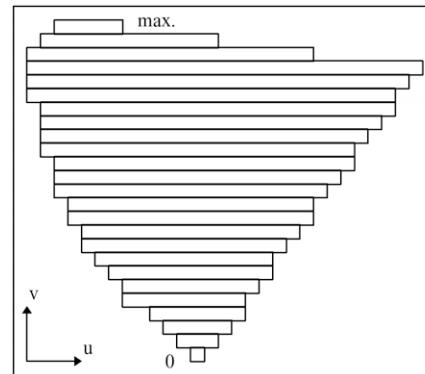
From Larson, CIC '98



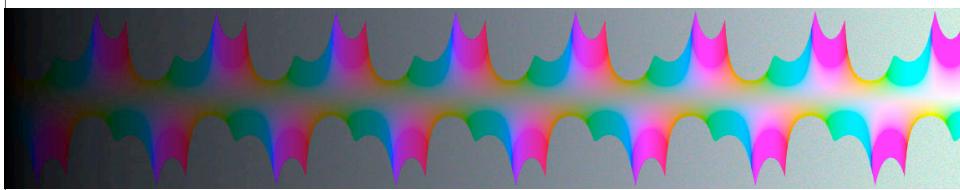
$$L_e = \lfloor 64(\log_2 L + 12) \rfloor$$

$$u' = \frac{4x}{-2x+12y+3}$$

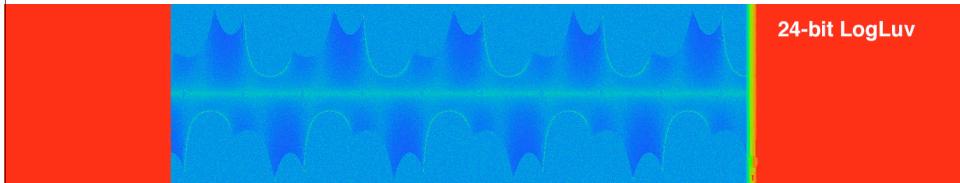
$$v' = \frac{9y}{-2x+12y+3}$$



TIFF LogLuv24 Accuracy



1:10⁸ dynamic range, covering visible gamut



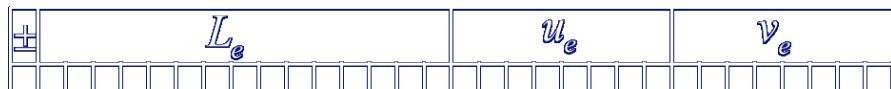
Visible error for 24-bit/pixel LogLuv encoding

SGI 32-bit LogLuv TIFF Codec

- Implemented in Leffler's TIFF library
- 16-bit LogL + 8 bits each for CIE (u',v')
- 38 orders of magnitude in 0.3% steps
- Run-length encoding (30% avg. compr.)
- Allows negative luminance value
- Amenable to tone-mapping as look-up

32-bit LogLuv Pixel

From Larson, JGT '98



$$L_e = \lfloor 256(\log_2 L + 64) \rfloor$$

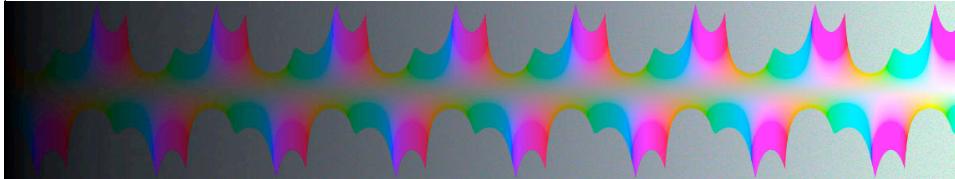
$$u_e = \lfloor 410u' \rfloor$$

$$u' = \frac{4x}{-2x+12y+3}$$

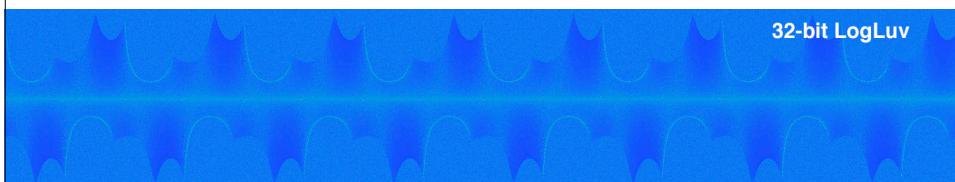
$$v_e = \lfloor 410v' \rfloor$$

$$v' = \frac{9y}{-2x+12y+3}$$

TIFF LogLuv32 Accuracy



1:10⁸ dynamic range, covering visible gamut



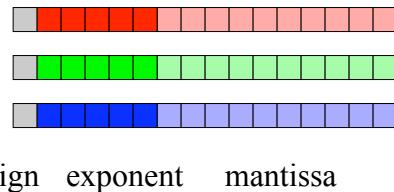
Visible error for 32-bit/pixel LogLuv encoding

ILM OpenEXR Format

- 16-bit/primary floating point (sign-e5-m10)
- 9.6 orders of magnitude in 0.1% steps
- Additional order of magnitude near black
- Wavelet compression of about 40%
- Negative colors and full gamut RGB
- Alpha and multichannel support
- Open Source I/O library released Fall 2002
- Slow to read and write

ILM's OpenEXR (.exr)

6 bytes per pixel, 2 for each channel, compressed



- Several lossless compression options, 2:1 typical
- Compatible with the “half” datatype in NVidia’s Cg
- Supported natively on GeForce FX and Quadro FX
- Available at www.openexr.com

OpenEXR Accuracy

1:10⁸ dynamic range, covering visible gamut

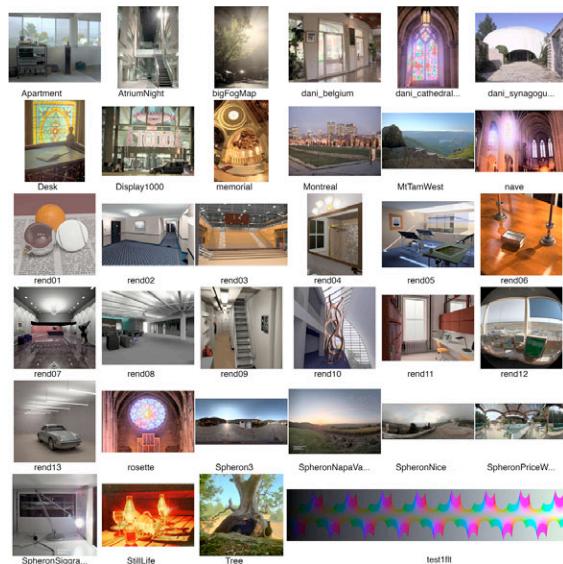
EXR half

Visible error for 48-bit/pixel EXR half encoding

Dolby Canada's JPEG-HDR Format

- Backwards-compatible JPEG extension for high dynamic range images
- Very compact: 1/10th size of other formats
- Naïve software displays tone-mapped sRGB
 - Different tone-mappings possible
- Desaturation can encompass visible gamut
- Lossy encoding so repeated read/write degrades
- Expensive (three pass) write process

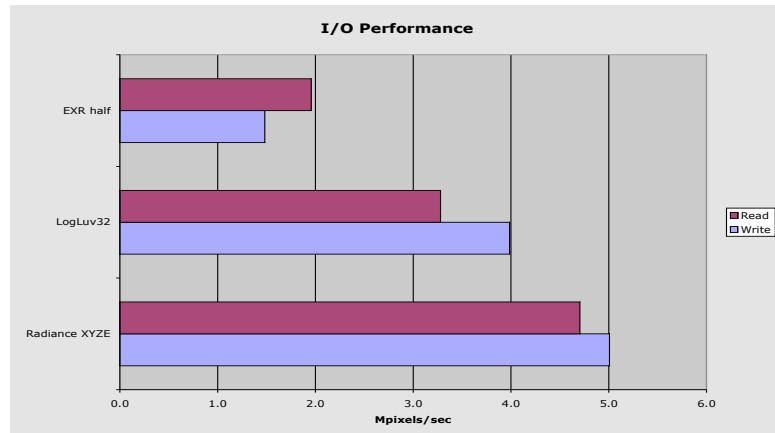
HDR Format Performance



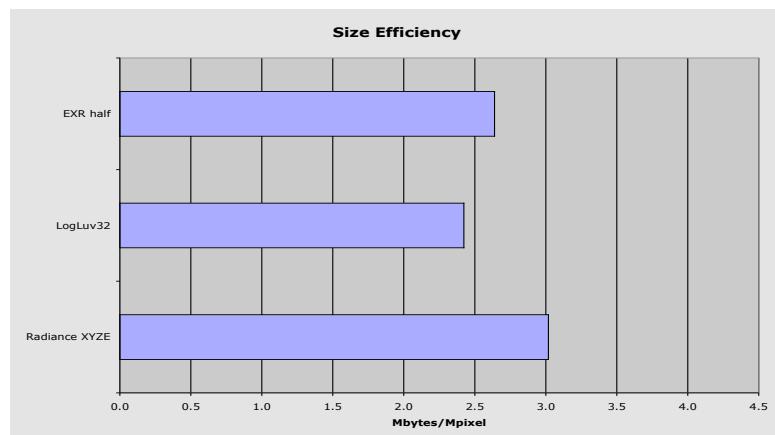
Test Set
(34 HDR images)

Measured i/o
speed and image
size

Read & Write Speed



Total Image Size



File Size & HDR Adoption

- Compression can match size of JPEG images + 20%
- Rationale for “lossy” HDR:
 - Lossy encodings are all about perception
 - Lossy HDR supports display to the limits of human vision
 - Required for digital photography & web applications
 - Mantiuk et al.’s [MPEG-4 extension \(SIGGRAPH 2004\)](#)
 - Xu et al.’s [JPEG-2000 extension \(CG+A 2005\)](#)
 - Microsoft’s HD Photo

File Size & HDR Adoption

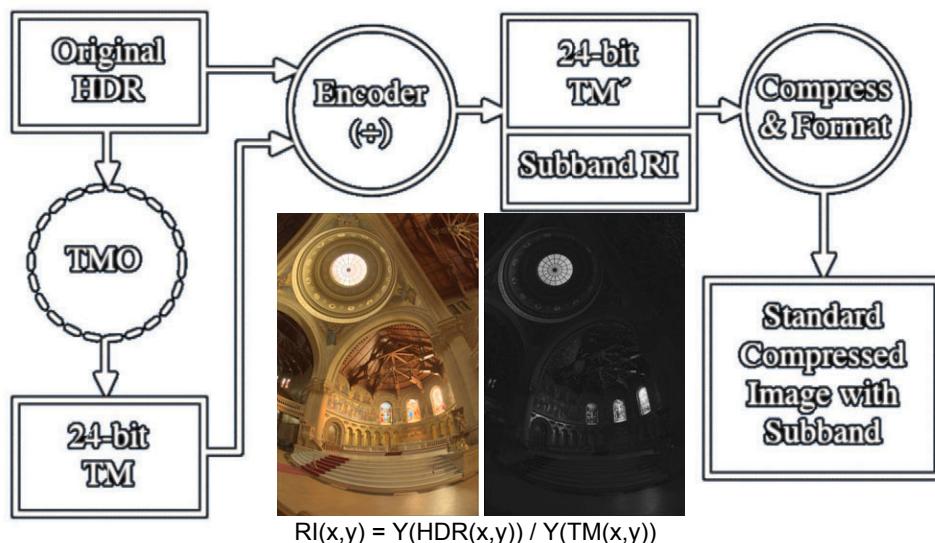
- What if HDR format was backwards-compatible?
 - Li et al. companding technique (SIGGRAPH 2005)
 - JPEG-HDR & new MPI technique (SIGGRAPH 2006)
 - XDepth

JPEG-HDR Format

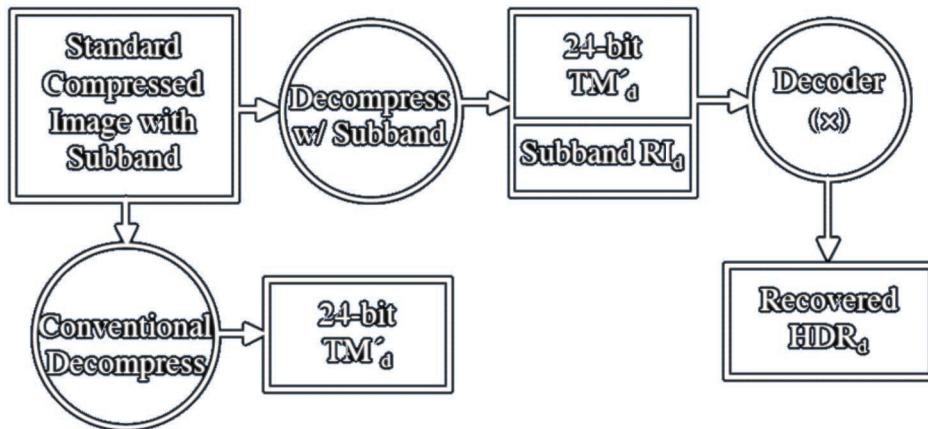
Ward & Simmons 2004 & 2005

1. Tone-map HDR input into 24-bit sRGB
2. Write as *output-referred* JPEG
3. Record restorative information as metadata (supplement)
 - Naïve applications see tone-mapped image
 - HDR applications use supplement to recover *scene-referred* original

JPEG-HDR Encoding Process



Decoding Process



Compressed JPEG-HDR size: 1-7 bits/pixel
(between 1/3 & 1/20 size of other HDR formats)

JPEG-HDR Software Availability

- Implemented as extension to Tom Lane's public JPEG library (www.ijg.org)
 - Tools & libraries included on HDRI book's DVD
 - Contact Dolby Canada for free license
- JPEG-HDR included in **Photosphere**
 - Handy export function for batch conversion and webpage creation
 - Download from www.anyhere.com

HDRI Encoding Conclusions

- Sufficient still formats to meet most needs:
 - Radiance RGBE for legacy systems
 - TIFF for greatest encoding variety
 - OpenEXR for good accuracy and support
 - JPEG-HDR for space efficiency
- HDR texture formats are being proposed
- HDR video formats are being demonstrated

HDR Capture Refinements

- Automatic exposure alignment
- “Ghost” removal
- Lens flare removal
- Implementing HDR in still & video cameras



LDR Exposure Registration

[Ward 2003, *Journal of Graphics Tools*, 8(2)]

The *median threshold bitmap* (MTB) allows us to quickly compare and align different images, because it is constant with respect to exposure for any camera with a monotonic response function

The same is not true for an edge map, which changes with exposure even with careful normalization and approximate response curves

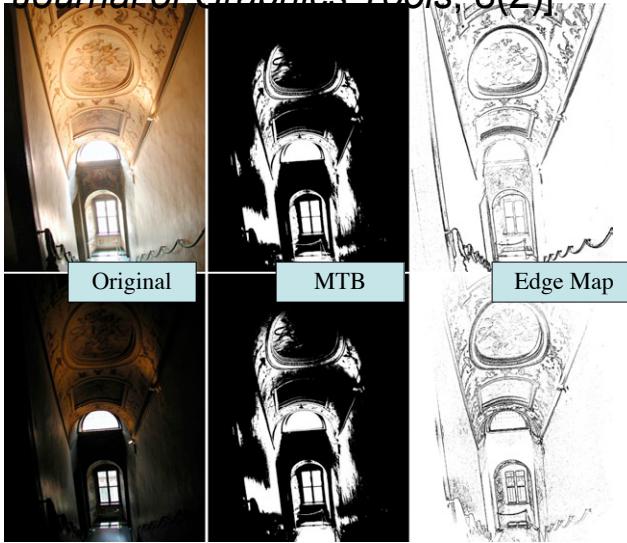


Image Pyramid Alignment

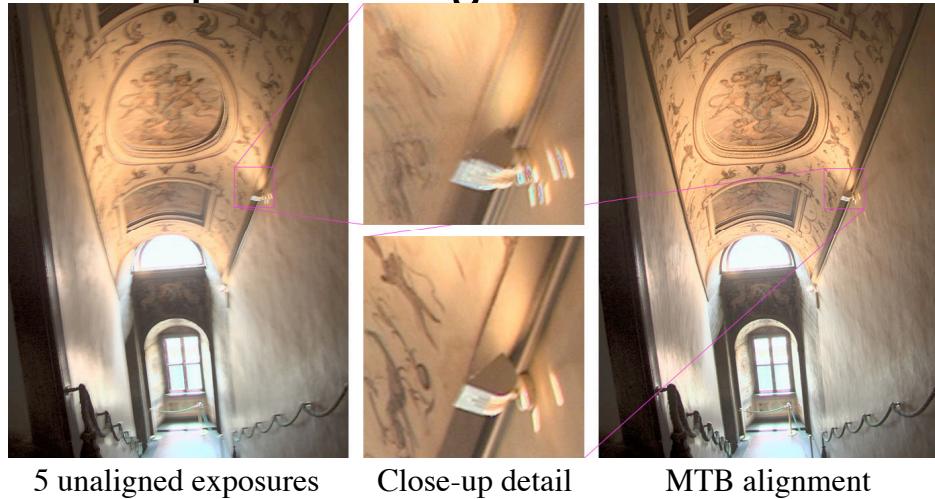


Grayscale images are scaled down repeatedly to create an image pyramid, which is then converted into MTBs for comparison

The smallest images are aligned first within a ± 1 pixel distance, which corresponds to a ± 32 pixel distance in the original

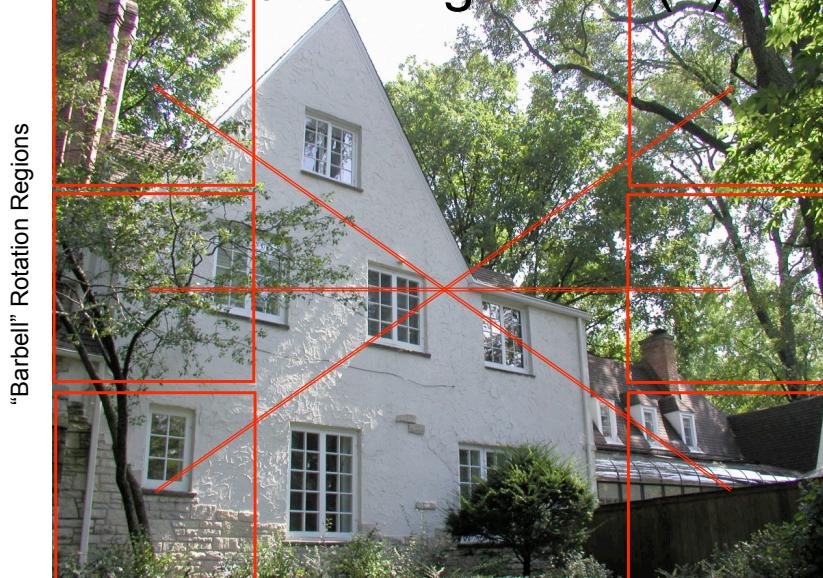
This becomes the MSB in the offset, which is shifted and used as the starting point for the next higher resolution alignment, and so on to the top

Mean Threshold Bitmap Exposure Alignment



Time: About 0.2 second/exposure for 5 MPixel image

Rotational Alignment (1)



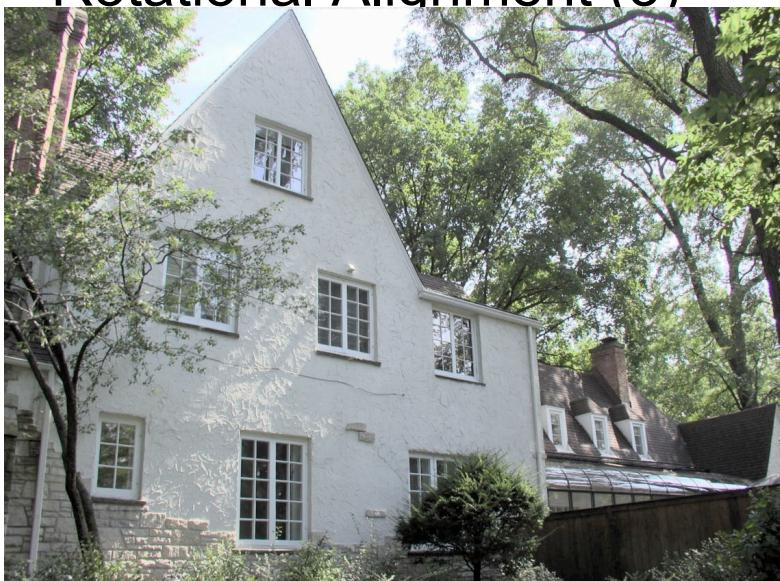
Rotational Alignment (2)

Old Translational Alignment



Rotational Alignment (3)

New Rotational Alignment



Automatic “Ghost” Removal



Before

After

Object Movement



Variance-based Detection



Region Masking



Best Exposure in Each Region



Lens Flare Removal

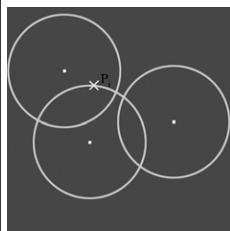


- From left image, we may directly measure the lens Point Spread Function (PSF)
- PSF is a function of focal length and aperture, so comprehensive measurement is impractical

More Usual Input



Estimate PSF from HDR Capture

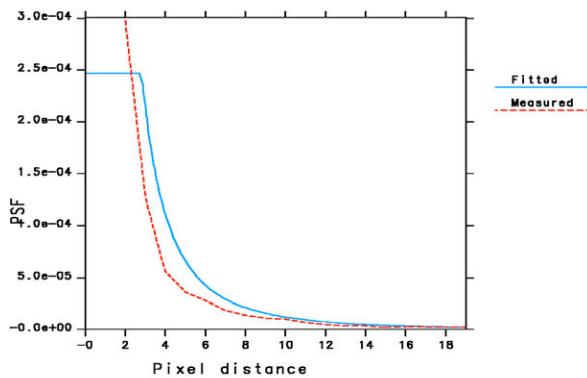


- For each “hot” pixel in input:
 - Find “coldest” relative pixel from at each radius
 - Consider overlapping hot pixel contributions
- PSF is minimal, monotonically decreasing function measured relative to cold pixels
 - Computed by fitting 3rd degree polynomial over all identified cold pixels:

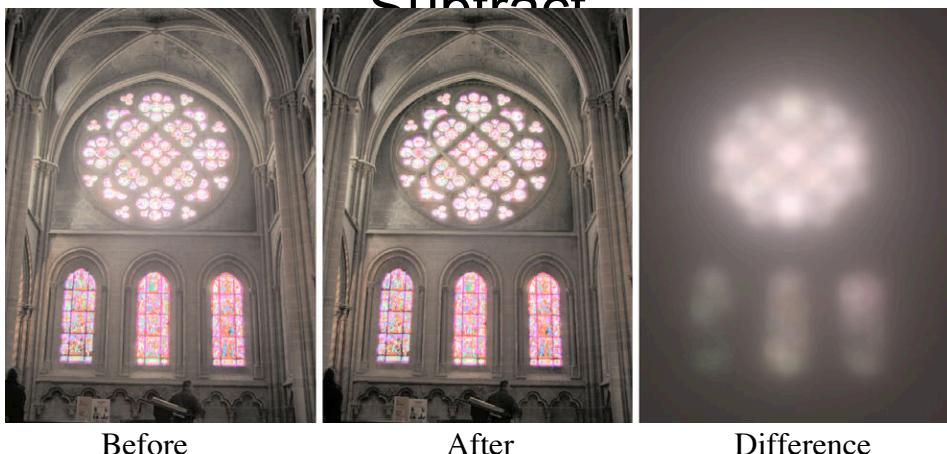
“Cold Pixel”
$$P_i = \sum_j P_j \left(C_0 + \frac{C_1}{r_{ij}} + \frac{C_2}{r_{ij}^2} + \frac{C_3}{r_{ij}^3} \right)$$

Fitted vs. Measured PSF

PSF estimate (apt. capture fit vs. tin foil spot)



Apply: Simulate Flare & Subtract

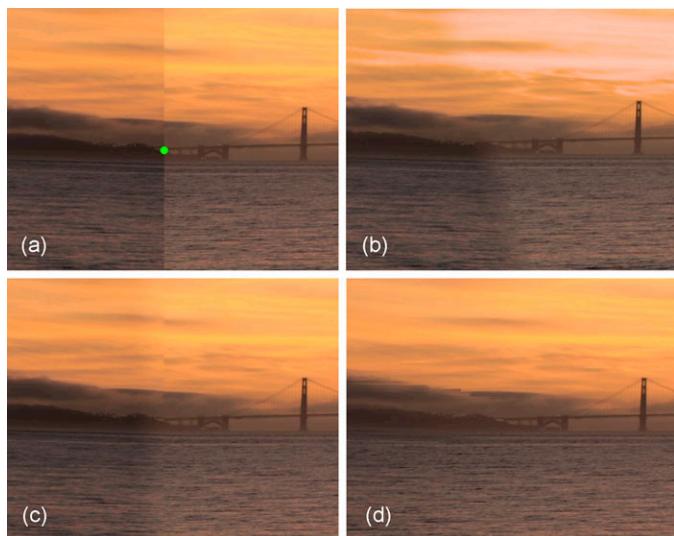


HDR Panorama Stitching



- Separate into high & low frequency bands
- Blend low frequencies using entire overlap region
- Splice high frequencies at detected edges
- If no edges, then blend high frequencies as well

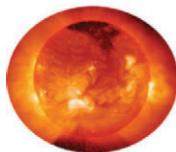
Comparison to Others



(a) Simple splice.
(b) Photoshop™
(c) Burt-Adelson
(d) New technique

Photosphere Demo

- HDRI Browsing & Cataloging Application
 - Also builds HDRI's from bracketed exposures
- Available from www.anyhere.com
 - Mac OS X app., Linux command-line tool



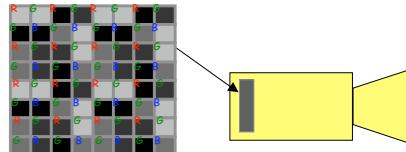
Launch Photosphere

Fast Forward to HDR Cameras

- Leverage CMOS Sensor Technology
 - Fuji has sensor with dual-sensitivity pixel
 - SMaL Camera has log sensor
 - Pixim sensor has local pixel exposure
- Alter camera/sensor design
 - Multi-image capture using modified scanout
 - Multiple sensors
 - Spatially varying filters for video mosaicing
 - Sensors with assorted pixels
 - Adaptive dynamic range system

Assorted Pixels

Nayar and Mitsunaga, IEEE CVPR 2000
Nayar and Narasimhan, ECCV 2002



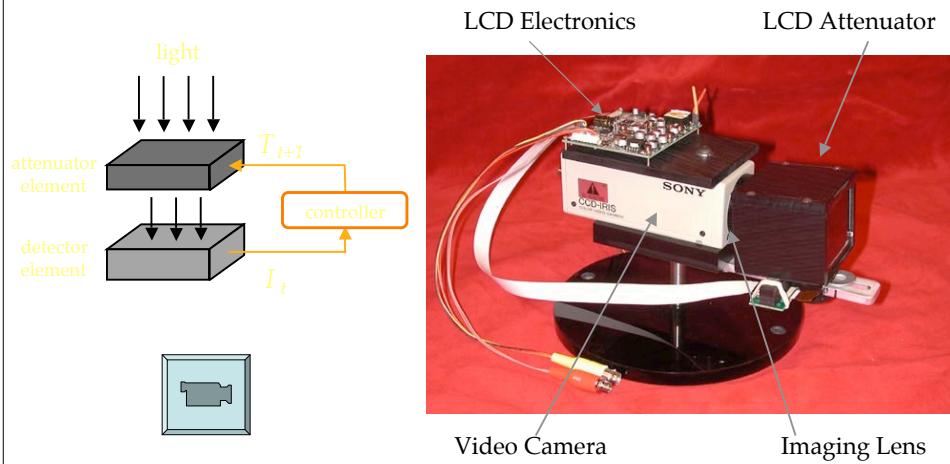
Digital Still Camera



Camera with Assorted Pixels

Adaptive Dynamic Range

Nayar and Branzoi, ICCV 2003



Conclusions

- HDR capture is under active development
- HDR video is challenging but has many potential benefits
- Movie industry is an early adopter
- Home entertainment market imminent

Tone Reproduction

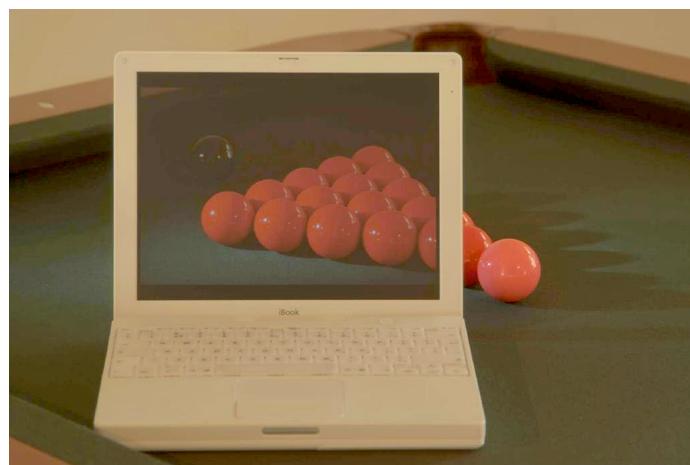
Erik Reinhard

[University of Bristol](#)

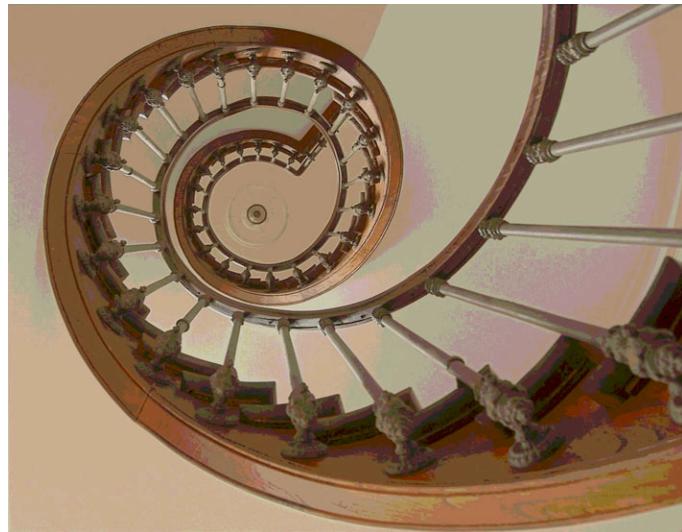
reinhard@cs.bris.ac.uk

Goals in HDRI

Capture, store, and display real-world luminance values



Quantization



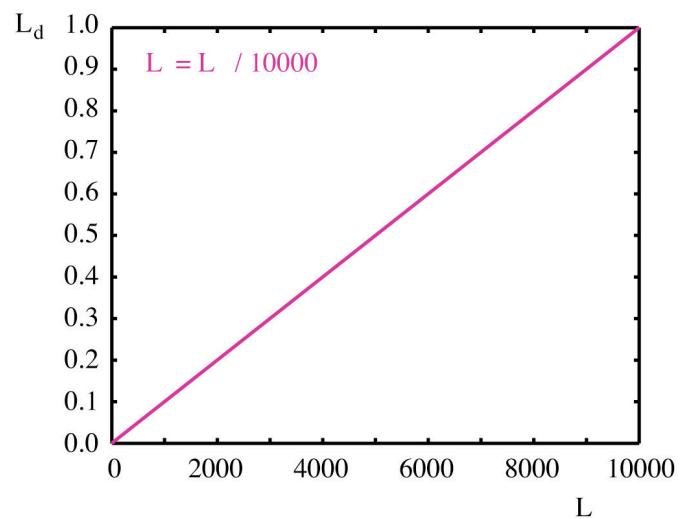
Dynamic Range



Tone Reproduction

- Match dynamic range to display capabilities
- Preserve some visual quality of the scene
 - Brightness
 - Contrast
 - Visibility
 - Appearance
 - ...

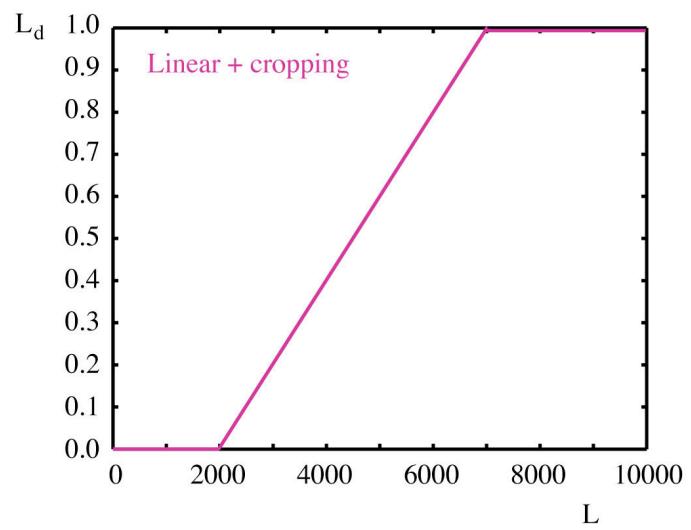
Linear Scaling



Linear scaling



Linear + Clamping



Linear + Clamping

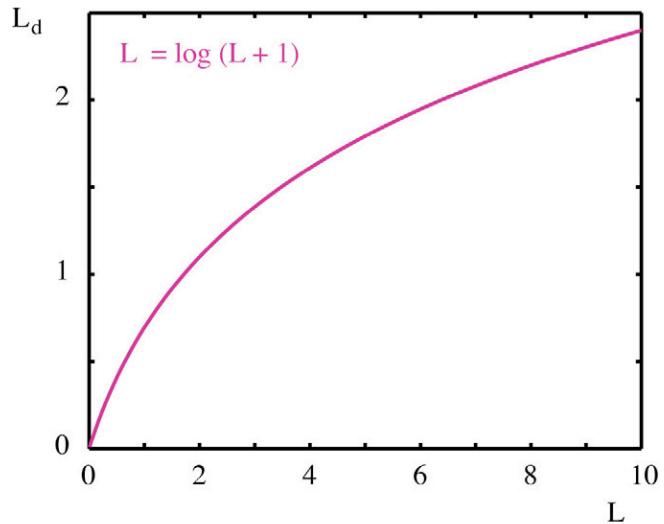


Logarithmic scaling

Simple version:

$$L_d = \frac{\log L_w - 1}{\log L_{w,\max} - 1}$$

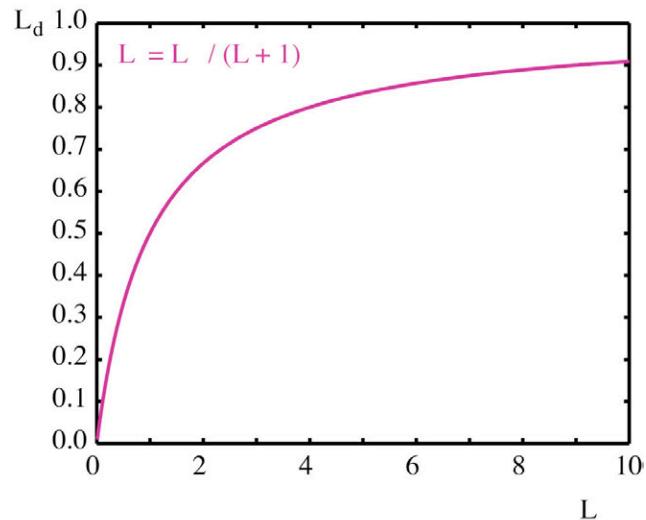
Logarithm



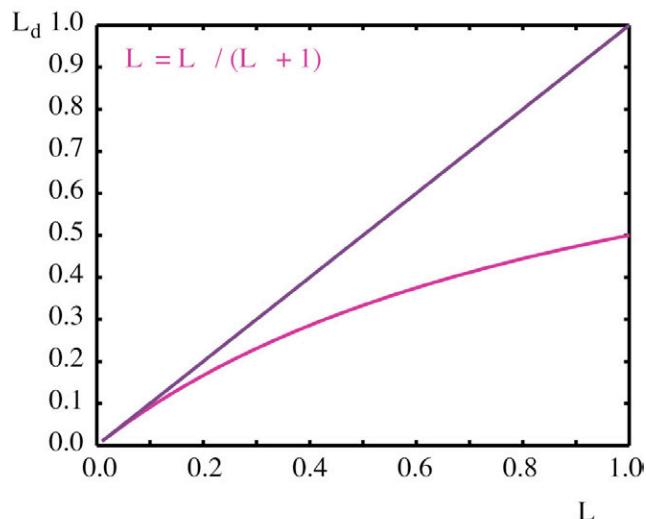
Logarithmic compression



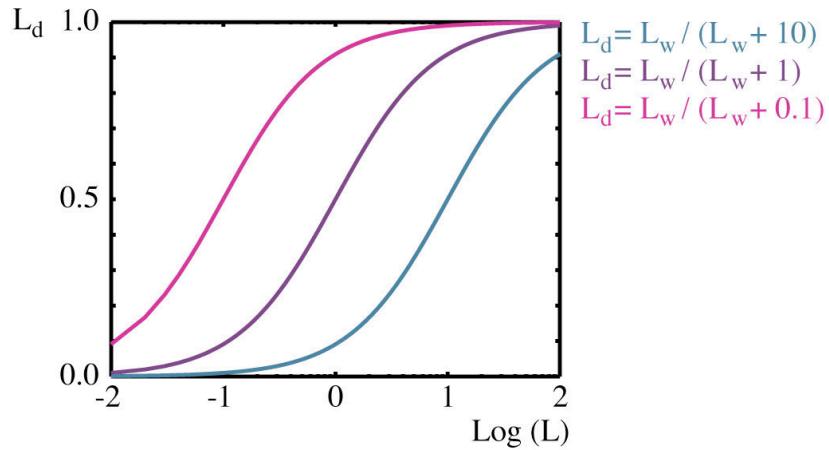
Sigmoid



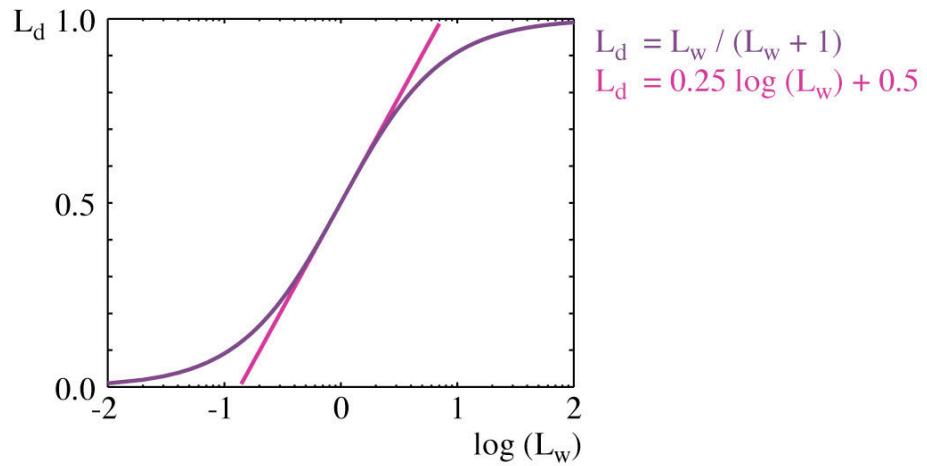
Sigmoid (Near Origin)



Sigmoids



Sigmoid vs Log



Sigmoid



Histogram adjustment

Ward et al (1997) use the shape of the image's histogram

- Compute histogram
- Compute cumulate histogram
- Result is a monotonically increasing function
- Reshape this function to avoid slopes greater than 1
- Remap luminances according to this function

Histogram Adjustment



Global operators

Recap:

- Easy to implement
- Fast / GPU implementations possible?
 - One pass over the image to compute log average
 - One pass to compress pixels
- Reasonable amounts of compression and plausible results are frequently obtained
- Useful for medium to high dynamic range images

Global to local

- Global average \longrightarrow local average
- Semi-saturation constant based on local neighborhood:

$$L_d[x, y] = \frac{L_w[x, y]^n}{L_w^{\text{LPF}}[x, y]^n + L_w[x, y]^n}$$

- Requires convolution with (small) kernel

Global vs. local



Sigmoid with
global adaptation

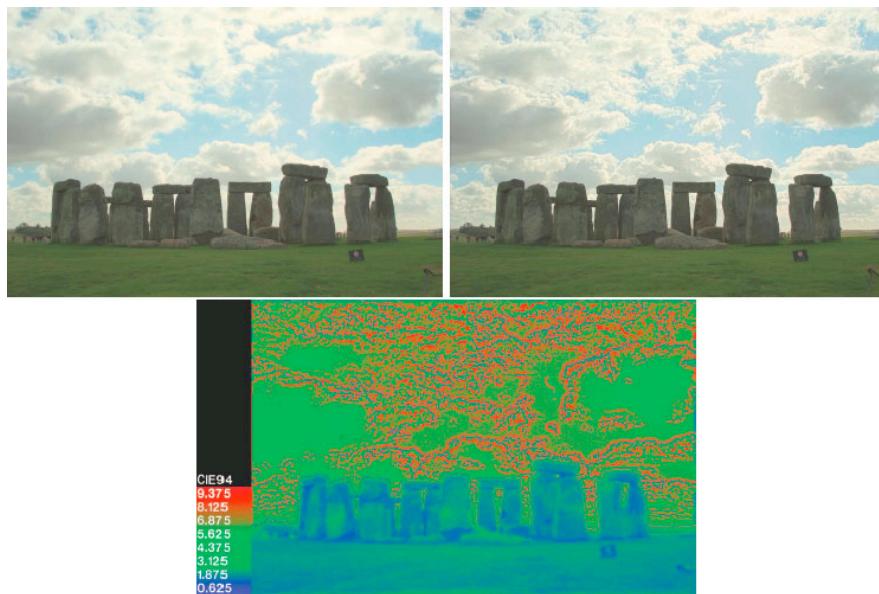


Sigmoid with
local adaptation

Local vs Global



Comparison



Local adaptation

There is a catch, though:

- The size of the low-pass filter kernel is important.
- Too large and haloing artifacts will occur
- Too small and compression will not be better than global operators

Halos



Low-pass filter kernel size

- For each pixel the size of the LPF kernel should be such that it does not overlap with sharp discontinuities
- But we still want to average over the largest spatial area for which the above is true (which may be different for each pixel)
- In practice often a small kernel size!

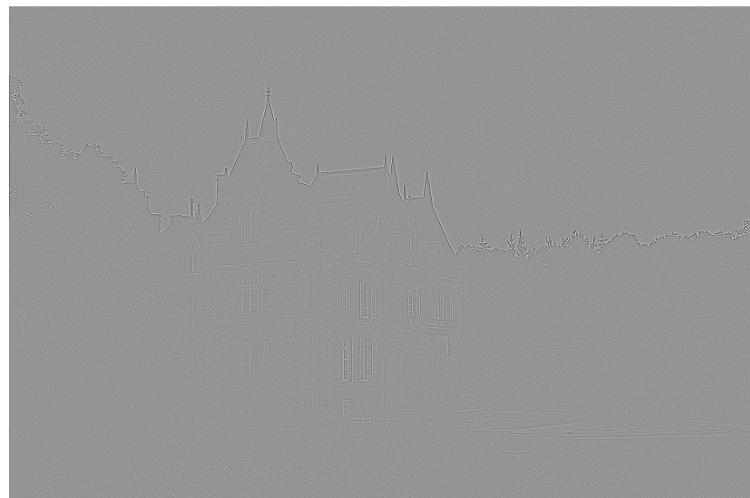
How to compute?

- Multi-scale techniques using difference of Gaussians and scale selection (Reinhard 2002, Ashikhmin 2002)
- Edge-preserving smoothing operator (bilateral filter, mean shift algorithm, LCIS...)
- Sub-band systems (Li 2005)

Difference of Gaussian



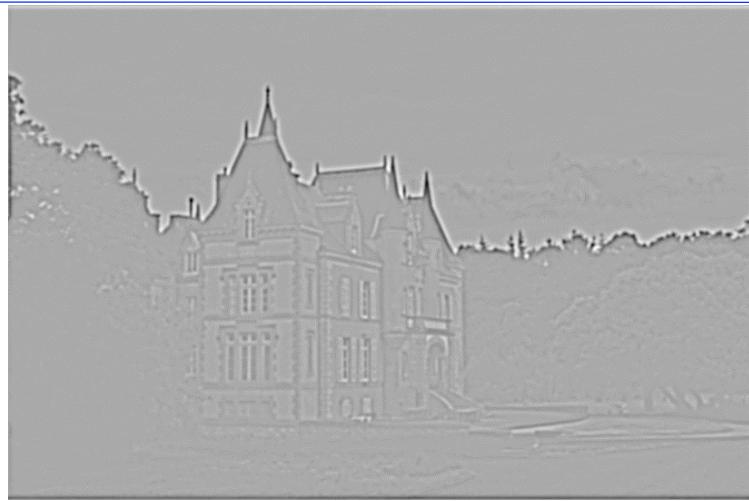
Difference of Gaussian



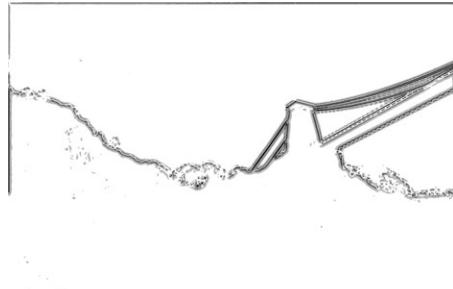
Scale-Space



Scale-Space



Scale Selection



Li et al

SIGGRAPH 2005
Sub-band encoding



Local Operators

Second approach:

- $L_d(x,y) = s(x,y) / L_w(x,y)$
- Replace global adaptation value with a per-pixel local adaptation value

Local operators

- Chiu et al, Rahman et al, iCAM model:

$$L_d(x,y) = s(x,y) / L_w(x,y)$$
$$s(x,y) = f \frac{1}{L_w^{\text{LPF}}(x,y)}$$

i.e. divide the image by a low-pass filtered version of itself

Division-based

To minimize
these artifacts:

- Use large filter kernel



Durand and Dorsey 2002

Split image into base- and detail layers:

- Use edge-preserving smoothing operator to filter in log space and call result ‘base layer’
- Divide image by base layer and call result ‘detail layer’

Durand and Dorsey 2002



Input image

Smoothed result

Bilateral (Base Layer)



Bilateral (Detail Layer)



Bilateral (Final Result)



Bilateral Filtering

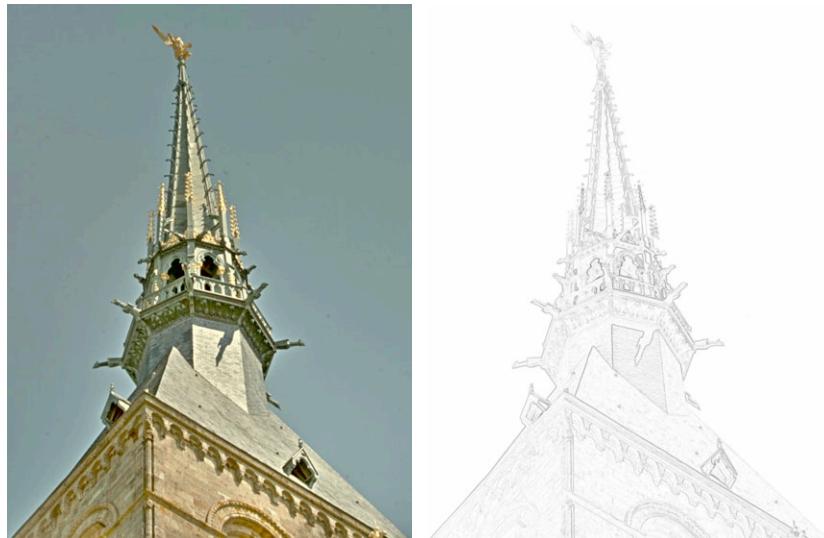


Fattal et al 2002

Gradient domain compression:

- Differentiate image in log space
- Attenuate large gradients
- Integrate
- Exponentiate

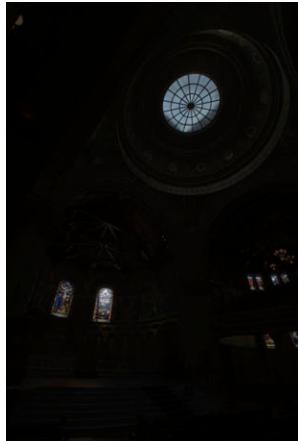
Gradient Domain Technique



Fattal et al 2002



Informal comparison



Linear scaling

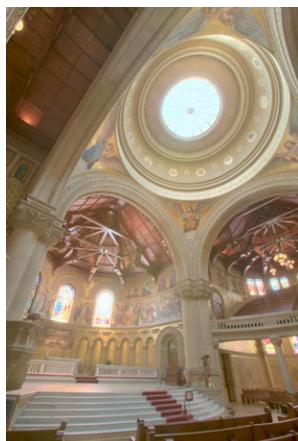


Linear with clamping



Linear with clamping (pct)

Informal comparison



Photographic operator



Ashikhmin



Li et al, SIGGRAPH 2005

Informal comparison



Bilateral filter



Gradient compression

Informal comparison



Histogram adjustment

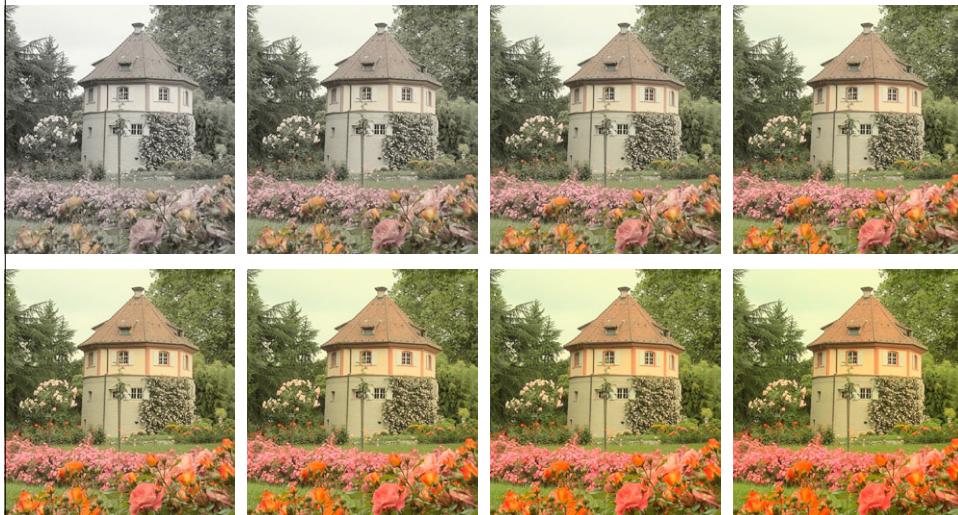


Photoreceptor-based



Tumblin-Rushmeier

Saturation $R=R_d \underline{R_w}/L_w$ ^s



Clamping (5% - 95%)



Clamping (7%-98%)



Model inversion

- More principled approach
 - Apply TMO with scene-referred parameters
 - Apply inverse TMO with display-referred parameters
 - Clamp
- Early TMOs took this approach
- Color appearance models do this too

Forward Model



Forward + Backward Model



Forward Model



Forward + Backward Model



Forward + Backward Sigmoid

$$V(x, y) = \frac{L_v^n(x, y)}{L_v^n(x, y) + g^n(x, y)}$$

$$L_d(x, y) = \left(\frac{V(x, y) L_{d,\text{mean}}^m}{1 - V(x, y)} \right)^{1/m}$$

n --> m
g --> L_{d,mean}, i.e. mean display luminance

Mean display luminance

$$L_{d,\text{mean}} = \frac{1}{2} (L_{d,\text{min}} + L_{d,\text{max}})$$

assuming linear data.

Combining forward + backward equations

$$\begin{aligned} L_d &= \left(\frac{\frac{L^n}{L^n + (\bar{L}/k)^n} L_{d,\text{mean}}^m}{1 - \frac{L^n}{L^n + (\bar{L}/k)^n}} \right)^{1/m} \\ &= \frac{L^{n/m} L_{d,\text{mean}}}{(\bar{L}/k)^{n/m}} \\ &= c L^{n/m} \\ c &= \frac{L_{d,\text{mean}}}{(\bar{L}/k)^{n/m}} \end{aligned}$$

Used $g = \Gamma/k$ here,
as in photographic
tone reproduction

Forward + backward sigmoid

- Amounts to applying a power function, i.e. gamma correction
- Ends up being similar to Tumblin-Rushmeier's operator
- Hence suitable for medium dynamic range applications
- Implications for tone reproduction as well as color appearance modeling!

Forward + Reverse



Figure 17: *Forward and reverse model with $n = 0.7$ and $k = 0.3$ for an HDR image with a relatively modest dynamic range 2.8 log units (left) and an image with a much higher dynamic range (right; $n = 0.7$, $k = 0.08$, 8.3 log units dynamic range).*

Forward



Figure 18: *Forward model only with $n = 0.7$ and $k = 0.3$ (left) and $n = 0.7$, $k = 0.08$ (right).*

Photographic Operator



Figure 19: *Photographic operator with key $k = 0.3$ (left) and $k = 0.08$ (right). By using the photographic operator, we have effectively changed the exponent to $n = 1$.*

Solutions?

- Maybe there exists a theoretical reason for why the backward transform should be a linear scaling, instead of an inverse sigmoid?
- Maybe sigmoids are not a good solution?
- Maybe a power function is the right answer?
- In any case, none of these solutions is very satisfying

Conclusions

Trade-offs exist between:

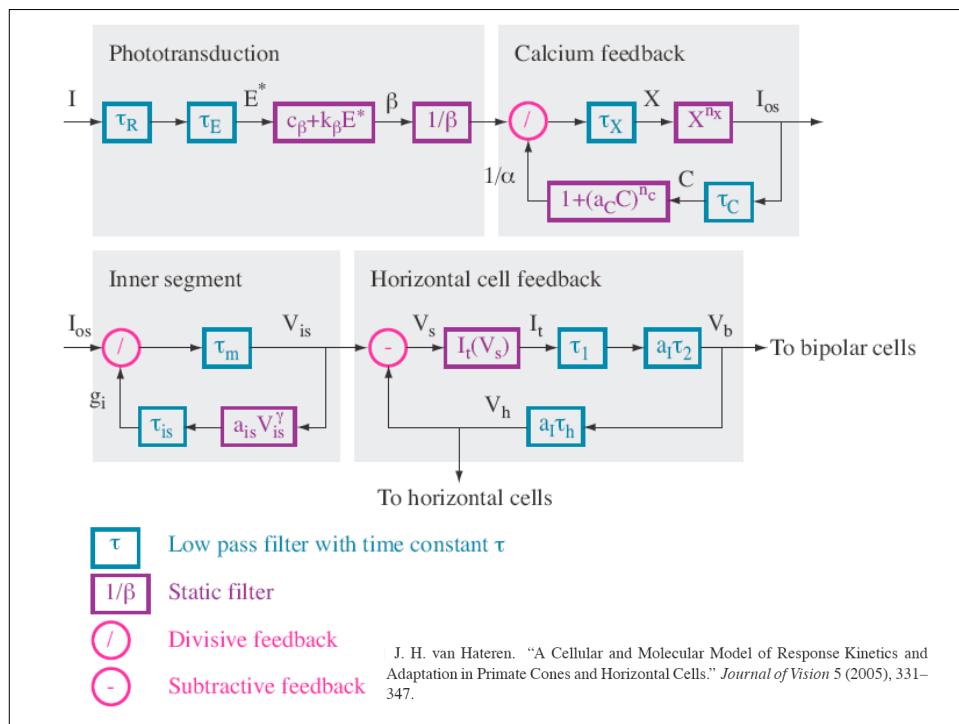
- Amount of compression
- Presence of artifacts
- Computation time

Need to reconsider applying TMOs in both forward and backward directions

Acknowledgments

Thanks to:

- Oguz Akyuz, Erum Khan, Timo Kunkel, Greg Ward, Helge Seetzen, Roland Fleming, Heinrich Buelthoff, Ted Adelson, Yuanzhen Li, Dani Lischinski, Ranaan Fattal, Karol Myszkowski, Grzegorz Krawczyk, Rafal Mantiuk, Charles Hughes, Kadi Bouatouch, Remi Cozot, Yoann Marion, Jonathan Brouillat
- And many others



SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)

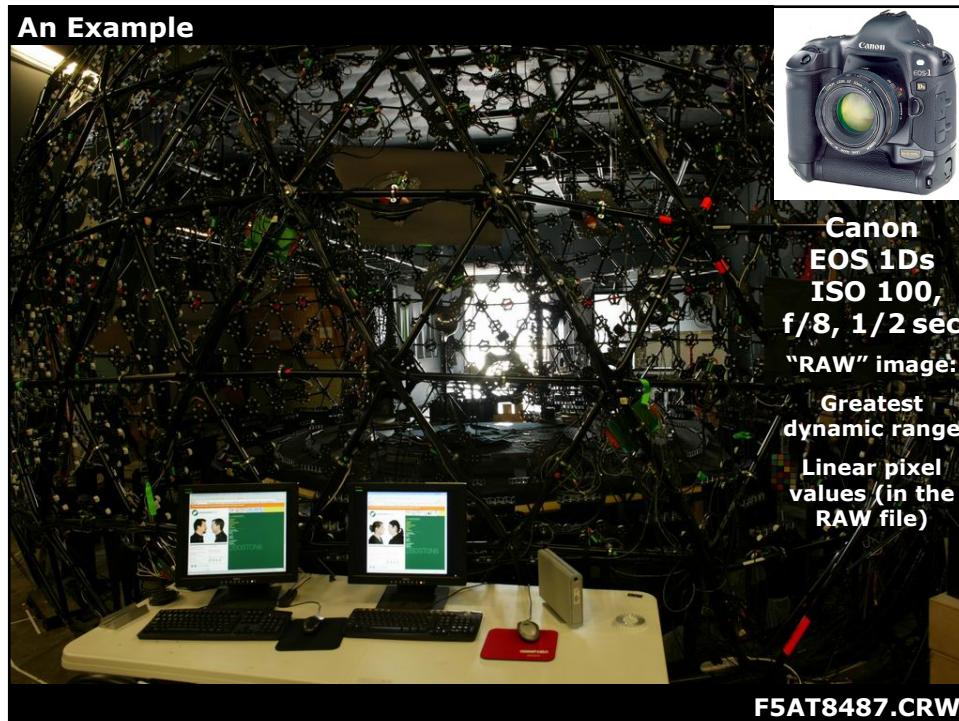
Taking HDR Images



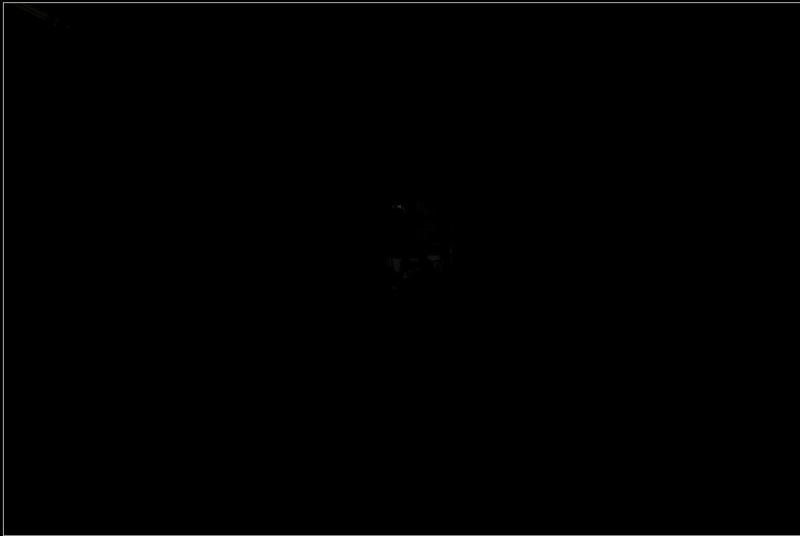
Paul Debevec
University of Southern California
Institute for Creative Technologies
Graphics Laboratory

SIGGRAPH 2008 Class
High-Dynamic-Range Imaging & Image-Based Lighting
August 2008

www.debevec.org/ / gl.ict.usc.edu



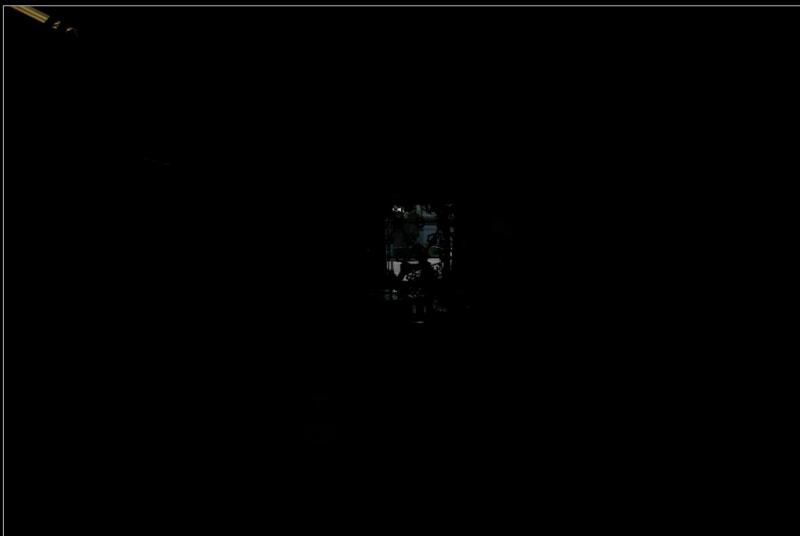
High Dynamic Range Imaging



ISO 100, f/8, 1/8000 sec



High Dynamic Range Imaging



ISO 100, f/8, 1/1000 sec



High Dynamic Range Imaging

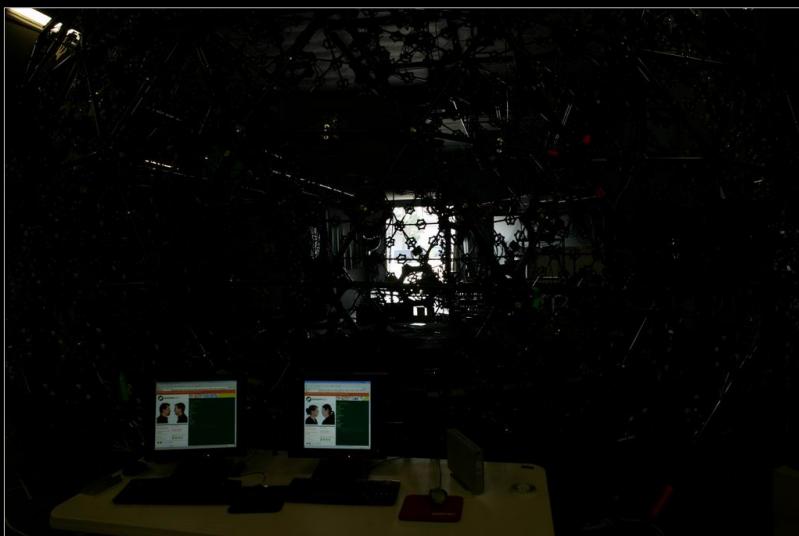


ISO 100, f/8, 1/125 sec

SIGGRAPH 2008



High Dynamic Range Imaging



ISO 100, f/8, 1/15 sec

SIGGRAPH 2008



High Dynamic Range Imaging

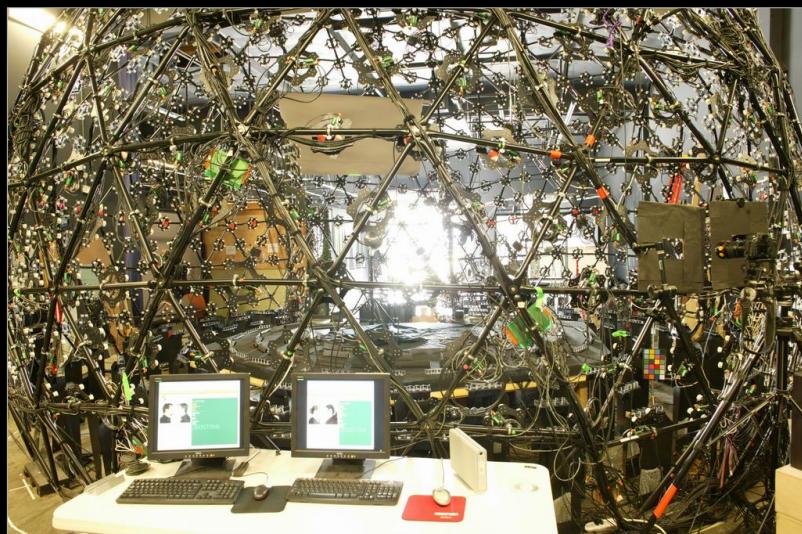


ISO 100, f/8, 1/2 sec

SIGGRAPH 2008



High Dynamic Range Imaging

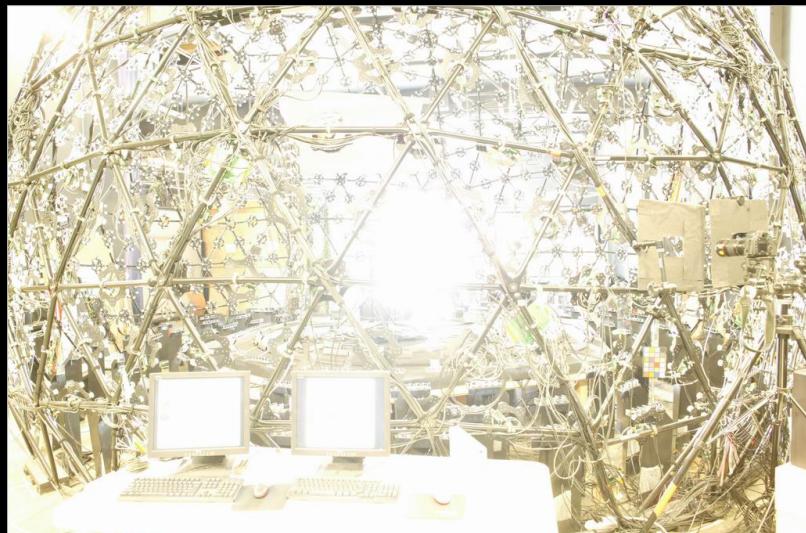


ISO 100, f/8, 4 sec

SIGGRAPH 2008



High Dynamic Range Imaging



ISO 100, f/8, 30 sec

SIGGRAPH2008



High Dynamic Range Imaging



f/8, 1/8000th sec



f/8, 1/1000th sec



f/8, 1/125th sec



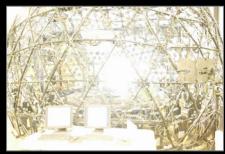
f/8, 1/15th sec



f/8, 1/2 sec



f/8, 4 sec



f/8, 30 sec



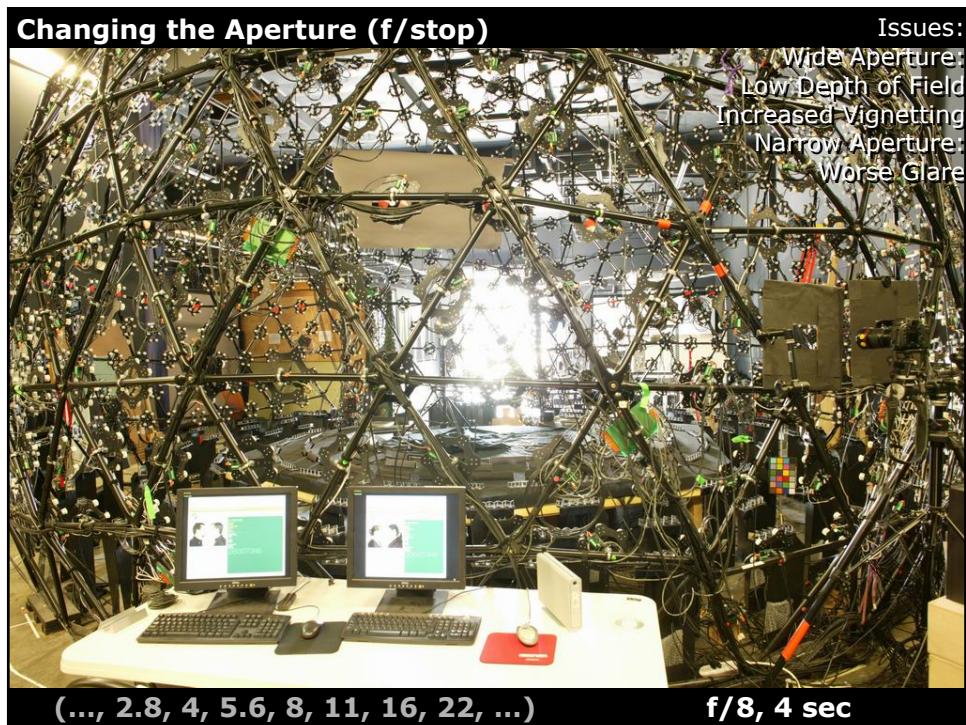
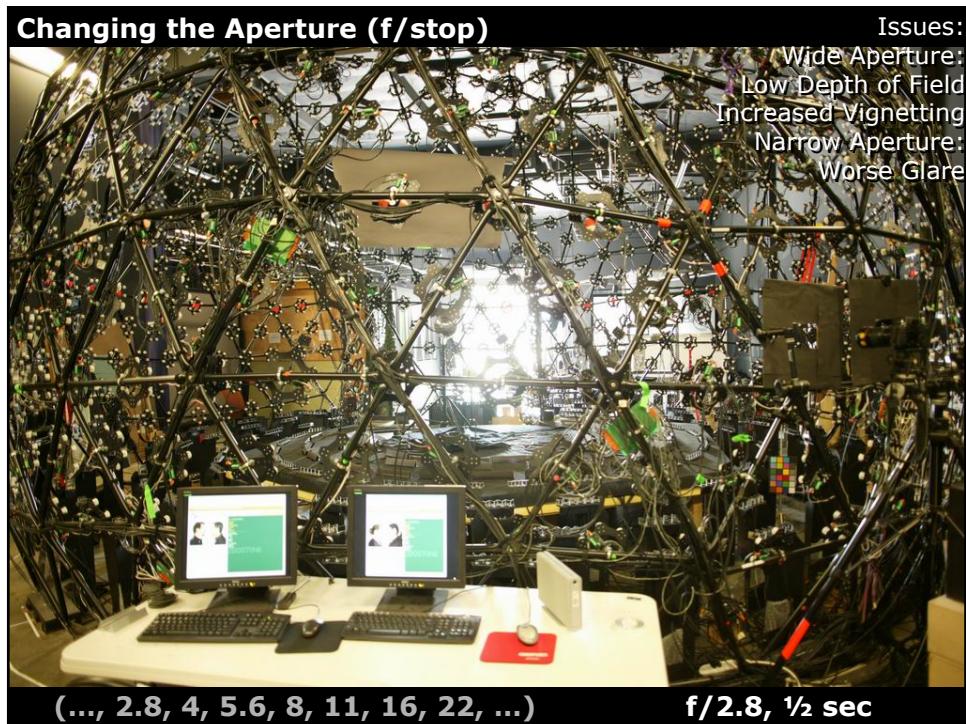
Sekonic Light Meter

5000 cd/m²

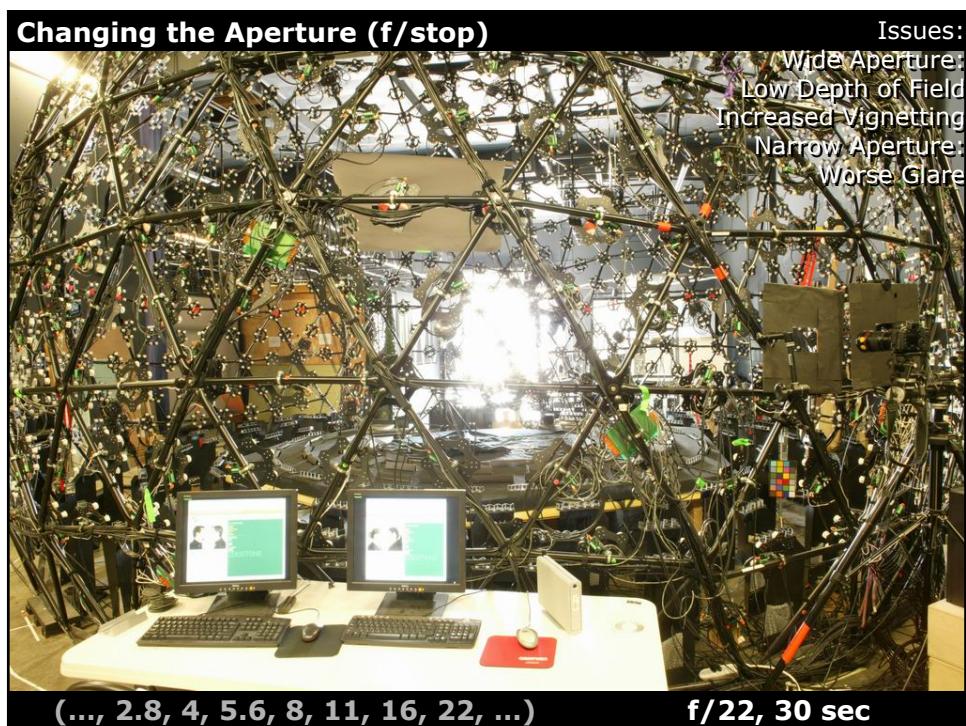
SIGGRAPH2008



SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)



SIGGRAPH 2008 Class: HDRI and Image-Based Lighting Image-Based Lighting (Paul Debevec)



Neutral Density (ND) Filters



Ranges: 0.1 to 4.0 density
(0.3, 0.6, 0.9 density = 1, 2, 3 stops common)

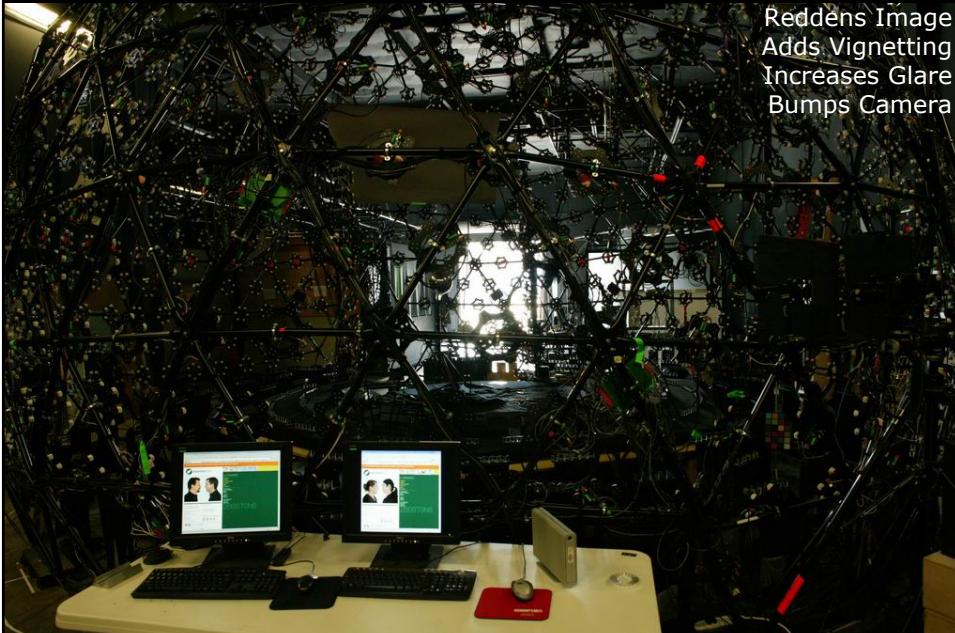
Log base 10 scale:

- Density of 0.3 = $\frac{1}{2}$ the light ('1 stop')
- Density of 1.0 = $\frac{1}{10}$ the light
- Density of 4.0 = $\frac{1}{10,000}$ the light

Drawback: Not perfectly neutral



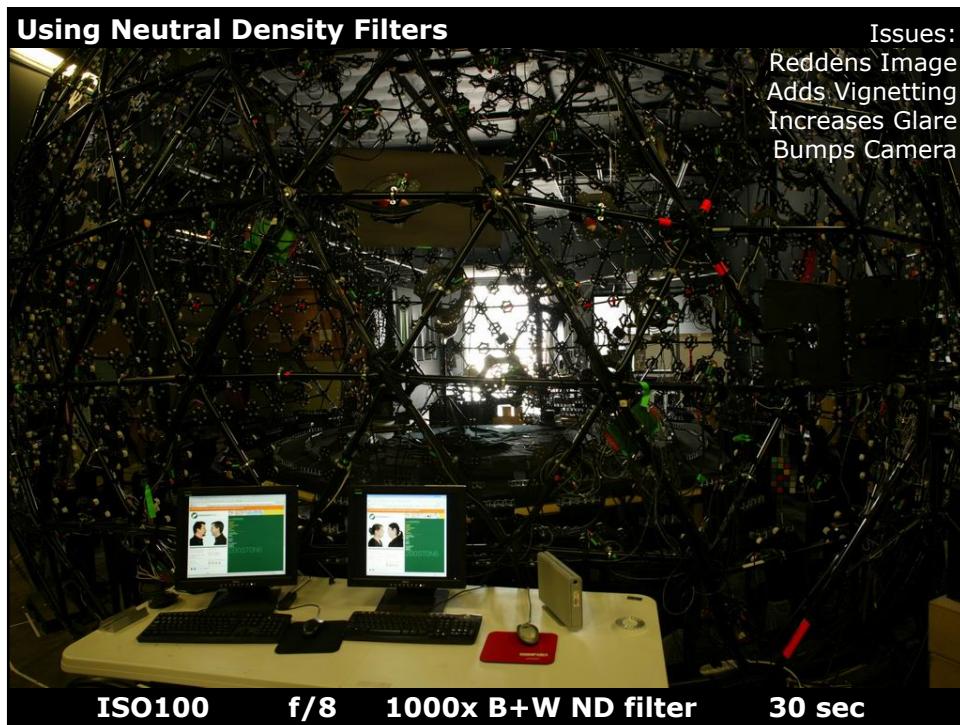
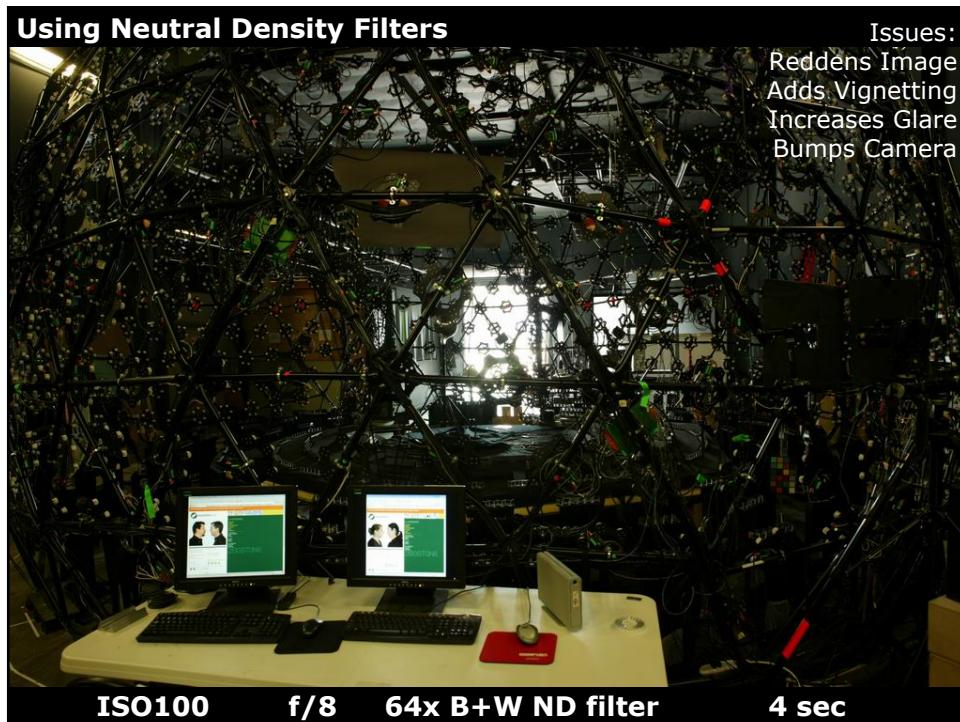
Using Neutral Density Filters

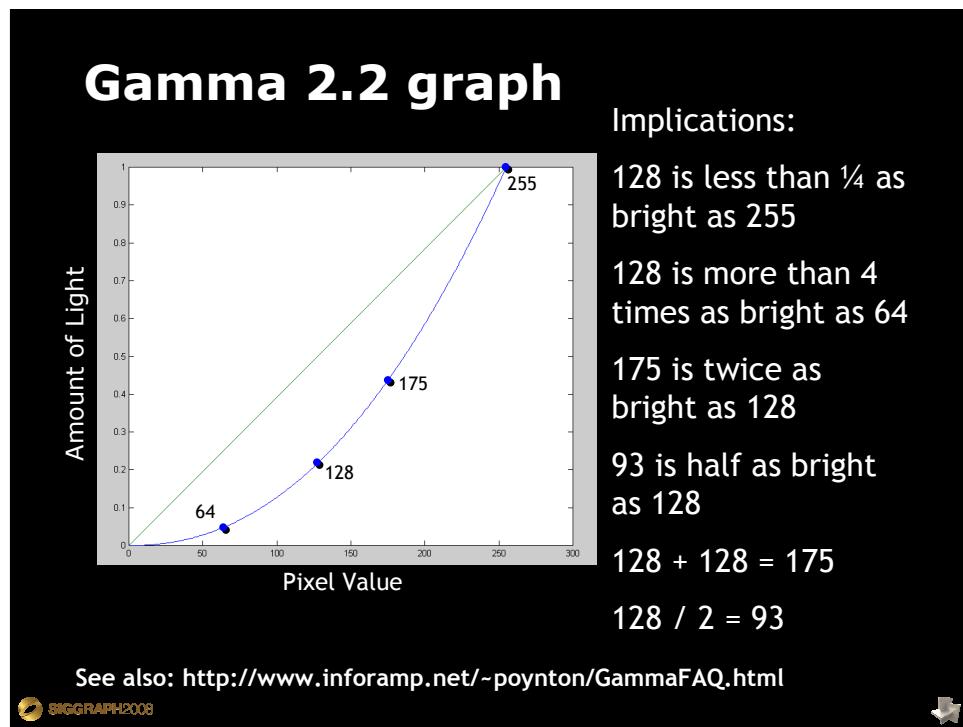
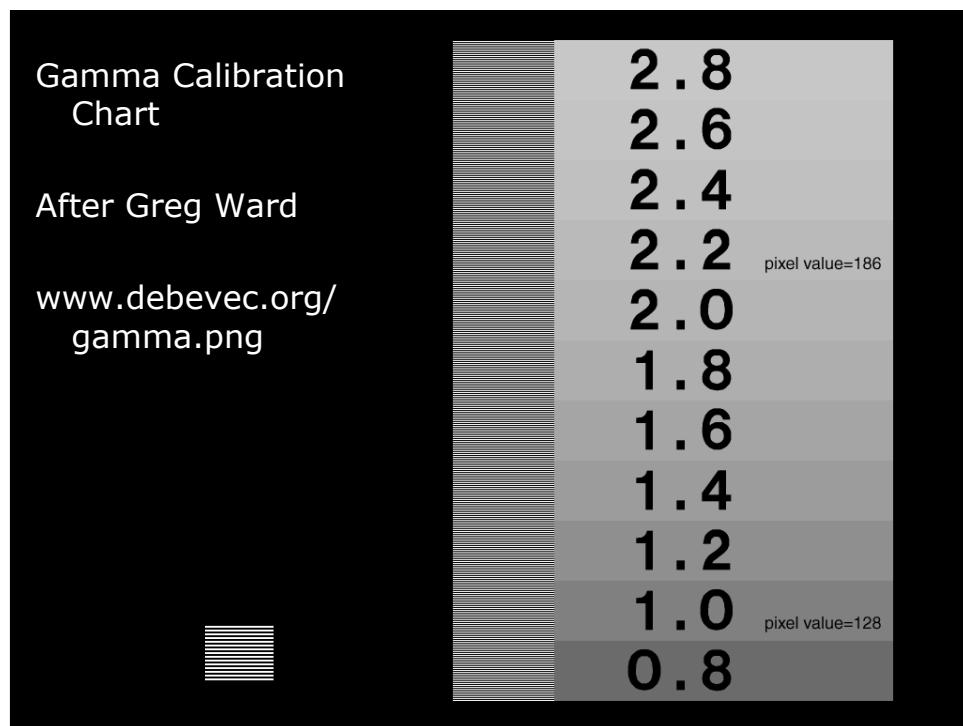


Issues:
Reddens Image
Adds Vignetting
Increases Glare
Bumps Camera

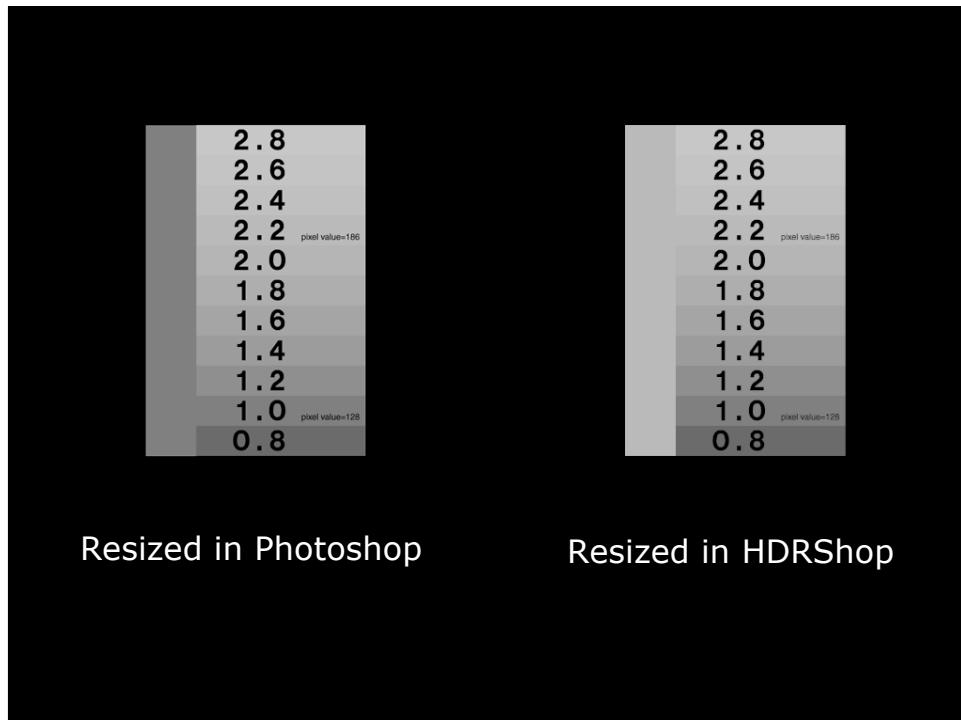
ISO100 f/8 No ND filter 1/15 sec

SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)





SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
 Image-Based Lighting (Paul Debevec)



Response Curve Recovery

Mann and Picard SPIE 95:
 Track one pixel value across series and fit a gamma-like curve

Debevec and Malik SIGGRAPH 97:
 Derive detailed curve from many pixels

$\Delta t = \frac{1}{64} \text{ sec}$

$\Delta t = \frac{1}{16} \text{ sec}$

$\Delta t = \frac{1}{4} \text{ sec}$

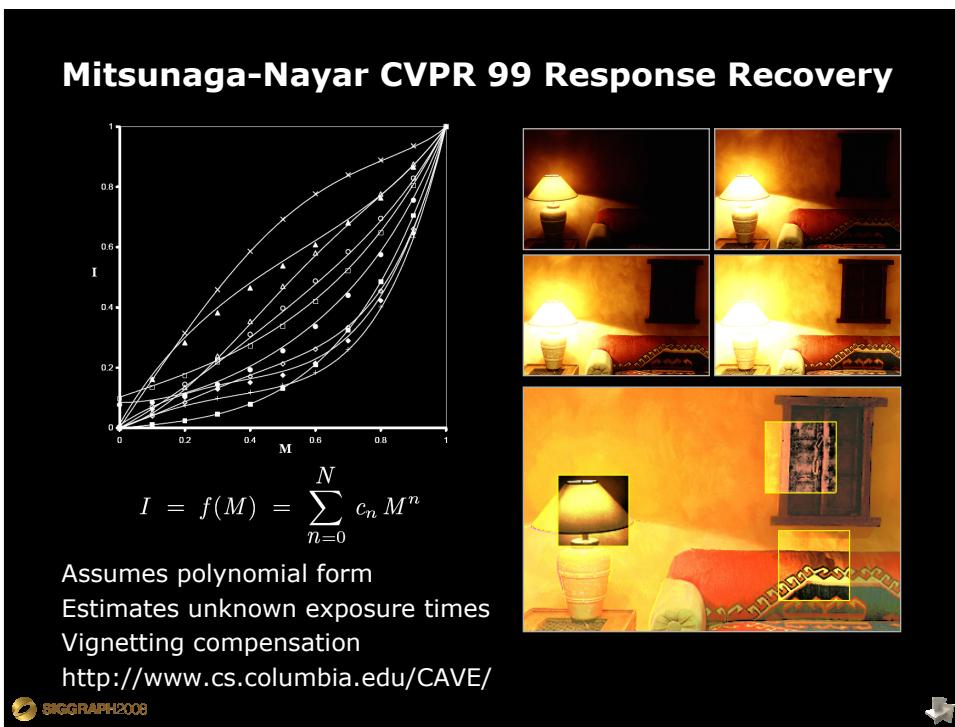
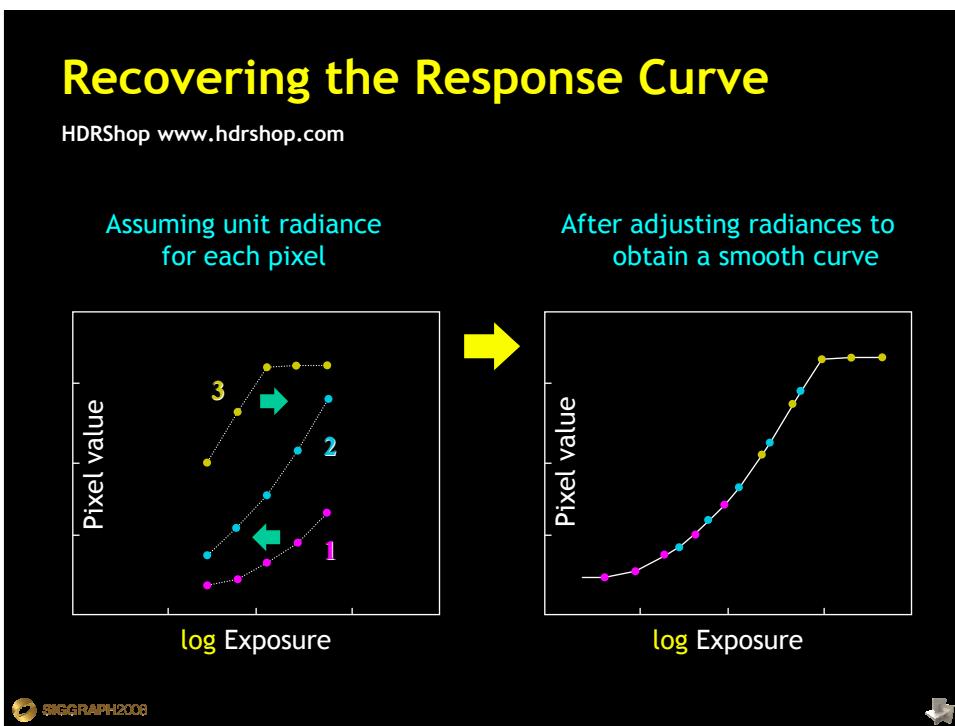
$\Delta t = 1 \text{ sec}$

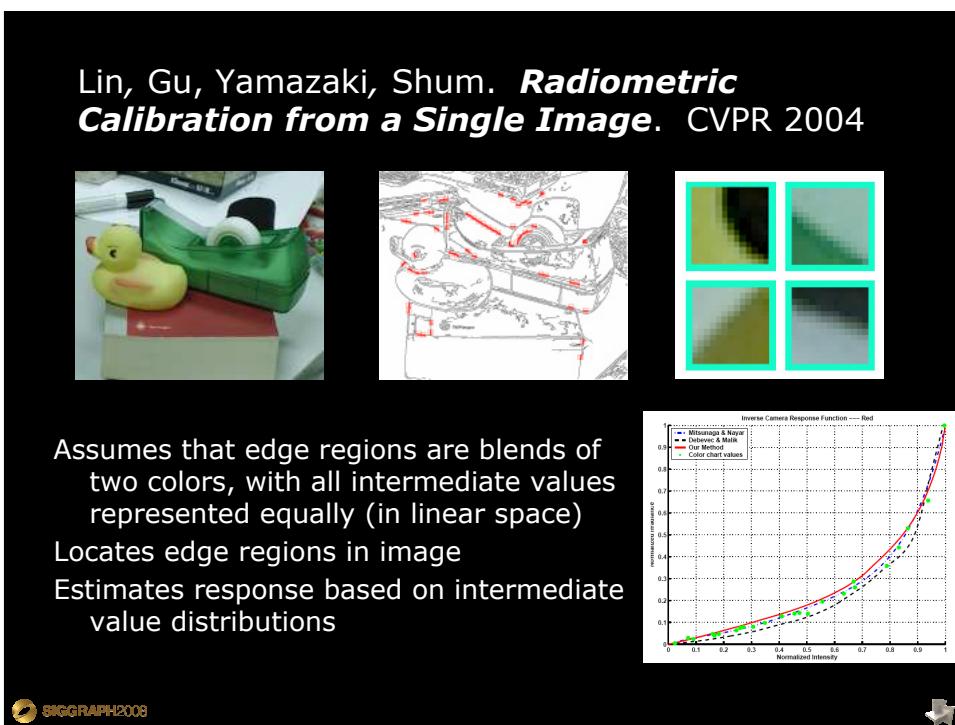
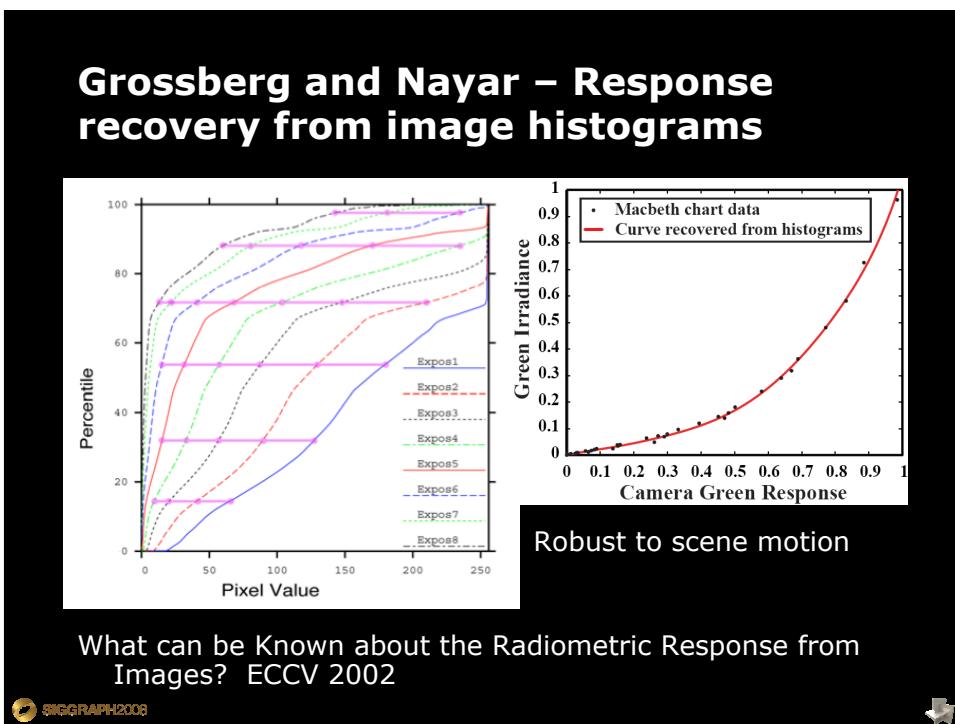
$\Delta t = 4 \text{ sec}$

$\text{Exposure} = \text{Radiance} \times \Delta t$

$\log \text{Exposure} = \log \text{Radiance} + \log \Delta t$

SIGGRAPH 2008

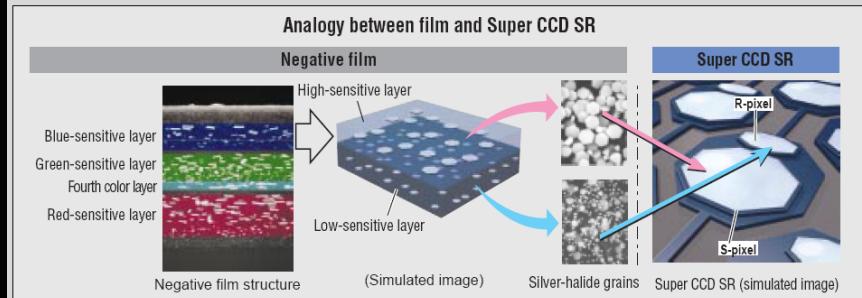




Emerging Sensor Technology

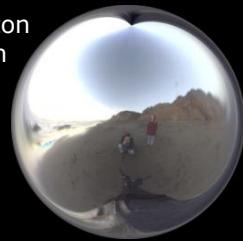
Spatially-varying pixels

How the Super CCD SR achieves greater dynamic range



Light Probe Images: High Dynamic Range Lighting Environments

Funston Beach



Eucalyptus Grove



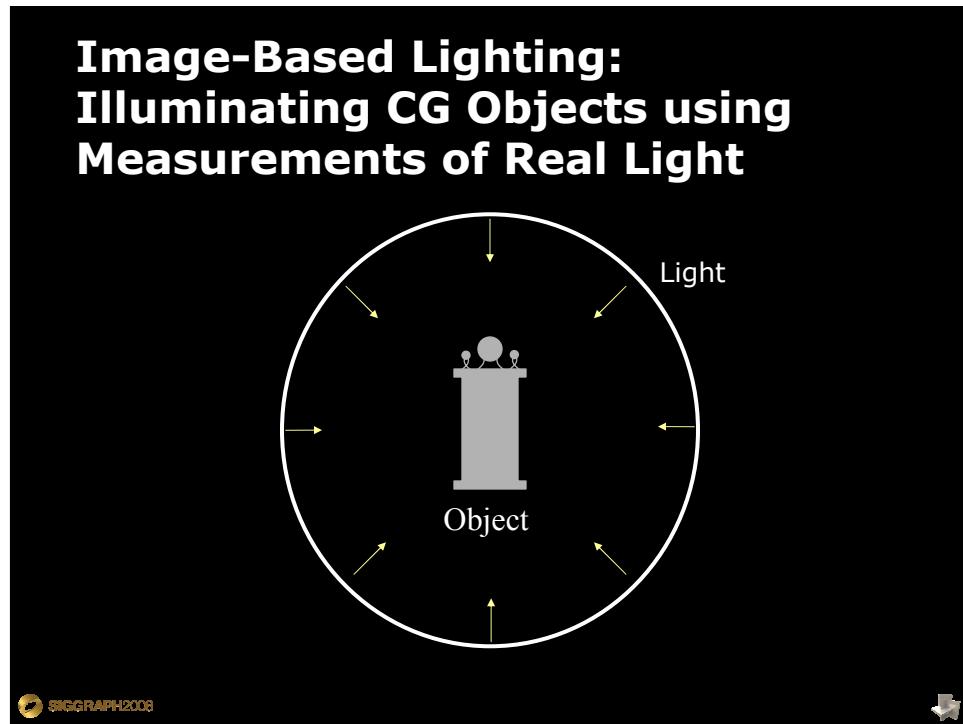
Uffizi Gallery



Grace Cathedral

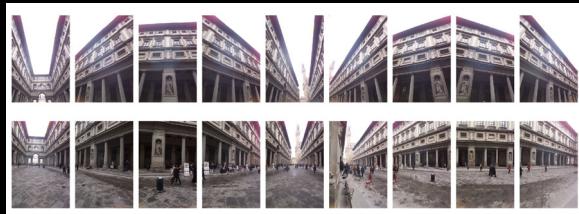


From the Light Probe Image Gallery: <http://www.debevec.org/Probes/>



P. Debevec, *Rendering Synthetic Objects into Real Scenes*, SIGGRAPH 98
P. Debevec, *A Tutorial on Image-Based Lighting*, IEEE CG&A, Jan/Feb 2002
Rendered in Greg Ward's RADIANCE

Mirrored ball, fisheyes, stitched photos, or scanning?



 SIGGRAPH 2008



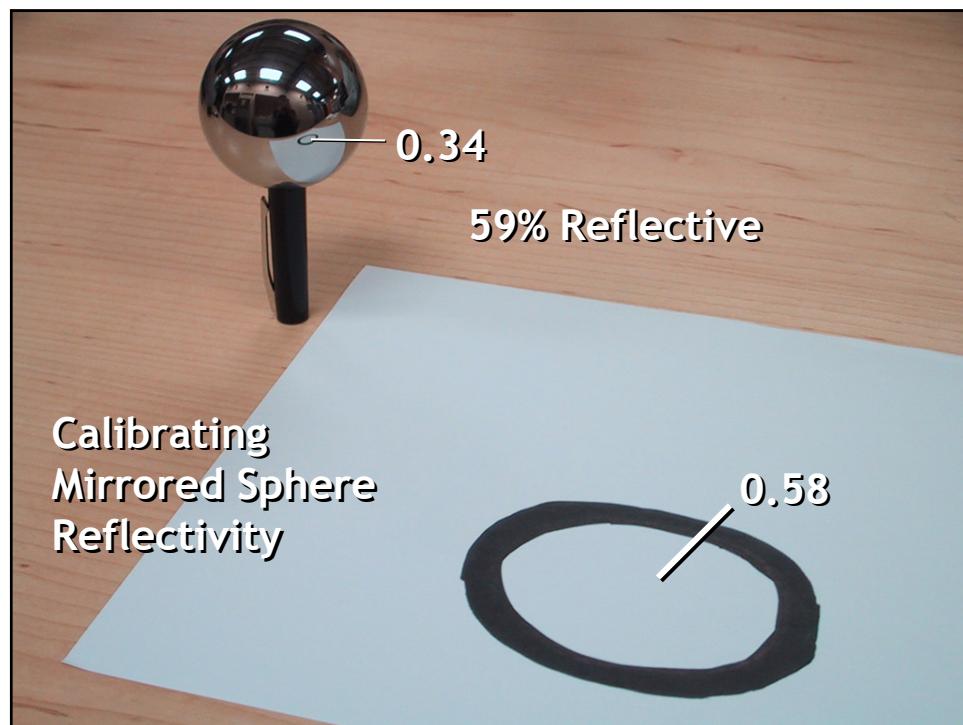
Sources of Mirrored Balls

- 2-inch chrome balls ~ \$20 ea.
 - McMaster-Carr Supply Company
www.mcmaster.com
- 6-12 inch large gazing balls
 - Baker's Lawn Ornaments
www.bakerslawnorn.com
- Hollow Spheres, 2in – 4in
 - Dube Juggling Equipment
www.dube.com
- **FAQ** on www.hdrshop.com



 SIGGRAPH 2008





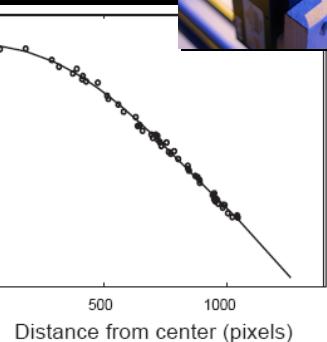
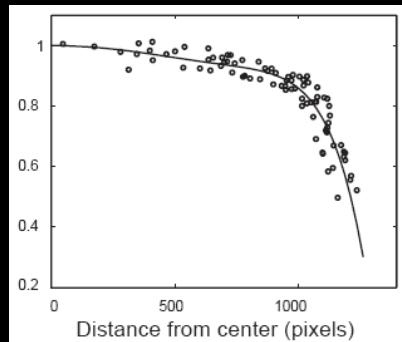
Mirrored ball, fisheyes, stitched photos, or scanning?



SIGGRAPH 2008



Fisheye Lens Radial Falloff Sigma 8mm Canon/Nikon



Sigma 8mm
fisheye lens

SIGGRAPH 2008



Mirrored ball, fisheyes, stitched photos, or scanning?



 SIGGRAPH 2008



Tiled Photographs – Nodal Acquisition Rig



See also www.kaidan.com

 SIGGRAPH 2008

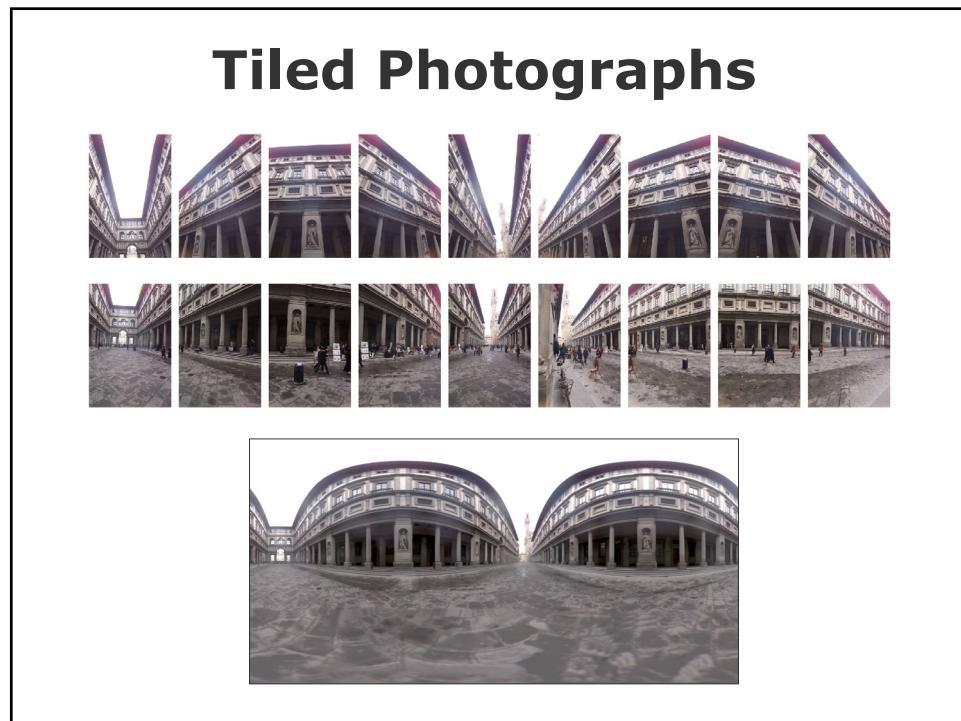


Stitching HDRI with Realviz Stitcher

<http://www.gregdowning.com/HDRI/stitched/>



The screenshot shows the Realviz Stitcher software interface. On the left, a 4x4 grid of images displays a scene of a restaurant with tables and chairs, with green lines indicating the stitching boundaries. Two circular insets provide close-up views: one labeled "Shadow Detail" showing a lamp fixture, and another labeled "Highlight Detail" showing a colorful building facade. On the right, the software's control panel is visible, featuring various tools and a preview window showing the stitched result.



Mirrored ball, fisheyes, stitched photos, or scanning?

The collage illustrates four different approaches to capturing panoramic or spherical images:

- A mirrored ball reflection showing a distorted view of a building.
- A wide-angle fisheye shot of a stone wall and a courtyard.
- A grid of 16 small images showing a building's facade from multiple angles, used for stitching.
- A professional scanning panoramic camera mounted on a tripod.

SIGGRAPH 2008

Scanning Panoramic Cameras (Panoscan, Spheron)

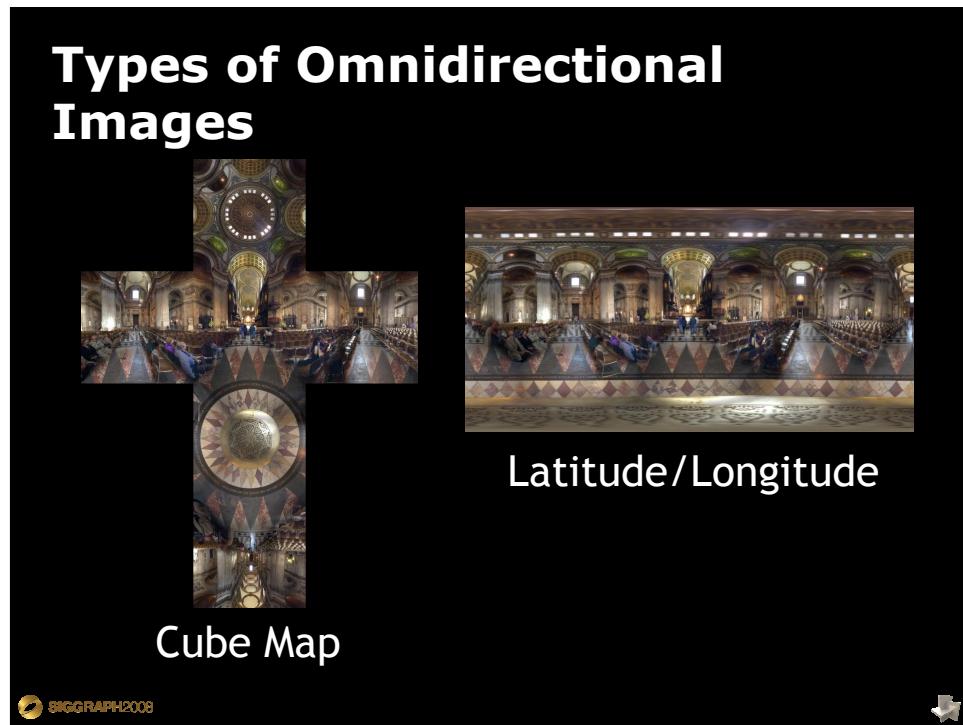
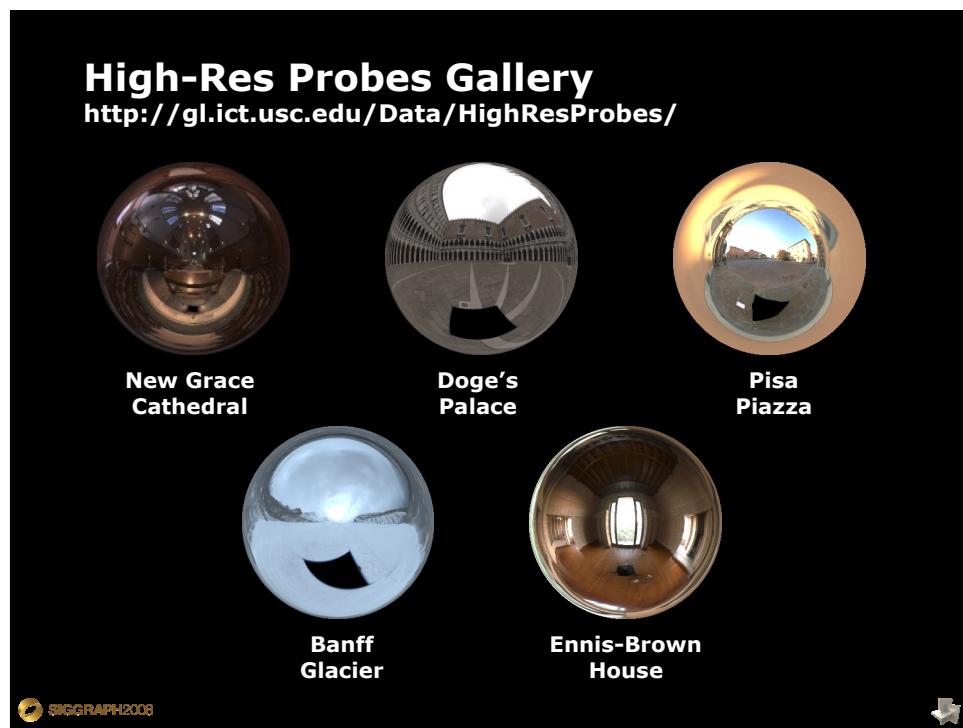
Pros:

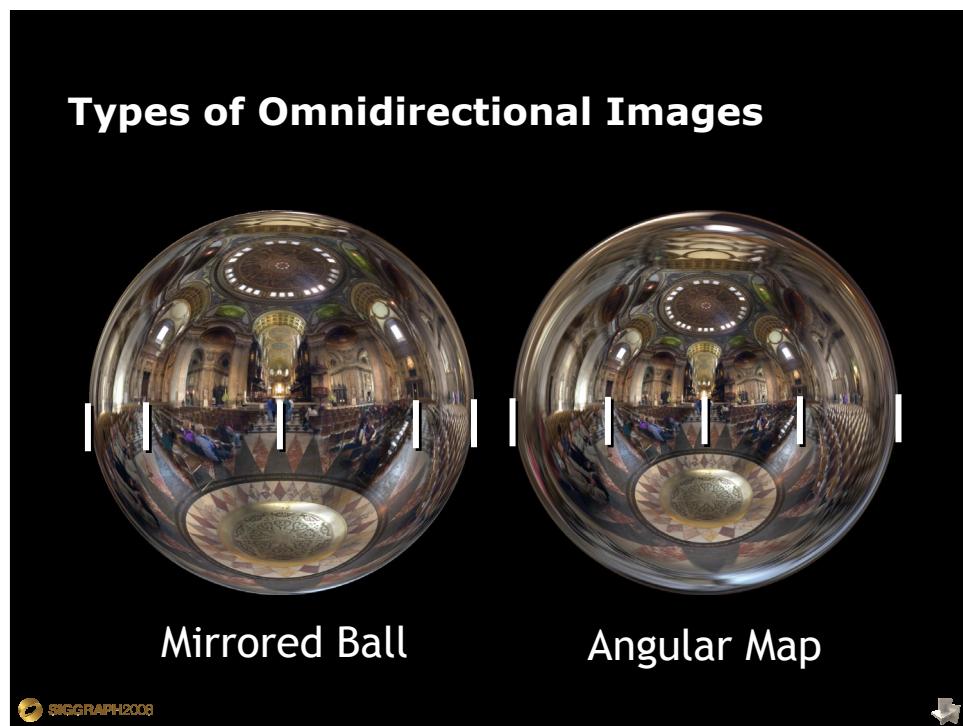
- very high res (10K x 7K+)
- Full sphere in one scan – no stitching
- Good dynamic range, some are HDR

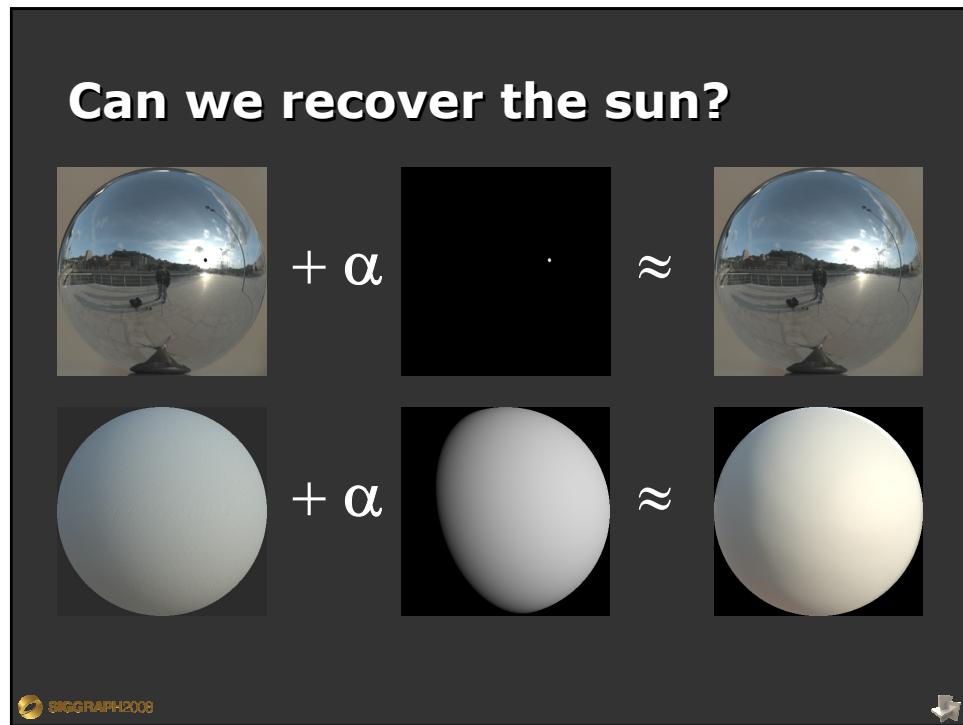
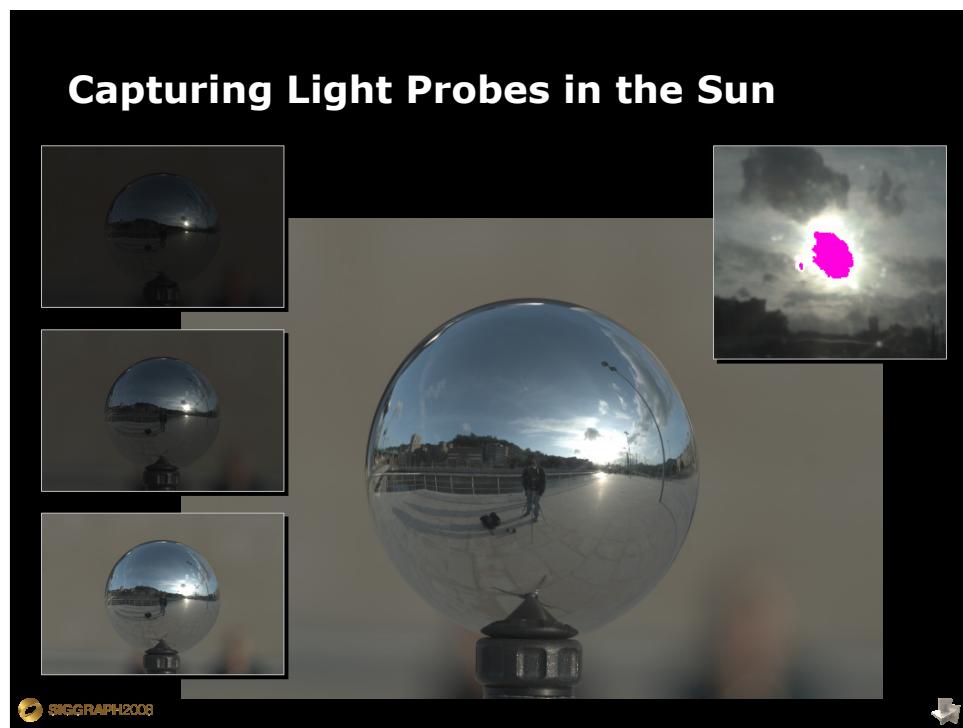
Issues:

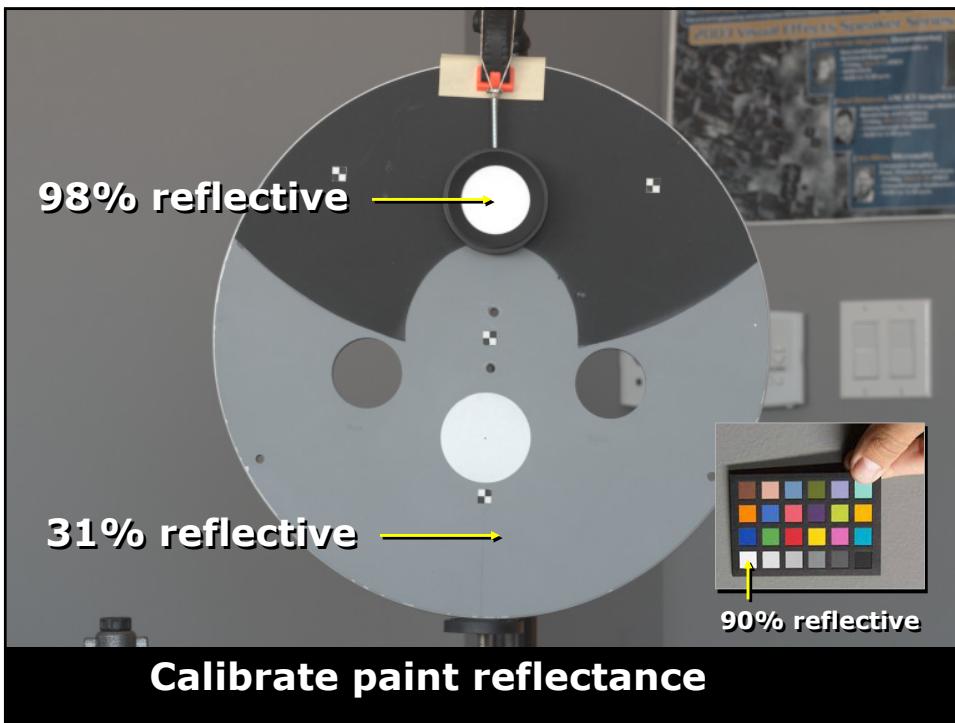
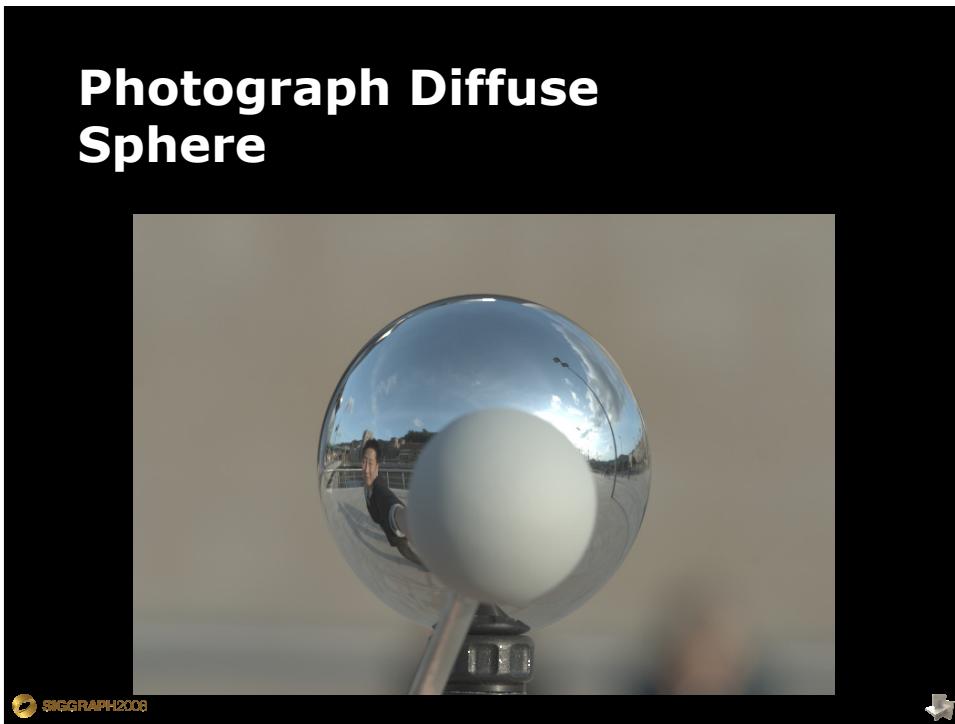
- More expensive
- Scans take a while

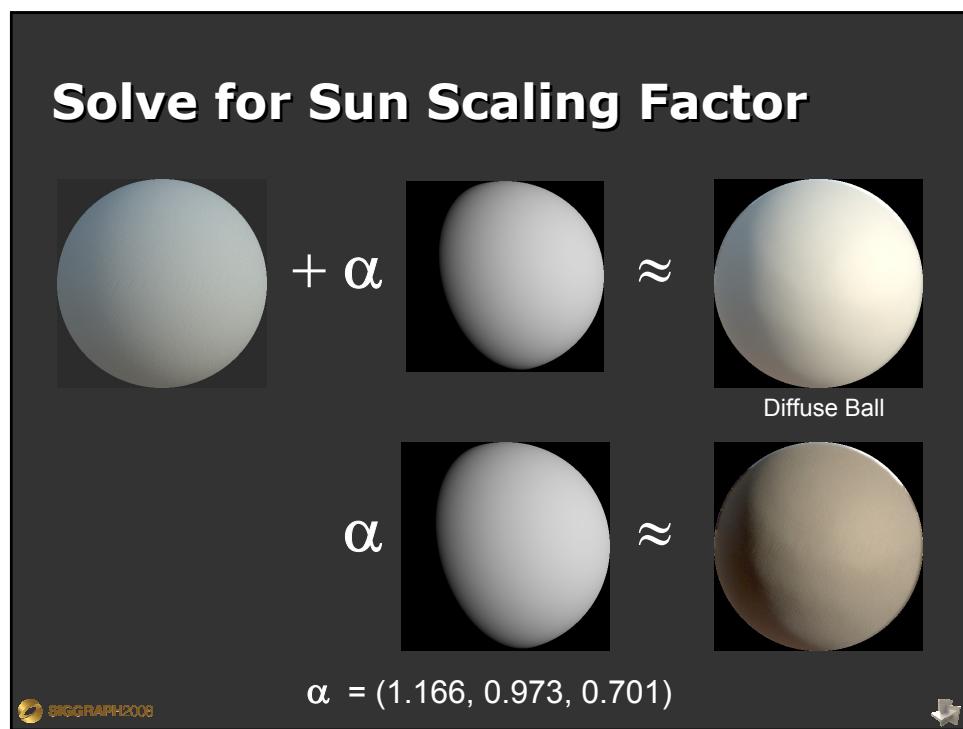
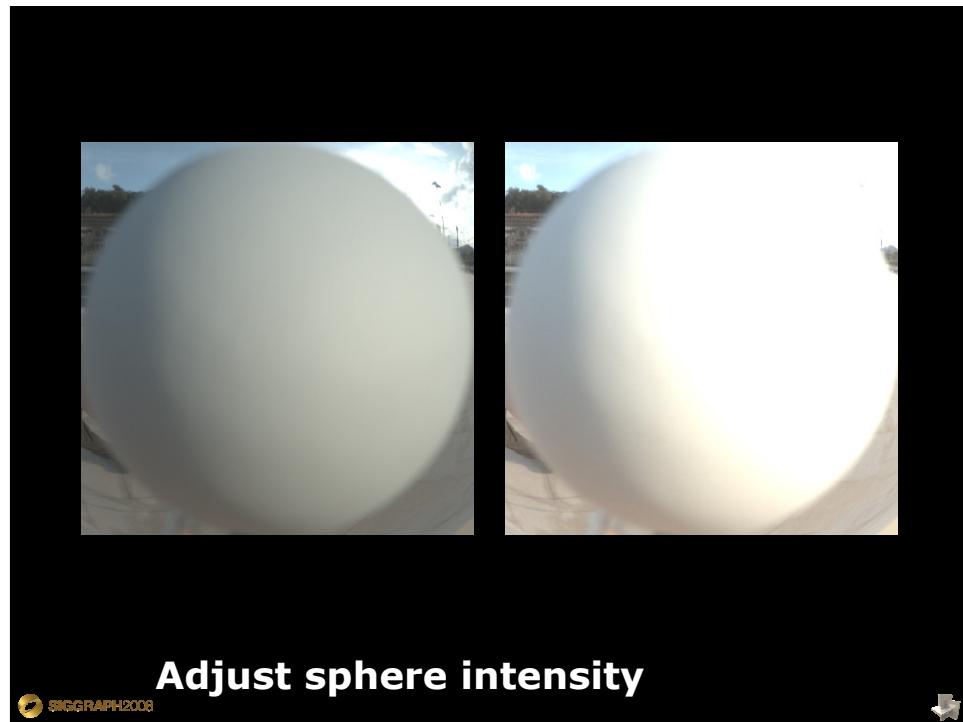
SIGGRAPH 2008

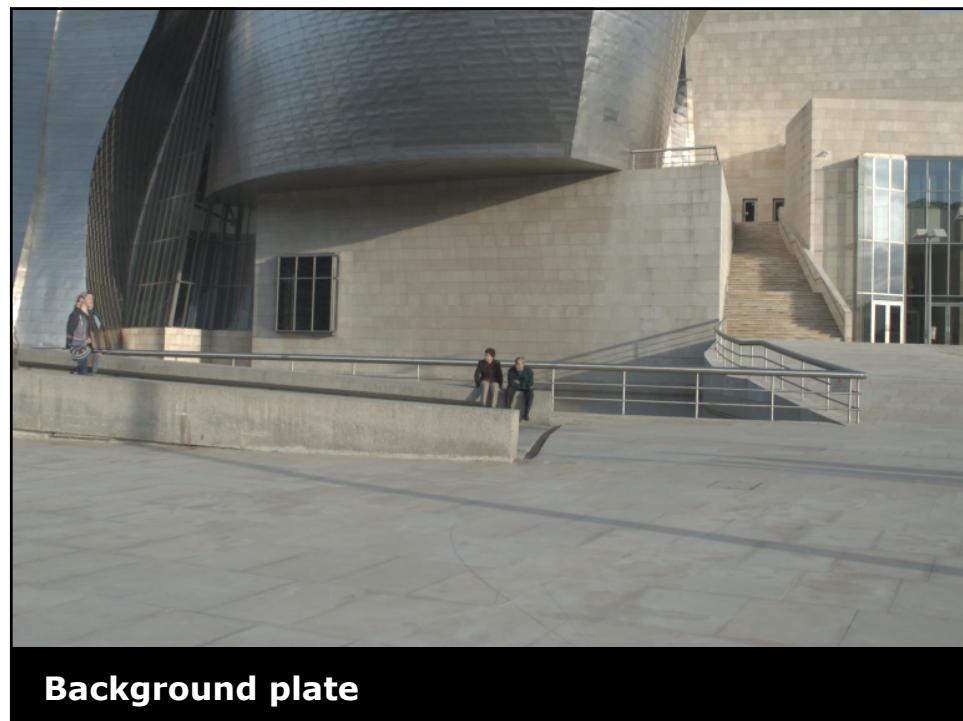
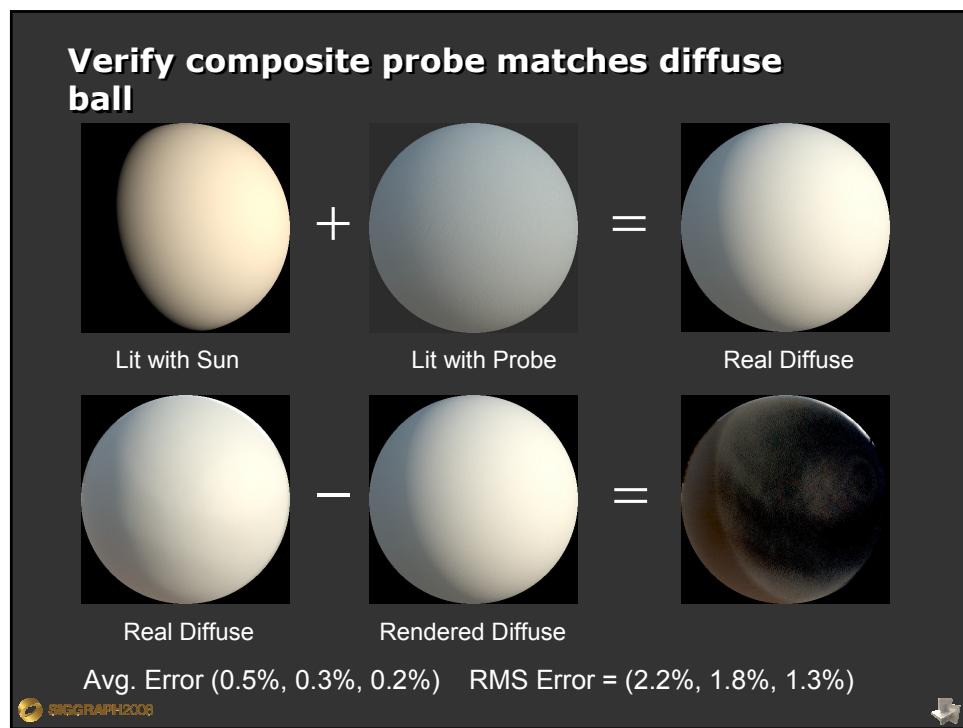




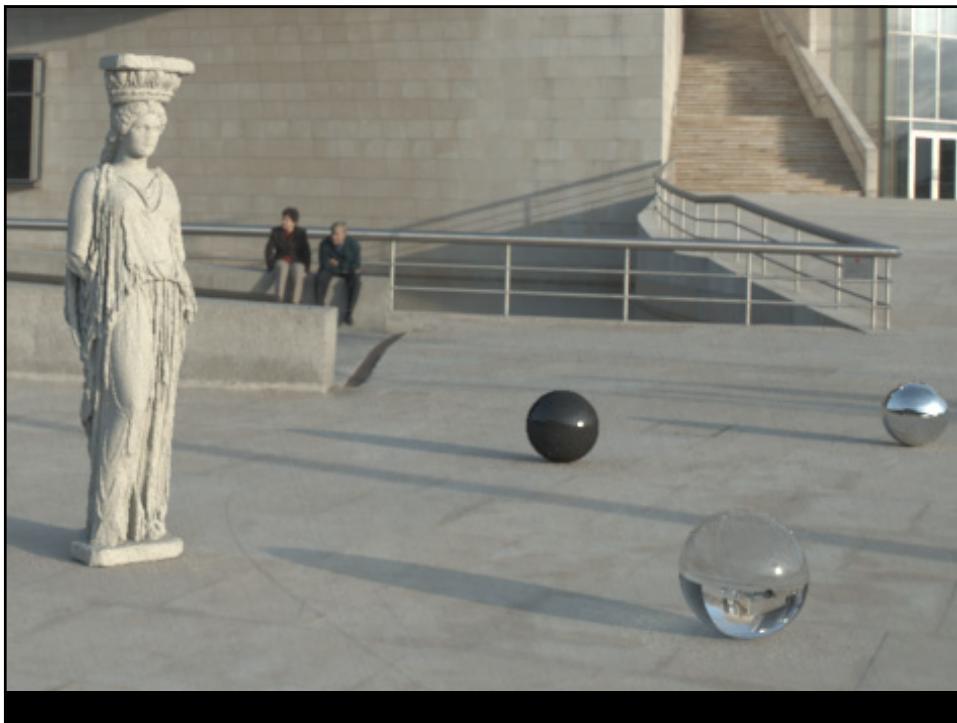
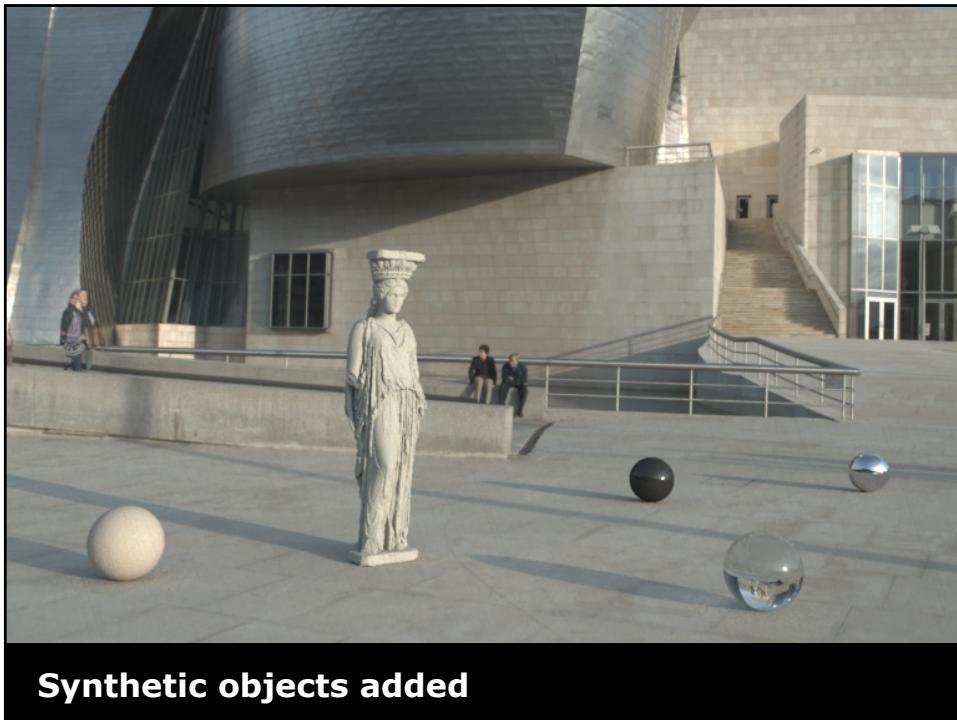








SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)





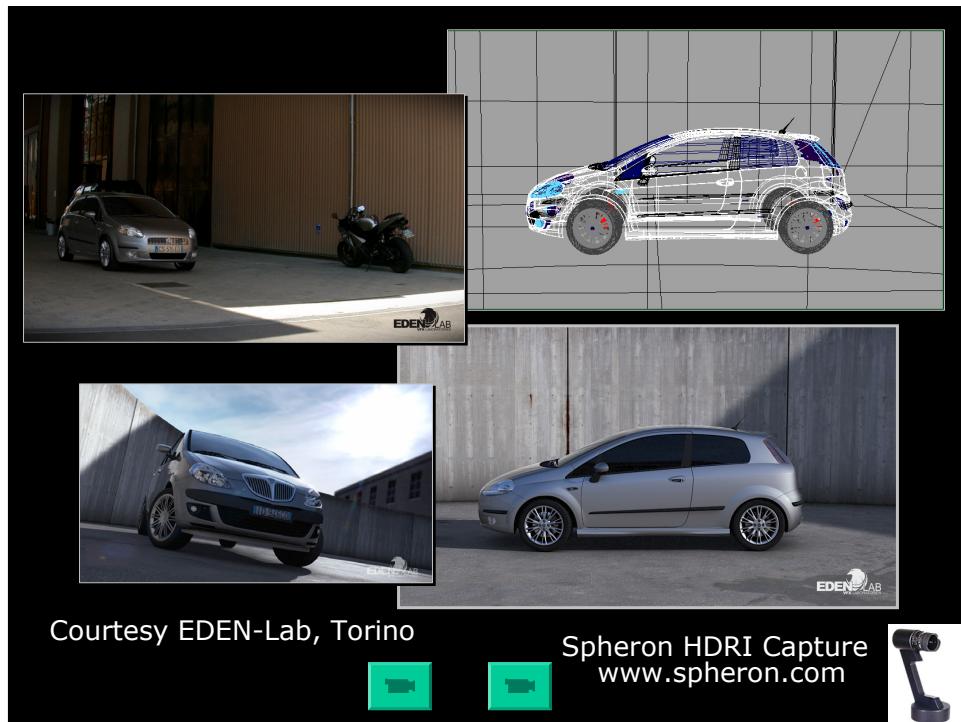
**Image-Based Lighting
and Relighting**

Paul Debevec
University of Southern California
Institute for Creative Technologies
Graphics Laboratory

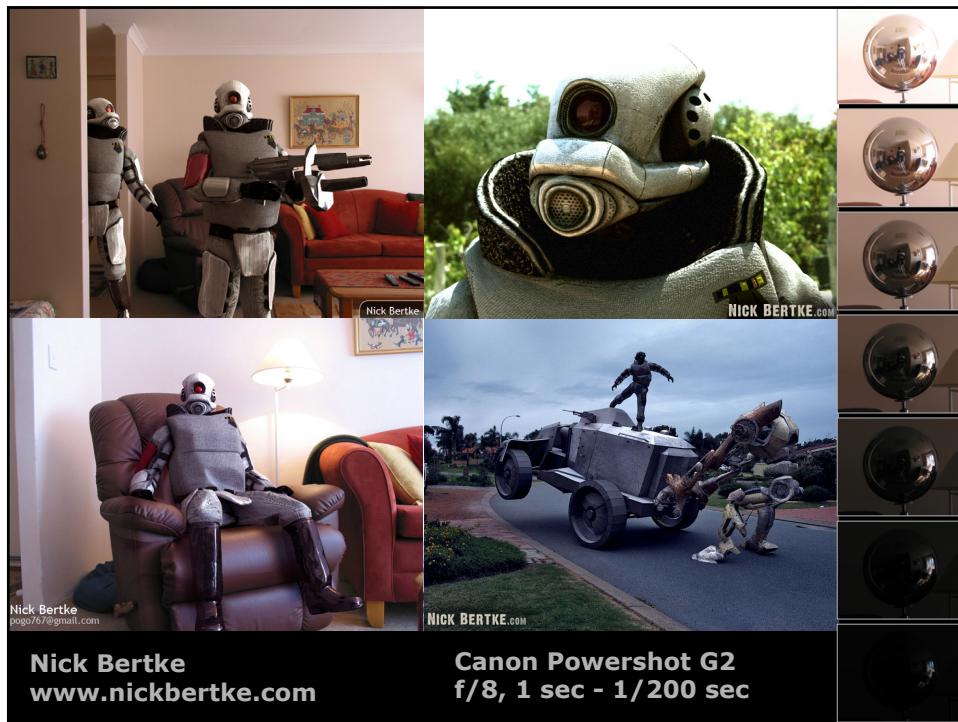
SIGGRAPH 2008 Class
High-Dynamic-Range Imaging & Image-Based Lighting
August 2008

www.debevec.org / gl.ict.usc.edu

SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)



SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)



Nick Bertke
www.nickbertke.com
HDRI using 3DsMax, VRay, Brazil

DaikenTana's
Guide to HDR Lighting
Written by Nick Bertke

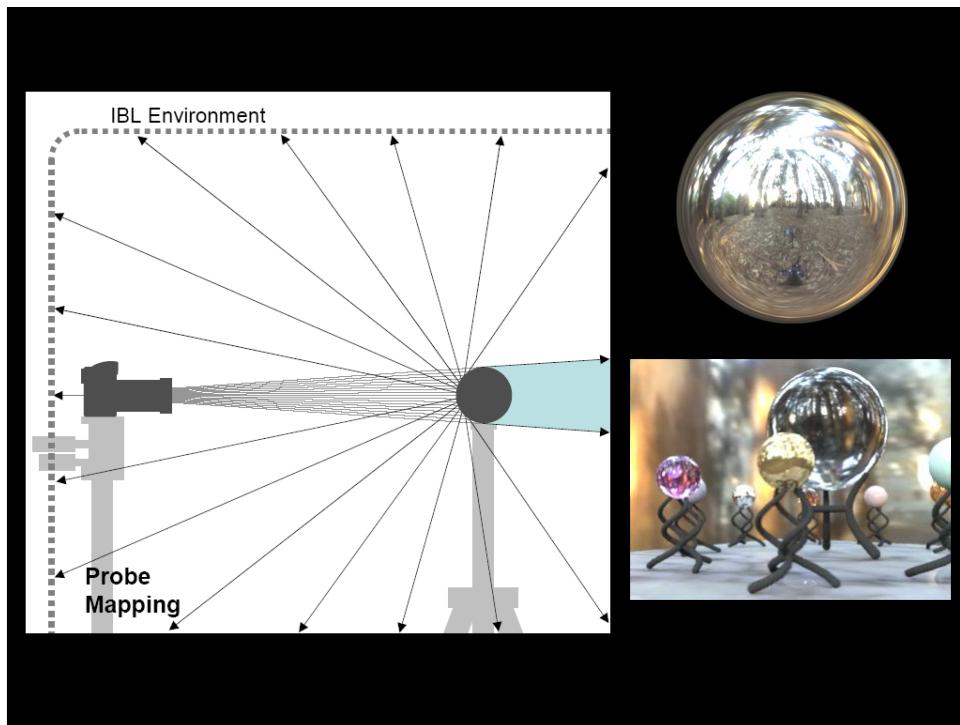
Nick Bertke
nick@bertke.com

HDR HIGH DYNAMIC RANGE (6 Photographs)						
LDR LOW DYNAMIC RANGE (1 Photograph)						

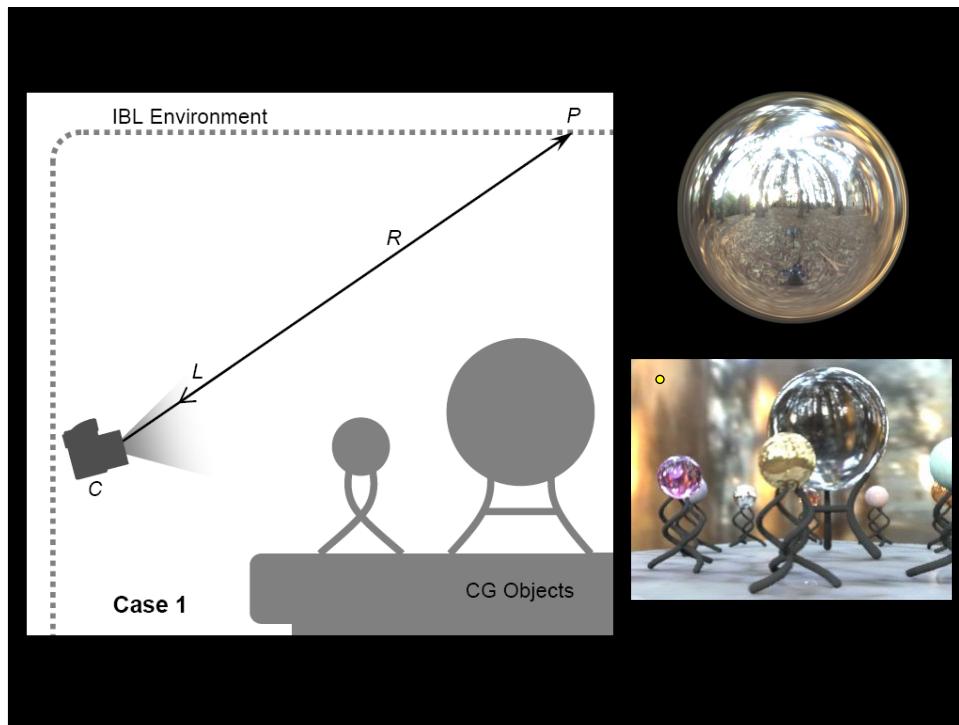
SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)



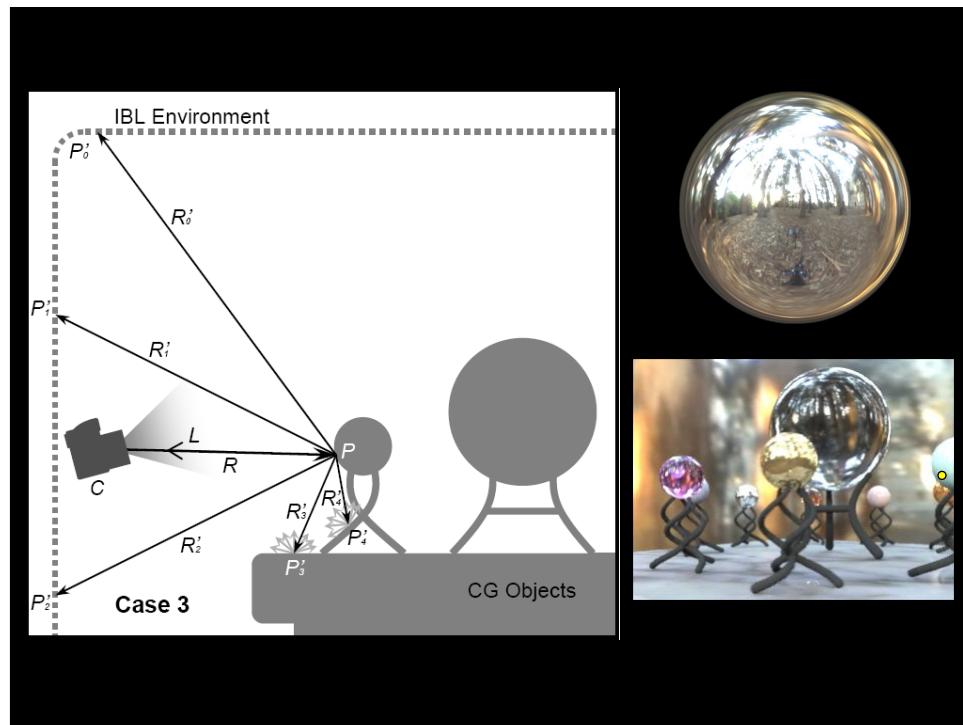
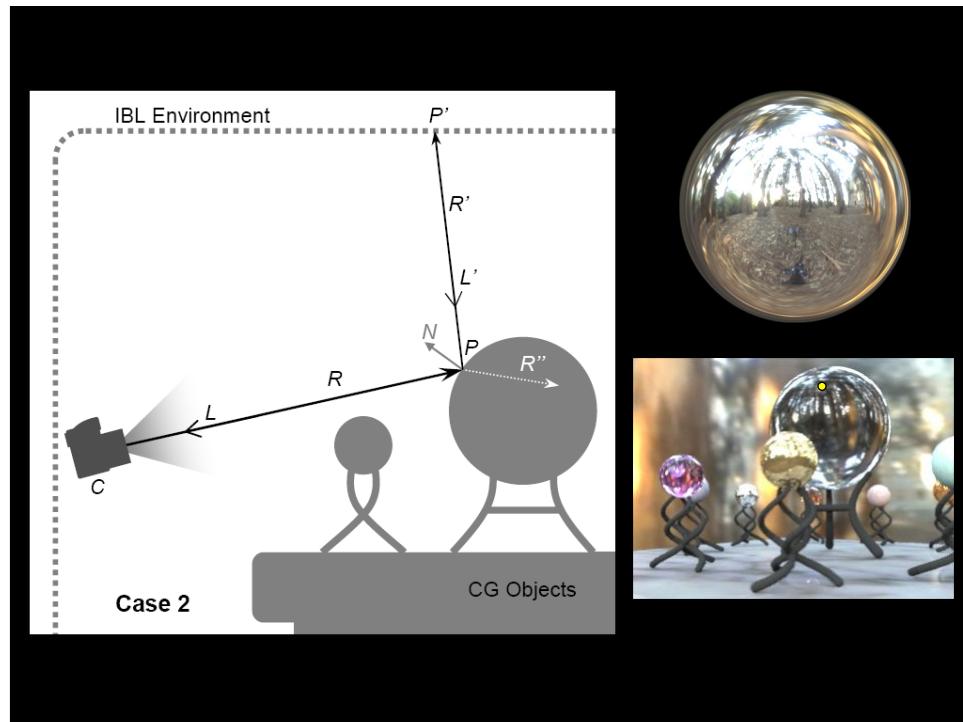
SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)



SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)



SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)



Direct HDR Capture of the Sun and Sky

Use Sigma 8mm fisheye lens and Canon EOS 1Ds to cover entire sky
Use 3.0 ND filter on lens back to cover full range of light



Stumpfel, Jones, Wenger, Tchou, Hawkins, and Debevec. "Direct HDR Capture of the Sun and Sky". To appear in Afrigraph 2004.

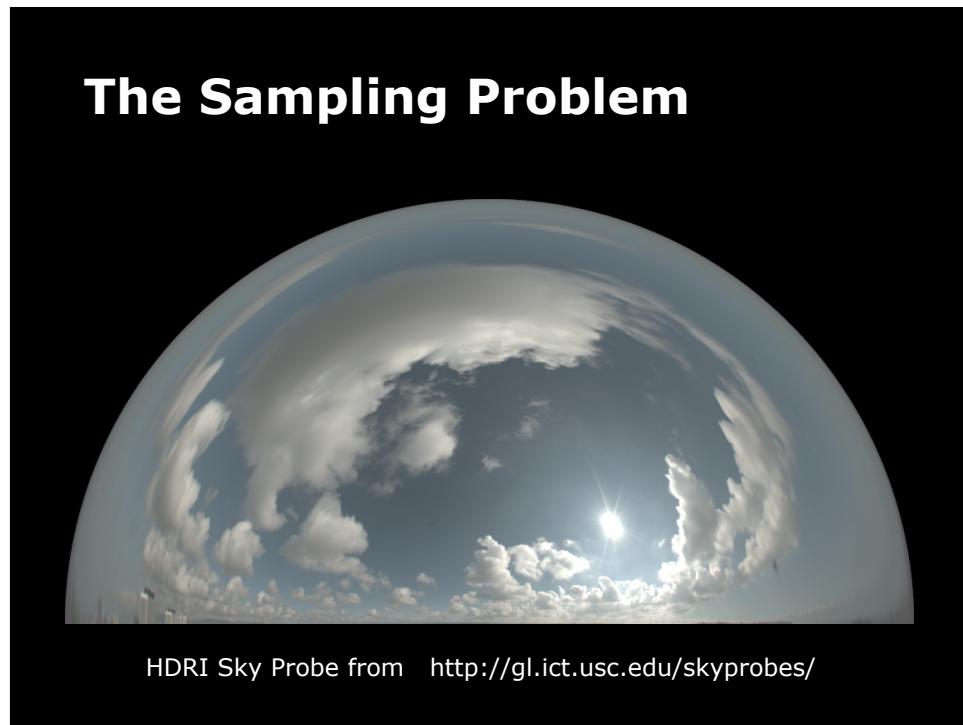
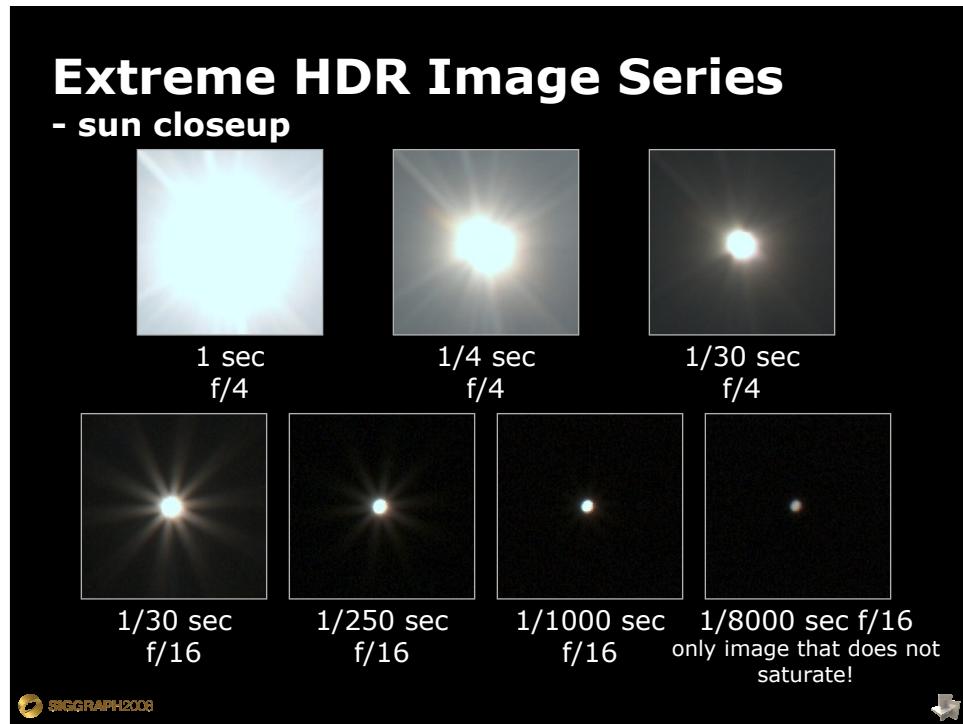
 SIGGRAPH 2008 

Extreme HDR Image Series

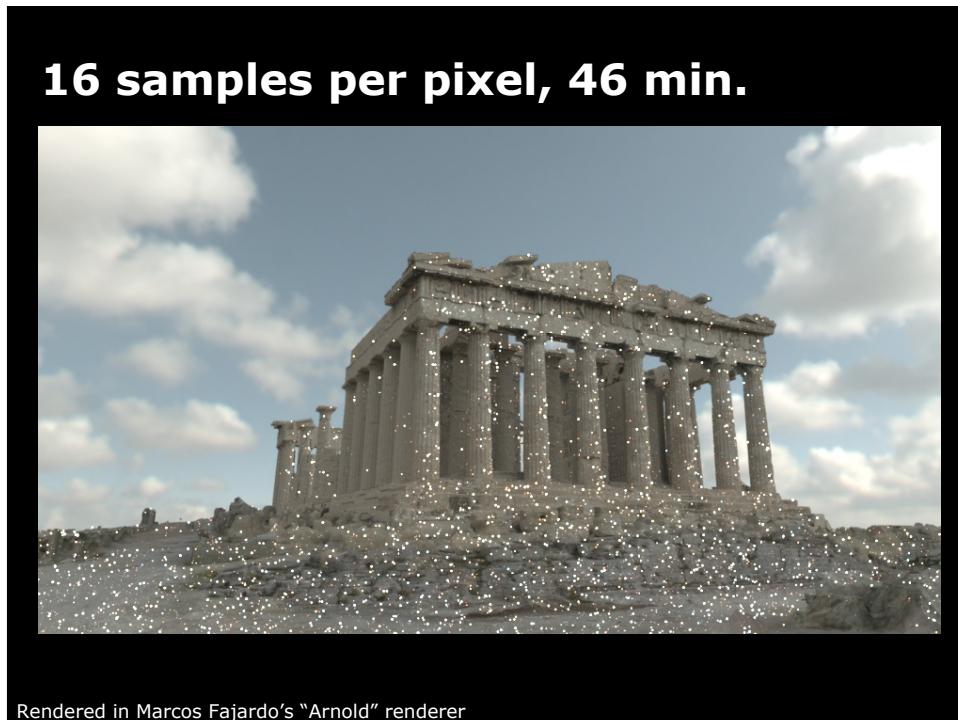


Exposure Time	Aperture
1 sec	f/4
1/4 sec	f/4
1/30 sec	f/4
1/30 sec	f/16
1/250 sec	f/16
1/1000 sec	f/16
1/8000 sec	f/16

 SIGGRAPH 2008 



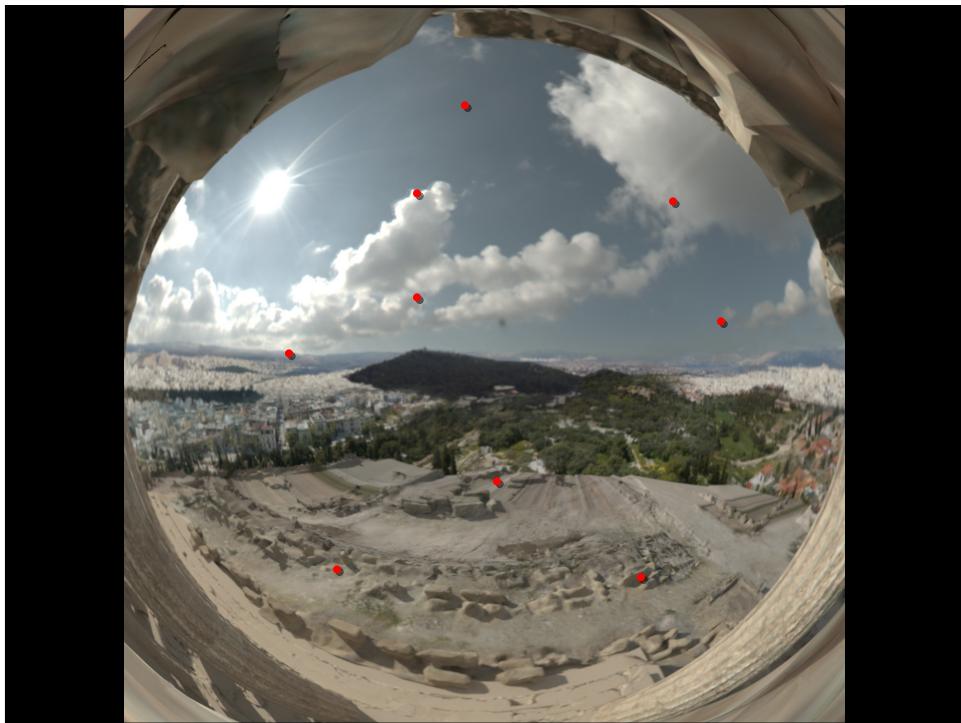
SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)



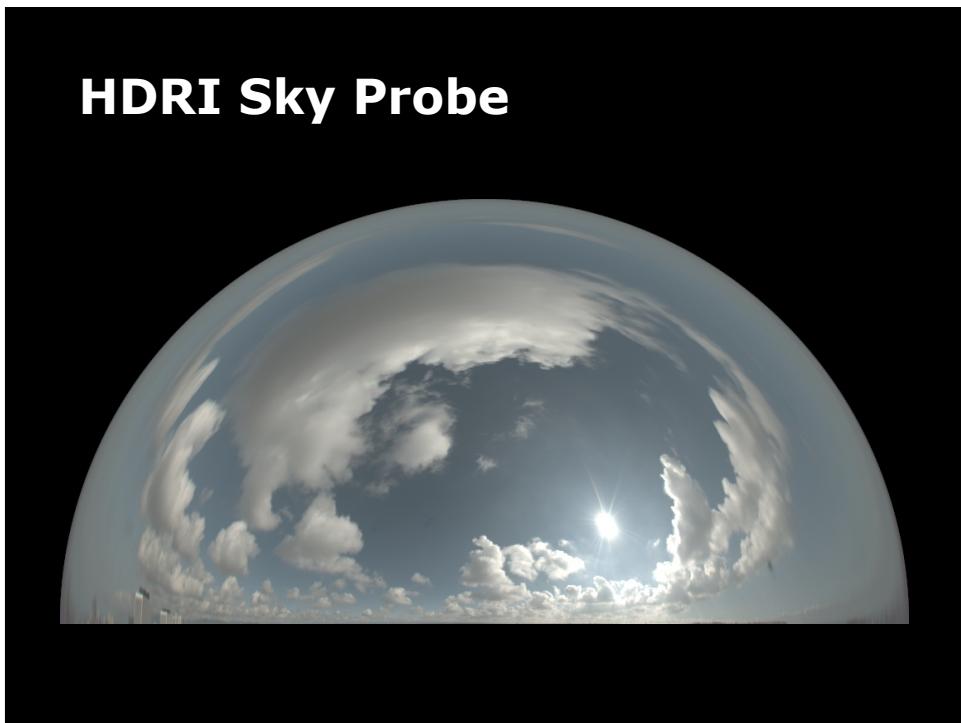
SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)

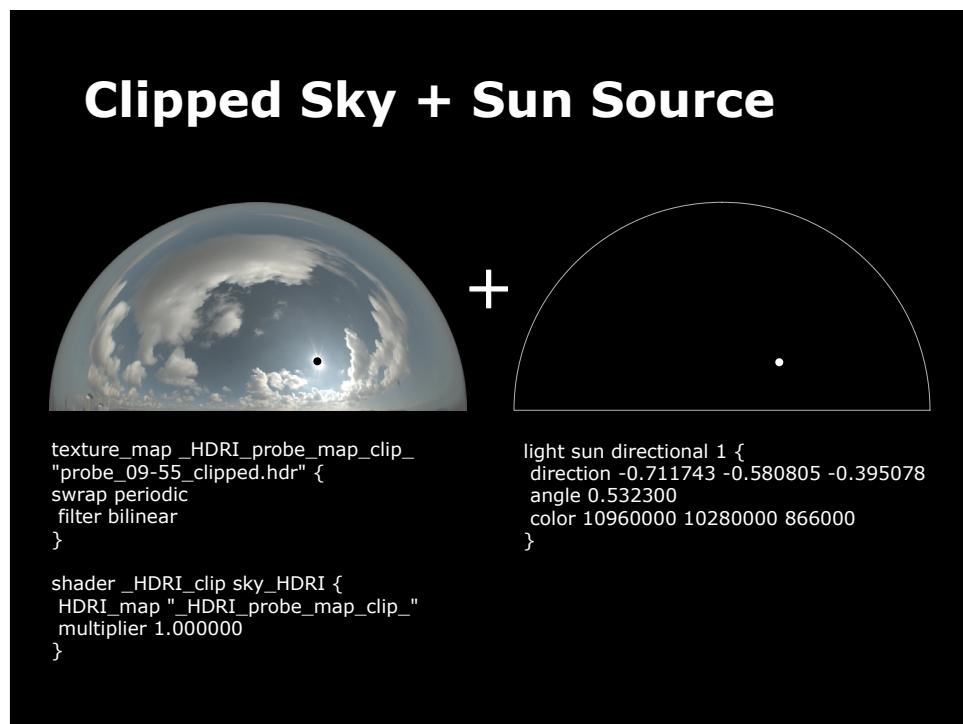


SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)

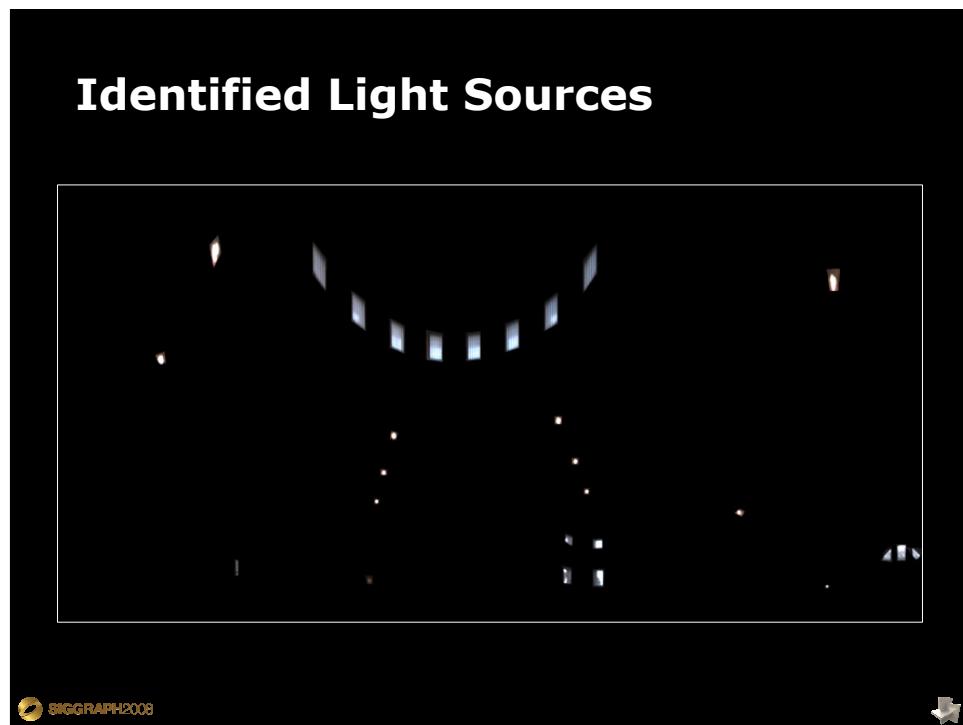


SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)

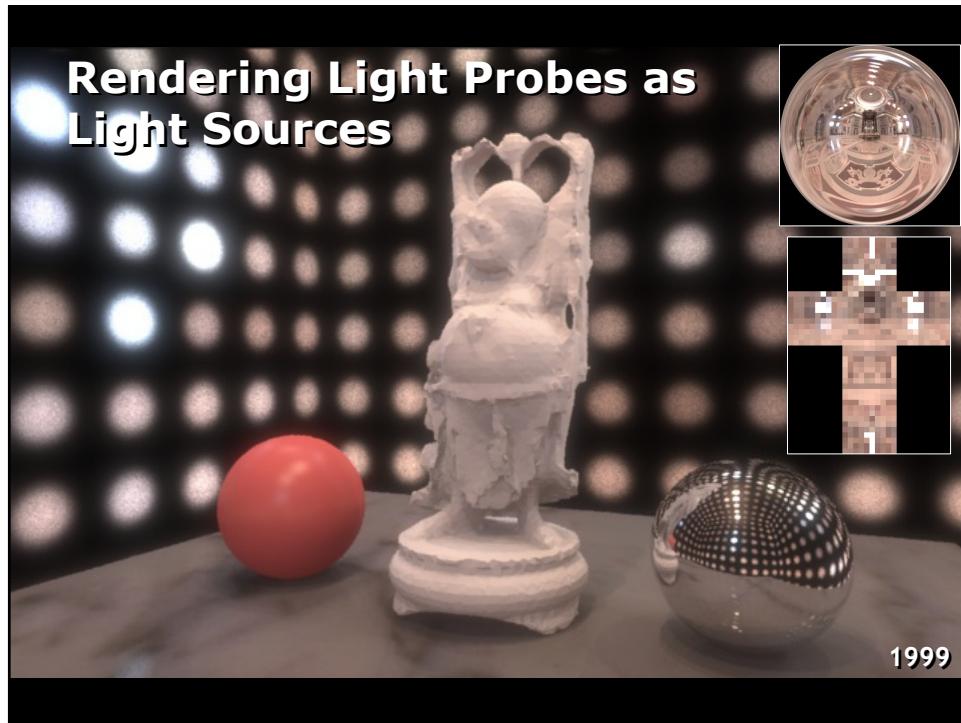
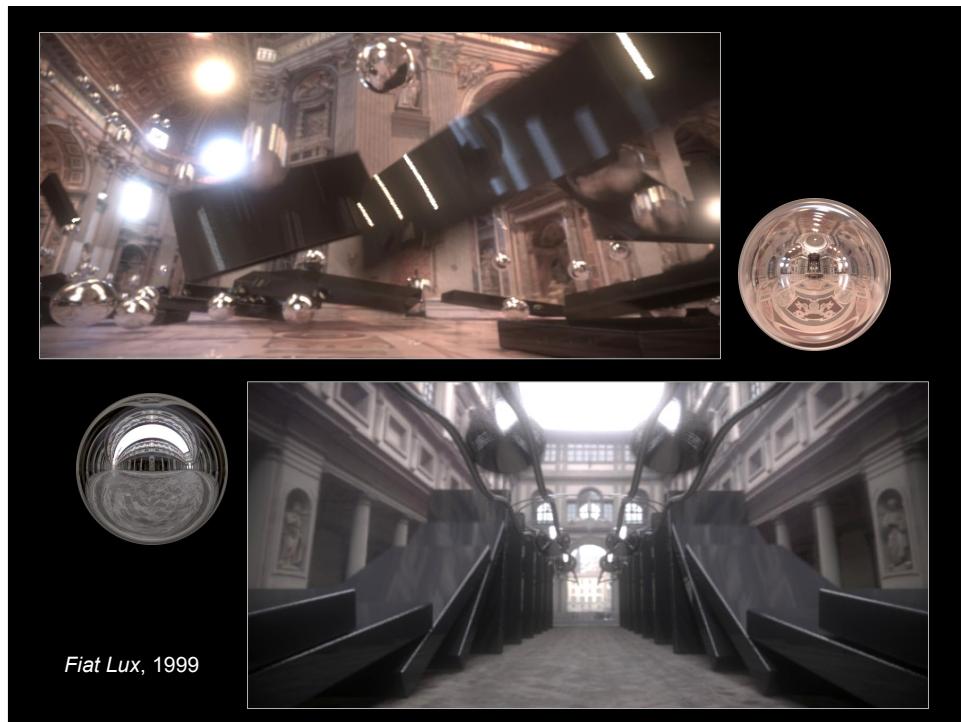








SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)



A Median Cut Algorithm for Light Source Placement

Inspired by Heckbert
"Color Image Quantization for Frame Buffer Display",
SIGGRAPH 82



1. Add the entire light probe image to the region list as a single region
2. For each region, subdivide along the longest dimension such that its light energy is divided evenly
3. If the number of iterations is less than n , return to step 2.

Paul Debevec, SIGGRAPH 2005 Poster

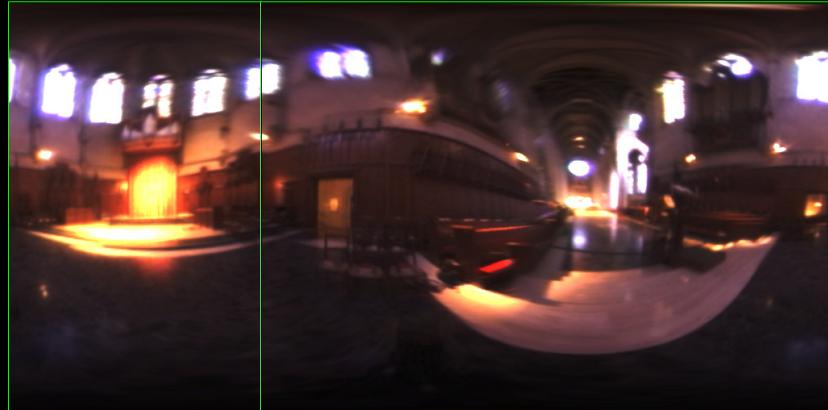
Median Cut Algorithm



1 region

Median Cut Algorithm

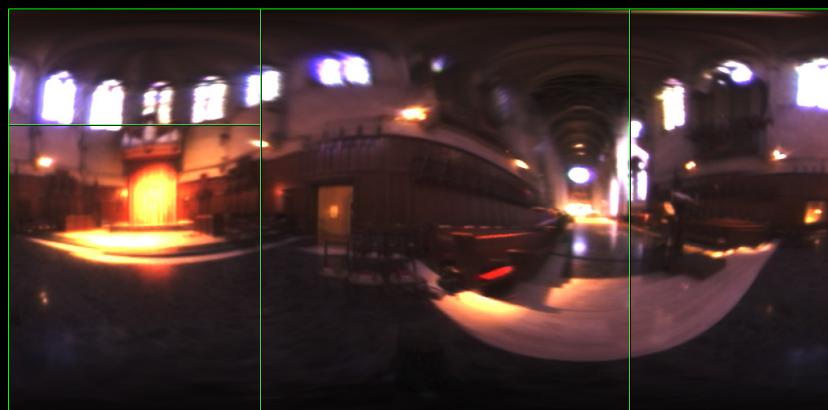


2 regions

 SIGGRAPH2008



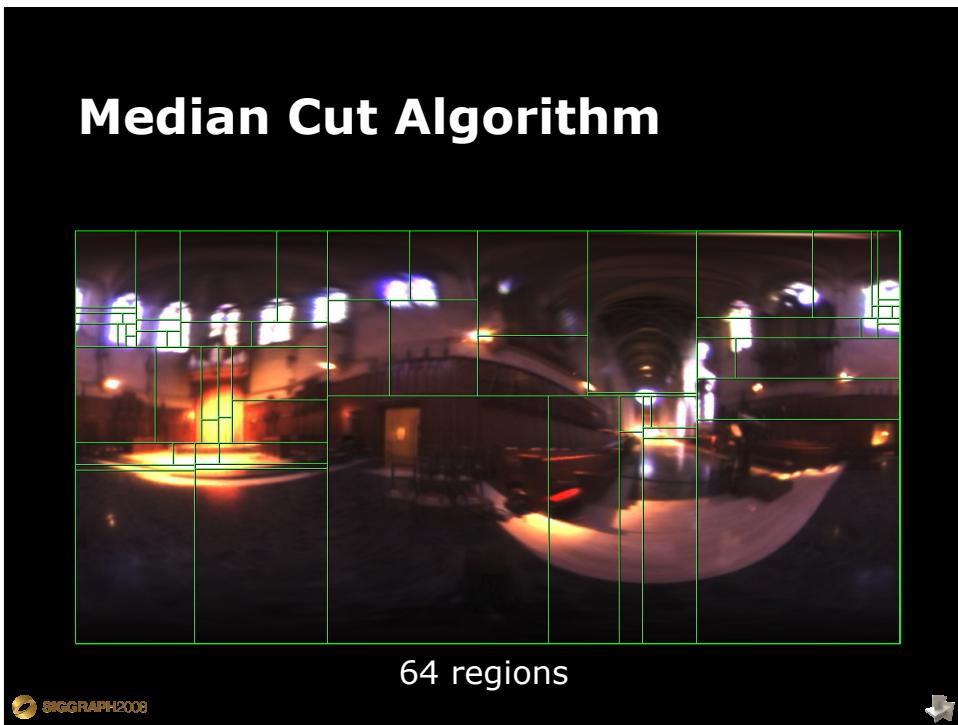
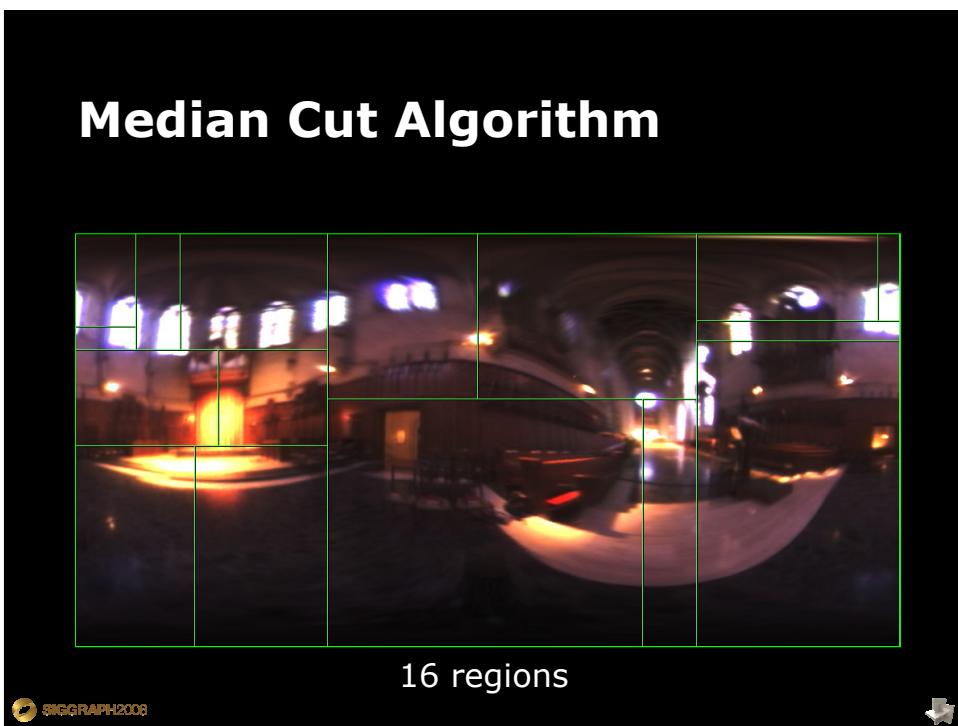
Median Cut Algorithm

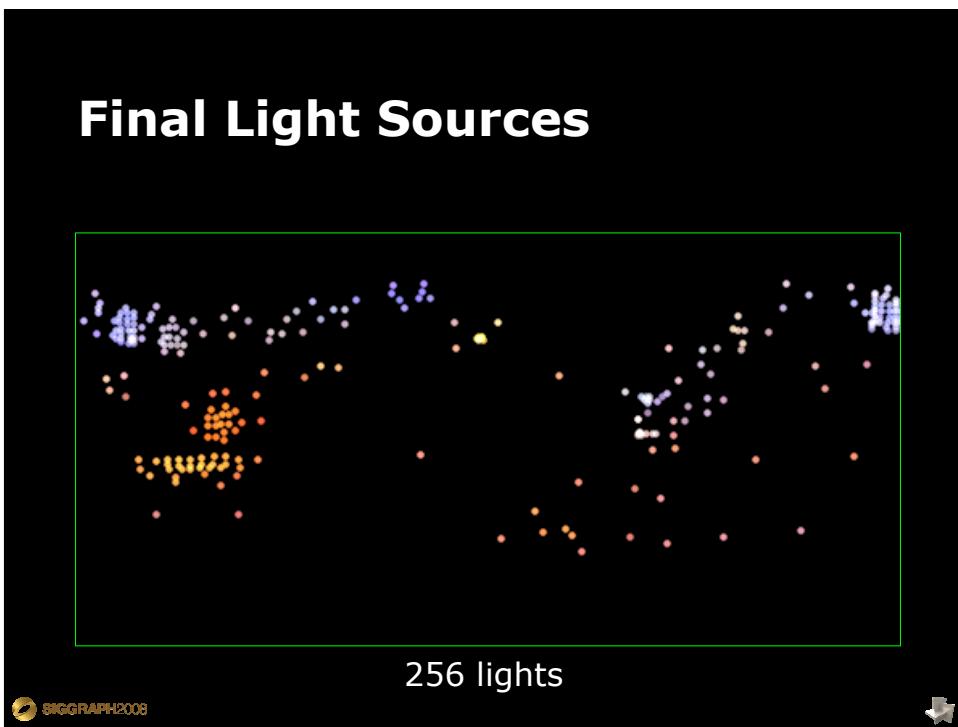
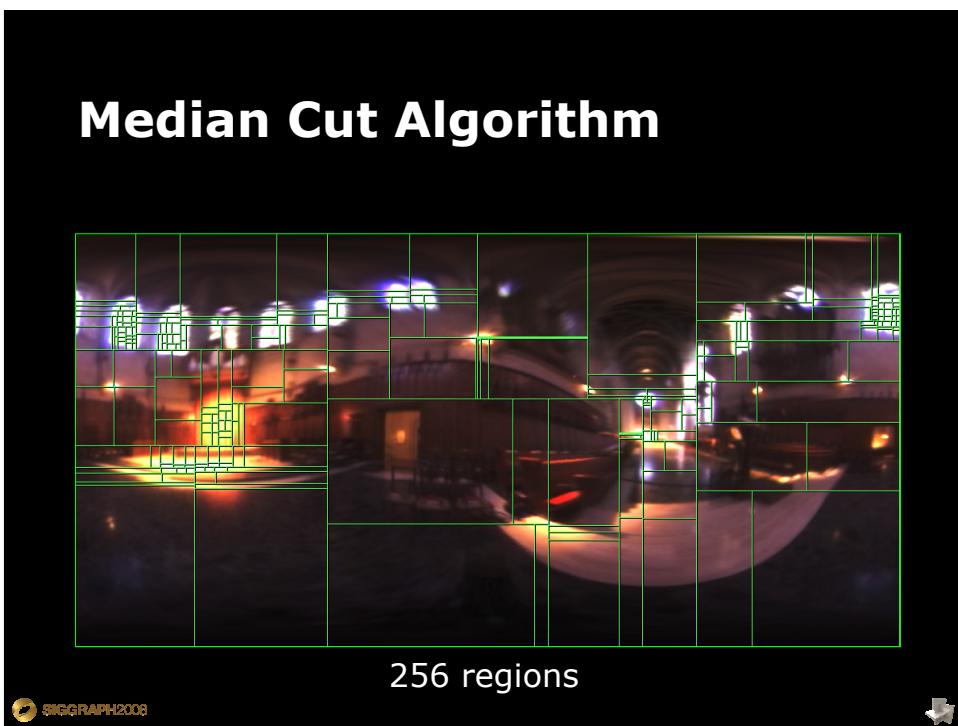


4 regions

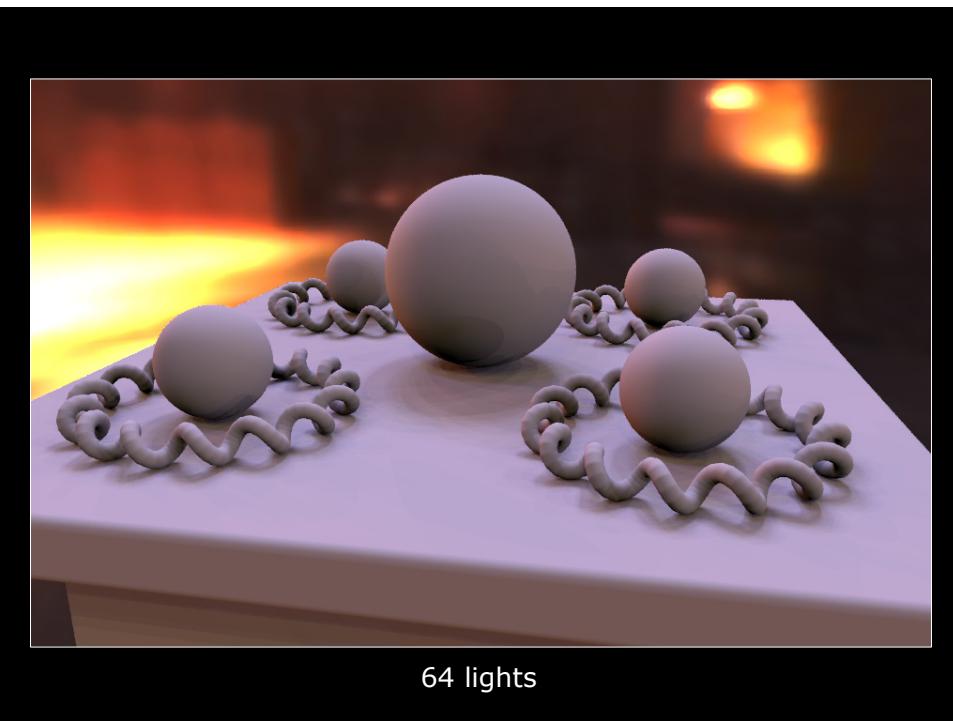
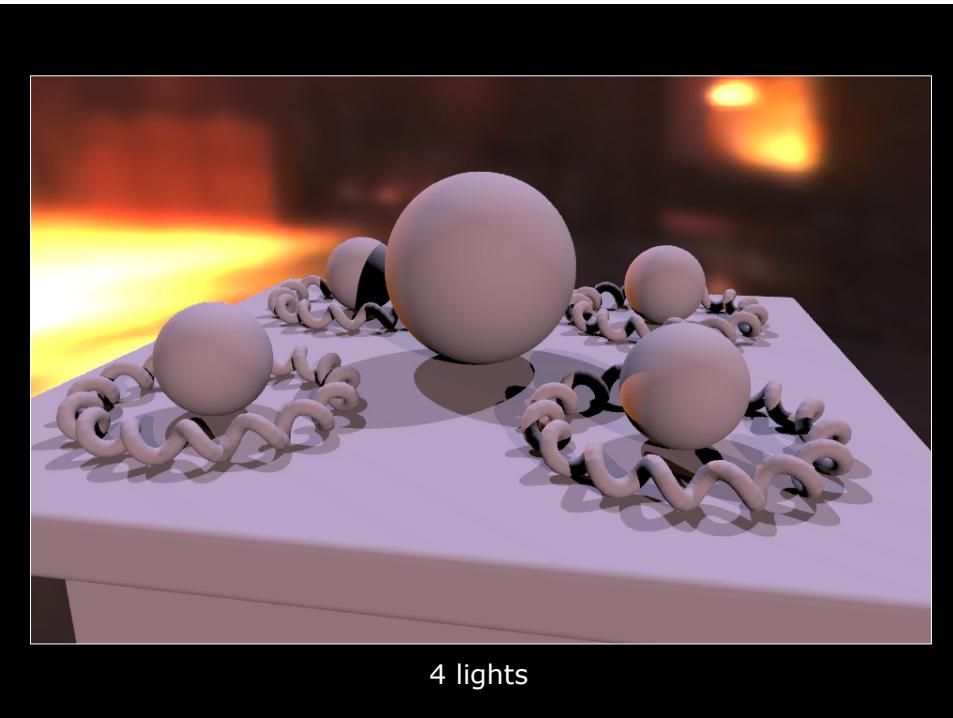
 SIGGRAPH2008



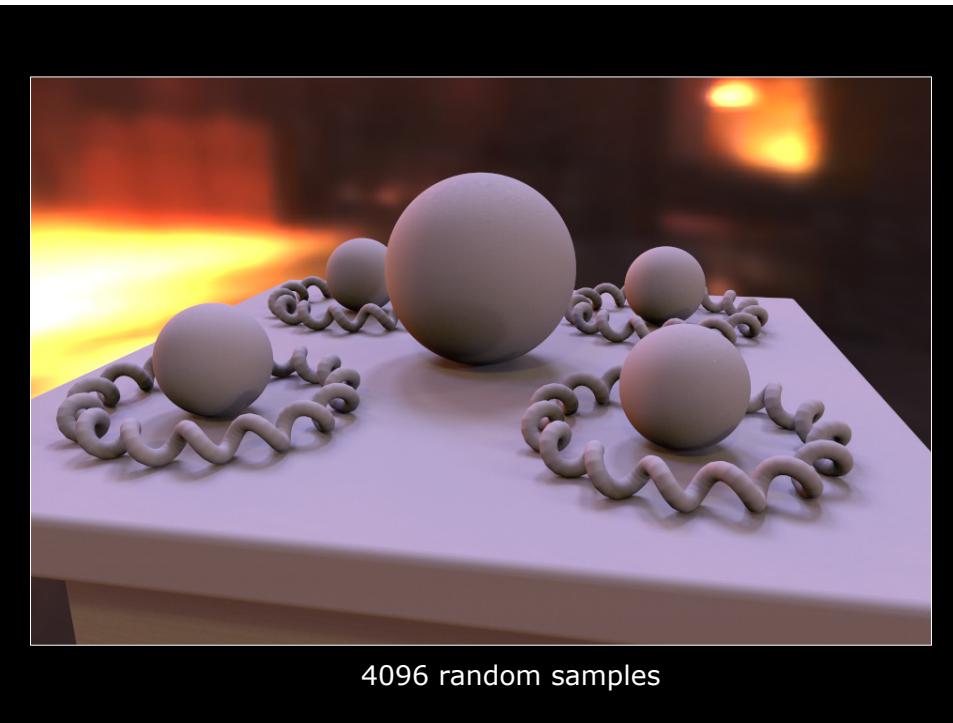
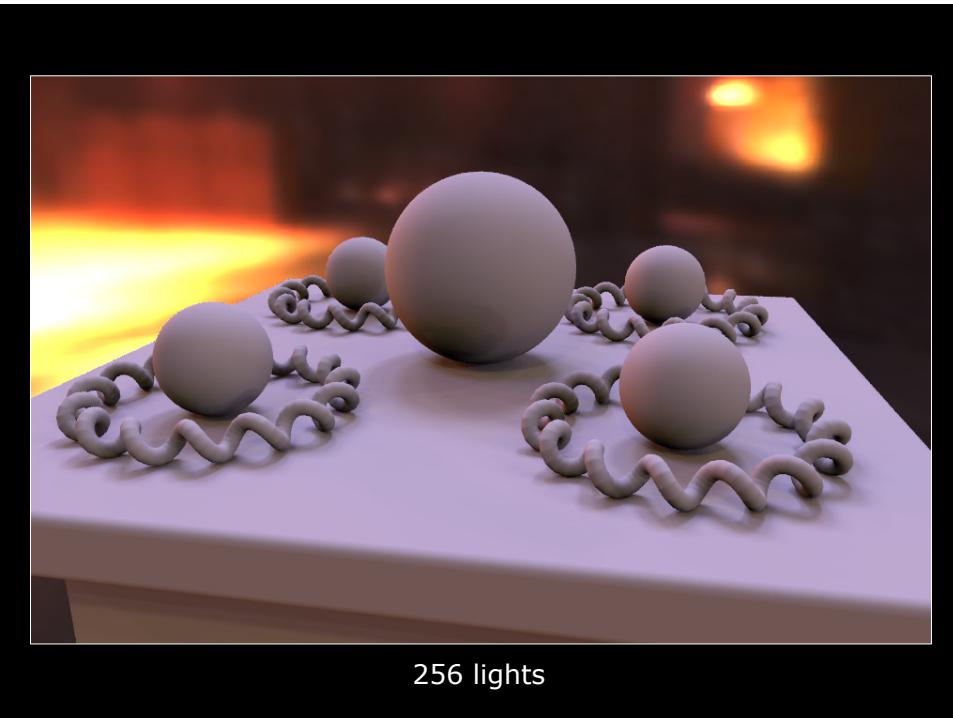


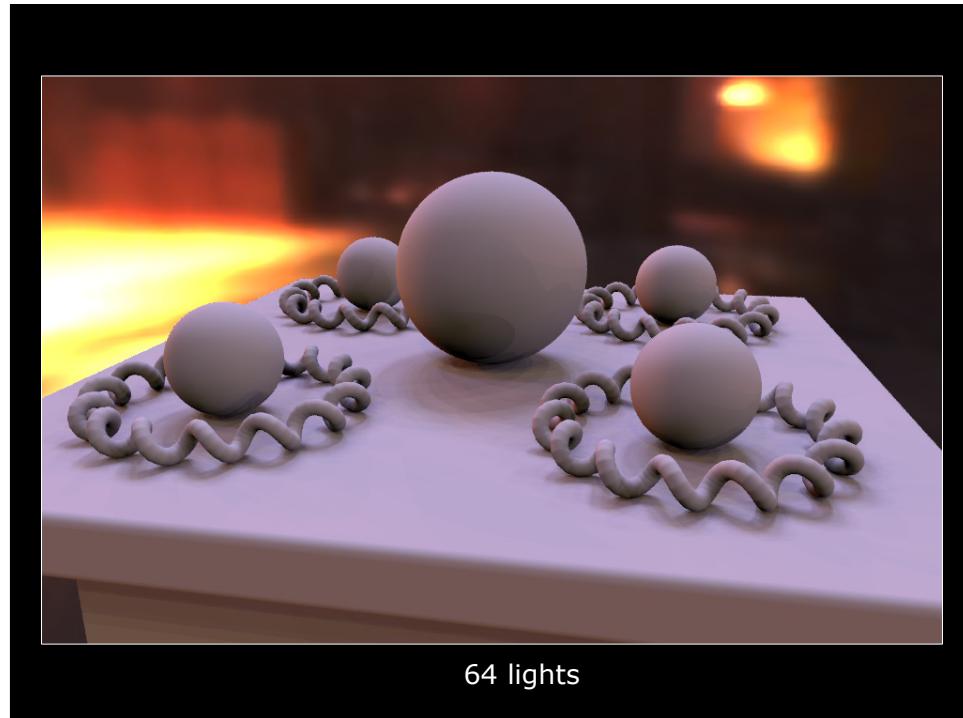


SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)



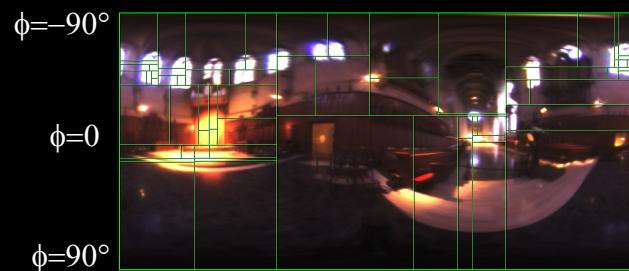
SIGGRAPH 2008 Class: HDRI and Image-Based Lighting
Image-Based Lighting (Paul Debevec)





Implementation Details

- Perform energy split decisions on a grayscale version of the image
- Can speed up region sums using summed area tables (Frank Crow, *Summed Area Tables for Texture Mapping*, SIGGRAPH 84)
- Must take into account area stretching near poles => multiply intensity and area widths by $\cos \phi$



Optional Improvement

Divide each region to minimize the sum of the variances within each region, rather than dividing the energy equally.

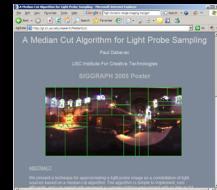
This can produce improved light clustering, but no longer keeps all lights at roughly equal energy.



Implementation Details at:
<http://gl.ict.usc.edu/Research/MedianCut/>

Links to implementations:

HDR Shop 1.0 plugin in Francesco Banterle's
Banty's Toolkit beta 1.1 => text file with
light parameters
<http://www.banterle.com/francesco/download.html>

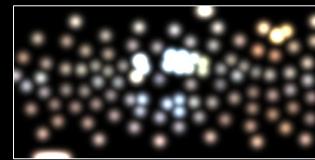


Jeremy Pronk's *Image Based Lighting Tools* for
Maya and Mental Ray
http://www.happiestdays.com/plugins_iblTools.php



Other Sampling Techniques

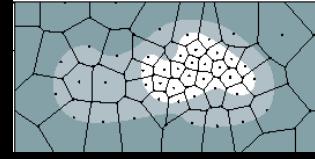
Cohen and Debevec
"LightGen" HDR Shop Plugin, 2001
www.hdrshop.com



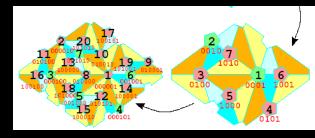
Kollig and Keller. *Efficient illumination by high dynamic range images.*
EGSR 2003
www.spheron.com



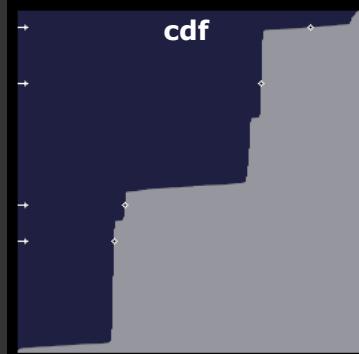
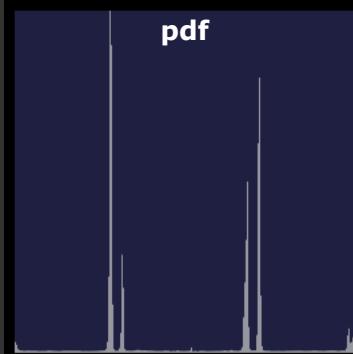
Agarwal, Ramamoorthi, Belongie, and Jensen. *Structured Importance Sampling of Environment Maps.*
SIGGRAPH 2003

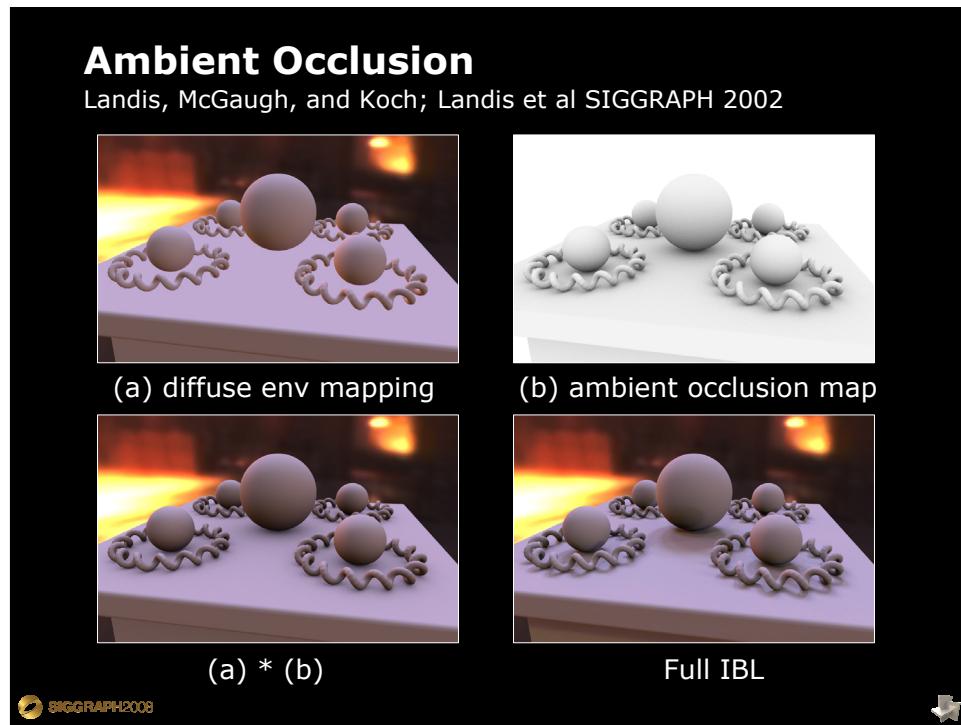
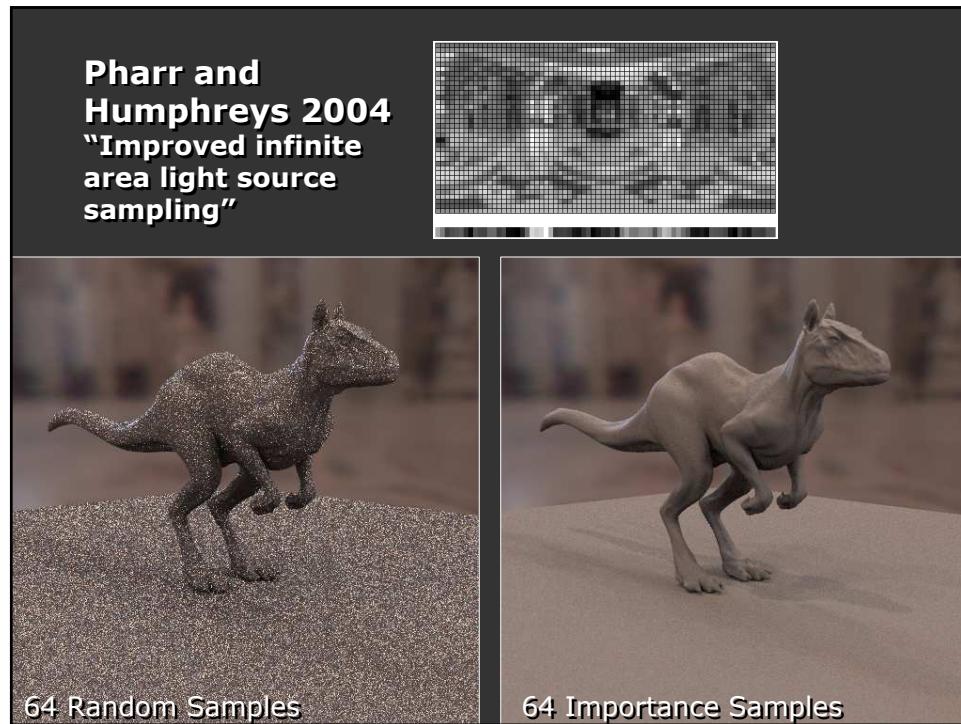


Ostromoukhov, Donohue, and Jodoin.
Fast Hierarchical Importance Sampling with Blue Noise Properties. SIGGRAPH 2004

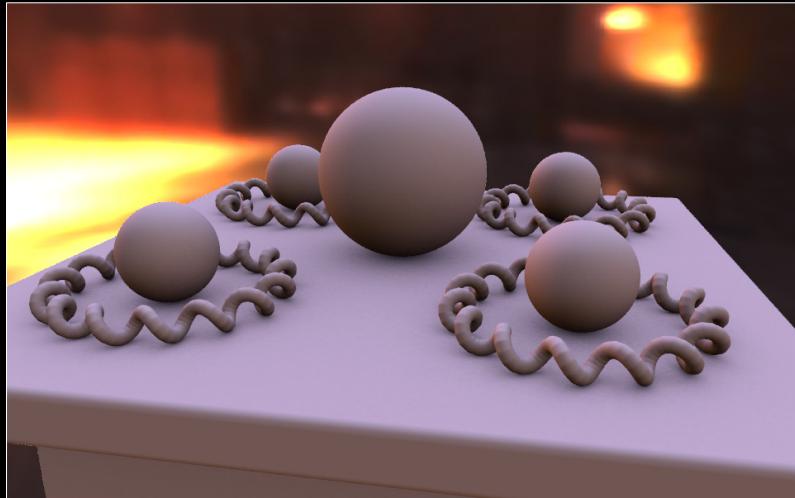


Importance Sampling





Ambient Occlusion

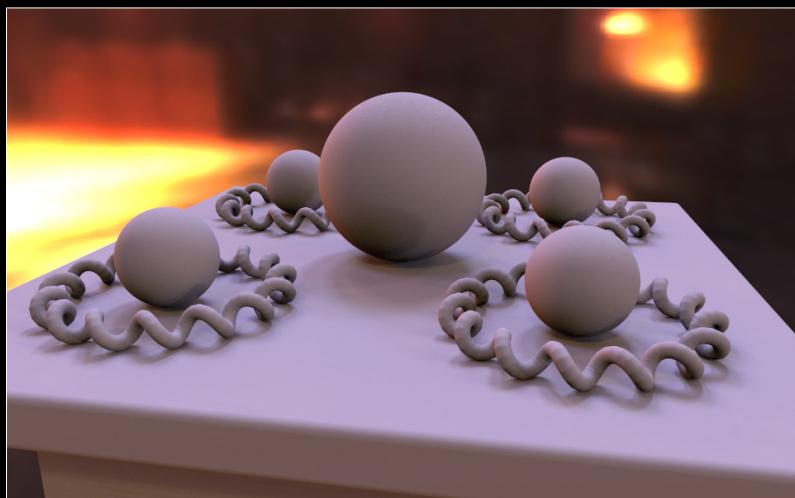


Ambient Occlusion Approximation

 SIGGRAPH 2008



Ambient Occlusion



Full IBL

 SIGGRAPH 2008



Optional Improvement

Compute multiple ambient occlusion maps for different regions of illumination. For example, compute ambient occlusion from just the sky (upper hemisphere) and from just the ground (lower hemisphere).

In compositing, modulate these occlusion maps by the average color of the indirect illumination from each region. This can simulate differently colored indirect illumination from the sky (e.g. blue) and the ground (e.g. green or tan).

Dividing the sphere into lighting regions yields a form of image-based relighting, and can converge to the full image-based lighting solution.



Ren et al, Real-time Soft Shadows in Dynamic Scenes using Spherical Harmonic Exponentiation, SIGGRAPH 2006



Sphere Blockers



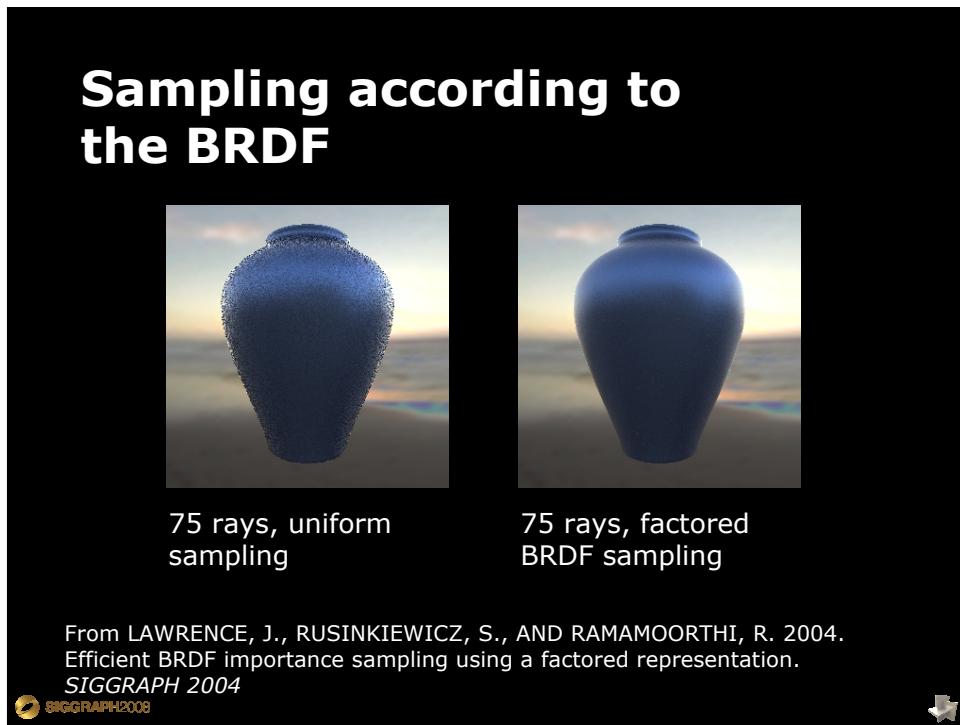
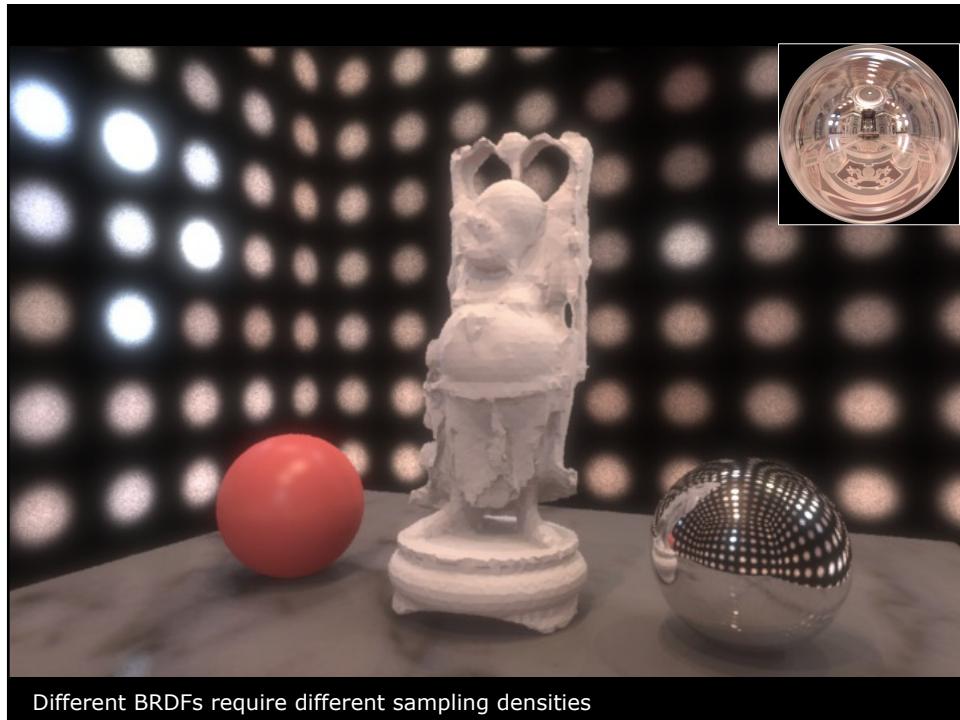
**Ren et al, Real-time Soft Shadows in
Dynamic Scenes using Spherical Harmonic
Exponentiation, SIGGRAPH 2006**



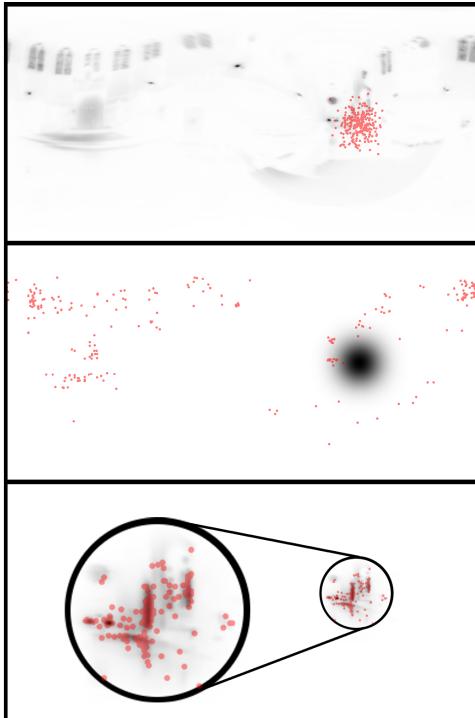
Actual Blockers

**Ren et al, Real-time Soft Shadows in
Dynamic Scenes using Spherical Harmonic
Exponentiation, SIGGRAPH 2006**





SIGGRAPH 2008 Class: HDRI and Image-Based Lighting Image-Based Lighting (Paul Debevec)



David Burke, Abhijeet Ghosh and Wolfgang Heidrich.
Bidirectional Importance Sampling for Direct Illumination. EGSR2005.

Uses *Structured Importance Resampling (SIR)*

See Also:

Talbot et al. *Importance Resampling for Global Illumination*, EGSR2005

Lawrence et al, *Adaptive Numerical Cumulative Distribution Functions for Efficient Importance Sampling*. EGSR2005.

Ghosh and Heidrich. *Correlated Visibility Sampling for Direct Illumination*. SIGGRAPH 2005 Sketch.

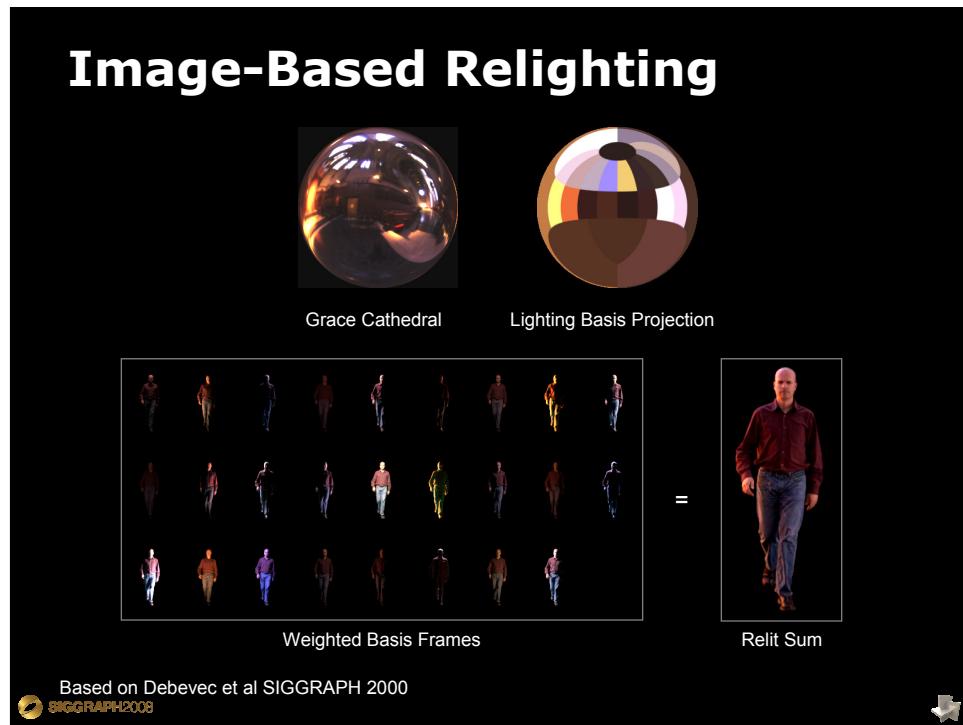
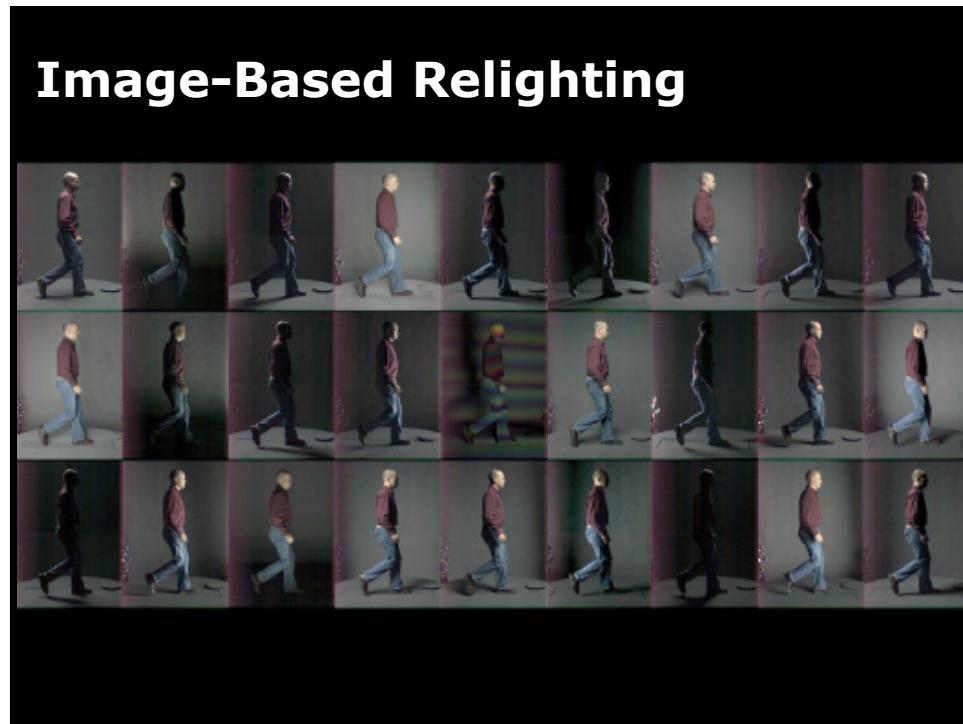
Wavelet Importance Sampling



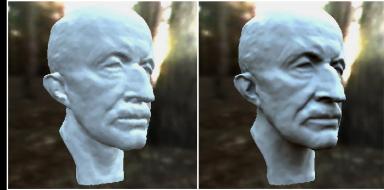
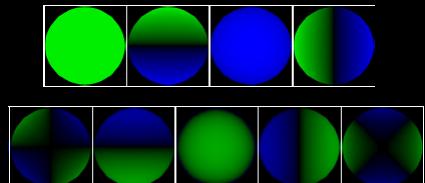
BRDF Sampling Lighting Sampling Product Sampling

Clarberg, Jarosz, Akenine-Möller, and Jensen. *Wavelet Importance Sampling: Efficiently Evaluating Products of Complex Functions*. SIGGRAPH 2005.





Real-Time IBL with Spherical Harmonics

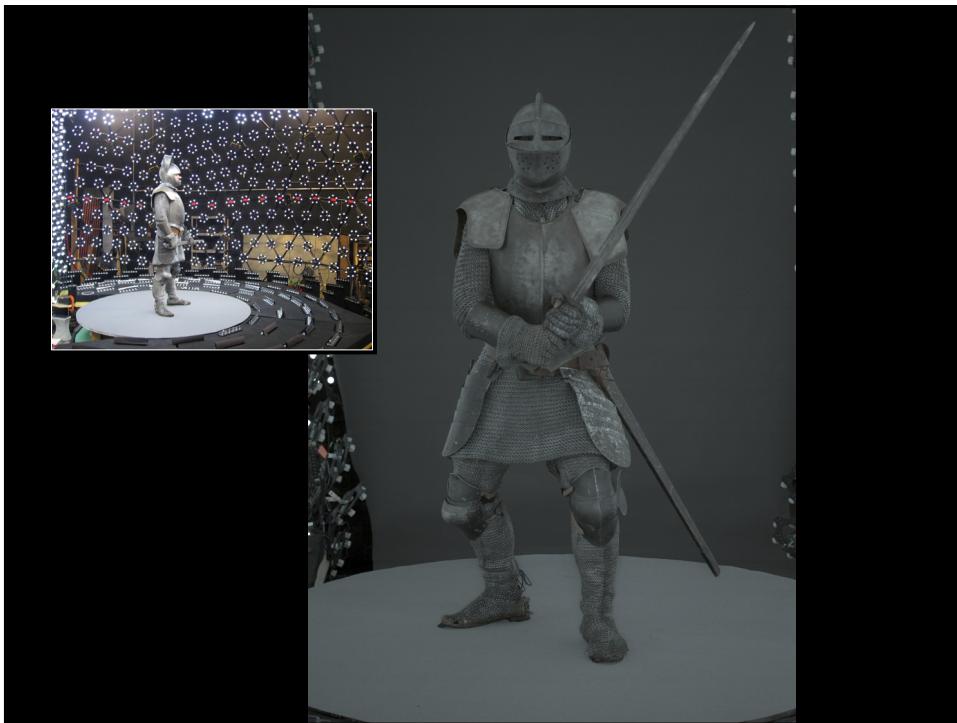


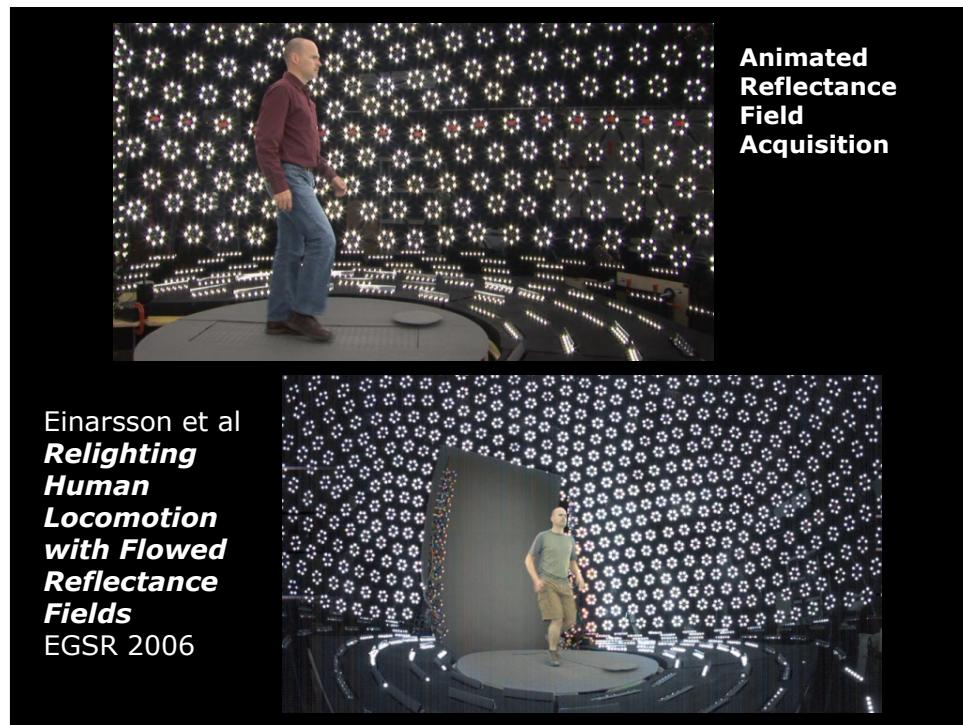
Frequency Space Environment Map Rendering
Ravi Ramamoorthi, Pat Hanrahan, SIGGRAPH 2002

Precomputed Radiance Transfer for Real-Time
Rendering in Dynamic, Low-Frequency Lighting
Environments

Peter-Pike Sloan, Jan Kautz, John Snyder,
SIGGRAPH 2002





SIGGRAPH 2008 Class: HDRI and Image-Based Lighting

Image-Based Lighting (Paul Debevec)



Acknowledgements

Nick Bertke, Marc Jacquier, Liam Kemp,
Brad Peebler, Tim Cooper
Greg Ward, Marcos Fajardo
Henrik Wann Jensen, UCSD
Hiro Matsuguma, Naomi Dennis, Hadi
Ogawa, Toppan Printing
ICT Graphics Lab: Andrew Jones, Charles-
Felix Chabert, Per Einarsson, Alex Ma,
Aimee Dozois, Jay Busch, Tom Pereira,
Naho Inamoto, Brian Emerson, Marc
Brownlow, Tim Hawkins, Andreas
Wenger, Andrew Gardner, Chris Tchou,
Jonas Unger, Frederik Gorranson, John
Lai, Tom Pereira, David Price
ICT Sponsors: USC Office of the Provost,
RDECOM, TOPPAN Printing Co, Ltd.
Bill Swartout, Randal Hall, USC ICT
Randal Hall, Max Nikias, USC
Course co-organizers and speakers

HIGH DYNAMIC RANGE IMAGING
DATA ACQUISITION, MANIPULATION, AND DISPLAY

ERIK REINHARD
GREG WARD
SUMANTA PATTANAIK
AND PAUL DEBEVEC

ERIK REINHARD
GREG WARD
SUMANTA PATTANAIK
AND PAUL DEBEVEC
Morgan Kaufman / Elsevier

www.debevec.org / gl.ict.usc.edu

SIGGRAPH 2008

21.1: Invited Paper: The Hopeful Future of High Dynamic Range Imaging

Greg Ward

BrightSide Technologies by Dolby

Abstract

This paper offers an overview of the challenges and opportunities presented by high dynamic range (HDR) imaging. We examine the length of the imaging pipeline, from creation and storage through image editing and viewing, and discuss how each stage is affected by a move to HDR.

Introduction

Since the first release of Photoshop™ in 1990, imaging has been well-grounded in an 8-bit/sample (i.e., 24-bit/pixel RGB) world. This is not without reason. An 8-bit integer provides enough levels on a standard display that banding is almost negligible using a *de facto* $\gamma=2.2$ encoding. Although considered hefty at its introduction, the 24-bit RGB standard is a reasonably efficient representation, which permits in-core editing of screen-sized displays, using less than 1 MByte for a VGA-resolution image. Years later, 24-bit color was also favored during the dissemination of color management systems, since it conveniently serves as an index in a 3-D lookup table.

As we adopt a wider conduit for imaging, many of the decisions that came before need to be re-examined, and some need to be reconsidered. Naturally, there are practical limits to the changes we can make. Much of the imaging pipeline is ingrained to the point where major changes would cause a technological upheaval whose short-term costs would undermine or even outweigh its long-term benefits. The purpose of this paper is to examine some of these technological trade-offs, compare an “ideal” HDR imaging pipeline to a “likely” one, and consider the need for backward-compatibility. We hope this exposition will spur additional ideas and solutions to the problems presented.

Image Creation

Most modes of digital image creation, including paint software, still cameras, video cameras, and low-end rendering and animation systems, work in a 24-bit “output-referred” color space, such as sRGB [Stokes et al. 96]. This is a convenient choice for viewing and editing on common video display devices, such as CRTs and LCDs, which have a limited dynamic range and color gamut. So long as the color depth resolution of the created imagery meets or exceeds the output device, the latter is unlikely to show up deficiencies in the former. However, as we graduate to higher bit depths and perceptual range in our display systems, digital cinema and high-end home theater will expose inadequate source materials for what they are.

The special effects industry was the first to recognize that greater bit depths were needed before computer graphics (CG) would blend seamlessly with live-action film footage. The greater color resolution, gamut, and dynamic range of film reveal the shortcomings of 24-bit output-referred encodings, which include the notion that a maximum value somehow corresponds to “white.” The real world presents highlights that are 1,000 to 10,000 times

brighter than the 18% gray level commonly used as a reference in cinematography, and these same highlights must be represented in special effects work to incorporate lens flare and similar cues that something is “brighter than white.” In the absence of HDR, special effects lack the depth and realism of live action.

An important driver for HDR in this context is the image-based lighting (IBL) technique introduced in [Debevec 98]. Using IBL, one may capture an HDR image of the environment reflected in a mirrored sphere, and use this to illuminate CG elements so they will blend convincingly with the captured film footage. Figure 1 outlines the basic application of this method.

Image-based lighting is now a principal practice employed in special effects for film, and its reliance on HDR imagery in certain parts of the pipeline have led to a general migration towards floating point representation throughout. The gaming industry is also pushing HDR applications, especially since the majority of graphics cards now support 16-bit/channel floating point natively in their texturing and rendering engines.

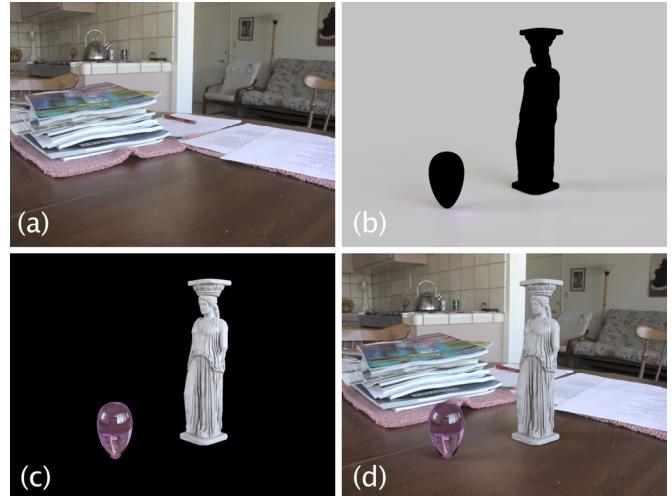


Figure 1. Image-based lighting [Debevec 98]: The background plate (a) is multiplied against a shadow and scattering image (b). The CG elements (c) are illuminated by an HDR *light probe* then composited into the final frame (d).

Capture Methods

While rendering and special effects applications have already made the transition to high dynamic range, capture technology appears to be lagging behind. The principal methods today for HDR moving and still image capture are (respectively) film scanning and multiple, bracketed exposures.

For movies, film offers a sufficient latitude that careful scanning can yield subtle shadows simultaneous with bright highlights, unlike today’s digital video cameras. Unfortunately, scanning film reels with the necessary bit depth is an expensive and time-

consuming operation, thus it is presently limited to big-budget productions. Furthermore, film scanning presents many challenges due to the carefully tuned and highly non-linear response of film to scene colors and intensities. Merging scanned film imagery with the linear world of computer graphics falls somewhere between science and art, and only a few houses have mastered it. If an HD-resolution digital video camera with true HDR output were to arrive on the market tomorrow, it might well change the entire industry.

For still capture, the method of multiple exposures popularized in [Debevec & Malik 97] ten years ago has yet to be replaced. Because a sequence of exposures necessarily span a substantial time frame, camera/tripod shake and scene movement are perennial problems. In the case of camera movement, fairly robust image alignment is available using either the Median Threshold Bitmap technique [Ward 03], Scale Invariant Image Feature Transforms [Lowe 04], or image-flow methods [Kang et al. 03]. Image-flow methods are also useful for rectifying scene movement. Other “ghost removal” techniques include variance-based segmentation [Reinhard et al. 05], and robust estimators [Khan et al. 06]. An HDR still camera would eliminate the need for such workarounds, and open up a world of new possibilities beyond 24-bit RGB.

Image Transmission

The first viable HDR image format was introduced as a little-known “exponent” extension to the Utah Raster Toolkit [Petersen 04]. Independently, the author developed a nearly identical 32-bit RGBE format as part of the *Radiance* rendering system [Ward 91] [Ward 94]. Several years later, 24-bit and 32-bit LogLuv extensions were added to the TIFF library, covering the full visible color gamut and dynamic range [Ward Larson 98]. More recently, the 48-bit/pixel EXR format was made public by Industrial Light and Magic in the form of the excellent OpenEXR™ C++ library [Kainz et al. 2002].

The current trend in HDR image transmission is towards better, customized compression algorithms. For example, ILM’s EXR image data starts out with 50% more bits/pixel, but ends up taking the same or less space than *Radiance* RGBE thanks to its lossless wavelet compression. (*Radiance* uses a basic run-length encoding scheme.)

In the realm of “lossy” compression algorithms, there is a trend towards backwards-compatible formats, such as JPEG-HDR [Ward & Simmons 04] and MPEG-HDR [Mantiuk et al. 06]. These formats have the dual advantage of taking up a small fraction of the space of the lossless HDR streams, while being displayable using conventional hardware and software. They are thus ideally suited to the internet, video, and perhaps a new generation of digital cameras.

Backwards-compatible HDR formats are a win-win for manufacturers and consumers alike. Easing the transition from 24-bit RGB to full dynamic-range and gamut capture, editing, and viewing has the potential to greatly accelerate market penetration. Early adopters would gain immediate access to an HDR world, while others would not be inconvenienced, and could even benefit from improved tone-mapping in their legacy content. A well-chosen strategy may even simplify the eventual retirement of low dynamic-range data years hence [Ward 06].

Image Editing

As we noted above in our discussion of special effects, HDR offers numerous advantages for combining disparate image sources such as film and CG. In particular, the freedom from a hard ceiling on values and the inclusion of the entire visible gamut avoids a host of mapping problems. Similarly, HDR offers numerous opportunities for image editing, though it presents some challenges as well.

The first challenge for image editors is to overcome the notion of a maximum value corresponding to “white.” This concept does not apply in the real world, until and unless you print something on paper. Being restricted to a maximum white during editing, even in cases where you do not intend to print, is unnecessarily constraining. One approach to this problem is to provide a slider to adjust the range of display values, or to edit locally exposed subregions. Until desktop systems are equipped with HDR displays, this may be a necessary compromise.

A parallel challenge is the appropriate on-screen representation of out-of-gamut colors. To some degree, we face this already with 24-bit RGB in the disparities between input, display, and hardcopy devices. The usual approach is to map colors to the current display, optionally highlighting pixels that will be out-of-gamut upon printing. The situation becomes even more interesting when we consider colors outside the visible gamut, which most HDR formats can represent. Overall, HDR offers the simplicity of working in a scene-referred linear color space, which streamlines color management enormously.

In Photoshop today, 32-bit/channel features are about where 16-bit/channel features were 10 years ago. Floating-point has been introduced, but most editing functions are currently disabled or restricted, and we expect the support to improve gradually with time. Cinematic image editing and compositing tools are pushing the envelope more forcefully, and there are other consumer-level image tools that are quite powerful in the HDR domain (e.g., Idruna’s Photogenics™ HDR, Artizen™ HDR, and Cinepaint).

Our fondest hope is that the current RAW format craze will give way to a more productive and long-term investment in HDR. In many ways, RAW is just a poor man’s HDR, where every camera has its own representation and each software application has to extract the image as best it can. Building a scene-referred HDR pipeline is like building an autobahn – suddenly there is a compelling reason to engineer better systems. In contrast, RAW is like 1000 gravel driveways, ultimately leading us nowhere.

Image Viewing

Ideally, everyone would have a high dynamic range display that would permit users to view and work effortlessly with HDR images [Seetzen et al. 2004]. While we’re at it, let’s give our displays four or five spectrally-spaced primaries so they cover the entire visible gamut. One and possibly both of these wishes will come true in the next 2-7 years, but in the meantime, we need some method to represent colors that are outside the range of our display. Regardless of display advances, printing will still require some gamut mapping, at least until someone invents self-luminous paper.

Tone-mapping is the general technique employed to reduce an image’s dynamic range and color gamut to fit on a monitor, projection system, or print. It has a long history, extending back to the invention of negative photography. Tone-mapping was first

introduced to computer graphics in 1993 [Tumblin & Rushmeier 93], and HDR tone-mapping has been explored extensively over the last five years [Reinhard et al. 05]. The best HDR tone-mapping algorithms also happen to be the most expensive, adjusting luminance values locally in a perceptually optimized way [Durand & Dorsey 02] [Fattal et al. 02] [Reinhard et al. 02]. By comparison, the older, global tone-mapping operators can be applied in real-time, providing convenient image editing and HDR video display [Krawczyk et al. 05]. Eventually, local TMOs may achieve real-time rates, but the path is not yet clear.

Digital cinema will probably be the first arena where medium-to-high dynamic range imagery will be presented. After all, cinematic prints already encompass a wider gamut and dynamic range than other media, and moviegoers have come to expect a certain richness. No theater owner wants to make an expensive upgrade that amounts to a backwards step in quality, so HDR could even be considered a prerequisite for digital cinema.

Once accustomed to a more exciting theater experience, consumers might start looking for equipment that offers similar dynamic range in their homes. In fact, the home theater market is currently moving towards HDR faster than either digital cinema or computer display equipment. This may be driven by vendors' desires to win a competitive advantage before large, flat-screen televisions saturate the market. If such a transition happens too quickly, we may find ourselves in the awkward position of having to synthesize all our HDR imagery from LDR content – essentially the inverse of the tone-mapping problem [Meylan et al. 06] [Rempel et al. 07].

Conclusion

It is clear that high dynamic range imaging will someday dominate the market. The question is, when? Our best estimate is between 2 and 7 years, and many things can influence its advance. It is preferable that HDR be introduced well rather than quickly. As engineers and imaging scientists, we are not powerless to affect this process. Through careful planning and intelligent standards-making, we can grow our businesses while delivering well-timed improvements in professional and consumer equipment and software.

The logical path is to introduce HDR to the high-end digital cinema market first, simultaneously with independent niche markets such as medical imaging, then allow a some years to pass before introducing HDR to the consumer. At that point, we will have settled the standards and worked out the kinks, the studios will have plenty of HDR content, and the job of consumer education will already be done.

References

- DEBEVEC, P. 1998. Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-Based Graphics with Global Illumination and High Dynamic Range Photography. In *Proceedings of ACM SIGGRAPH 1998*.
- DEBEVEC, P., and MALIK, J. 1997. Recovering High Dynamic Range Radiance Maps from Photographs. In *Proceedings of ACM SIGGRAPH 1997*.
- DURAND, F., and DORSEY, J. 2002. Fast Bilateral Filtering for the Display of High-Dynamic Range Images. *ACM Transactions on Graphics*, 21, 3.
- FATTAL, R., LISCHINSKI, D., and WERMAN, M. 2002. Gradient Domain High Dynamic Range Compression. *ACM Transactions on Graphics*, 21, 3.
- KAINS, F., BOGART, R., HESS, D., SCHNEIDER, P., ANDERSON, B., 2002. *OpenEXR*. www.openexr.org/.
- KHAN, E.A., AKYUZ, A.O., REINHARD, E. 2006. Ghost Removal in High Dynamic Range Images, *IEEE International Conference on Image Processing* (accepted for publication)
- KANG, S.B., UYTENDAELE, M., WINDER, S., SZELISKI, R. 2003. "High Dynamic Range Video," *Proceedings of ACM SIGGRAPH*.
- KRAWCZYK, G., MYSZKOWSKI, K., SEIDEL, H.P. 2005. Perceptual Effects in Real-time Tone Mapping, *Proc. of Spring Conf. on Computer Graphics*.
- LOWE D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, 2.
- MANIUK, R. EFREMOV, A., MYSZKOWSKI, K., SEIDEL, H-P. 2006. Backward Compatible High Dynamic Range MPEG Video Compression, *Proc. of SIGGRAPH '06 (Special issue of ACM Transactions on Graphics)*.
- MEYLAN, L., DALY, S., SÜSSTRUNK, S. 2006. The Reproduction of Specular Highlights on High Dynamic Range Displays, *IS&T SID 14th Color Imaging Conf.*
- PETERSEN, J. 2004. float_to_exp(3) man page. www.cs.utah.edu/gdc/projects/urt/help/man3/float_to_exp.html
- REINHARD, E., WARD, G., PATTANAIK, S., DEBEVEC, P. 2005. *High Dynamic Range Imaging: Acquisition, Display, and Image-based Lighting*, Morgan Kaufmann Publishers, San Francisco.
- REINHARD, E., STARK, M., SHIRLEY, P., FERWERDA, J. 2002. "Photographic Tone Reproduction for Digital Images," *ACM Transactions on Graphics*, 21,3.
- REMPEL, A., TRENTACOSTE, M., SEETZEN, H., YOUNG, D., HEIDRICH, W., WHITEHEAD, L., WARD, G. 2007. "Ldr2Hdr: On-the-fly Reverse Tone Mapping of Legacy Video and Photographs," *Proc. of SIGGRAPH '07 (Special issue of ACM Transactions on Graphics)*.
- SEETZEN, H., HEIDRICH, W., STUEZLINGER, W., WARD, G., WHITEHEAD, L., TRENTACOSTE, M., GHOSH, A., VOROZCOVS, A. 2004. "High Dynamic Range Display Systems," *Proc. of SIGGRAPH '04 (Special issue of ACM Transactions on Graphics)*.
- STOKES, M., ANDERSON, M., CHANDRASEKAR, S., and MOTTA, R. 1996. Standard Default Color Space for the Internet. www.w3.org/Graphics/Color/sRGB.
- TUMBLIN, J. and RUSHMEIER, H. 1993. Tone Reproduction for Realistic Images, *IEEE Computer Graphics and Applications*, 13(6).
- WARD, G. 2006. A General Approach to Backwards-Compatible Delivery of High Dynamic Range Images and Video, *Proceedings of the Fourteenth Color Imaging Conference*.
- WARD, G., AND SIMMONS, M. 2004. JPEG-HDR: A Backwards-Compatible, High Dynamic Range Extension to JPEG, *Proceedings of the Thirteenth Color Imaging Conference*.
- WARD, G. 2003. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures, *Journal of Graphics Tools*, 8(2):17-30.
- WARD LARSON, G. 1998. Overcoming Gamut and Dynamic Range Limitations in Digital Images. *Proc. of IS&T 6th Color Imaging Conf.*
- WARD, G. 1994. The RADIANCE Lighting Simulation and Rendering System. , In *Proc. of ACM SIGGRAPH*.
- WARD, G. 1991. Real Pixels. In *Graphics Gems II*, edited by James Arvo, Academic Press.

A General Approach to Backwards-Compatible Delivery of High Dynamic Range Images and Video

Greg Ward, BrightSide Technologies

Abstract

We propose a general solution to the problem of decoding high dynamic range (HDR) information stored as a supplement to low dynamic range (LDR) images or video. Each LDR frame is paired with a lower resolution HDR version, and these are compressed separately using any of the existing methods appropriate to the task. On decode, the low-resolution HDR image is upsampled to match the resolution of the LDR version, and the high frequencies are transferred from the LDR to the HDR frame. The recovery process places no constraints on the color space or tone-mapping of the backwards-compatible LDR content, and is thus ideally suited to applications such as DVD movies that target legacy equipment while building in forward-compatibility with emerging HDR systems. A fast and simple recovery algorithm is demonstrated, followed by a more sophisticated and accurate technique. Examples are shown on computer-generated video frames as well as HDR captured video.

Background

Recent work in high dynamic range image encoding has focused on “lossy” compression, with a particular emphasis on backwards compatibility with existing formats [Spaulding et al. 2003] [Ward & Simmons 2004, 2005] [Li et al. 2005] [Mantiuk et al. 2006]. This is an important trend for the practical adoption of scene-referred high dynamic range (HDR) digital imagery in a world where lossy, output-referred formats such as JPEG and MPEG dominate. The consumer market will not accept HDR formats that take 12 times as much space as JPEG and 80 times as much space as MPEG, especially if they cannot be displayed with standard viewers and players. Such is the case for the most common HDR formats in use today: *Radiance* RGBE (.hdr), OpenEXR, and TIFF [Reinhard et al. 2005]. Though adoption of such lossless formats is taking place in high end tools such as Adobe Photoshop™ and systems such as Mac OS X and (hopefully) the next version of Windows, the space requirements of lossless HDR will preclude its use in digital photography, video, and web applications for the foreseeable future. Lossy HDR encodings that are not backwards-compatible [Mantiuk et al. 2004] [Xu et al. 2005] will eventually make it to the marketplace, but currently offer no practical transition path.

Methods for backwards-compatible HDR image encoding can be divided into two categories: reversible tone-mapping [Li et al. 2005] and supplemental images [Spaulding et al. 2003] [Ward & Simmons 2004, 2005] [Mantiuk et al. 2006]. Reversible tone-mapping presents a challenge to efficient encoding, because JPEG and MPEG tend to degrade information that is important for proper reconstruction. Thus, compression performance is much worse than with supplemental methods. Supplemental methods encode additional image data to recover the HDR original from

the recorded low dynamic range (LDR) information, storing the extra data in an auxiliary stream that is ignored by naïve viewers and players. The key is to minimize the size of this auxiliary stream, and existing methods add between 5% and 30% to the LDR image size, depending on the method and settings.

The problem with supplemental encoding schemes is their inherent complexity. Kodak’s ERI format [Spaulding et al. 2003] uses a residual image with sophisticated color and bit manipulations to minimize the size of the auxiliary stream, achieving good compression but only modest gains in dynamic range. The backwards-compatible HDR version of MPEG introduced by Mantiuk et al. [2006] follows a similar approach, with a residual image storing the difference between a predictor function on the LDR data and an perceptual HDR color space. In both methods, the residual tends to be small and noise-like where the LDR image is within its output-referred gamut, but jumps abruptly wherever the LDR image saturates at the top end. This challenges standard lossy image compression techniques, which must be tailored to encode only the perceptually important information without introducing false contours at the gamut boundaries. The different bit sizes between the LDR and HDR data pose additional difficulties during encoding and decoding, and care must be taken not to introduce new quantization errors in the process. To avoid these complications, the JPEG-HDR encoding of Ward & Simmons [2004, 2005] employs a ratio image in place of a residual, which can be multiplied against the decompressed LDR image to recover the HDR original. This simplifies the process by allowing a single 8-bit log channel to store the ratio between HDR and LDR pixels, but complexities creep back in when the ratio image is downsampled to reduce the size of the auxiliary stream. With ratio image downsampling, the LDR image must either be “precorrected” against lost resolution, or “postcorrected” using a resolution enhancement technique. Also, the LDR image must encode all the necessary color information, since the ratio image is only for the luminance channel. This places important restrictions on the tone-mapping operator and color space of the LDR data. Such restrictions are avoided in the method of Mantiuk et al. [2006], which encodes color information as well as HDR luminance in its residual image.

The method we propose for backwards-compatible encoding of HDR imagery is inspired by previous work and motivated by the following observations:

- In today’s applications, the LDR data is more important than the HDR data, and should not be compromised.
- HDR imagery takes longer to decompress because hardware and software are tailored to 8-bit streams.
- Both of these conditions will change in the next 5 years.

The main reason that no one has implemented the obvious solution of storing a full-blown HDR image in an auxiliary stream is that it would more than double the data size, incorporating largely redundant information. However, storing a low-resolution version of the HDR original has a number of advantages relative to our three observations. First, it would not affect the LDR data in any way. Second, it would decompress quickly because most algorithms are proportional to the number of pixels – using $\frac{1}{4}$ the resolution in each dimension would speed decompression by a factor of 16. Third, as more applications come to rely on HDR imagery and the associated hardware and software get faster, the resolution of the auxiliary stream can be increased over time, eventually reversing the roles and importance of the LDR and HDR data streams.

This is the essence of our proposal: along with the conventional LDR image stream (still image or video sequence), we store a corresponding HDR image stream at a reduced resolution. No special preparation is made on either stream, and no restrictions are placed on the color space or tone-mapping of the LDR data. Storing an HDR image with every LDR image may seem redundant, but only the low frequency information is repeated, and from this we can derive a correlation to better recover the high frequencies; so it is not redundant, but necessary. We retain the nicer features of the supplemental encoding methods, with none of the complexity – at least none on the encoding side. Decoding is another matter, and the subject of this paper.

We start by describing a basic recovery method that is fast but depends on a global tone-mapping operator, then describe a more advanced method that estimates the local tone-mapping response automatically. In our results section, we evaluate the performance of our two algorithms on a variety of source images and video, both synthetic and captured, and mapped to different color spaces using different tone-mapping operators. We conclude with some final observations and suggestions for future work.

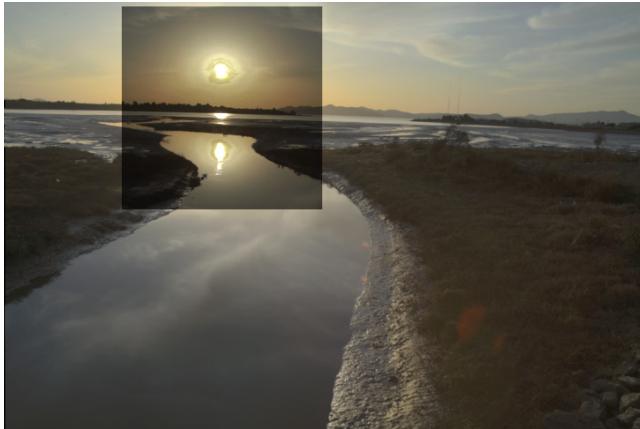


Figure 1. Tone curve inversion from 8-bit/channel original, showing quantization artifacts in bright region.

Method

Our decoder is given two versions of our image, one that is high resolution but low dynamic range, and one that is low resolution but high dynamic range. From these, we wish to derive an image that is high resolution and high dynamic range. If a global tone-mapping operator were used to generate the LDR image, we could try inverting this mapping to arrive at an HDR version, ignoring the low-resolution HDR information provided. However, we would run out of color resolution in places where the image values were clipped (out of gamut) or the tone curve underwent a large degree of compression. Such a result is shown in Figure 1. The sky region surrounding the sun shows quantization artifacts as a result of the expansion of LDR data. Clearly, we need the HDR data to supplement our results where the LDR image is inadequate.

Basic Method

Since we have a low resolution version of our HDR image, we can take a different approach. Rather than inverting the tone curve, we can take the high frequency data above the quantization threshold from the LDR image to augment the low frequency information in the HDR image. Specifically, we:

1. Convert our LDR image color space to approximately match the primaries of the HDR data.
2. Extract high frequencies from the LDR image between HDR and LDR image resolutions and apply quantization threshold.
3. Upsample the HDR image to LDR resolution and apply the high frequencies from Step 2.

The high frequency ratio image can be computed using a rational convolution filter, or with equal efficiency via a downsampling-upsampling-divide process:

- A. Downsample the image to lower resolution limit.
- B. Upsample again to original resolution.
- C. Divide the original image by resampled image from Step B.

This results in a scaling image whose pixels have an average value of 1.0, which can be multiplied by an upsampled image to recover the missing high frequencies. It is critical that the upsampling method not introduce spurious high frequencies, therefore bilinear interpolation is preferred over the more usual bicubic basis in Step B. (We employed a separated Gaussian kernel with a radius of 0.6 pixels in a square footprint of 5x5 pixels for downsampling.) Figure 2 shows a high frequency image before and after a threshold of 1.5 quantization steps is applied.

If we can invert our tone-mapping curve prior to Step 1 above, our results will be fairly accurate because we only take high frequency edge information from the LDR version, avoiding the quantization artifacts that were showing up in the smooth gradient regions. Figure 3 shows the comparison results, recovered from a 480x640 pixel LDR and a 120x160 HDR image.

Unfortunately, we do not always know the tone-mapping that was applied to arrive at the LDR image, and even when we do, it may be too difficult or too expensive to invert, leaving us with inaccuracies in our high frequency data. Globally, our recovered

images will look about right, but details may be softened or lost, as shown in Figure 4. In general, tone-mapping operators may be globally or locally determined, and may preserve visibility over the entire image or may allow some regions to saturate to black or to white. In saturated regions, we must fall back on the low frequency information in our HDR image. This is acceptable in most cases, as this information was selected out of the LDR version by the content creators.

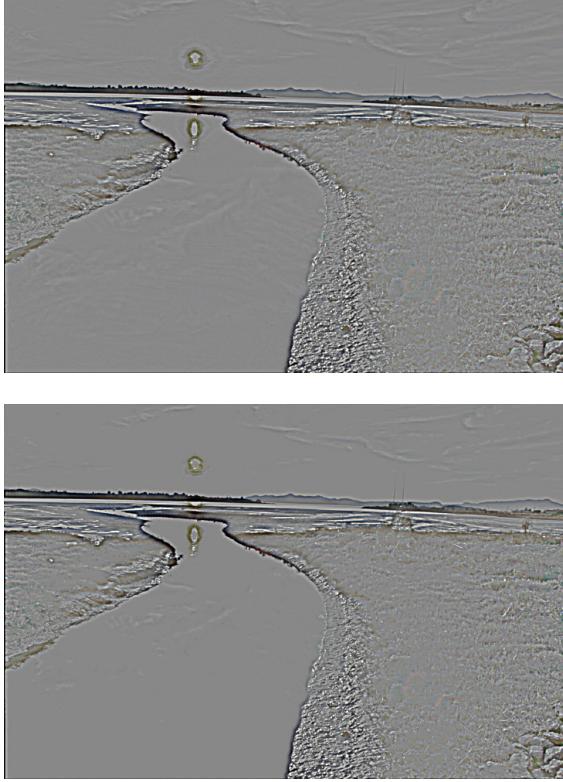


Figure 2. High frequency image before and after quantization threshold is applied.

Advanced Method

The basic method just described is very similar to the “postcorrection” technique introduced by Ward & Simmons [2004], which is known to be a crude approximation due to its ignorance of the tone-mapping function. When the LDR tone-mapping is complex or unknown, we still wish to make the best use of high frequencies possible. To accomplish this, we need to approximate the impulse response in the LDR data, which is the important part of the tone-mapping for our purposes. Because the tone-mapping operator can vary over the image, we need an approximation of the impulse function that is also allowed to vary. But how do we derive such an approximation? We need to relate our LDR and HDR data, but we cannot compare the frequencies we wish to recover, since they are missing from our HDR image. So, we settle for the closest information available – the top frequencies in the lower resolution image, assuming the impulse response does not change dramatically from one band to the next.¹

¹ This assumption may be violated in tone-mapping operators that incorporate a sharpening stage. In such cases, we can either undo



Figure 3. Basic method of recovering high frequencies using a priori knowledge of the tone-mapping function.



Figure 4. Basic method without knowledge of tone-mapping curve, showing loss of detail relative to Figure 3.

The steps are listed below:

1. Convert our LDR image to match the color space of the HDR image, \mathbf{I}_{SR} . Call this \mathbf{I}_{SOR} .
2. Reduce to the resolution of \mathbf{I}_{SOR} to match that of \mathbf{I}_{SR} . Call this \mathbf{I}'_{SOR} .
3. Extract high frequencies for grayscale versions of \mathbf{I}_{SR} and \mathbf{I}'_{SOR} using a bandwidth equal to the resolution difference between \mathbf{I}_{SR} and \mathbf{I}_{SOR} .
4. Compute a spatially varying impulse response function between the \mathbf{I}'_{SOR} upper band image and the \mathbf{I}_{SR} upper band image. (Estimating the impulse response is explained in the following section.)
5. Compute the highest frequencies of full-resolution image \mathbf{I}_{SOR} and apply the impulse response from Step 4 to recover the HDR high frequencies.
6. Upsample \mathbf{I}_{SR} to match the full LDR image resolution.
7. Multiply high frequency data from Step 5 to get our full-resolution HDR output.

the sharpening filter prior to recovery or accept that our final HDR output will exhibit a similarly sharpened result.

Applying our more advanced method to the same tone-mapped image as before, we see improved sharpness in the unsaturated regions of Figure 5, based on the same input used to generate Figure 4. The additional high frequency and mapping calculations increase the processing time by a factor of 4, from 0.45 to 1.8 seconds on a 1.5 GHz G4 processor, with both calculations proportional to the number of pixels.



Figure 5. Advanced method estimates the high frequency impulse response over the image, obtaining a sharp result without assumptions about the tone-mapping operator.

Figure 6 shows a comparison between the simple method and the advanced method on a 2704x4064 LDR image mapped using Durand & Dorsey’s bilateral filter [2002], paired with a 676x1016 HDR version. The blue inset shows the red area recovered using the simple method without knowledge of the TMO. The green inset shows the same area mapped using the advanced method to estimate the TMO. Although neither reconstruction is perfect, the advanced method retains greater sharpness, at the expense of some over-shooting, visible as slight discolorations at high contrast edges. For comparison, the white inset shows the HDR data upsampled with a bicubic filter on the left, and the full-resolution original on the right. (Insets were tone-mapped using a histogram operator [Larson et al. 1997].)

Estimating the Impulse Response Function

In order to correctly map the high frequencies in the tone-mapped LDR image into the HDR domain, we need to estimate the local impulse response. We start by assuming that the impulse response function is monotonically increasing, at least locally. This is a reasonable assumption for any tone-mapping operator, since a decreasing impulse response would imply that larger gradients in the original yield smaller gradients in the tone-mapped image, which would show up as edges with reversed contrast. Even with this assumption, it is difficult to estimate a continuously changing function, so we further assume that the function is constant within small local regions of the image. In our implementation, we use overlapping blocks of roughly 64x64 pixels in the subsampled HDR input, regardless of the input image resolution. We found this to be a reasonable size to obtain a sampling of the impulse response.

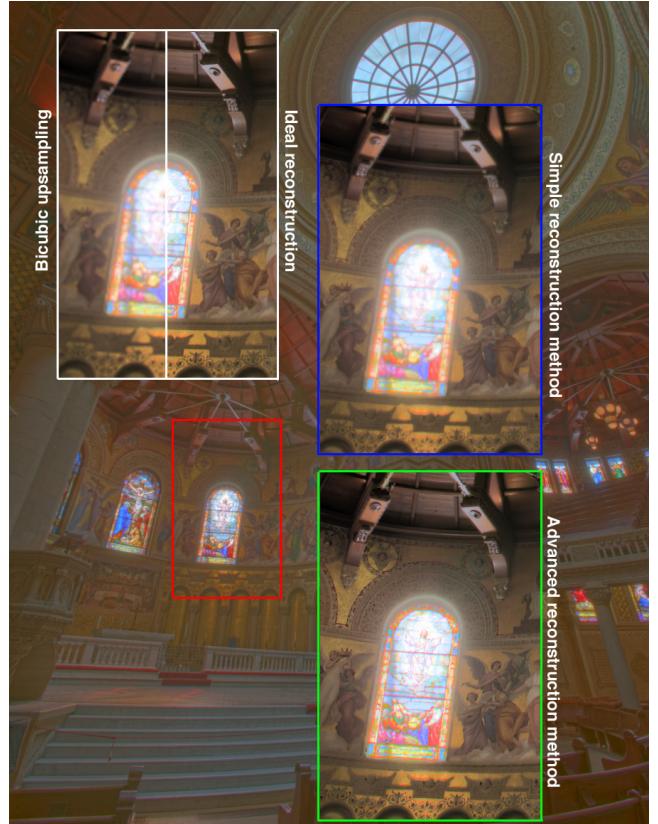


Figure 6. An image of the Stanford Memorial Church mapped with a bilateral filter, and recovered using the simple and advanced methods from a 4x downsampled HDR version. (Image courtesy Chris Cox of Adobe Systems.)

Within each block, we put the luminance (gray) values from the LDR upper band image in one 4096-entry data array corresponding to 64x64 pixels, and the luminance values from the HDR upper band image in a second data array. We independently sort the two arrays, then extract the input-output pairs corresponding to every 164th entry. By construction, this creates a set of 25 monotonically increasing coordinates, evenly spaced in the data population, which we can use in a linear or cubic interpolation of the impulse response function. The graph in Figure 7 show the original scatter of values from an example HDR/LDR mid-frequency block pair in Figure 6, and the sorted interpolation points. In regions such as the one chosen here, we may get diverging impulse responses that we combine into a single curve, but this is necessary in order to derive a smooth function of one variable.² Besides enforcing monotonicity, independent sorting avoids outliers caused by minor misalignments between the LDR and HDR data. Computing a single impulse response function based on luminance further minimizes color shifts when we apply it independently to each channel during reconstruction.

² Better correlations could be obtained by adding a second variable to our scheme, the original HDR luminance. This adds a level of complexity, but could improve the results for some tone-mapping operators.

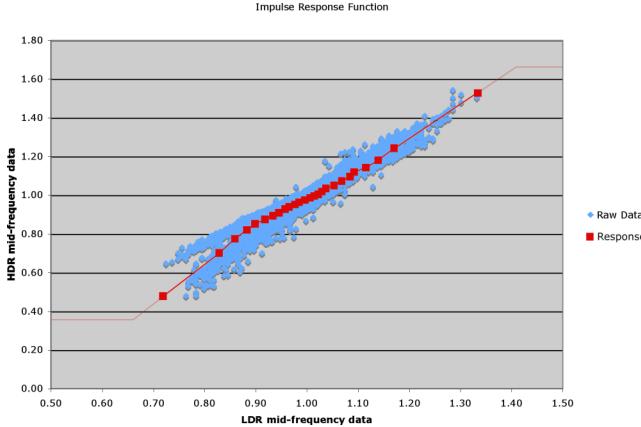


Figure 7. Mid-frequency data from a 64x64 pixel region and the derived impulse response function to be applied to the highest frequencies.

The blocks we use to derive the local impulse response functions overlap, enabling us to smoothly interpolate the results over the image. This scheme is shown diagrammatically in Figure 8. Consider the four neighboring 64x64 blocks: A, B, C, & D. Each has its center at the intersection of four 32x32 pixel regions, and characterizes the response in the surrounding square using the method just described.³ To map the response of the impulse for pixel \mathbf{P} , we pass it through each of the four response functions, and linearly interpolate the results based on the position of \mathbf{P} .

To avoid excessive quantization noise in high gradient areas, we place a restriction on the maximum overall gain for the impulse response within each block. If the difference between the maximum and minimum output values is greater than 5 times the difference between the maximum and minimum input values, the response is scaled to reduce the average slope to fit the maximum 1:5 ratio. This limit is rarely reached in practice.

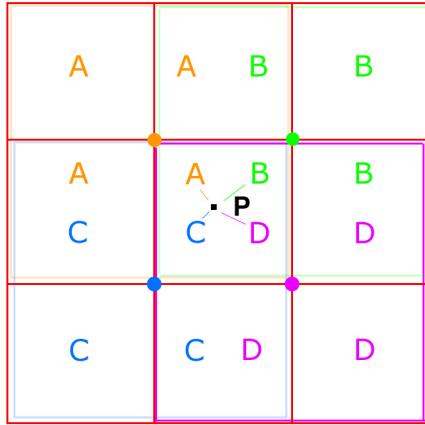


Figure 8. Overlapping impulse response blocks permit smooth interpolation of output.

³ The 50% overlap we have chosen with each neighbor is adjustable. We recommend at least 25% neighbor overlap for smooth results.

It is also important to consider values outside the range of the interpolated response pairs. We extrapolate the last value pair on top and bottom out to one half the distance between the last two points. After that, we cap the impulse response function, effectively cutting off high frequencies this far above the observed gradient values. This response extension is shown as the faded ends on the curve in Figure 7. Limiting extrapolation in this way reduces ringing artifacts caused by sharp edges in the LDR image.

Results

We tested three variants of our algorithm each on a computer-generated HDR animation and two captured HDR video sequences. The first variant of our algorithm assumes a linear mapping between HDR and LDR (tone-mapped) versions. The second variant assumes a non-linear, global tone-mapping over the image. The third variant is our “advanced” method, which allows the tone-mapping to change locally and non-linearly. Figure 9 shows stills from our lighting simulation of an air traffic control tower, a beach at sunset, and a trip through a tunnel.



Figure 9. Our three test scenes.: an HD-resolution animation and two VGA-resolution video sequences.

Not surprisingly, we found that the linear variant worked well enough on the linear tone-mapping operator, but there was considerable detail lost in the brighter regions where the linear operator had clipped (Figure 10). The linear assumption was not appropriate for any of the other tone-mapping operators, and tended to produce excessive sharpness in the highlights due to over-shooting (Figure 11). Also not a surprise, we found that the global, non-linear algorithm variant was acceptable for the global versions of the Reinhard [Reinhard et al. 2002] and histogram [Larson et al. 1997] operators, but degenerated with the bilateral filter [Durand & Dorsey 2002] and the gradient domain operator [Fattal et al. 2002], due to their local behavior.

On the synthetic control tower animation, the advanced algorithm performed acceptably for every tone-mapping operator we tried, though clamping in the linear operator still lost high frequency

information. In the HDR captured sequences, we noticed a few problems at the boundaries of bright objects, which appeared as colorful outlines. Even within some low-gradient fields, such as the orange sky of the sunset, patterns would occasionally emerge (Figure 12). These are due to the Bayer pattern of the color image sensor, and the best solution is to improve the demosaicing filter. Barring that, a comb filter could be applied during HDR recovery to reduce the appearance of Bayer mosaic remnants.

Conclusion

We have proposed a solution for backwards-compatible HDR imagery that stores a low-resolution HDR version of each frame as a supplement to the LDR data. This places the burden on the decompression engine to recover high-resolution HDR frames by combining the two streams. This approach is not necessarily better or faster than previous backwards-compatible solutions. The real benefit to our approach is the evolutionary path it offers, especially for video.

Any backwards-compatible format is a stop-gap solution that requires compromises in encoding efficiency. This is illustrated by the lower performance of the backwards-compatible HDR extension to MPEG by Mantiuk et al. [2006] relative to their original proposal, which was not backwards-compatible [Mantiuk et al. 2004]. Unfortunately, backwards-compatibility is considered essential to market adoption, especially for video. Furthermore, once we settle on a new encoding standard, we are committed to it for about 10 years. This seems like a long time to be using an ornate, stop-gap format, and in the end we might prefer a simpler solution that provides a smooth transition to native HDR video.

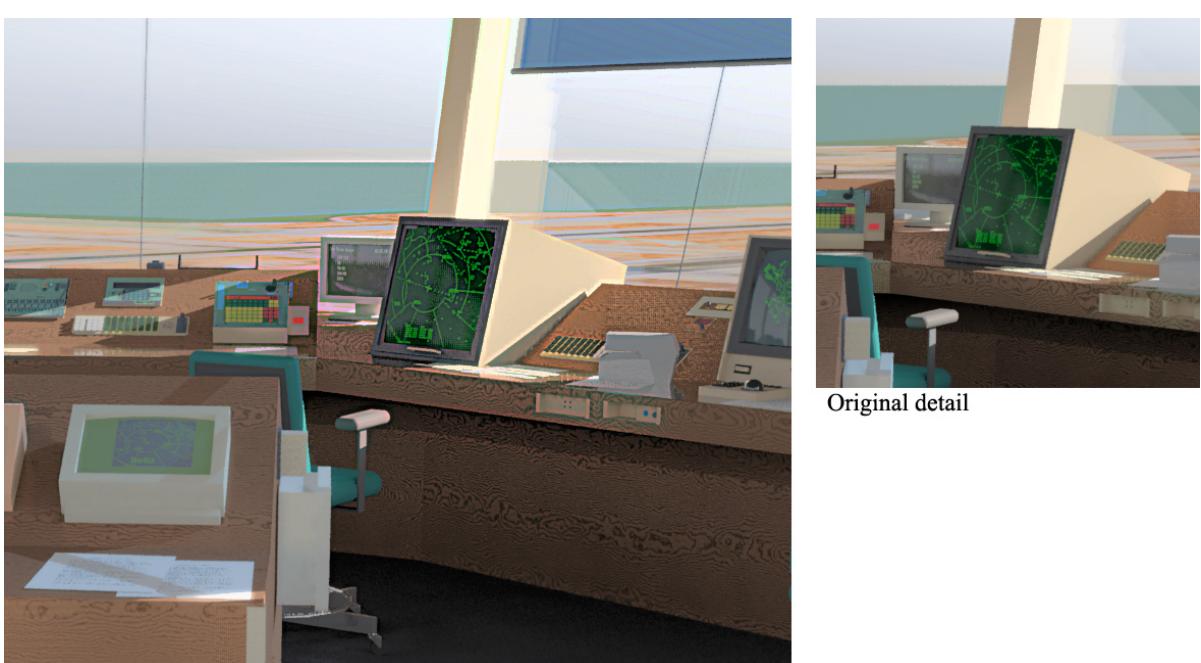
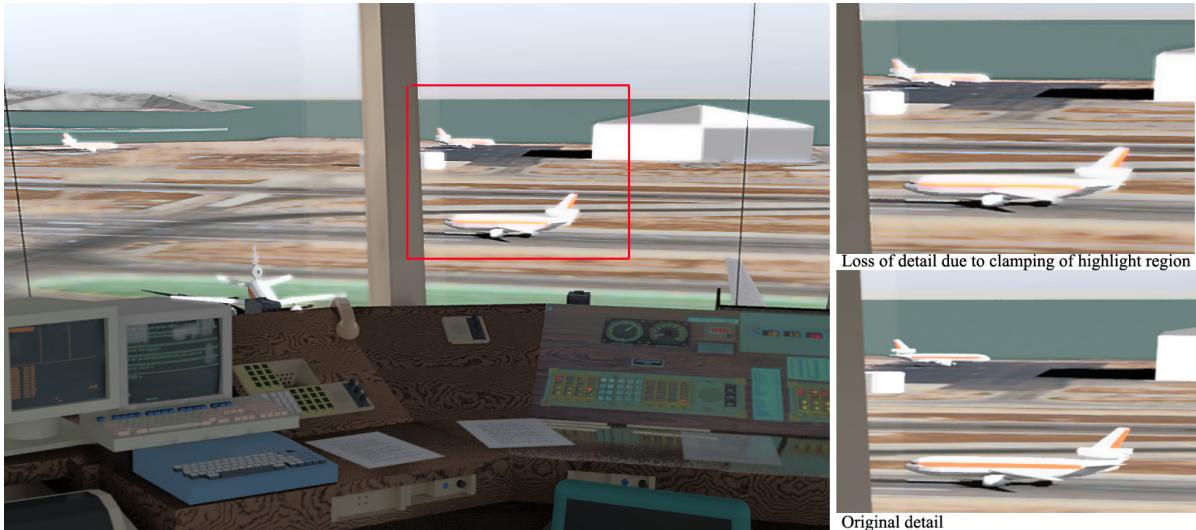
This is exactly what we offer. By logically separating the HDR and LDR data streams, tying the quality of each to its resolution alone, we provide a seamless upgrade path from the LDR world of today to the HDR world of tomorrow. We are free to standardize on the most efficient HDR encoding we can devise, with no compromises for backwards-compatibility. We can then incorporate this standard in new hardware and software, coupling it with a method to extract resolution from the legacy LDR stream. As time goes on, hardware and software will continue to improve, enabling real-time decoding of higher resolution HDR streams, simultaneously obviating the need for LDR data. Eventually, the LDR stream will become subservient to HDR, offering little more than a low-resolution tone-mapping suggestion for legacy devices. Color management will then move into display devices, and a high dynamic range profile connection space will become the preferred delivery medium.

While the basic recovery method described can be implemented efficiently on the GPU and works well enough for global tone-mapping operators, the advanced method relies on accumulation and sorting operations that are more conveniently carried out on the CPU, and is currently too slow for real-time playback. It should be possible to implement the advanced method in a more GPU-friendly way, by replacing our derivation of the impulse response function or off-loading this efficiently. This is left as future work, along with the reverse problem of recovering high frequencies in an LDR image given a high resolution HDR frame. We expect the solution to look very similar to the current one,

with better results thanks to the greater bit depth of the HDR stream.

References

- [Durand & Dorsey 2002] FREDO DURAND & JULIE DORSEY, “Fast Bilateral Filtering for the Display of High-Dynamic Range Images,” *ACM Transactions on Graphics*, 21, 3.
- [Fattal et al. 2002] RAANAN FATTAL, D. LISCHINSKI, M. WERMAN, “Gradient Domain High Dynamic Range Compression,” *ACM Transactions on Graphics*, 21, 3.
- [Larson et al. 1997] GREG WARD LARSON, HOLLY RUSHMEIER, CHRISTINE PIATKO, “A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes,” *IEEE Trans. on Visualization and Computer Graphics*, 3, 4.
- [Li et al. 2005] YUANZHEN LI, LAVANYA SHARAN, EDWARD ADELSON, “Compressing and Companding High Dynamic Range Images with Subband Architectures,” *Proc. of SIGGRAPH ’05 (Special issue of ACM Transactions on Graphics)*.
- [Mantiuk et al. 2004] RAFAL MANTIUK, GRZEGORZ KRAWCZYK, KAROL MYSZKOWSKI, HANS-PETER SEIDEL, “Perception-motivated High Dynamic Range Video Encoding,” *Proc. of SIGGRAPH ’04 (Special issue of ACM Transactions on Graphics)*.
- [Mantiuk et al. 2006] RAFAL MANTIUK, ALEXANDER EFREMOV, KAROL MYSZKOWSKI, HANS-PETER SEIDEL, “Backward Compatible High Dynamic Range MPEG Video Compression,” *Proc. of SIGGRAPH ’06 (Special issue of ACM Transactions on Graphics)*.
- [Reinhard et al. 2002] ERIK REINHARD, M. STARK, P. SHIRLEY, J. FERWERDA, “Photographic Tone Reproduction for Digital Images,” *ACM Transactions on Graphics*, 21, 3.
- [Reinhard et al. 2005] ERIK REINHARD, GREG WARD, SUMANTA PATTANAIK, PAUL DEBEVEC, *High Dynamic Range Imaging: Acquisition, Display, and Image-based Lighting*, Morgan Kaufmann Publishers, San Francisco.
- [Spaulding et al. 2003] KEVIN SPAULDING, G. J. WOOLFE AND R. L. JOSHI, “Extending the Color Gamut and Dynamic Range of an sRGB Image using a Residual Image,” *Color Res. Appl.* 28.
- [Ward & Simmons 2004] GREG WARD & MARYANN SIMMONS, “Subband Encoding of High Dynamic Range Imagery,” *First Symposium on Applied Perception in Graphics and Visualization (APGV)*.
- [Ward & Simmons 2005] GREG WARD & MARYANN SIMMONS, “JPEG-HDR: A Backwards-Compatible, High Dynamic Range Extension to JPEG,” *Proceedings of the Thirteenth Color Imaging Conference*.
- [Xu et al. 2005] RULIFENG XU, SUMANTA PATTANAIK, CHARLES HUGHES, “High-dynamic range still-image encoding in JPEG 2000,” *IEEE Comp. Graph. and Appl.* 26, 6.



59.2: Defining Dynamic Range

Greg Ward

Dolby Canada, Vancouver, BC, Canada

Abstract

We introduce a new metric for dynamic range of displays that closely corresponds with human perception in practical settings. The Number of Distinguishable Grays is a count of the visible luminance steps from the deepest black to the brightest white a display produces under known ambient conditions.

1 Introduction

Progress in high dynamic range displays relies on adequate methods for comparing and understanding differences between competing technologies and products. This is currently hampered by the lack of a clear definition of dynamic range, usually represented as the maximum contrast ratio or CR. The Video Electronics Standards Assoc. (VESA) 2.0 standard [1,2] specifies measurement conditions for full-screen white and full-screen black. The ratio of these two values is most often used by manufacturers to denote maximum contrast for their displays, despite complaints by both consumers and manufacturers that the specification permits excessive tampering with display settings prior to such measurements to obtain unrealistic CR numbers [3].

A more critical problem is the precise meaning of “black” in such measurements, which ignores the effect of ambient lighting and the importance of visible quantization in determining whether the measured minimum is useful in any practical context. Emerging high dynamic range displays may have 10, 12, and even 16-bit/channel inputs, and a metric that quantifies the effect of bit depth and gamma response on perception is essential. We address each of these issues in our recommended approach, and suggest a new dynamic range metric that incorporates ambient lighting and quantization, which we call the *Number of Distinguishable Grays* (NDG).

The question NDG answers is, “How many gray levels will a standard observer be able to distinguish between the darkest black and the brightest white in a given ambient environment?” We submit that this is an easily understood concept from a buyer’s perspective and a meaningful metric from a vendor’s standpoint, which defines dynamic range more reliably than the contrast ratio. Furthermore, NDG values are perceptually well-spaced by design; i.e., a display with twice the NDG value has roughly twice the apparent dynamic range.

2 NDG Method

Determining the Number of Distinguishable Grays for a display requires the following three pieces of information, comprised of two measurements and a specified condition:

1. The gray level luminance response of the display for identical RGB inputs, $L_d(i)$, including the range of i (e.g., 0-1023 for 10-bit/channel input).
2. The average hemispherical reflectance of the display surface.
3. The presumed ambient lighting for this NDG evaluation.

If the display is assumed to be in a completely darkened environment, the hemispherical reflectance is not needed. However, we strongly recommend that any new standard require at least one NDG evaluation under typical indoor lighting conditions, either home lighting (around 50 lux vertical illuminance) for television units, or office lighting (around 200 lux) for video display terminals. This may be given in addition to the “best case” darkened viewing environment. Even for televisions, the worst case for front reflections may be dozens of times greater than the proposed 50 lux level.

The NDG value may then be computed with the following formula:

$$NDG = \sum_{i=1}^{i_{\max}} \min\left(\frac{L_d(i) - L_d(i-1)}{D(L_d(i)) + L_a}, 1\right)$$

where L_a is the reflected display luminance due to ambient lighting, computed as the vertical ambient illuminance times the hemispherical front reflectance, divided by π . The $D(L)$ function is an agreed-upon *threshold versus intensity* or *t.v.i.* curve based on psychophysical data. A few such curves have been proposed, such as Barten’s formula employed in the DICOM standard [4], Ferwerda et al.’s perceptually-based tone-mapping [5], and Lubin and Pica’s perceptual quantizer [6]. The choice of t.v.i. curve is not critical, so long as it is well-specified and adhered to by the standard.

The $\min(f, 1)$ function in the NDG formula serves an important purpose, which is to count fractional steps for sub-threshold jumps, while never giving more than one NDG per control quanta. In other words, we will not give an NDG of 4096 to a 12-bit display unless an observer can see every one of those steps. Nor will we give an NDG value of 256 to an 8-bit display if the quanta near white take supra-threshold steps while the bottom 20 quanta are lost in unwanted screen reflections.

Display	Screen Reflectance	Maximum Luminance	Ambient Level (incident illuminance)	Contrast Ratio	Number of Distinguishable Grays
17" LCD A 8-bit input	2%	200 cd/m ²	0 50 200	400 244 113	240
					237
					234
17" LCD B 8-bit input	1%	200 cd/m ²	0 50 200	400 303 176	240
					238
					236
17" LCD C 8-bit input	1%	400 cd/m ²	0 50 200	400 345 244	242
					242
					241
17" LCD D 10-bit input	1%	400 cd/m ²	0 50 200	400 345 244	351
					348
					338

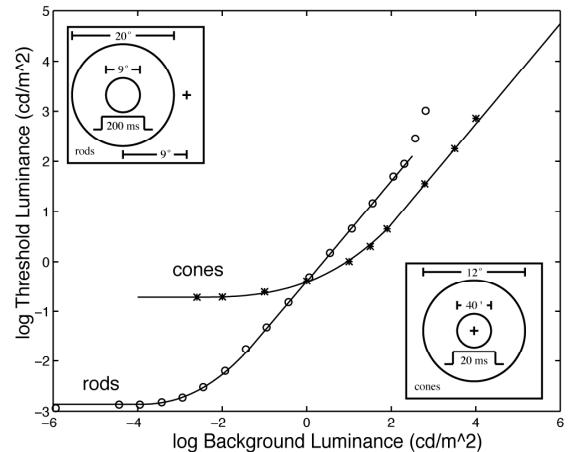
Table 1. The CRs and NDGs of four hypothetical displays under three ambient conditions.

In a simple illustration of the NDG metric, we compared four hypothetical displays, measured at different ambient lighting levels. We assume an sRGB response from each, and a VESA CR of 400. LCD **A** has a screen reflectance of 2% and LCD **B** has a screen reflectance of 1%. Both **A** and **B** have the same maximum luminance of 200 cd/m². In addition to having a lower screen reflectance, LCD **C** has a boosted maximum luminance of 400 cd/m², and LCD **D** provides a 10-bit input. (These specifications are chosen for illustrative purposes rather than to match any actual display.) Our $D(L)$ curve was adapted from [5], as described in the following subsection. The “results” of our virtual test are summarized in Table 1.

As we can see from our table, the NDG values for displays A, B, and C are limited primarily by their 8-bit inputs, whose delta between levels is greater than the visible threshold over most of the range. Ambient lighting and display reflectance have only a minor influence by comparison, and it is only when we move to the 10-bit input of display D that the NDG becomes an actual count of visible thresholds over the display luminance range. It is important that the 10-bit response follows the sRGB curve. If we had switched to a linear response when we increased the number of bits to 10, we would have only have achieved a NDG of 298 at $L_a=0$. This is because the linear step sizes would exceed the $D(L_a)$ value at the bottom end of the range.

2.1 Threshold vs. Intensity

As noted earlier, calculating NDG from display output requires a human threshold vs. intensity curve relating threshold response to adaptation luminance. Psychophysical studies have established this non-linear relationship, shown for rods and cones in Figure 1 [5]. The piecewise fit to the combined curve is written in Table 2 [7].

**Figure 1.** Plot of threshold vs. intensity (t.v.i.) for human vision.

log10 of Just Noticeable Difference tvi(L_a)	Applicable Adapted Luminance Range
-2.86	$\log_{10}(L_a) < -3.94$
$(0.405 \log_{10}(L_a) + 1.6)^{2.18} - 2.86$	$-3.94 \leq \log_{10}(L_a) < -1.44$
$\log_{10}(L_a) - 0.395$	$-1.44 \leq \log_{10}(L_a) < -0.0184$
$(0.249 \log_{10}(L_a) + 0.65)^{2.7} - 0.72$	$-0.0184 \leq \log_{10}(L_a) < 1.9$
$\log_{10}(L_a) - 1.255$	$\log_{10}(L_a) \geq 1.9$

Table 2. Numerical fit to t.v.i. shown in Figure 1.

There is general agreement that Ferwerda's thresholds are on the high side, and banding may be visible in gradient images at thresholds as low as $1/10^{\text{th}}$ of the reported t.v.i. curve (corresponding to 0.6% steps in the photopic region). The probable reason for this discrepancy is that the experiments from which this function is derived use a pulsing target on a constant background, and the thresholds for such transient stimuli are higher than those for static stimuli. We therefore recommend subtracting 0.95 from the formulae on the left side of Table 2, which is equivalent to dividing the JND threshold luminances by a factor of 9. This brings our t.v.i. function into better agreement with the Barten model recommended for radiology by the Digital Imaging and Communications in Medicine (DICOM) standards body, as shown in Figure 2. We did not follow the DICOM recommendation directly because their fit to Barten's data only covers part of the visible luminance range. This range, $0.05:4000 \text{ cd/m}^2$, while greater than any conventional display, is on par with dual modulator displays just entering the market, and may be exceeded in the near future. It is therefore preferable to either extend the Barten fit to cover the full range of human vision, or adjust Ferwerda's model to match a more realistic threshold. We follow the latter strategy, although either approach will work.

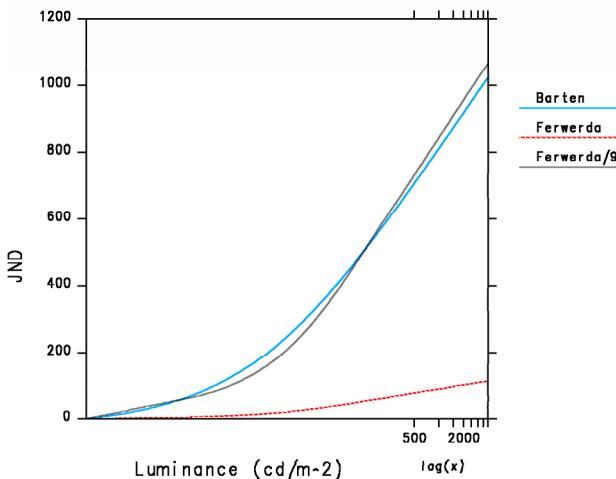


Figure 2. The DICOM fit to Barten's model, and our adjustment of Ferwerda's.

We use this modified Ferwerda model for $D(L)$:

$$D(L) = 10^{tvi(L)-0.95}$$

Using a threshold model places an upper limit on the acceptable quantization error, and offers a staircase scale corresponding to visible luminance differences over a given display's dynamic range independent of quantization. This allows us to distinguish between a display that covers a wide dynamic range, but in a region where humans can barely see (e.g., below 10^{-2} cd/m^2) versus a display that covers a range where we see well.

3 Conclusion

Knowledgeable consumers and many manufacturers agree that the current practice for measuring maximum contrast ratio, if not the VESA 2.0 standard itself, is suspect to the point of being meaningless for comparisons. One solution is to shore up the VESA standard by requiring that manufacturers fix their monitor settings to their defaults prior to measurements. Product specifications should also include the effects of ambient illumination on delivered contrast, perhaps offering a few different levels so the consumer may judge appropriately for their intended application.

However, fixing the contrast ratio measurement alone does not provide for intelligent evaluation of displays with greater bit depth and controllable dynamic range. For emerging high dynamic range displays, we need a sensible way to quantify the delivered range of visible luminances. In this paper, we have suggested the *Number of Distinguishable Grays* specification. By measuring the gray output levels of a display for each $\text{RGB}=(i,i,i)$ input and applying the given formula, one arrives at a quantity with good correspondence to the actual visible differences, accounting simultaneously for the effects of brightness, contrast, ambient lighting, and quantization. With a modest education effort, we believe the NDG concept is intuitive and will be appreciated by consumers and professionals alike.

4 References

- [1] "Flat Panel Display Measurements" 2.0 avail. from www.vesa.org
- [2] "Flat Panel Display Measurements Update" avail. from www.vesa.org
- [3] Tom Mainelli, "Tested contrast ratios rarely conform to vendors' specs," *PC World e-zine* www.pcworld.com/news/article/0,aid,110483,pg,1,00.a sp, May 2003.
- [4] "Part 14: Grayscale Standard Display Function," *Digital Imaging and Communications in Medicine (DICOM)*, National Electrical Manufacturer's Assoc.
- [5] James Ferwerda, S. Pattanaik, P. Shirley and D. Greenberg, "A Model of Visual Adaptation for Realistic Image Synthesis," *Proceedings of ACM SIGGRAPH '96*
- [6] Jeffrey Lubin, A. Pica, "A Non-Uniform Quantizer Matched to Human Visual Performance," *Soc. for Information Display Digest*, 1991.
- [7] Greg Ward Larson, H. Rushmeier, C. Piatko, "A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 3, No. 4, December 1997. radsite.lbl.gov/radiance/papers/lbnl39882/tonemap.pdf



Recovering High Dynamic Range Radiance Maps from Photographs

Paul E. Debevec

Jitendra Malik

University of California at Berkeley¹

ABSTRACT

We present a method of recovering high dynamic range radiance maps from photographs taken with conventional imaging equipment. In our method, multiple photographs of the scene are taken with different amounts of exposure. Our algorithm uses these differently exposed photographs to recover the response function of the imaging process, up to a factor of scale, using the assumption of reciprocity. With the known response function, the algorithm can fuse the multiple photographs into a single, high dynamic range radiance map whose pixel values are proportional to the true radiance values in the scene. We demonstrate our method on images acquired with both photochemical and digital imaging processes. We discuss how this work is applicable in many areas of computer graphics involving digitized photographs, including image-based modeling, image compositing, and image processing. Lastly, we demonstrate a few applications of having high dynamic range radiance maps, such as synthesizing realistic motion blur and simulating the response of the human visual system.

CR Descriptors: I.2.10 [Artificial Intelligence]: Vision and Scene Understanding - *Intensity, color, photometry and thresholding*; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism - *Color, shading, shadowing, and texture*; I.4.1 [Image Processing]: Digitization - *Scanning*; I.4.8 [Image Processing]: Scene Analysis - *Photometry, Sensor Fusion*.

1 Introduction

Digitized photographs are becoming increasingly important in computer graphics. More than ever, scanned images are used as texture maps for geometric models, and recent work in image-based modeling and rendering uses images as the fundamental modeling primitive. Furthermore, many of today's graphics applications require computer-generated images to mesh seamlessly with real photographic imagery. Properly using photographically acquired imagery in these applications can greatly benefit from an accurate model of the photographic process.

When we photograph a scene, either with film or an electronic imaging array, and digitize the photograph to obtain a two-dimensional array of "brightness" values, these values are rarely

true measurements of relative radiance in the scene. For example, if one pixel has twice the value of another, it is unlikely that it observed twice the radiance. Instead, there is usually an unknown, nonlinear mapping that determines how radiance in the scene becomes pixel values in the image.

This nonlinear mapping is hard to know beforehand because it is actually the composition of several nonlinear mappings that occur in the photographic process. In a conventional camera (see Fig. 1), the film is first exposed to light to form a latent image. The film is then developed to change this latent image into variations in transparency, or *density*, on the film. The film can then be digitized using a film scanner, which projects light through the film onto an electronic light-sensitive array, converting the image to electrical voltages. These voltages are digitized, and then manipulated before finally being written to the storage medium. If prints of the film are scanned rather than the film itself, then the printing process can also introduce nonlinear mappings.

In the first stage of the process, the film response to variations in exposure X (which is $E\Delta t$, the product of the irradiance E the film receives and the exposure time Δt) is a non-linear function, called the "characteristic curve" of the film. Noteworthy in the typical characteristic curve is the presence of a small response with no exposure and saturation at high exposures. The development, scanning and digitization processes usually introduce their own nonlinearities which compose to give the aggregate nonlinear relationship between the image pixel exposures X and their values Z .

Digital cameras, which use charge coupled device (CCD) arrays to image the scene, are prone to the same difficulties. Although the charge collected by a CCD element is proportional to its irradiance, most digital cameras apply a nonlinear mapping to the CCD outputs before they are written to the storage medium. This nonlinear mapping is used in various ways to mimic the response characteristics of film, anticipate nonlinear responses in the display device, and often to convert 12-bit output from the CCD's analog-to-digital converters to 8-bit values commonly used to store images. As with film, the most significant nonlinearity in the response curve is at its saturation point, where any pixel with a radiance above a certain level is mapped to the same maximum image value.

Why is this any problem at all? The most obvious difficulty, as any amateur or professional photographer knows, is that of limited dynamic range—one has to choose the range of radiance values that are of interest and determine the exposure time suitably. Sunlit scenes, and scenes with shiny materials and artificial light sources, often have extreme differences in radiance values that are impossible to capture without either under-exposing or saturating the film. To cover the full dynamic range in such a scene, one can take a series of photographs with different exposures. This then poses a problem: how can we combine these separate images into a composite radiance map? Here the fact that the mapping from scene radiance to pixel values is unknown and nonlinear begins to haunt us. The purpose of this paper is to present a simple technique for recovering this response function, up to a scale factor, using nothing more than a set of photographs taken with varying, known exposure durations. With this mapping, we then use the pixel values from all available photographs to construct an accurate map of the radiance in the scene, up to a factor of scale. This radiance map will cover

¹Computer Science Division, University of California at Berkeley, Berkeley, CA 94720-1776. Email: debevec@cs.berkeley.edu, malik@cs.berkeley.edu. More information and additional results may be found at: <http://www.cs.berkeley.edu/~debevec/Research>

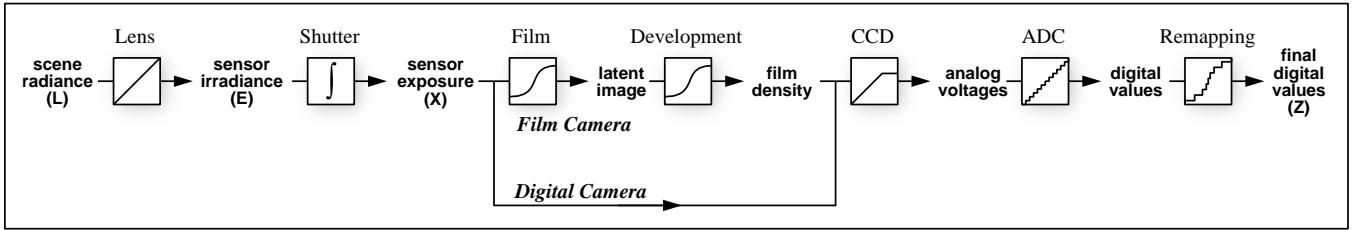


Figure 1: **Image Acquisition Pipeline** shows how scene radiance becomes pixel values for both film and digital cameras. Unknown nonlinear mappings can occur during exposure, development, scanning, digitization, and remapping. The algorithm in this paper determines the aggregate mapping from scene radiance L to pixel values Z from a set of differently exposed images.

the entire dynamic range captured by the original photographs.

1.1 Applications

Our technique of deriving imaging response functions and recovering high dynamic range radiance maps has many possible applications in computer graphics:

Image-based modeling and rendering

Image-based modeling and rendering systems to date (e.g. [11, 15, 2, 3, 12, 6, 17]) make the assumption that all the images are taken with the same exposure settings and film response functions. However, almost any large-scale environment will have some areas that are much brighter than others, making it impossible to adequately photograph the scene using a single exposure setting. In indoor scenes with windows, this situation often arises within the field of view of a single photograph, since the areas visible through the windows can be far brighter than the areas inside the building.

By determining the response functions of the imaging device, the method presented here allows one to correctly fuse pixel data from photographs taken at different exposure settings. As a result, one can properly photograph outdoor areas with short exposures, and indoor areas with longer exposures, without creating inconsistencies in the data set. Furthermore, knowing the response functions can be helpful in merging photographs taken with different imaging systems, such as video cameras, digital cameras, and film cameras with various film stocks and digitization processes.

The area of image-based modeling and rendering is working toward recovering more advanced reflection models (up to complete BRDF's) of the surfaces in the scene (e.g. [21]). These methods, which involve observing surface radiance in various directions under various lighting conditions, require absolute radiance values rather than the nonlinearly mapped pixel values found in conventional images. Just as important, the recovery of high dynamic range images will allow these methods to obtain accurate radiance values from surface specularities and from incident light sources. Such higher radiance values usually become clamped in conventional images.

Image processing

Most image processing operations, such as blurring, edge detection, color correction, and image correspondence, expect pixel values to be proportional to the scene radiance. Because of nonlinear image response, especially at the point of saturation, these operations can produce incorrect results for conventional images.

In computer graphics, one common image processing operation is the application of synthetic motion blur to images. In our results (Section 3), we will show that using true radiance maps produces significantly more realistic motion blur effects for high dynamic range scenes.

Image compositing

Many applications in computer graphics involve compositing image data from images obtained by different processes. For example, a background matte might be shot with a still camera, live action might be shot with a different film stock or scanning process, and CG elements would be produced by rendering algorithms. When there are significant differences in the response curves of these imaging processes, the composite image can be visually unconvincing. The technique presented in this paper provides a convenient and robust method of determining the overall response curve of any imaging process, allowing images from different processes to be used consistently as radiance maps. Furthermore, the recovered response curves can be inverted to render the composite radiance map as if it had been photographed with any of the original imaging processes, or a different imaging process entirely.

A research tool

One goal of computer graphics is to simulate the image formation process in a way that produces results that are consistent with what happens in the real world. Recovering radiance maps of real-world scenes should allow more quantitative evaluations of rendering algorithms to be made in addition to the qualitative scrutiny they traditionally receive. In particular, the method should be useful for developing reflectance and illumination models, and comparing global illumination solutions against ground truth data.

Rendering high dynamic range scenes on conventional display devices is the subject of considerable previous work, including [20, 16, 5, 23]. The work presented in this paper will allow such methods to be tested on real radiance maps in addition to synthetically computed radiance solutions.

1.2 Background

The photochemical processes involved in silver halide photography have been the subject of continued innovation and research ever since the invention of the daguerreotype in 1839. [18] and [8] provide a comprehensive treatment of the theory and mechanisms involved. For the newer technology of solid-state imaging with charge coupled devices, [19] is an excellent reference. The technical and artistic problem of representing the dynamic range of a natural scene on the limited range of film has concerned photographers from the early days – [1] presents one of the best known systems to choose shutter speeds, lens apertures, and developing conditions to best coerce the dynamic range of a scene to fit into what is possible on a print. In scientific applications of photography, such as in astronomy, the nonlinear film response has been addressed by suitable calibration procedures. It is our objective instead to develop a simple self-calibrating procedure not requiring calibration charts or photometric measuring devices.

In previous work, [13] used multiple flux integration times of a CCD array to acquire extended dynamic range images. Since direct CCD outputs were available, the work did not need to deal with the

problem of nonlinear pixel value response. [14] addressed the problem of nonlinear response but provide a rather limited method of recovering the response curve. Specifically, a parametric form of the response curve is arbitrarily assumed, there is no satisfactory treatment of image noise, and the recovery process makes only partial use of the available data.

2 The Algorithm

This section presents our algorithm for recovering the film response function, and then presents our method of reconstructing the high dynamic range radiance image from the multiple photographs. We describe the algorithm assuming a grayscale imaging device. We discuss how to deal with color in Section 2.6.

2.1 Film Response Recovery

Our algorithm is based on exploiting a physical property of imaging systems, both photochemical and electronic, known as *reciprocity*.

Let us consider photographic film first. The response of a film to variations in exposure is summarized by the characteristic curve (or Hurter-Driffield curve). This is a graph of the optical density D of the processed film against the logarithm of the exposure X to which it has been subjected. The exposure X is defined as the product of the irradiance E at the film and exposure time, Δt , so that its units are J m^{-2} . Key to the very concept of the characteristic curve is the assumption that only the product $E\Delta t$ is important, that halving E and doubling Δt will not change the resulting optical density D . Under extreme conditions (very large or very low Δt), the reciprocity assumption can break down, a situation described as reciprocity failure. In typical print films, reciprocity holds to within $\frac{1}{3}$ stop¹ for exposure times of 10 seconds to 1/10,000 of a second.² In the case of charge coupled arrays, reciprocity holds under the assumption that each site measures the total number of photons it absorbs during the integration time.

After the development, scanning and digitization processes, we obtain a digital number Z , which is a nonlinear function of the original exposure X at the pixel. Let us call this function f , which is the composition of the characteristic curve of the film as well as all the nonlinearities introduced by the later processing steps. Our first goal will be to recover this function f . Once we have that, we can compute the exposure X at each pixel, as $X = f^{-1}(Z)$. We make the reasonable assumption that the function f is monotonically increasing, so its inverse f^{-1} is well defined. Knowing the exposure X and the exposure time Δt , the irradiance E is recovered as $E = X/\Delta t$, which we will take to be proportional to the radiance L in the scene.³

Before proceeding further, we should discuss the consequences of the spectral response of the sensor. The exposure X should be thought of as a function of wavelength $X(\lambda)$, and the abscissa on the characteristic curve should be the integral $\int X(\lambda)R(\lambda)d\lambda$ where $R(\lambda)$ is the spectral response of the sensing element at the pixel location. Strictly speaking, our use of irradiance, a radiometric quantity, is not justified. However, the spectral response of the sensor site may not be the photopic luminosity function V_λ , so the photometric term *illuminance* is not justified either. In what follows, we will use the term irradiance, while urging the reader to remember that the

quantities we will be dealing with are weighted by the spectral response at the sensor site. For color photography, the color channels may be treated separately.

The input to our algorithm is a number of digitized photographs taken from the same vantage point with different known exposure durations Δt_j .⁴ We will assume that the scene is static and that this process is completed quickly enough that lighting changes can be safely ignored. It can then be assumed that the film irradiance values E_i for each pixel i are constant. We will denote pixel values by Z_{ij} where i is a spatial index over pixels and j indexes over exposure times Δt_j . We may now write down the film reciprocity equation as:

$$Z_{ij} = f(E_i \Delta t_j) \quad (1)$$

Since we assume f is monotonic, it is invertible, and we can rewrite (1) as:

$$f^{-1}(Z_{ij}) = E_i \Delta t_j$$

Taking the natural logarithm of both sides, we have:

$$\ln f^{-1}(Z_{ij}) = \ln E_i + \ln \Delta t_j$$

To simplify notation, let us define function $g = \ln f^{-1}$. We then have the set of equations:

$$g(Z_{ij}) = \ln E_i + \ln \Delta t_j \quad (2)$$

where i ranges over pixels and j ranges over exposure durations. In this set of equations, the Z_{ij} are known, as are the Δt_j . The unknowns are the irradiances E_i , as well as the function g , although we assume that g is smooth and monotonic.

We wish to recover the function g and the irradiances E_i that best satisfy the set of equations arising from Equation 2 in a least-squared error sense. We note that recovering g only requires recovering the finite number of values that $g(z)$ can take since the domain of Z , pixel brightness values, is finite. Letting Z_{min} and Z_{max} be the least and greatest pixel values (integers), N be the number of pixel locations and P be the number of photographs, we formulate the problem as one of finding the $(Z_{max} - Z_{min} + 1)$ values of $g(Z)$ and the N values of $\ln E_i$ that minimize the following quadratic objective function:

$$\mathcal{O} = \sum_{i=1}^N \sum_{j=1}^P [g(Z_{ij}) - \ln E_i - \ln \Delta t_j]^2 + \lambda \sum_{z=Z_{min}+1}^{Z_{max}-1} g''(z)^2 \quad (3)$$

The first term ensures that the solution satisfies the set of equations arising from Equation 2 in a least squares sense. The second term is a smoothness term on the sum of squared values of the second derivative of g to ensure that the function g is smooth; in this discrete setting we use $g''(z) = g(z-1) - 2g(z) + g(z+1)$. This smoothness term is essential to the formulation in that it provides coupling between the values $g(z)$ in the minimization. The scalar λ weights the smoothness term relative to the data fitting term, and should be chosen appropriately for the amount of noise expected in the Z_{ij} measurements.

Because it is quadratic in the E_i 's and $g(z)$'s, minimizing \mathcal{O} is a straightforward linear least squares problem. The overdetermined

¹1 stop is a photographic term for a factor of two; $\frac{1}{3}$ stop is thus $2^{\frac{1}{3}}$

²An even larger dynamic range can be covered by using neutral density filters to lessen the amount of light reaching the film for a given exposure time. A discussion of the modes of reciprocity failure may be found in [18], ch. 4.

³ L is proportional to E for any particular pixel, but it is possible for the proportionality factor to be different at different places on the sensor. One formula for this variance, given in [7], is $E = L \frac{\pi}{4} \left(\frac{d}{f} \right)^2 \cos^4 \alpha$, where α measures the pixel's angle from the lens' optical axis. However, most modern camera lenses are designed to compensate for this effect, and provide a nearly constant mapping between radiance and irradiance at f/8 and smaller apertures. See also [10].

⁴Most modern SLR cameras have electronically controlled shutters which give extremely accurate and reproducible exposure times. We tested our Canon EOS Elan camera by using a Macintosh to make digital audio recordings of the shutter. By analyzing these recordings we were able to verify the accuracy of the exposure times to within a thousandth of a second. Conveniently, we determined that the actual exposure times varied by powers of two between stops ($\frac{1}{64}, \frac{1}{32}, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8, 16, 32$), rather than the rounded numbers displayed on the camera readout ($\frac{1}{60}, \frac{1}{30}, \frac{1}{15}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8, 15, 30$). Because of problems associated with vignetting, varying the aperture is not recommended.

system of linear equations is robustly solved using the singular value decomposition (SVD) method. An intuitive explanation of the procedure may be found in Fig. 2.

We need to make three additional points to complete our description of the algorithm:

First, the solution for the $g(z)$ and E_i values can only be up to a single scale factor α . If each log irradiance value $\ln E_i$ were replaced by $\ln E_i + \alpha$, and the function g replaced by $g + \alpha$, the system of equations 2 and also the objective function \mathcal{O} would remain unchanged. To establish a scale factor, we introduce the additional constraint $g(Z_{mid}) = 0$, where $Z_{mid} = \frac{1}{2}(Z_{min} + Z_{max})$, simply by adding this as an equation in the linear system. The meaning of this constraint is that a pixel with value midway between Z_{min} and Z_{max} will be assumed to have unit exposure.

Second, the solution can be made to have a much better fit by anticipating the basic shape of the response function. Since $g(z)$ will typically have a steep slope near Z_{min} and Z_{max} , we should expect that $g(z)$ will be less smooth and will fit the data more poorly near these extremes. To recognize this, we can introduce a weighting function $w(z)$ to emphasize the smoothness and fitting terms toward the middle of the curve. A sensible choice of w is a simple hat function:

$$w(z) = \begin{cases} z - Z_{min} & \text{for } z \leq \frac{1}{2}(Z_{min} + Z_{max}) \\ Z_{max} - z & \text{for } z > \frac{1}{2}(Z_{min} + Z_{max}) \end{cases} \quad (4)$$

Equation 3 now becomes:

$$\mathcal{O} = \sum_{i=1}^N \sum_{j=1}^P \{w(Z_{ij})[g(Z_{ij}) - \ln E_i - \ln \Delta t_j]\}^2 + \lambda \sum_{z=Z_{min}+1}^{Z_{max}-1} [w(z)g''(z)]^2$$

Finally, we need not use every available pixel site in this solution procedure. Given measurements of N pixels in P photographs, we have to solve for N values of $\ln E_i$ and $(Z_{max} - Z_{min})$ samples of g . To ensure a sufficiently overdetermined system, we want $N(P-1) > (Z_{max} - Z_{min})$. For the pixel value range $(Z_{max} - Z_{min}) = 255$, $P = 11$ photographs, a choice of N on the order of 50 pixels is more than adequate. Since the size of the system of linear equations arising from Equation 3 is on the order of $N \times P + Z_{max} - Z_{min}$, computational complexity considerations make it impractical to use every pixel location in this algorithm. Clearly, the pixel locations should be chosen so that they have a reasonably even distribution of pixel values from Z_{min} to Z_{max} , and so that they are spatially well distributed in the image. Furthermore, the pixels are best sampled from regions of the image with low intensity variance so that radiance can be assumed to be constant across the area of the pixel, and the effect of optical blur of the imaging system is minimized. So far we have performed this task by hand, though it could easily be automated.

Note that we have not explicitly enforced the constraint that g must be a monotonic function. If desired, this can be done by transforming the problem to a non-negative least squares problem. We have not found it necessary because, in our experience, the smoothness penalty term is enough to make the estimated g monotonic in addition to being smooth.

To show its simplicity, the MATLAB routine we used to minimize Equation 5 is included in the Appendix. Running times are on the order of a few seconds.

2.2 Constructing the High Dynamic Range Radiance Map

Once the response curve g is recovered, it can be used to quickly convert pixel values to relative radiance values, assuming the exposure Δt_j is known. Note that the curve can be used to determine radiance values in any image(s) acquired by the imaging process associated with g , not just the images used to recover the response function.

From Equation 2, we obtain:

$$\ln E_i = g(Z_{ij}) - \ln \Delta t_j \quad (5)$$

For robustness, and to recover high dynamic range radiance values, we should use all the available exposures for a particular pixel to compute its radiance. For this, we reuse the weighting function in Equation 4 to give higher weight to exposures in which the pixel's value is closer to the middle of the response function:

$$\ln E_i = \frac{\sum_{j=1}^P w(Z_{ij})(g(Z_{ij}) - \ln \Delta t_j)}{\sum_{j=1}^P w(Z_{ij})} \quad (6)$$

Combining the multiple exposures has the effect of reducing noise in the recovered radiance values. It also reduces the effects of imaging artifacts such as film grain. Since the weighting function ignores saturated pixel values, “blooming” artifacts⁵ have little impact on the reconstructed radiance values.

2.2.1 Storage

In our implementation the recovered radiance map is computed as an array of single-precision floating point values. For efficiency, the map can be converted to the image format used in the RADIANCE [22] simulation and rendering system, which uses just eight bits for each of the mantissa and exponent. This format is particularly compact for color radiance maps, since it stores just one exponent value for all three color values at each pixel. Thus, in this format, a high dynamic range radiance map requires just one third more storage than a conventional RGB image.

2.3 How many images are necessary?

To decide on the number of images needed for the technique, it is convenient to consider the two aspects of the process:

- 1. Recovering the film response curve:* This requires a minimum of two photographs. Whether two photographs are enough can be understood in terms of the heuristic explanation of the process of film response curve recovery shown in Fig. 2. If the scene has sufficiently many different radiance values, the entire curve can, in principle, be assembled by sliding together the sampled curve segments, each with only two samples. Note that the photos must be similar enough in their exposure amounts that some pixels fall into the working range⁶ of the film in both images; otherwise, there is no information to relate the exposures to each other. Obviously, using more than two images with differing exposure times improves performance with respect to noise sensitivity.
- 2. Recovering a radiance map given the film response curve:* The number of photographs needed here is a function of the dynamic range of radiance values in the scene. Suppose the range of maximum to minimum radiance values that we are

⁵Blooming occurs when charge or light at highly saturated sites on the imaging surface spills over and affects values at neighboring sites.

⁶The working range of the film corresponds to the middle section of the response curve. The ends of the curve, in which large changes in exposure cause only small changes in density (or pixel value), are called the *toe* and the *shoulder*.

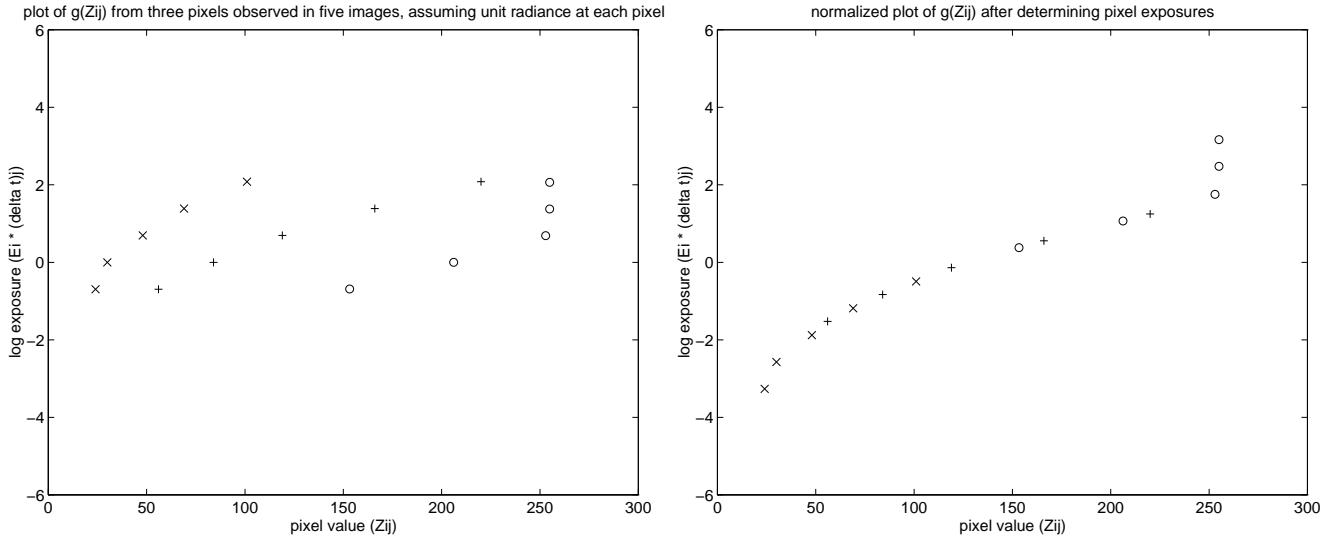


Figure 2: In the figure on the left, the \times symbols represent samples of the g curve derived from the digital values at one pixel for 5 different known exposures using Equation 2. The unknown log irradiance $\ln E_i$ has been arbitrarily assumed to be 0. Note that the shape of the g curve is correct, though its position on the vertical scale is arbitrary corresponding to the unknown $\ln E_i$. The $+$ and \circ symbols show samples of g curve segments derived by consideration of two other pixels; again the vertical position of each segment is arbitrary. Essentially, what we want to achieve in the optimization process is to slide the 3 sampled curve segments up and down (by adjusting their $\ln E_i$'s) until they "line up" into a single smooth, monotonic curve, as shown in the right figure. The vertical position of the composite curve will remain arbitrary.

interested in recovering accurately is R , and the film is capable of representing in its working range a dynamic range of F . Then the minimum number of photographs needed is $\lceil \frac{R}{F} \rceil$ to ensure that every part of the scene is imaged in at least one photograph at an exposure duration that puts it in the working range of the film response curve. As in recovering the response curve, using more photographs than strictly necessary will result in better noise sensitivity.

If one wanted to use as few photographs as possible, one might first recover the response curve of the imaging process by photographing a scene containing a diverse range of radiance values at three or four different exposures, differing by perhaps one or two stops. This response curve could be used to determine the working range of the imaging process, which for the processes we have seen would be as many as five or six stops. For the remainder of the shoot, the photographer could decide for any particular scene the number of shots necessary to cover its entire dynamic range. For diffuse indoor scenes, only one exposure might be necessary; for scenes with high dynamic range, several would be necessary. By recording the exposure amount for each shot, the images could then be converted to radiance maps using the pre-computed response curve.

2.4 Recovering extended dynamic range from single exposures

Most commercially available film scanners can detect reasonably close to the full range of useful densities present in film. However, many of these scanners (as well as the Kodak PhotoCD process) produce 8-bit-per-channel images designed to be viewed on a screen or printed on paper. Print film, however, records a significantly greater dynamic range than can be displayed with either of these media. As a result, such scanners deliver only a portion of the detected dynamic range of print film in a single scan, discarding information in either high or low density regions. The portion of the detected dynamic range that is delivered can usually be influenced by "brightness" or "density adjustment" controls.

The method presented in this paper enables two methods for recovering the full dynamic range of print film which we will briefly

outline⁷. In the first method, the print negative is scanned with the scanner set to scan slide film. Most scanners will then record the entire detectable dynamic range of the film in the resulting image. As before, a series of differently exposed images of the same scene can be used to recover the response function of the imaging system with each of these scanner settings. This response function can then be used to convert individual exposures to radiance maps. Unfortunately, since the resulting image is still 8-bits-per-channel, this results in increased quantization.

In the second method, the film can be scanned twice with the scanner set to different density adjustment settings. A series of differently exposed images of the same scene can then be used to recover the response function of the imaging system at each of these density adjustment settings. These two response functions can then be used to combine two scans of any single negative using a similar technique as in Section 2.2.

2.5 Obtaining Absolute Radiance

For many applications, such as image processing and image compositing, the relative radiance values computed by our method are all that are necessary. If needed, an approximation to the scaling term necessary to convert to absolute radiance can be derived using the ASA of the film⁸ and the shutter speeds and exposure amounts in the photographs. With these numbers, formulas that give an approximate prediction of film response can be found in [9]. Such an approximation can be adequate for simulating visual artifacts such as glare, and predicting areas of scotopic retinal response. If desired, one could recover the scaling factor precisely by photographing a calibration luminaire of known radiance, and scaling the radiance values to agree with the known radiance of the luminaire.

2.6 Color

Color images, consisting of red, green, and blue channels, can be processed by reconstructing the imaging system response curve for

⁷This work was done in collaboration with Gregory Ward Larson

⁸Conveniently, most digital cameras also specify their sensitivity in terms of ASA.

each channel independently. Unfortunately, there will be three unknown scaling factors relating relative radiance to absolute radiance, one for each channel. As a result, different choices of these scaling factors will change the color balance of the radiance map.

By default, the algorithm chooses the scaling factor such that a pixel with value Z_{mid} will have unit exposure. Thus, any pixel with the RGB value $(Z_{mid}, Z_{mid}, Z_{mid})$ will have equal radiance values for R, G, and B, meaning that the pixel is achromatic. If the three channels of the imaging system actually do respond equally to achromatic light in the neighborhood of Z_{mid} , then our procedure correctly reconstructs the relative radiances.

However, films are usually calibrated to respond achromatically to a particular color of light C , such as sunlight or fluorescent light. In this case, the radiance values of the three channels should be scaled so that the pixel value $(Z_{mid}, Z_{mid}, Z_{mid})$ maps to a radiance with the same color ratios as C . To properly model the color response of the entire imaging process rather than just the film response, the scaling terms can be adjusted by photographing a calibration luminaire of known color.

2.7 Taking virtual photographs

The recovered response functions can also be used to map radiance values back to pixel values for a given exposure Δt using Equation 1. This process can be thought of as taking a virtual photograph of the radiance map, in that the resulting image will exhibit the response qualities of the modeled imaging system. Note that the response functions used need not be the same response functions used to construct the original radiance map, which allows photographs acquired with one imaging process to be rendered as if they were acquired with another.⁹

3 Results

Figures 3-5 show the results of using our algorithm to determine the response curve of a DCS460 digital camera. Eleven grayscale photographs filtered down to 765×509 resolution (Fig. 3) were taken at f/8 with exposure times ranging from $\frac{1}{30}$ of a second to 30 seconds, with each image receiving twice the exposure of the previous one. The film curve recovered by our algorithm from 45 pixel locations observed across the image sequence is shown in Fig. 4. Note that although CCD image arrays naturally produce linear output, from the curve it is evident that the camera nonlinearly remaps the data, presumably to mimic the response curves found in film. The underlying registered $(E_i \Delta t_j, Z_{ij})$ data are shown as light circles underneath the curve; some outliers are due to sensor artifacts (light horizontal bands across some of the darker images.)

Fig. 5 shows the reconstructed high dynamic range radiance map. To display this map, we have taken the logarithm of the radiance values and mapped the range of these values into the range of the display. In this representation, the pixels at the light regions do not saturate, and detail in the shadow regions can be made out, indicating that all of the information from the original image sequence is present in the radiance map. The large range of values present in the radiance map (over four orders of magnitude of useful dynamic range) is shown by the values at the marked pixel locations.

Figure 6 shows sixteen photographs taken inside a church with a Canon 35mm SLR camera on Fuji 100 ASA color print film. A fish-eye 15mm lens set at f/8 was used, with exposure times ranging from 30 seconds to $\frac{1}{1000}$ of a second in 1-stop increments. The film was developed professionally and scanned in using a Kodak PhotoCD film scanner. The scanner was set so that it would not individually

⁹Note that here we are assuming that the spectral response functions for each channel of the two imaging processes is the same. Also, this technique does not model many significant qualities of an imaging system such as film grain, chromatic aberration, blooming, and the modulation transfer function.



Figure 3: (a) Eleven grayscale photographs of an indoor scene acquired with a Kodak DCS460 digital camera, with shutter speeds progressing in 1-stop increments from $\frac{1}{30}$ of a second to 30 seconds.

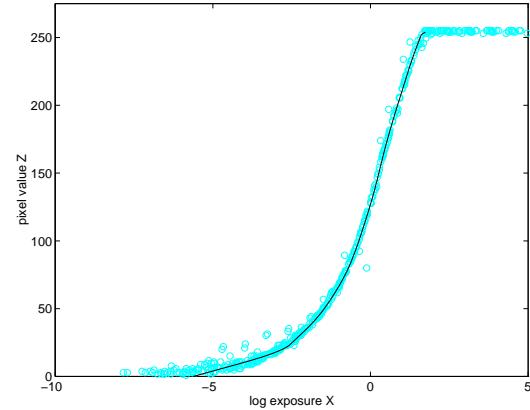


Figure 4: The response function of the DCS460 recovered by our algorithm, with the underlying $(E_i \Delta t_j, Z_{ij})$ data shown as light circles. The logarithm is base e .



Figure 5: The reconstructed high dynamic range radiance map, mapped into a grayscale image by taking the logarithm of the radiance values. The relative radiance values of the marked pixel locations, clockwise from lower left: 1.0, 46.2, 1907.1, 15116.0, and 18.0.



Figure 6: Sixteen photographs of a church taken at 1-stop increments from 30 sec to $\frac{1}{1000}$ sec. The sun is directly behind the rightmost stained glass window, making it especially bright. The blue borders seen in some of the image margins are induced by the image registration process.

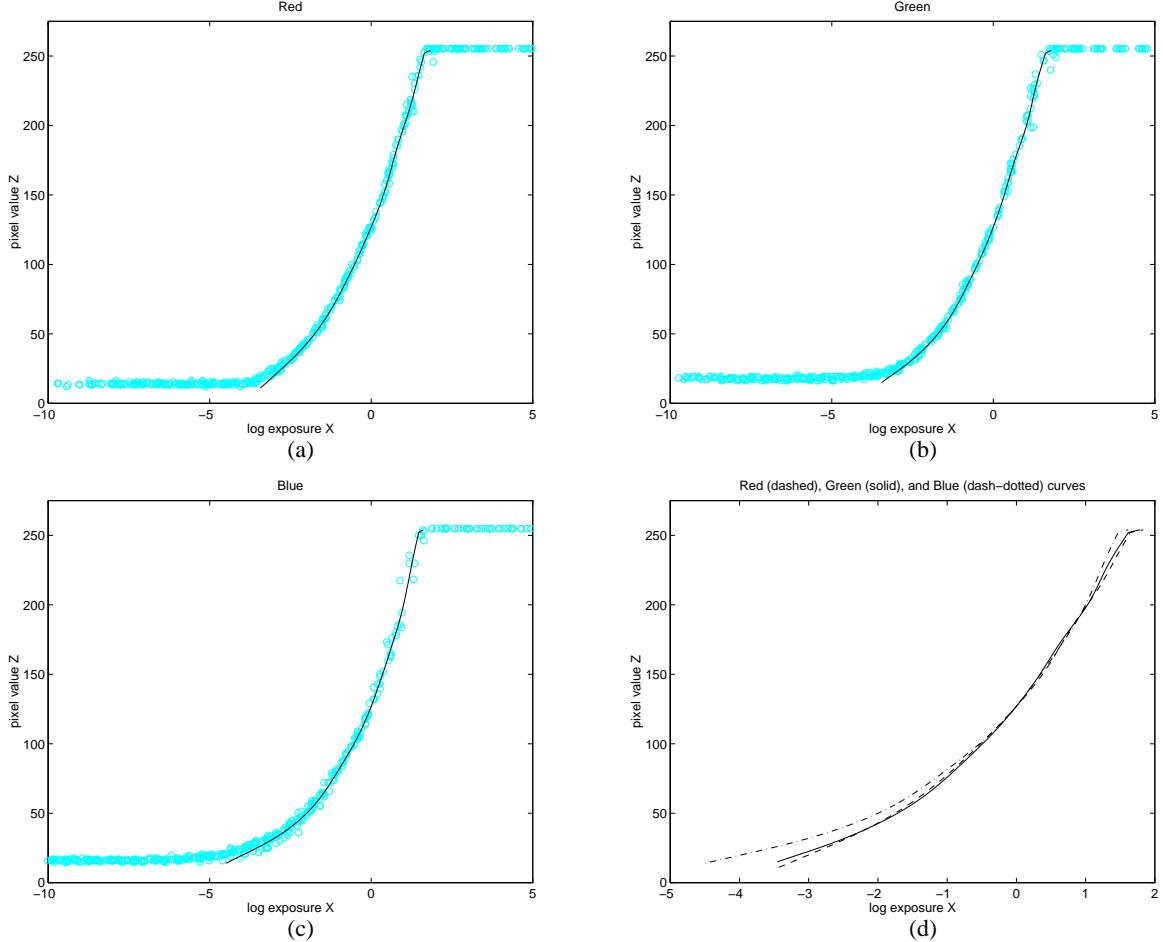


Figure 7: Recovered response curves for the imaging system used in the church photographs in Fig. 8. (a-c) Response functions for the red, green, and blue channels, plotted with the underlying $(E_i \Delta t_j, Z_{ij})$ data shown as light circles. (d) The response functions for red, green, and blue plotted on the same axes. Note that while the red and green curves are very consistent, the blue curve rises significantly above the others for low exposure values. This indicates that dark regions in the images exhibit a slight blue cast. Since this artifact is recovered by the response curves, it does not affect the relative radiance values.

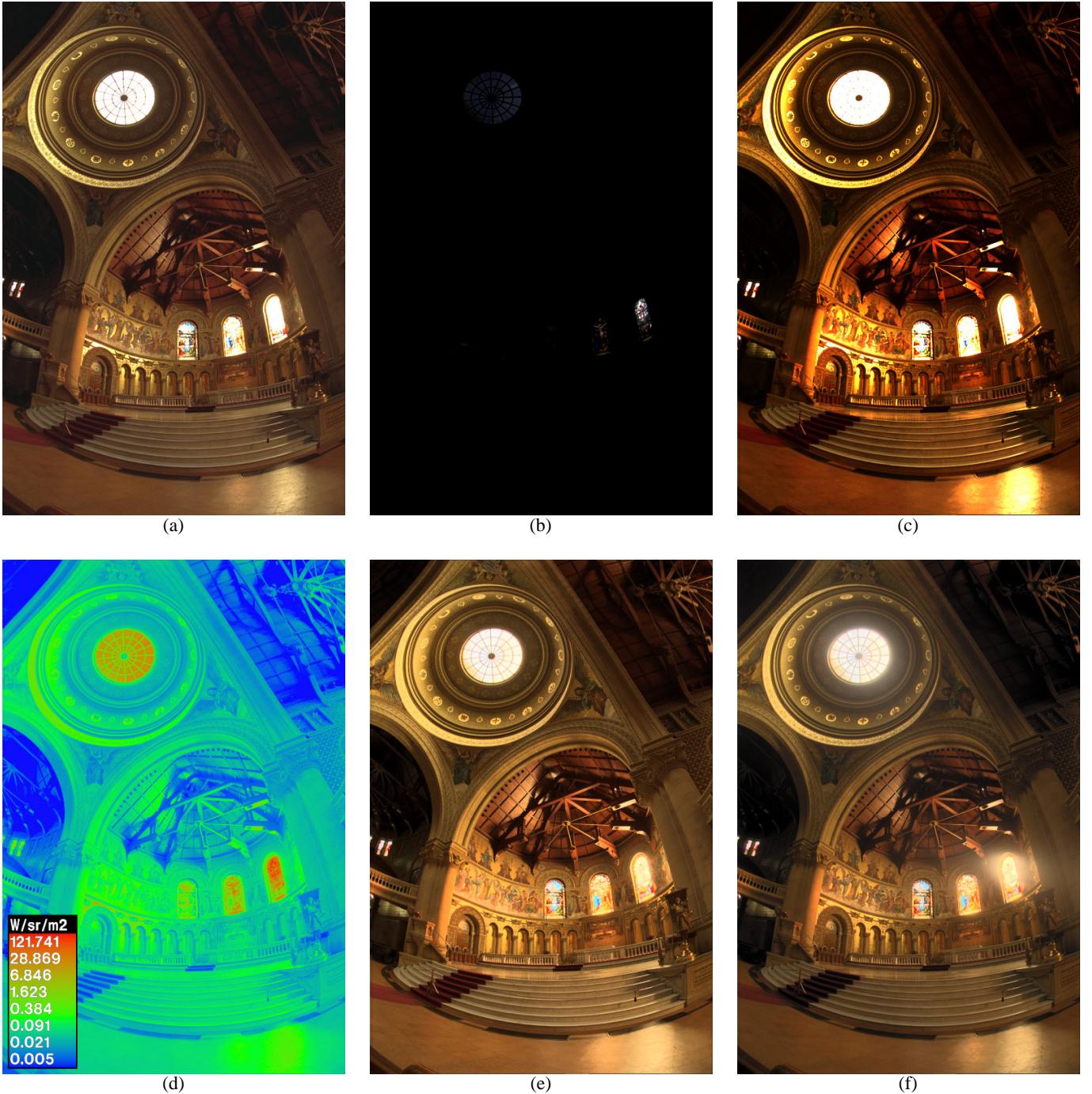


Figure 8: (a) An actual photograph, taken with conventional print film at two seconds and scanned to PhotoCD. (b) The high dynamic range radiance map, displayed by linearly mapping its entire dynamic range into the dynamic range of the display device. (c) The radiance map, displayed by linearly mapping the lower 0.1% of its dynamic range to the display device. (d) A false-color image showing relative radiance values for a grayscale version of the radiance map, indicating that the map contains over five orders of magnitude of useful dynamic range. (e) A rendering of the radiance map using adaptive histogram compression. (f) A rendering of the radiance map using histogram compression and also simulating various properties of the human visual system, such as glare, contrast sensitivity, and scotopic retinal response. Images (e) and (f) were generated by a method described in [23]. Images (d-f) courtesy of Gregory Ward Larson.

adjust the brightness and contrast of the images¹⁰ to guarantee that each image would be digitized using the same response function.

An unfortunate aspect of the PhotoCD process is that it does not scan precisely the same area of each negative relative to the extents of the image.¹¹ To counteract this effect, we geometrically registered the images to each other using a using normalized correlation (see [4]) to determine, with sub-pixel accuracy, corresponding pixels between pairs of images.

Fig. 7(a-c) shows the response functions for the red, green, and blue channels of the church sequence recovered from 28 pixel locations. Fig. 7(d) shows the recovered red, green, and blue response curves plotted on the same set of axes. From this plot, we can see that while the red and green curves are very consistent, the blue curve rises significantly above the others for low exposure values. This indicates that dark regions in the images exhibit a slight blue cast. Since this artifact is modeled by the response curves, it will not affect the relative radiance values.

Fig. 8 interprets the recovered high dynamic range radiance map in a variety of ways. Fig. 8(a) is one of the actual photographs, which lacks detail in its darker regions at the same time that many values within the two rightmost stained glass windows are saturated. Figs. 8(b,c) show the radiance map, linearly scaled to the display device using two different scaling factors. Although one scaling factor is one thousand times the other, there is useful detail in both images. Fig. 8(d) is a false-color image showing radiance values for a grayscale version of the radiance map; the highest listed radiance value is nearly 250,000 times that of the lowest. Figs. 8(e,f) show two renderings of the radiance map using a new tone reproduction algorithm [23]. Although the rightmost stained glass window has radiance values over a thousand times higher than the darker areas in the rafters, these renderings exhibit detail in both areas.

Figure 9 demonstrates two applications of the techniques presented in this paper: accurate signal processing and virtual photography. The task is to simulate the effects of motion blur caused by moving the camera during the exposure. Fig. 9(a) shows the results of convolving an actual, low-dynamic range photograph with a 37×1 pixel box filter to simulate horizontal motion blur. Fig. 9(b) shows the results of applying this same filter to the high dynamic range radiance map, and then sending this filtered radiance map back through the recovered film response functions using the same exposure time Δt as in the actual photograph. Because we are seeing this image through the actual image response curves, the two left images are tonally consistent with each other. However, there is a large difference between these two images near the bright spots. In the photograph, the bright radiance values have been clamped to the maximum pixel values by the response function. As a result, these clamped values blur with lower neighboring values and fail to saturate the image in the final result, giving a muddy appearance.

In Fig. 9(b), the extremely high pixel values were represented properly in the radiance map and thus remained at values above the level of the response function's saturation point within most of the blurred region. As a result, the resulting virtual photograph exhibits several crisply-defined saturated regions.

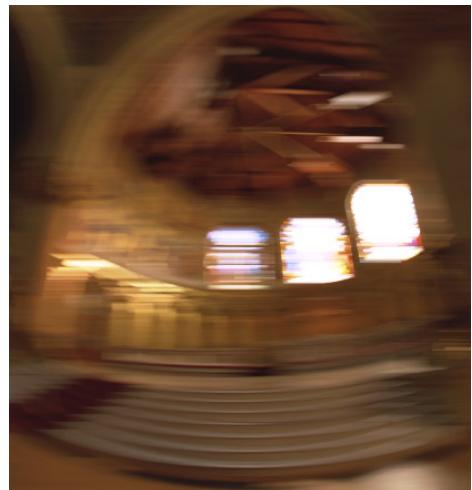
Fig. 9(c) is an actual photograph with real motion blur induced by spinning the camera on the tripod during the exposure, which is equal in duration to Fig. 9(a) and the exposure simulated in Fig. 9(b). Clearly, in the bright regions, the blurring effect is qualitatively similar to the synthetic blur in 9(b) but not 9(a). The precise shape of the real motion blur is curved and was not modeled for this demonstration.

¹⁰This feature of the PhotoCD process is called “Scene Balance Adjustment”, or SBA.

¹¹This is far less of a problem for cinematic applications, in which the film sprocket holes are used to expose and scan precisely the same area of each frame.



(a) Synthetically blurred digital image



(b) Synthetically blurred radiance map



(c) Actual blurred photograph

Figure 9: (a) Synthetic motion blur applied to one of the original digitized photographs. The bright values in the windows are clamped before the processing, producing mostly unsaturated values in the blurred regions. (b) Synthetic motion blur applied to a recovered high-dynamic range radiance map, then virtually rephotographed through the recovered film response curves. The radiance values are clamped to the display device after the processing, allowing pixels to remain saturated in the window regions. (c) Real motion blur created by rotating the camera on the tripod during the exposure, which is much more consistent with (b) than (a).

4 Conclusion

We have presented a simple, practical, robust and accurate method of recovering high dynamic range radiance maps from ordinary photographs. Our method uses the constraint of sensor reciprocity to derive the response function and relative radiance values directly from a set of images taken with different exposures. This work has a wide variety of applications in the areas of image-based modeling and rendering, image processing, and image compositing, a few of which we have demonstrated. It is our hope that this work will be able to help both researchers and practitioners of computer graphics make much more effective use of digitized photographs.

Acknowledgments

The authors wish to thank Tim Hawkins, Carlo Séquin, David Forsyth, Steve Chenney, Chris Healey, and our reviewers for their valuable help in revising this paper. This research was supported by a Multidisciplinary University Research Initiative on three dimensional direct visualization from ONR and BMDO, grant FDN00014-96-1-1200.

References

- [1] ADAMS, A. *Basic Photo*, 1st ed. Morgan & Morgan, Hastings-on-Hudson, New York, 1970.
- [2] CHEN, E. QuickTime VR - an image-based approach to virtual environment navigation. In *SIGGRAPH '95* (1995).
- [3] DEBEVEC, P. E., TAYLOR, C. J., AND MALIK, J. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *SIGGRAPH '96* (August 1996), pp. 11–20.
- [4] FAUGERAS, O. *Three-Dimensional Computer Vision*. MIT Press, 1993.
- [5] FERWERDA, J. A., PATTANAIK, S. N., SHIRLEY, P., AND GREENBERG, D. P. A model of visual adaptation for realistic image synthesis. In *SIGGRAPH '96* (1996), pp. 249–258.
- [6] GORTLER, S. J., GRZESZCZUK, R., SZELISKI, R., AND COHEN, M. F. The Lumigraph. In *SIGGRAPH '96* (1996), pp. 43–54.
- [7] HORN, B. K. P. *Robot Vision*. MIT Press, Cambridge, Mass., 1986, ch. 10, pp. 206–208.
- [8] JAMES, T., Ed. *The Theory of the Photographic Process*. Macmillan, New York, 1977.
- [9] KAUFMAN, J. E., Ed. *IES Lighting Handbook; the standard lighting guide*, 7th ed. Illuminating Engineering Society, New York, 1987, p. 24.
- [10] KOLB, C., MITCHELL, D., AND HANRAHAN, P. A realistic camera model for computer graphics. In *SIGGRAPH '95* (1995).
- [11] LAVEAU, S., AND FAUGERAS, O. 3-D scene representation as a collection of images. In *Proceedings of 12th International Conference on Pattern Recognition* (1994), vol. 1, pp. 689–691.
- [12] LEVOY, M., AND HANRAHAN, P. Light field rendering. In *SIGGRAPH '96* (1996), pp. 31–42.
- [13] MADDEN, B. C. Extended intensity range imaging. Tech. rep., GRASP Laboratory, University of Pennsylvania, 1993.
- [14] MANN, S., AND PICARD, R. W. Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proceedings of IS&T 46th annual conference* (May 1995), pp. 422–428.
- [15] McMILLAN, L., AND BISHOP, G. Plenoptic Modeling: An image-based rendering system. In *SIGGRAPH '95* (1995).
- [16] SCHLICK, C. Quantization techniques for visualization of high dynamic range pictures. In *Fifth Eurographics Workshop on Rendering* (Darmstadt, Germany) (June 1994), pp. 7–18.
- [17] SZELISKI, R. Image mosaicing for tele-reality applications. In *IEEE Computer Graphics and Applications* (1996).
- [18] TANI, T. *Photographic sensitivity : theory and mechanisms*. Oxford University Press, New York, 1995.
- [19] THEUWISSEN, A. J. P. *Solid-state imaging with charge-coupled devices*. Kluwer Academic Publishers, Dordrecht; Boston, 1995.
- [20] TUMBLIN, J., AND RUSHMEIER, H. Tone reproduction for realistic images. *IEEE Computer Graphics and Applications* 13, 6 (1993), 42–48.
- [21] WARD, G. J. Measuring and modeling anisotropic reflection. In *SIGGRAPH '92* (July 1992), pp. 265–272.
- [22] WARD, G. J. The radiance lighting simulation and rendering system. In *SIGGRAPH '94* (July 1994), pp. 459–472.
- [23] WARD, G. J., RUSHMEIER, H., AND PIATKO, C. A visibility matching tone reproduction operator for high dynamic range scenes. Tech. Rep. LBNL-39882, Lawrence Berkeley National Laboratory, March 1997.

A Matlab Code

Here is the MATLAB code used to solve the linear system that minimizes the objective function \mathcal{O} in Equation 3. Given a set of observed pixel values in a set of images with known exposures, this routine reconstructs the imaging response curve and the radiance values for the given pixels. The weighting function $w(z)$ is found in Equation 4.

```
% gsolve.m - Solve for imaging system response function
%
% Given a set of pixel values observed for several pixels in several
% images with different exposure times, this function returns the
% imaging system's response function g as well as the log film irradiance
% values for the observed pixels.
%
% Assumes:
%
% Zmin = 0
% Zmax = 255
%
% Arguments:
%
% Z(i,j) is the pixel values of pixel location number i in image j
% B(j) is the log delta t, or log shutter speed, for image j
% l is lambda, the constant that determines the amount of smoothness
% w(z) is the weighting function value for pixel value z
%
% Returns:
%
% g(z) is the log exposure corresponding to pixel value z
% IE(i) is the log film irradiance at pixel location i
%
function [g,IE]=gsolve(Z,B,l,w)
n = 256;
A = zeros(size(Z,1)*size(Z,2)+n+1,n+size(Z,1));
b = zeros(size(A,1),1);
%
% include the data-fitting equations
k = 1;
for j=1:size(Z,1)
    for i=1:size(Z,2)
        wij = w(Z(i,j)+1);
        A(k,Z(i,j)+1) = wij; A(k,n+i) = -wij; b(k,1) = wij * B(i,j);
        k=k+1;
    end
end
%
% Fix the curve by setting its middle value to 0
A(k,129) = 1;
k=k+1;
%
% include the smoothness equations
for i=1:n-2
    A(k,i)=l*w(i+1); A(k,i+1)=-2*l*w(i+1); A(k,i+2)=l*w(i+1);
    k=k+1;
end
%
% solve the system using SVD
x = A\b;
g = x(1:n);
IE = x(n+1:size(x,1));
```

Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-based Graphics with Global Illumination and High Dynamic Range Photography

Paul Debevec

University of California at Berkeley¹

ABSTRACT

We present a method that uses measured scene radiance and global illumination in order to add new objects to light-based models with correct lighting. The method uses a high dynamic range image-based model of the scene, rather than synthetic light sources, to illuminate the new objects. To compute the illumination, the scene is considered as three components: the distant scene, the local scene, and the synthetic objects. The distant scene is assumed to be photometrically unaffected by the objects, obviating the need for reflectance model information. The local scene is endowed with estimated reflectance model information so that it can catch shadows and receive reflected light from the new objects. Renderings are created with a standard global illumination method by simulating the interaction of light amongst the three components. A differential rendering technique allows for good results to be obtained when only an estimate of the local scene reflectance properties is known.

We apply the general method to the problem of rendering synthetic objects into real scenes. The light-based model is constructed from an approximate geometric model of the scene and by using a light probe to measure the incident illumination at the location of the synthetic objects. The global illumination solution is then composited into a photograph of the scene using the differential rendering technique. We conclude by discussing the relevance of the technique to recovering surface reflectance properties in uncontrolled lighting situations. Applications of the method include visual effects, interior design, and architectural visualization.

CR Descriptors: I.2.10 [Artificial Intelligence]: Vision and Scene Understanding - *Intensity, color, photometry and thresholding*; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism - *Color, shading, shadowing, and texture*; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism - *Radiosity*; I.4.1 [Image Processing]: Digitization - *Scanning*; I.4.8 [Image Processing]: Scene Analysis - *Photometry, Sensor Fusion*.

¹Computer Science Division, University of California at Berkeley, Berkeley, CA 94720-1776. Email: debevec@cs.berkeley.edu. More information and additional results may be found at: <http://www.cs.berkeley.edu/~debevec/Research>

1 Introduction

Rendering synthetic objects into real-world scenes is an important application of computer graphics, particularly in architectural and visual effects domains. Oftentimes, a piece of furniture, a prop, or a digital creature or actor needs to be rendered seamlessly into a real scene. This difficult task requires that the objects be lit consistently with the surfaces in their vicinity, and that the interplay of light between the objects and their surroundings be properly simulated. Specifically, the objects should cast shadows, appear in reflections, and refract, focus, and emit light just as real objects would.

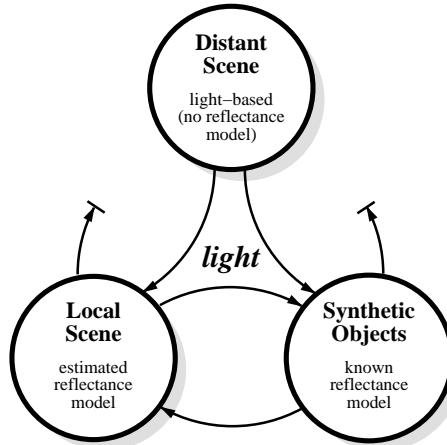


Figure 1: **The General Method** In our method for adding synthetic objects into light-based scenes, the scene is partitioned into three components: the distant scene, the local scene, and the synthetic objects. Global illumination is used to simulate the interplay of light amongst all three components, except that light reflected back at the distant scene is ignored. As a result, BRDF information for the distant scene is unnecessary. Estimates of the geometry and material properties of the local scene are used to simulate the interaction of light between it and the synthetic objects.

Currently available techniques for realistically rendering synthetic objects into scenes are labor intensive and not always successful. A common technique is to manually survey the positions of the light sources, and to instantiate a virtual light of equal color and intensity for each real light to illuminate the synthetic objects. Another technique is to photograph a reference object (such as a gray sphere) in the scene where the new object is to be rendered, and use its appearance as a qualitative guide in manually configuring the lighting environment. Lastly, the technique of reflection mapping is useful for mirror-like reflections. These methods typically require considerable hand-refinement and none of them easily simulates the effects of indirect illumination from the environment.

Accurately simulating the effects of both direct and indirect lighting has been the subject of research in global illumination. With a global illumination algorithm, if the entire scene were modeled with its full geometric and reflectance (BRDF) characteristics, one could correctly render a synthetic object into the scene simply by adding it to the model and recomputing the global illumination solution. Unfortunately, obtaining a full geometric and reflectance model of a large environment is extremely difficult. Furthermore, global illumination solutions for large complex environments are extremely computationally intensive.

Moreover, it seems that having a full reflectance model of the large-scale scene should be unnecessary: under most circumstances, a new object will have no significant effect on the appearance of most of the distant scene. Thus, for such distant areas, knowing just its radiance (under the desired lighting conditions) should suffice.

Recently, [9] introduced a high dynamic range photographic technique that allows accurate measurements of scene radiance to be derived from a set of differently exposed photographs. This technique allows both low levels of indirect radiance from surfaces and high levels of direct radiance from light sources to be accurately recorded. When combined with image-based modeling techniques (e.g. [22, 24, 4, 10, 23, 17, 29]), and possibly active techniques for measuring geometry (e.g. [35, 30, 7, 27]) these derived radiance maps can be used to construct spatial representations of scene radiance.

We will use the term **light-based model** to refer to a representation of a scene that consists of radiance information, possibly with specific reference to light leaving surfaces, but not necessarily containing material property (BRDF) information. A light-based model can be used to evaluate the 5D plenoptic function [1] $P(\theta, \phi, V_x, V_y, V_z)$ for a given virtual or real subset of space¹. A material-based model is converted to a light-based model by computing an illumination solution for it. A light-based model is differentiated from an image-based model in that its light values are actual measures of radiance², whereas image-based models may contain pixel values already transformed and truncated by the response function of an image acquisition or synthesis process.

In this paper, we present a general method for using accurate measurements of scene radiance in conjunction with global illumination to realistically add new objects to light-based models. The synthetic objects may have arbitrary material properties and can be rendered with appropriate illumination in arbitrary lighting environments. Furthermore, the objects can correctly interact with the environment around them: they cast the appropriate shadows, they are properly reflected, they can reflect and focus light, and they exhibit appropriate diffuse interreflection. The method can be carried out with commonly available equipment and software.

In this method (see Fig. 1), the scene is partitioned into three components. The first is the distant scene, which is the visible part of the environment too remote to be perceptibly affected by the synthetic object. The second is the local scene, which is the part of the environment which will be significantly affected by the presence of the objects. The third component is the synthetic objects. Our approach uses global illumination to correctly simulate the interaction of light amongst these three elements, with the exception that light radiated toward the distant environment will not be considered in the calculation. As a result, the BRDF of the distant environment need not be known — the technique uses BRDF information only for the local scene and the synthetic objects. We discuss the challenges in estimating the BRDF of the local scene, and methods for obtaining usable approximations. We also present a differential rendering

¹Time and wavelength dependence can be included to represent the general 7D plenoptic function as appropriate.

²In practice, the measures of radiance are with respect to a discrete set of spectral distributions such as the standard tristimulus model.

technique that produces perceptually accurate results even when the estimated BRDF is somewhat inaccurate.

We demonstrate the general method for the specific case of rendering synthetic objects into particular views of a scene (such as background plates) rather than into a general image-based model. In this method, a light probe is used to acquire a high dynamic range panoramic radiance map near the location where the object will be rendered. A simple example of a light probe is a camera aimed at a mirrored sphere, a configuration commonly used for acquiring environment maps. An approximate geometric model of the scene is created (via surveying, photogrammetry, or 3D scanning) and mapped with radiance values measured with the light probe. The distant scene, local scene, and synthetic objects are rendered with global illumination from the same point of view as the background plate, and the results are composited into the background plate with a differential rendering technique.

1.1 Overview

The rest of this paper is organized as follows. In the next section we discuss work related to this paper. Section 3 introduces the basic technique of using acquired maps of scene radiance to illuminate synthetic objects. Section 4 presents the general method we will use to render synthetic objects into real scenes. Section 5 describes a practical technique based on this method using a *light probe* to measure incident illumination. Section 6 presents a differential rendering technique for rendering the local environment with only an approximate description of its reflectance. Section 7 presents a simple method to approximately recover the diffuse reflectance characteristics of the local environment. Section 8 presents results obtained with the technique. Section 9 discusses future directions for this work, and we conclude in Section 10.

2 Background and Related Work

The practice of adding new objects to photographs dates to the early days of photography in the simple form of pasting a cut-out from one picture onto another. While the technique conveys the idea of the new object being in the scene, it usually fails to produce an image that as a whole is a believable photograph. Attaining such realism requires a number of aspects of the two images to match. First, the camera projections should be consistent, otherwise the object may seem too foreshortened or skewed relative to the rest of the picture. Second, the patterns of film grain and film response should match. Third, the lighting on the object needs to be consistent with other objects in the environment. Lastly, the object needs to cast realistic shadows and reflections on the scene. Skilled artists found that by giving these considerations due attention, synthetic objects could be painted into still photographs convincingly.

In optical film compositing, the use of object mattes to prevent particular sections of film from being exposed made the same sort of cut-and-paste compositing possible for moving images. However, the increased demands of realism imposed by the dynamic nature of film made matching camera positions and lighting even more critical. As a result, care was taken to light the objects appropriately for the scene into which they were to be composited. This would still not account for the objects casting shadows onto the scene, so often these were painted in by an artist frame by frame [13, 2, 28]. Digital film scanning and compositing [26] helped make this process far more efficient.

Work in global illumination [16, 19] has recently produced algorithms (e.g. [31]) and software (e.g. [33]) to realistically simulate lighting in synthetic scenes, including indirect lighting with both specular and diffuse reflections. We leverage this work in order to create realistic renderings.

Some work has been done on the specific problem of compositing objects into photography. [25] presented a procedure for ren-

dering architecture into background photographs using knowledge of the sun position and measurements or approximations of the local ambient light. For diffuse buildings in diffuse scenes, the technique is effective. The technique of *reflection mapping* (also called *environment mapping*) [3, 18] produces realistic results for mirror-like objects. In reflection mapping, a panoramic image is rendered or photographed from the location of the object. Then, the surface normals of the object are used to index into the panoramic image by reflecting rays from the desired viewpoint. As a result, the shiny object appears to properly reflect the desired environment³. However, the technique is limited to mirror-like reflection and does not account for objects casting light or shadows on the environment.

A common visual effects technique for having synthetic objects cast shadows on an existing environment is to create an approximate geometric model of the environment local to the object, and then compute the shadows from the various light sources. The shadows can then be subtracted from the background image. In the hands of professional artists this technique can produce excellent results, but it requires knowing the position, size, shape, color, and intensity of each of the scene's light sources. Furthermore, it does not account for diffuse reflection from the scene, and light reflected by the objects onto the scene must be handled specially.

To properly model the interaction of light between the objects and the local scene, we pose the compositing problem as a global illumination computation as in [14] and [12]. As in this work, we apply the effect of the synthetic objects in the lighting solution as a differential update to the original appearance of the scene. In the previous work an approximate model of the entire scene and its original light sources is constructed; the positions and sizes of the light sources are measured manually. Rough methods are used to estimate diffuse-only reflectance characteristics of the scene, which are then used to estimate the intensities of the light sources. [12] additionally presents a method for performing fast updates of the illumination solution in the case of moving objects. As in the previous work, we leverage the basic result from incremental radiosity [6, 5] that making a small change to a scene does not require recomputing the entire solution.

3 Illuminating synthetic objects with real light

In this section we propose that computer-generated objects be lit by actual recordings of light from the scene, using global illumination. Performing the lighting in this manner provides a unified and physically accurate alternative to manually attempting to replicate incident illumination conditions.

Accurately recording light in a scene is difficult because of the high dynamic range that scenes typically exhibit; this wide range of brightness is the result of light sources being relatively concentrated. As a result, the intensity of a source is often two to six orders of magnitude larger than the intensity of the non-emissive parts of an environment. However, it is necessary to accurately record both the large areas of indirect light from the environment and the concentrated areas of direct light from the sources since both are significant parts of the illumination solution.

Using the technique introduced in [9], we can acquire correct measures of scene radiance using conventional imaging equipment. The images, called *radiance maps*, are derived from a series of images with different sensor integration times and a technique for computing and accounting for the imaging system response function f . We can use these measures to illuminate synthetic objects exhibiting arbitrary material properties.

Fig. 2 shows a high-dynamic range lighting environment with electric, natural, and indirect lighting. This environment was

recorded by taking a full dynamic range photograph of a mirrored ball on a table (see Section 5). A digital camera was used to acquire a series images in one-stop exposure increments from $\frac{1}{4}$ to $\frac{1}{1000}$ second. The images were fused using the technique in [9].

The environment is displayed at three exposure levels (-0, -3.5, and -7.0 stops) to show its full dynamic range. Recovered RGB radiance values for several points in the scene and on the two major light sources are indicated; the color difference between the tungsten lamp and the sky is evident. A single low-dynamic range photograph would be unable to record the correct colors and intensities over the entire scene.

Fig. 3(a-e) shows the results of using this panoramic radiance map to synthetically light a variety of materials using the RADIANCE global illumination algorithm [33]. The materials are: (a) perfectly reflective, (b) rough gold, (c) perfectly diffuse gray material, (d) shiny green plastic, and (e) dull orange plastic. Since we are computing a full illumination solution, the objects exhibit self-reflection and shadows from the light sources as appropriate. Note that in (c) the protrusions produce two noticeable shadows of slightly different colors, one corresponding to the ceiling light and a softer shadow corresponding to the window.

The shiny plastic object in (d) has a 4 percent specular component with a Gaussian roughness of 0.04 [32]. Since the object's surface both blurs and attenuates the light with its rough specular component, the reflections fall within the dynamic range of our display device and the different colors of the light sources can be seen. In (e) the rough plastic diffuses the incident light over a much larger area.

To illustrate the importance of using high dynamic range radiance maps, the same renderings were produced using just one of the original photographs as the lighting environment. In this single image, similar in appearance to Fig. 2(a), the brightest regions had been truncated to approximately 2 percent of their true values. The rendering of the mirrored surface (f) appears similar to (a) since it is displayed in low-dynamic range printed form. Significant errors are noticeable in (g-j) since these materials blur the incident light. In (g), the blurring of the rough material darkens the light sources, whereas in (b) they remain saturated. Renderings (h-j) are very dark due to the missed light; thus we have brightened by a factor of eight on the right in order to make qualitative comparisons to (c-e) possible. In each it can be seen that the low-dynamic range image of the lighting environment fails to capture the information necessary to simulate correct color balance, shadows, and highlights.

Fig. 4 shows a collection of objects with different material properties illuminated by two different environments. A wide variety of light interaction between the objects and the environment can be seen. The (synthetic) mirrored ball reflects both the synthetic objects as well as the environment. The floating diffuse ball shows a subtle color shift along its right edge as it shadows itself from the windows and is lit primarily by the incandescent lamp in Fig. 4(a). The reflection of the environment in the black ball (which has a specular intensity of 0.04) shows the colors of the light sources, which are too bright to be seen in the mirrored ball. A variety of shadows, reflections, and focused light can be observed on the resting surface.

The next section describes how the technique of using radiance maps to illuminate synthetic objects can be extended to compute the proper photometric interaction of the objects with the scene. It also describes how high dynamic range photography and image-based modeling combine in a natural manner to allow the simulation of arbitrary (non-infinite) lighting environments.

4 The General Method

This section explains our method for adding new objects to light-based scene representations. As in Fig. 1, we partition our scene into three parts: the distant scene, the local scene, and the synthetic

³Using the surface normal indexing method, the object will not reflect itself. Correct self-reflection can be obtained through ray tracing.



Figure 2: An omnidirectional radiance map This full dynamic range lighting environment was acquired by photographing a mirrored ball balanced on the cap of a pen sitting on a table. The environment contains natural, electric, and indirect light. The three views of this image adjusted to (a) +0 stops, (b) -3.5 stops, and (c) -7.0 stops show that the full dynamic range of the scene has been captured without saturation. As a result, the image usefully records the direction, color, and intensity of all forms of incident light.

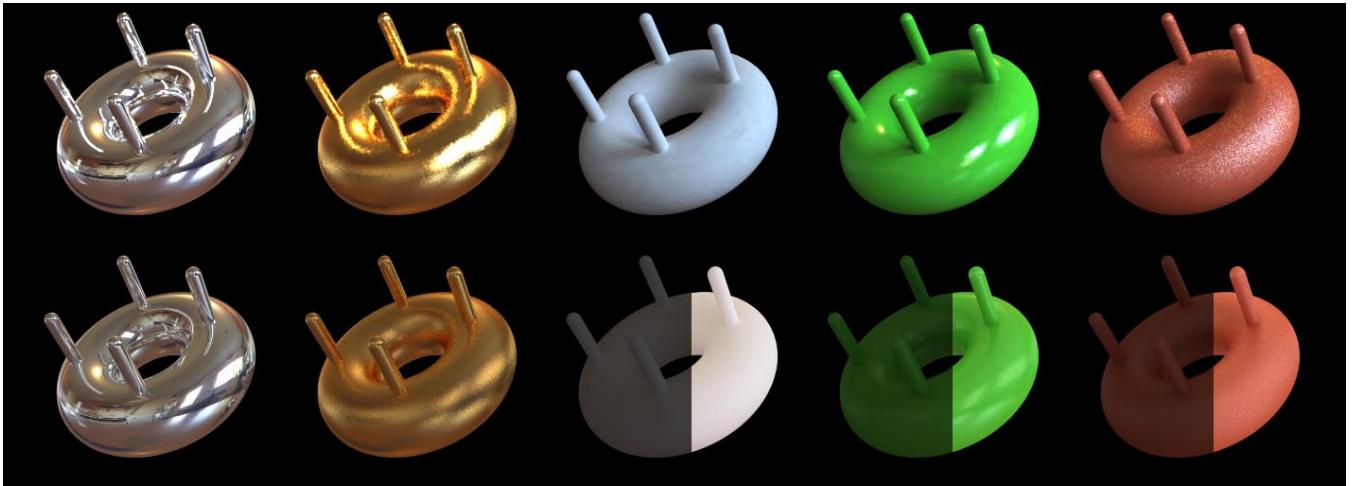


Figure 3: Illuminating synthetic objects with real light (Top row: a,b,c,d,e) With full dynamic range measurements of scene radiance from Fig. 2. (Bottom row: f,g,h,i,j) With low dynamic range information from a single photograph of the ball. The right sides of images (h,i,j) have been brightened by a factor of six to allow qualitative comparison to (c,d,e). The high dynamic range measurements of scene radiance are necessary to produce proper lighting on the objects.

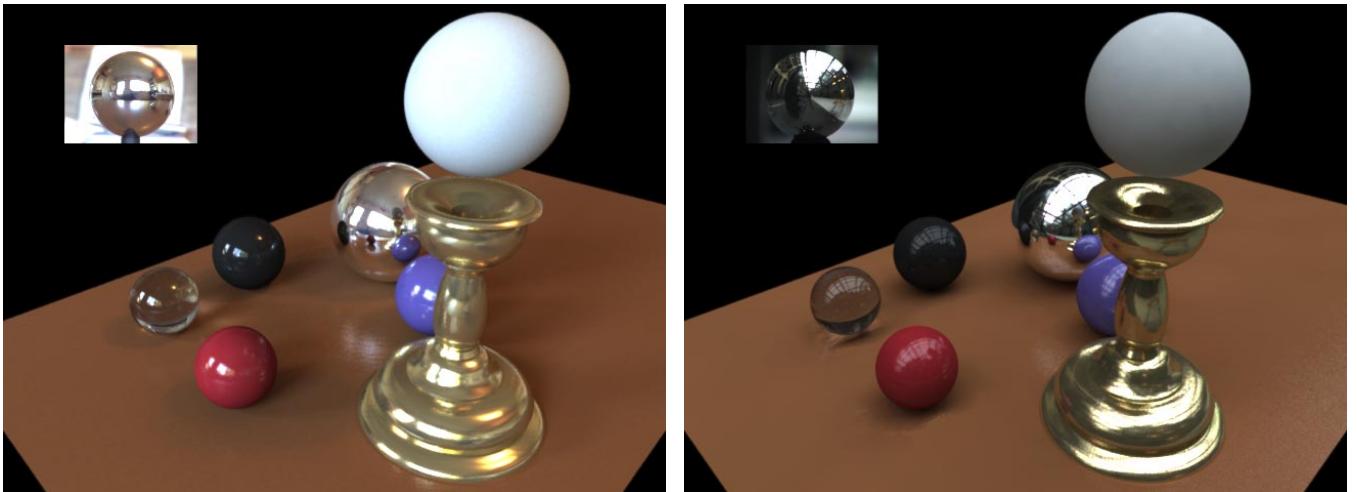


Figure 4: Synthetic objects lit by two different environments (a) A collection of objects is illuminated by the radiance information in 2. The objects exhibit appropriate interreflection. (b) The same objects are illuminated by different radiance information obtained in an outdoor urban environment on an overcast day. The radiance map used for the illumination is shown in the upper left of each image. Candle holder model courtesy of Gregory Ward Larson.

objects. We describe the geometric and photometric requirements for each of these components.

1. A light-based model of the distant scene

The distant scene is constructed as a light-based model. The synthetic objects will receive light from this model, so it is necessary that the model store true measures of radiance rather than low dynamic range pixel values from conventional images. The light-based model can take on any form, using very little explicit geometry [23, 17], some geometry [24], moderate geometry [10], or be a full 3D scan of an environment with view-dependent texture-mapped [11] radiance. What is important is for the model to provide accurate measures of incident illumination in the vicinity of the objects, as well as from the desired viewpoint. In the next section we will present a convenient procedure for constructing a minimal model that meets these requirements.

In the global illumination computation, the distant scene radiates light toward the local scene and the synthetic objects, but ignores light reflected back to it. We assume that no area of the distant scene will be significantly affected by light reflecting from the synthetic objects; if that were the case, the area should instead belong to the local scene, which contains the BRDF information necessary to interact with light. In the RADIANCE [33] system, this exclusively emissive behavior can be specified with the "glow" material property.

2. An approximate material-based model of the local scene

The local scene consists of the surfaces that will photometrically interact with the synthetic objects. It is this geometry onto which the objects will cast shadows and reflect light. Since the local scene needs to fully participate in the illumination solution, both its geometry and reflectance characteristics should be known, at least approximately. If the geometry of the local scene is not readily available with sufficient accuracy from the light-based model of the distant scene, there are various techniques available for determining its geometry through active or passive methods. In the common case where the local scene is a flat surface that supports the synthetic objects, its geometry is determined easily from the camera pose. Methods for estimating the BRDF of the local scene are discussed in Section 7.

Usually, the local scene will be the part of the scene that is geometrically close to the synthetic objects. When the local scene is mostly diffuse, the rendering equation shows that the visible effect of the objects on the local scene decreases as the inverse square of the distance between the two. Nonetheless, there is a variety of circumstances in which synthetic objects can significantly affect areas of the scene not in the immediate vicinity. Some common circumstances are:

- If there are concentrated light sources illuminating the object, then the object can cast a significant shadow on a distant surface collinear with it and the light source.
- If there are concentrated light sources and the object is flat and specular, it can focus a significant amount of light onto a distant part of the scene.
- If a part of the distant scene is flat and specular (e.g. a mirror on a wall), its appearance can be significantly affected by a synthetic object.
- If the synthetic object emits light (e.g. a synthetic laser), it can affect the appearance of the distant scene significantly.

These situations should be considered in choosing which parts of the scene should be considered local and which parts distant. Any part of the scene that will be significantly affected in

its appearance from the desired viewpoint should be included as part of the local scene.

Since the local scene is a full BRDF model, it can be added to the global illumination problem as would any other object. The local scene may consist of any number of surfaces and objects with different material properties. For example, the local scene could consist of a patch of floor beneath the synthetic object to catch shadows as well as a mirror surface hanging on the opposite wall to catch a reflection. The local scene replaces the corresponding part of the light-based model of the distant scene.

Since it can be difficult to determine the precise BRDF characteristics of the local scene, it is often desirable to have only the *change* in the local scene's appearance be computed with the BRDF estimate; its appearance due to illumination from the distant scene is taken from the original light-based model. This differential rendering method is presented in Section 6.

3. Complete material-based models of the objects

The synthetic objects themselves may consist of any variety of shapes and materials supported by the global illumination software, including plastics, metals, emitters, and dielectrics such as glass and water. They should be placed in their desired geometric correspondence to the local scene.

Once the distant scene, local scene, and synthetic objects are properly modeled and positioned, the global illumination software can be used in the normal fashion to produce renderings from the desired viewpoints.

5 Compositing using a light probe

This section presents a particular technique for constructing a light-based model of a real scene suitable for adding synthetic objects at a particular location. This technique is useful for compositing objects into actual photography of a scene.

In Section 4, we mentioned that the light-based model of the distant scene needs to appear correctly in the vicinity of the synthetic objects as well as from the desired viewpoints. This latter requirement can be satisfied if it is possible to directly acquire radiance maps of the scene from the desired viewpoints. The former requirement, that the appear photometrically correct in all directions in the vicinity of the synthetic objects, arises because this information comprises the incident light which will illuminate the objects.

To obtain this part of the light-based model, we acquire a full dynamic range omnidirectional radiance map near the location of the synthetic object or objects. One technique for acquiring this radiance map is to photograph a spherical first-surface mirror, such as a polished steel ball, placed at or near the desired location of the synthetic object⁴. This procedure is illustrated in Fig. 7(a). An actual radiance map obtained using this method is shown in Fig. 2.

The radiance measurements observed in the ball are mapped onto the geometry of the distant scene. In many circumstances this model can be very simple. In particular, if the objects are small and resting on a flat surface, one can model the scene as a horizontal plane for the resting surface and a large dome for the rest of the environment. Fig. 7(c) illustrates the ball image being mapped onto a table surface and the walls and ceiling of a finite room; 5 shows the resulting light-based model.

5.1 Mapping from the probe to the scene model

To precisely determine the mapping between coordinates on the ball and rays in the world, one needs to record the position of the ball

⁴Parabolic mirrors combined with telecentric lenses [34] can be used to obtain hemispherical fields of view with a consistent principal point, if so desired.

relative to the camera, the size of the ball, and the camera parameters such as its location in the scene and focal length. With this information, it is straightforward to trace rays from the camera center through the pixels of the image, and reflect rays off the ball into the environment. Often a good approximation results from assuming the ball is small relative to the environment and that the camera's view is orthographic.

The data acquired from a single ball image will exhibit a number of artifacts. First, the camera (and possibly the photographer) will be visible. The ball, in observing the scene, interacts with it: the ball (and its support) can appear in reflections, cast shadows, and can reflect light back onto surfaces. Lastly, the ball will not reflect the scene directly behind it, and will poorly sample the area nearby. If care is taken in positioning the ball and camera, these effects can be minimized and will have a negligible effect on the final renderings. If the artifacts are significant, the images can be fixed manually in image editing program or by selectively combining images of the ball taken from different directions; Fig. 6 shows a relatively artifact-free environment constructed using the latter method. We have found that combining two images of the ball taken ninety degrees apart from each other allows us to eliminate the camera's appearance and to avoid poor sampling.

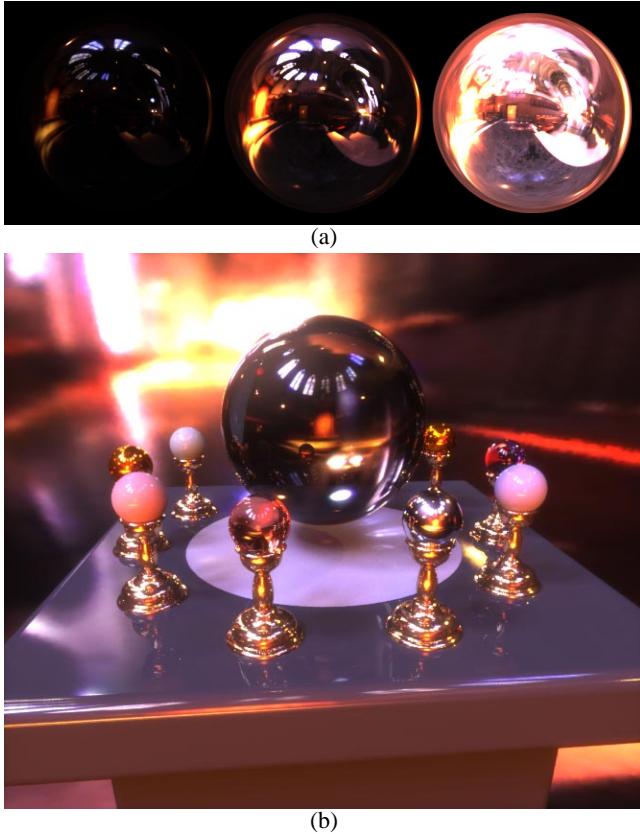


Figure 6: Rendering with a Combined Probe Image The full dynamic range environment map shown at the top was assembled from two light probe images taken ninety degrees apart from each other. As a result, the only visible artifact is a small amount of the probe support visible on the floor. The map is shown at -4.5, 0, and +4.5 stops. The bottom rendering was produced using this lighting information, and exhibits diffuse and specular reflections, shadows from different sources of light, reflections, and caustics.

5.2 Creating renderings

To render the objects into the scene, a synthetic local scene model is created as described in Section 4. Images of the scene from the desired viewpoint(s) are taken (Fig. 7(a)), and their position relative to the scene is recorded through pose-instrumented cameras or (as in our work) photogrammetry. The location of the ball in the scene is also recorded at this time. The global illumination software is then run to render the objects, local scene, and distant scene from the desired viewpoint (Fig. 7(d)).

The objects and local scene are then composited onto the background image. To perform this compositing, a mask is created by rendering the objects and local scene in white and the distant scene in black. If objects in the distant scene (which may appear in front of the objects or local scene from certain viewpoints) are geometrically modeled, they will properly obscure the local scene and the objects as necessary. This compositing can be considered as a subset of the general method (Section 4) wherein the light-based model of the distant scene acts as follows: if (V_x, V_y, V_z) corresponds to an actual view of the scene, return the radiance value looking in direction (θ, ϕ) . Otherwise, return the radiance value obtained by casting the ray $(\theta, \phi, V_x, V_y, V_z)$ onto the radiance-mapped distant scene model.

In the next section we describe a more robust method of compositing the local scene into the background image.

6 Improving quality with differential rendering

The method we have presented so far requires that the local scene be modeled accurately in both its geometry and its spatially varying material properties. If the model is inaccurate, the appearance of the local scene will not be consistent with the appearance of adjacent distant scene. Such a border is readily apparent in Fig. 8(c), since the local scene was modeled with a homogeneous BRDF when in reality it exhibits a patterned albedo (see [21]). In this section we describe a method for greatly reducing such effects.

Suppose that we compute a global illumination solution for the local and distant scene models without including the synthetic objects. If the BRDF and geometry of the local scene model were perfectly accurate, then one would expect the appearance of the rendered local scene to be consistent with its appearance in the light-based model of the entire scene. Let us call the appearance of the local scene from the desired viewpoint in the light-based model LS_b . In the context of the method described in Section 5, LS_b is simply the background image. We will let LS_{noobj} denote the appearance of the local scene, without the synthetic objects, as calculated by the global illumination solution. The error in the rendered local scene (without the objects) is thus: $Err_{ls} = LS_{noobj} - LS_b$. This error results from the difference between the BRDF characteristics of the actual local scene as compared to the modeled local scene.

Let LS_{obj} denote the appearance of the local environment as calculated by the global illumination solution with the synthetic objects in place. We can compensate for the error if we compute our final rendering LS_{final} as:

$$LS_{final} = LS_{obj} - Err_{ls}$$

Equivalently, we can write:

$$LS_{final} = LS_b + (LS_{obj} - LS_{noobj})$$

In this form, we see that whenever LS_{obj} and LS_{noobj} are the same (i.e. the addition of the objects to the scene had no effect on the local scene) the final rendering of the local scene is equivalent to LS_b (e.g. the background plate). When LS_{obj} is darker than LS_{noobj} , light is subtracted from the background to form shadows,

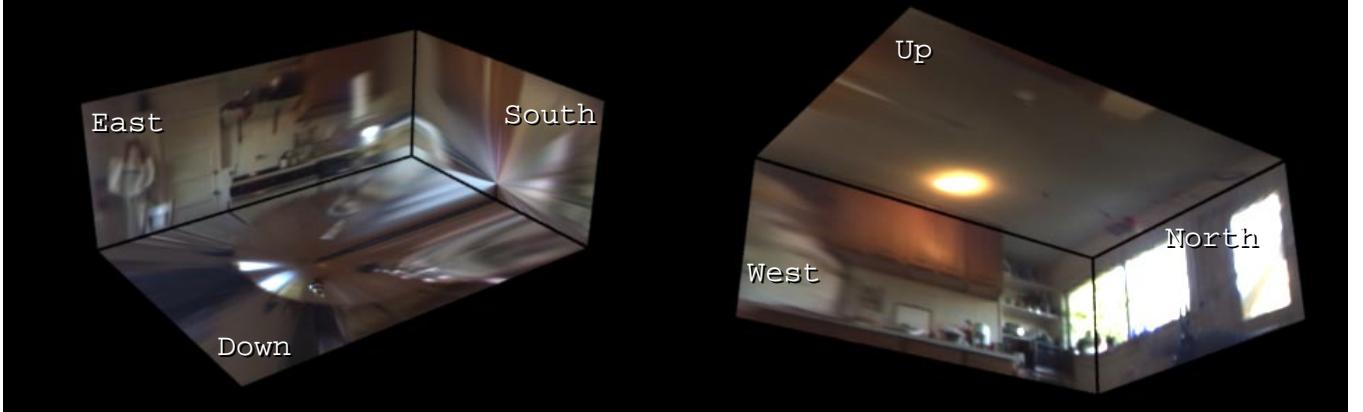


Figure 5: A Light-Based Model A simple light-based model of a room is constructed by mapping the image from a light probe onto a box. The box corresponds to the upper half of the room, with the bottom face of the box being coincident with the top of the table. The model contains the full dynamic range of the original scene, which is not reproduced in its entirety in this figure.

and when LS_{obj} is lighter than LS_{noobj} light is added to the background to produce reflections and caustics.

Stated more generally, the appearance of the local scene without the objects is computed with the correct reflectance characteristics lit by the correct environment, and the change in appearance due to the presence of the synthetic objects is computed with the modeled reflectance characteristics as lit by the modeled environment. While the realism of LS_{final} still benefits from having a good model of the reflectance characteristics of the local scene, the perceptual effect of small errors in albedo or specular properties is considerably reduced. Fig. 8(g) shows a final rendering in which the local environment is computed using this differential rendering technique. The objects are composited into the image directly from the LS_{obj} solution shown in Fig. 8(c).

It is important to stress that this technique can still produce arbitrarily wrong results depending on the amount of error in the estimated local scene BRDF and the inaccuracies in the light-based model of the distance scene. In fact, Err_{ls} may be larger than LS_{obj} , causing LS_{final} to be negative. An alternate approach is to compensate for the *relative* error in the appearance of the local scene: $LS_{final} = LS_b(LS_{obj}/LS_{noobj})$. Inaccuracies in the local scene BRDF will also be reflected in the objects.

In the next section we discuss techniques for estimating the BRDF of the local scene.

7 Estimating the local scene BRDF

Simulating the interaction of light between the local scene and the synthetic objects requires a model of the reflectance characteristics of the local scene. Considerable recent work [32, 20, 8, 27] has presented methods for measuring the reflectance properties of materials through observation under controlled lighting configurations. Furthermore, reflectance characteristics can also be measured with commercial radiometric devices.

It would be more convenient if the local scene reflectance could be estimated directly from observation. Since the light-based model contains information about the radiance of the local scene as well as its irradiance, it actually contains information about the local scene reflectance. If we hypothesize reflectance characteristics for the local scene, we can illuminate the local scene with its known irradiance from the light-based model. If our hypothesis is correct, then the appearance should be consistent with the measured appearance. This suggests the following iterative method for recovering the reflectance properties of the local scene:

1. Assume a reflectance model for the local scene (e.g. diffuse only, diffuse + specular, metallic, or arbitrary BRDF, including

spatial variation)

2. Choose approximate initial values for the parameters of the reflectance model
3. Compute a global illumination solution for the local scene with the current parameters using the observed lighting configuration or configurations.
4. Compare the appearance of the rendered local scene to its actual appearance in one or more views.
5. If the renderings are not consistent, adjust the parameters of the reflectance model and return to step 3.

Efficient methods of performing the adjustment in step 5 that exploit the properties of particular reflectance models are left as future work. However, assuming a diffuse-only model of the local scene in step 1 makes the adjustment in step 5 straightforward. We have:

$$L_{r1}(\theta_r, \phi_r) = \int_0^{2\pi} \int_0^{\pi/2} \rho_d L_i(\theta_i, \phi_i) \cos \theta_i \sin \theta_i d\theta_i d\phi_i = \\ \rho_d \int_0^{2\pi} \int_0^{\pi/2} L_i(\theta_i, \phi_i) \cos \theta_i \sin \theta_i d\theta_i d\phi_i$$

If we initialize the local scene to be perfectly diffuse ($\rho_d = 1$) everywhere, we have:

$$L_{r2}(\theta_r, \phi_r) = \int_0^{2\pi} \int_0^{\pi/2} L_i(\theta_i, \phi_i) \cos \theta_i \sin \theta_i d\theta_i d\phi_i$$

The updated diffuse reflectance coefficient for each part of the local scene can be computed as:

$$\rho'_d = \frac{L_{r1}(\theta_r, \phi_r)}{L_{r2}(\theta_r, \phi_r)}$$

In this manner, we use the global illumination calculation to render each patch as a perfectly diffuse reflector, and compare the resulting radiance to the observed value. Dividing the two quantities yields the next estimate of the diffuse reflection coefficient ρ'_d . If there is no interreflection within the local scene, then the ρ'_d estimates will make the renderings consistent. If there is interreflection, then the algorithm should be iterated until there is convergence.

For a trichromatic image, the red, green, and blue diffuse reflectance values are computed independently. The diffuse characteristics of the background material used to produce Fig. 8(c) were

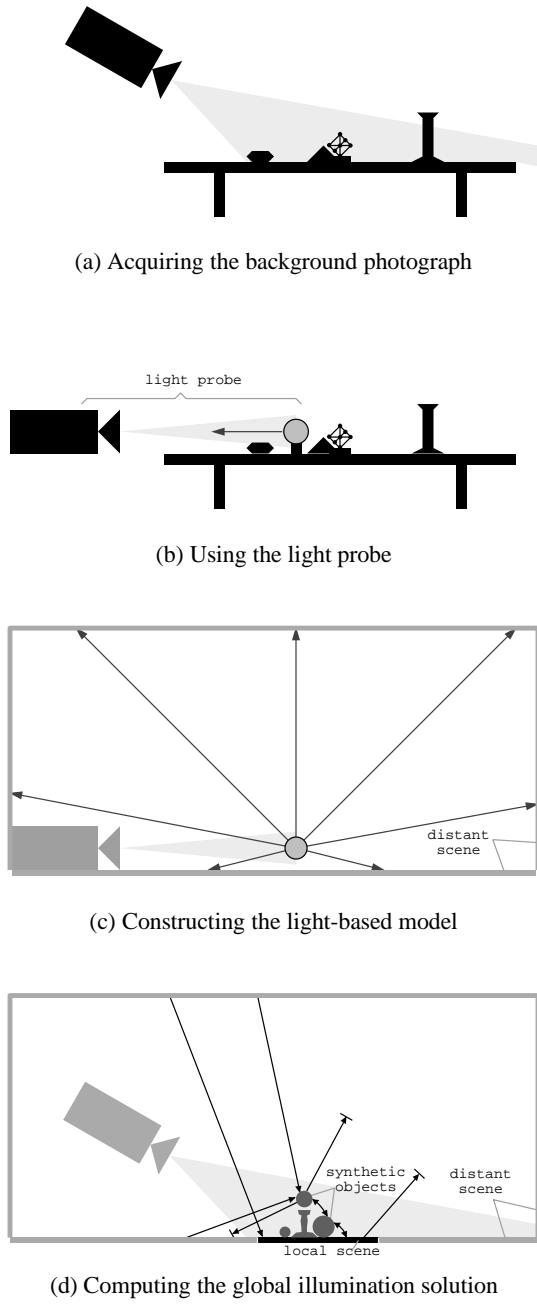


Figure 7: Using a light probe (a) The background plate of the scene (some objects on a table) is taken. (b) A light probe (in this case, the camera photographing a steel ball) records the incident radiance near the location of where the synthetic objects are to be placed. (c) A simplified light-based model of the distant scene is created as a planar surface for the table and a finite box to represent the rest of the room. The scene is texture-mapped in high dynamic range with the radiance map from the light probe. The objects on the table, which were not explicitly modeled, become projected onto the table. (d) Synthetic objects and a BRDF model of the local scene are added to the light-based model of the distant scene. A global illumination solution of this configuration is computed with light coming from the distant scene and interacting with the local scene and synthetic objects. Light reflected back to the distant scene is ignored. The results of this rendering are composited (possibly with differential rendering) into the background plate from (a) to achieve the final result.

computed using this method, although it was assumed that the entire local scene had the same diffuse reflectance.

In the standard “plastic” illumination model, just two more coefficients – those for specular intensity and roughness – need to be specified. In Fig. 8, the specular coefficients for the local scene were estimated manually based on the specular reflection of the window in the table in Fig. 2.

8 Compositing Results

Fig. 5 shows a simple light-based model of a room constructed using the panoramic radiance map from Fig. 2. The room model begins at the height of the table and continues to the ceiling; its measurements and the position of the ball within it were measured manually. The table surface is visible on the bottom face. Since the room model is finite in size, the light sources are effectively local rather than infinite. The stretching on the south wall is due to the poor sampling toward the silhouette edge of the ball.

Figs. 4 and 6 show complex arrangements of synthetic objects lit entirely by a variety of light-based models. The selection and composition of the objects in the scene was chosen to exhibit a wide variety of light interactions, including diffuse and specular reflectance, multiple soft shadows, and reflected and focused light. Each rendering was produced using the RADIANCE system with two diffuse light bounces and a relatively high density of ambient sample points.

Fig. 8(a) is a background plate image into which the synthetic objects will be rendered. In 8(b) a calibration grid was placed on the table in order to determine the camera pose relative to the scene and to the mirrored ball, which can also be seen. The poses were determined using the photogrammetric method in [10]. In 8(c), a model of the local scene as well as the synthetic objects is geometrically matched and composited onto the background image. Note that the local scene, while the same average color as the table, is readily distinguishable at its edges and because it lacks the correct variations in albedo.

Fig. 8(d) shows the results of lighting the local scene model with the light-based model of the room, without the objects. This image will be compared to 8(c) in order to determine the effect the synthetic objects have on the local scene. Fig. 8(e) is a mask image in which the white areas indicate the location of the synthetic objects. If the distant or local scene were to occlude the objects, such regions would be dark in this image.

Fig. 8(f) shows the difference between the appearance of the local scene rendered with (8(c)) and without (8(d)) the objects. For illustration purposes, the difference in radiance values have been offset so that zero difference is shown in gray. The objects have been masked out using image 8(e). This difference image encodes both the shadowing (dark areas) and reflected and focussed light (light areas) imposed on the local scene by the addition of the synthetic objects.

Fig. 8(g) shows the final result using the differential rendering method described in Section 6. The synthetic objects are copied directly from the global illumination solution 8(c) using the object mask 8(e). The effects the objects have on the local scene are included by adding the difference image 8(f) (without offset) to the background image. The remainder of the scene is copied directly from the background image 8(a). Note that in the mirror ball’s reflection, the modeled local scene can be observed without the effects of differential rendering — a limitation of the compositing technique.

In this final rendering, the synthetic objects exhibit a consistent appearance with the real objects present in the background image 8(a) in both their diffuse and specular shading, as well as the direction and coloration of their shadows. The somewhat speckled nature of the object reflections seen in the table surface is due to



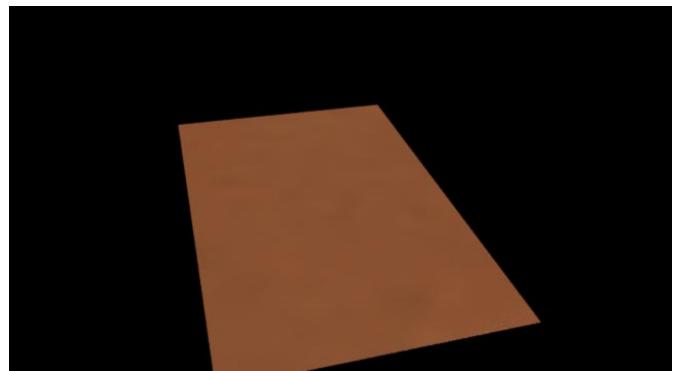
(a) Background photograph



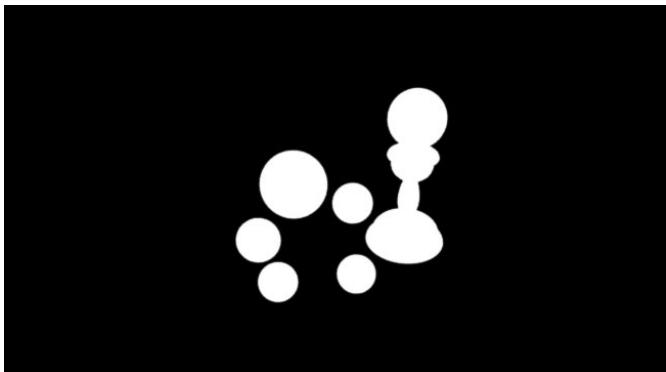
(b) Camera calibration grid and light probe



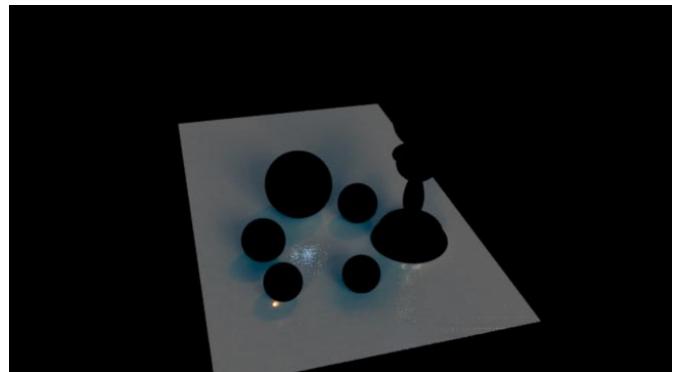
(c) Objects and local scene matched to background



(d) Local scene, without objects, lit by the model



(e) Object matte



(f) Difference in local scene between c and d



(g) Final result with differential rendering

Figure 8: Compositing synthetic objects into a real scene using a light probe and differential rendering

the stochastic nature of the particular global illumination algorithm used.

The differential rendering technique successfully eliminates the border between the local scene and the background image seen in 8(c). Note that the albedo texture of the table in the local scene area is preserved, and that a specular reflection of a background object on the table (appearing just to the left of the floating sphere) is correctly preserved in the final rendering. The local scene also exhibits reflections from the synthetic objects. A caustic from the glass ball focusing the light of the ceiling lamp onto the table is evident.

9 Future work

The method proposed here suggests a number of areas for future work. One area is to investigate methods of automatically recovering more general reflectance models for the local scene geometry, as proposed in Section 7. With such information available, the program might also also be able to suggest which areas of the scene should be considered as part of the local scene and which can safely be considered distant, given the position and reflectance characteristics of the desired synthetic objects.

Some additional work could be done to allow the global illumination algorithm to compute the illumination solution more efficiently. One technique would be to have an algorithm automatically locate and identify concentrated light sources in the light-based model of the scene. With such knowledge, the algorithm could compute most of the direct illumination in a forward manner, which could dramatically increase the efficiency with which an accurate solution could be calculated. To the same end, use of the method presented in [15] to expedite the solution could be investigated. For the case of compositing moving objects into scenes, greatly increased efficiency could be obtained by adapting incremental radiosity methods to the current framework.

10 Conclusion

We have presented a general framework for adding new objects to light-based models with correct illumination. The method leverages a technique of using high dynamic range images of real scene radiance to synthetically illuminate new objects with arbitrary reflectance characteristics. We leverage this technique in a general method to simulate interplay of light between synthetic objects and the light-based environment, including shadows, reflections, and caustics. The method can be implemented with standard global illumination techniques.

For the particular case of rendering synthetic objects into real scenes (rather than general light-based models), we have presented a practical instance of the method that uses a light probe to record incident illumination in the vicinity of the synthetic objects. In addition, we have described a differential rendering technique that can convincingly render the interplay of light between objects and the local scene when only approximate reflectance information for the local scene is available. Lastly, we presented an iterative approach for determining reflectance characteristics of the local scene based on measured geometry and observed radiance in uncontrolled lighting conditions. It is our hope that the techniques presented here will be useful in practice as well as comprise a useful framework for combining material-based and light-based graphics.

Acknowledgments

The author wishes to thank Chris Bregler, David Forsyth, Jianbo Shi, Charles Ying, Steve Chenney, and Andrean Kalemis for the various forms of help and advice they provided. Special gratitude is also due to Jitendra Malik for helping make this work possible. Discussions with Michael Naimark and Steve Saunders helped motivate this work. Tim Hawkins provided extensive assistance on improving and revising this paper and provided invaluable assistance with image acquisition. Gregory Ward Larson deserves great thanks for the RADIANCE lighting simulation system and his invaluable assistance and advice in using RADIANCE in this research, for assisting with reflectance measurements, and for very helpful comments and suggestions on the paper. This research was supported by a Multidisciplinary University Research Initiative on

three dimensional direct visualization from ONR and BMDO, grant FDN00014-96-1-1200.

References

- [1] ADELSON, E. H., AND BERGEN, J. R. *Computational Models of Visual Processing*. MIT Press, Cambridge, Mass., 1991, ch. 1. The Plenoptic Function and the Elements of Early Vision.
- [2] AZARMI, M. *Optical Effects Cinematography: Its Development, Methods, and Techniques*. University Microfilms International, Ann Arbor, Michigan, 1973.
- [3] BLINN, J. F. Texture and reflection in computer generated images. *Communications of the ACM* 19, 10 (October 1976), 542–547.
- [4] CHEN, E. QuickTime VR - an image-based approach to virtual environment navigation. In *SIGGRAPH '95* (1995).
- [5] CHEN, S. E. Incremental radiosity: An extension of progressive radiosity to an interactive synthesis system. In *SIGGRAPH '90* (1990), pp. 135–144.
- [6] COHEN, M. F., CHEN, S. E., WALLACE, J. R., AND GREENBERG, D. P. A progressive refinement approach to fast radiosity image generation. In *SIGGRAPH '88* (1988), pp. 75–84.
- [7] CURLESS, B., AND LEVOY, M. A volumetric method for building complex models from range images. In *SIGGRAPH '96* (1996), pp. 303–312.
- [8] DANA, K. J., GINNEKEN, B., NAYAR, S. K., AND KOENDERINK, J. J. Reflectance and texture of real-world surfaces. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.* (1997), pp. 151–157.
- [9] DEBEVEC, P. E., AND MALIK, J. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH '97* (August 1997), pp. 369–378.
- [10] DEBEVEC, P. E., TAYLOR, C. J., AND MALIK, J. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *SIGGRAPH '96* (August 1996), pp. 11–20.
- [11] DEBEVEC, P. E., YU, Y., AND BORSHUKOV, G. D. Efficient view-dependent image-based rendering with projective texture-mapping. Tech. Rep. UCB//CSD-98-1003, University of California at Berkeley, 1998.
- [12] DRETTAKIS, G., ROBERT, L., AND BOUGNOUX, S. Interactive common illumination for computer augmented reality. In *8th Eurographics workshop on Rendering, St. Etienne, France* (May 1997), J. Dorsey and P. Slusallek, Eds., pp. 45–57.
- [13] FIELDING, R. *The Technique of Special Effects Cinematography*. Hastings House, New York, 1968.
- [14] FOURNIER, A., GUNAWAN, A., AND ROMANZIN, C. Common illumination between real and computer generated scenes. In *Graphics Interface* (May 1993), pp. 254–262.
- [15] GERSHBEIN, R., SCHRODER, P., AND HANRAHAN, P. Textures and radiosity: Controlling emission and reflection with texture maps. In *SIGGRAPH '94* (1994).
- [16] GORAL, C. M., TORRANCE, K. E., GREENBERG, D. P., AND BATTAILLE, B. Modeling the interaction of light between diffuse surfaces. In *SIGGRAPH '84* (1984), pp. 213–222.
- [17] GORTLER, S. J., GRZESZCZUK, R., SZELISKI, R., AND COHEN, M. F. The Lumigraph. In *SIGGRAPH '96* (1996), pp. 43–54.
- [18] HECKBERT, P. S. Survey of texture mapping. *IEEE Computer Graphics and Applications* 6, 11 (November 1986), 56–67.
- [19] KAJIYA, J. The rendering equation. In *SIGGRAPH '86* (1986), pp. 143–150.
- [20] KARNER, K. F., MAYER, H., AND GERVAUTZ, M. An image based measurement system for anisotropic reflection. In *EUROGRAPHICS Annual Conference Proceedings* (1996).
- [21] KOENDERINK, J. J., AND VAN DOORN, A. J. Illuminant texture due to surface mesostructure. *J. Opt. Soc. Am. 13*, 3 (1996).
- [22] LAVEAU, S., AND FAUGERAS, O. 3-D scene representation as a collection of images. In *Proceedings of 12th International Conference on Pattern Recognition* (1994), vol. 1, pp. 689–691.
- [23] LEVOY, M., AND HANRAHAN, P. Light field rendering. In *SIGGRAPH '96* (1996), pp. 31–42.
- [24] MCMILLAN, L., AND BISHOP, G. Plenoptic Modeling: An image-based rendering system. In *SIGGRAPH '95* (1995).
- [25] NAKAMAE, E., HARADA, K., AND ISHIZAKI, T. A montage method: The overlaying of the computer generated images onto a background photograph. In *SIGGRAPH '86* (1986), pp. 207–214.
- [26] PORTER, T., AND DUFF, T. Compositing digital images. In *SIGGRAPH 84* (July 1984), pp. 253–259.
- [27] SATO, Y., WHEELER, M. D., AND IKEUCHI, K. Object shape and reflectance modeling from observation. In *SIGGRAPH '97* (1997), pp. 379–387.
- [28] SMITH, T. G. *Industrial Light and Magic: The Art of Special Effects*. Ballantine Books, New York, 1986.
- [29] SZELISKI, R. Image mosaicing for tele-reality applications. In *IEEE Computer Graphics and Applications* (1996).
- [30] TURK, G., AND LEVOY, M. Zippered polygon meshes from range images. In *SIGGRAPH '94* (1994), pp. 311–318.
- [31] VEECH, E., AND GUIBAS, L. J. Metropolis light transport. In *SIGGRAPH '97* (August 1997), pp. 65–76.
- [32] WARD, G. J. Measuring and modeling anisotropic reflection. In *SIGGRAPH '92* (July 1992), pp. 265–272.
- [33] WARD, G. J. The radiance lighting simulation and rendering system. In *SIGGRAPH '94* (July 1994), pp. 459–472.
- [34] WATANABE, M., AND NAYAR, S. K. Telecentric optics for computational vision. In *Proceedings of Image Understanding Workshop (IUW 96)* (February 1996).
- [35] Y.CHEN, AND MEDIONI, G. Object modeling from multiple range images. *Image and Vision Computing* 10, 3 (April 1992), 145–155.

A Median Cut Algorithm for Light Probe Sampling

Paul Debevec* USC Institute for Creative Technologies

ABSTRACT We present a technique for approximating a light probe image as a constellation of light sources based on a median cut algorithm. The algorithm is efficient, simple to implement, and can realistically represent a complex lighting environment with as few as 64 point light sources.

Introduction The quality of approximating an image-based lighting (IBL) environment as a finite number of point lights is increased if the light positions are chosen to follow the distribution of the incident illumination; this has been a goal of previous stratified sampling approaches [Cohen and Debevec 2001; Kollig and Keller 2003; Agarwal et al. 2003; Ostromoukhov et al. 2004]. In this work, we show that subdividing the image into regions of equal energy achieves this property and yields a well-conditioned and easy to implement static sampling algorithm.



Figure 1: The Grace Cathedral light probe subdivided into 64 regions of equal light energy using the median cut algorithm. The small circles are the 64 light sources chosen as the energy centroids of each region; the lights are all approximately equal in energy.

Algorithm Taking inspiration from Paul Heckbert's median-cut color quantization algorithm [Heckbert 1982], we can partition a light probe image in the rectangular latitude-longitude format into 2^n regions of similar light energy as follows:

1. Add the entire light probe image to the region list as a single region.
2. For each region in the list, subdivide along the longest dimension such that its light energy is divided evenly.
3. If the number of iterations is less than n , return to step 2.
4. Place a light source at the center or centroid of each region, and set the light source color to the sum of pixel values within the region.

Implementation Calculating the total energy within regions of the image can be accelerated using a summed area table [Crow 1984]. Computing the total light energy is most naturally performed on a monochrome version of the lighting environment rather than the RGB pixel colors; such an image can be formed as a weighted average of the color channels of the light probe image, e.g. $Y = 0.2125R + 0.7154G + 0.0721B$ following ITU-R Recommendation BT.709. While the partitioning decisions are made on the monochrome image, the light source colors are computed using the corresponding regions in the original RGB image.

The latitude-longitude mapping over-represents regions near the poles. To compensate, the pixels of the probe image should first be scaled by $\cos \phi$ where ϕ is the pixel's angle of inclination. Determining the longest dimension of a region should also take the over-representation into account; this can be accomplished by weighting a regions width by $\cos \phi$ for an inclination ϕ at center of the region.

Results Fig. 1 shows the Grace Cathedral lighting environment partitioned into 64 light source regions, and Fig. 2 shows a small diffuse scene rendered with 16, 64, and 256 light sources chosen in this manner. Using 64 lights produces a close approximation to a computationally expensive Monte Carlo solution, and the 256-light approximation is nearly indistinguishable.

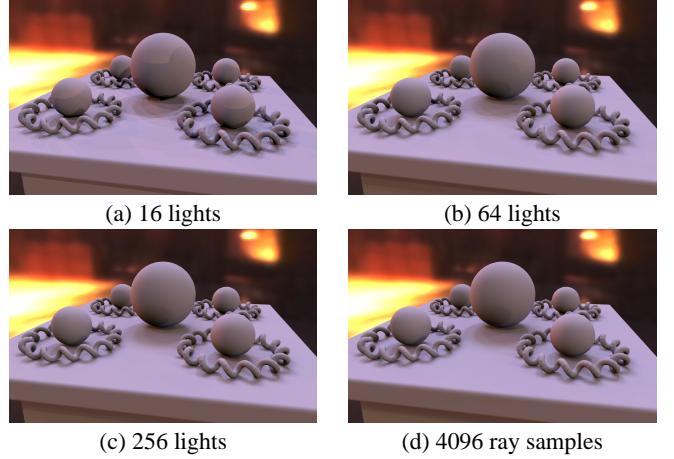


Figure 2: (a-c) Noise-free renderings in the Grace Cathedral environment approximated by 16, 64, and 256 light sources. (d) A not quite noise-free Monte Carlo rendering using 4096 randomly chosen rays per pixel.

Conclusion The median cut technique is extremely fast compared to most other sampling techniques and produces noise-free renderings at the expense of bias inversely proportional to the number of light sources used. In future work we will investigate the stability of the technique for animated lighting environments and explore adaptations for scenes with general BRDFs.

References

- AGARWAL, S., RAMAMOORTHI, R., BELONGIE, S., AND JENSEN, H. W. 2003. Structured importance sampling of environment maps. *ACM Transactions on Graphics* 22, 3 (July), 605–612.
- COHEN, J. M., AND DEBEVEC, P. 2001. The LightGen HDRShop plugin. <http://www.hdrshop.com/main-pages/plugins.html>.
- CROW, F. C. 1984. Summed-area tables for texture mapping. In *Computer Graphics (Proceedings of SIGGRAPH 84)*, vol. 18, 207–212.
- HECKBERT, P. 1982. Color image quantization for frame buffer display. In *SIGGRAPH '82: Proceedings of the 9th annual conference on Computer graphics and interactive techniques*, ACM Press, New York, NY, USA, 297–307.
- KOLLIG, T., AND KELLER, A. 2003. Efficient illumination by high dynamic range images. In *Eurographics Symposium on Rendering: 14th Eurographics Workshop on Rendering*, 45–51.
- OSTROMOUKHOV, V., DONOHUE, C., AND JODOIN, P.-M. 2004. Fast hierarchical importance sampling with blue noise properties. *ACM Transactions on Graphics* 23, 3 (Aug.), 488–495.

*Email: debevec@ict.usc.edu Web: www.debevec.org/MedianCut/

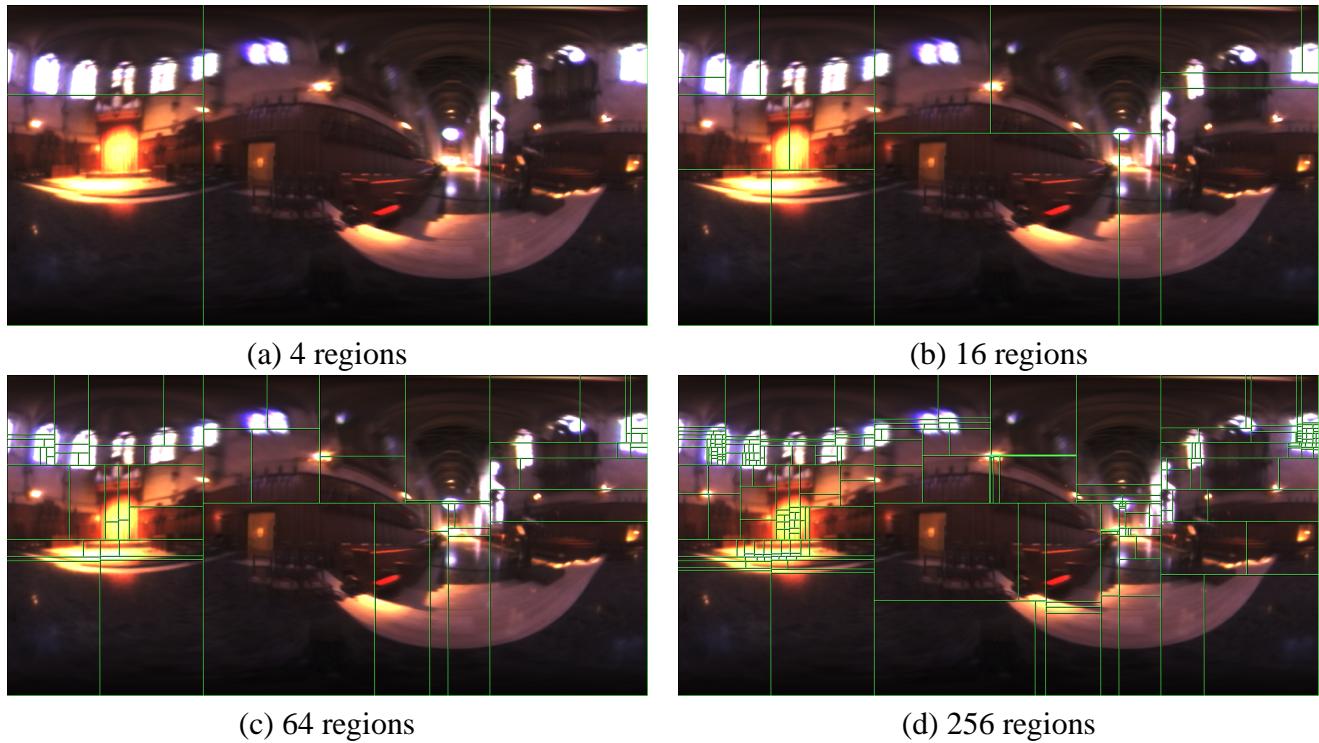


Figure 1: The Grace Cathedral light probe subdivided into 4, 16, 64, and 256 regions of equal light energy using the median cut algorithm.

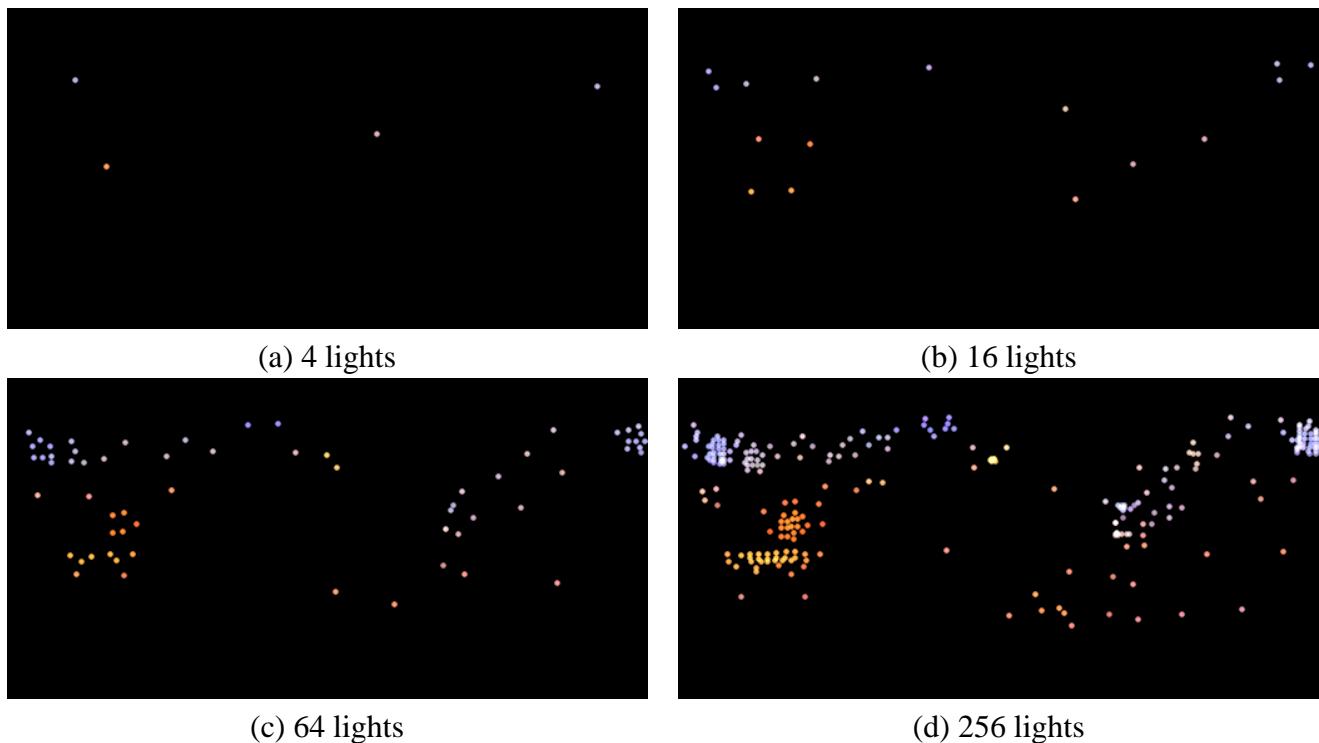
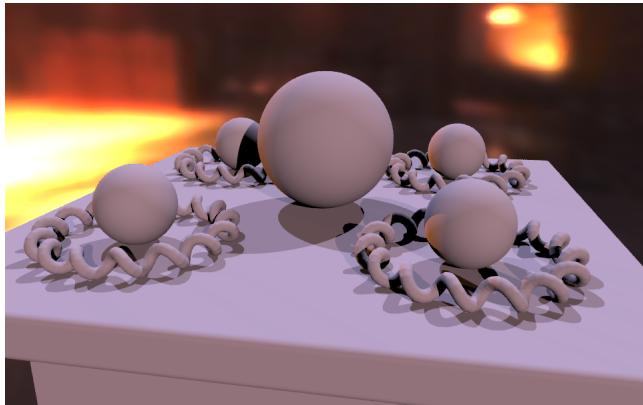
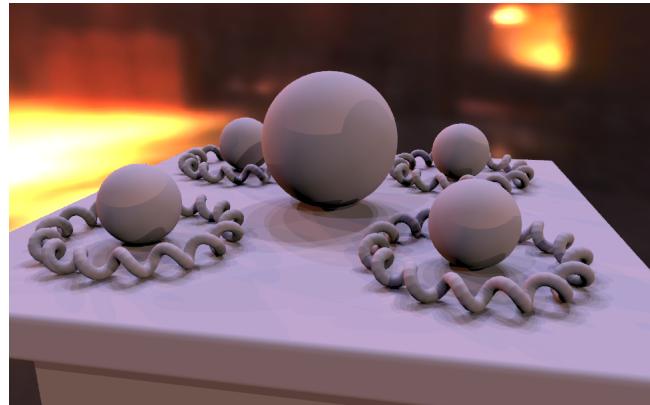


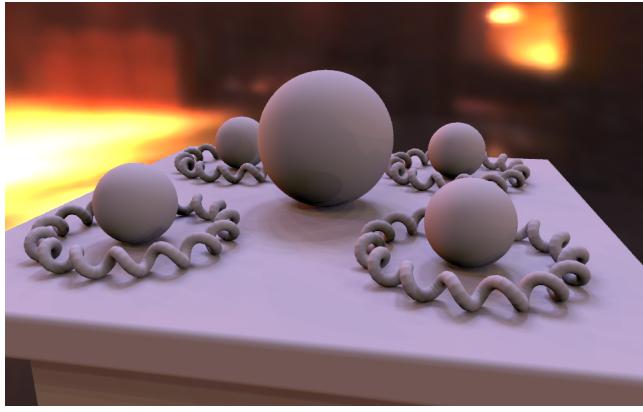
Figure 2: The Grace Cathedral light probe represented as 4, 16, 64, and 256 light sources chosen as the energy centroids of each region; each light is approximately equal energy.



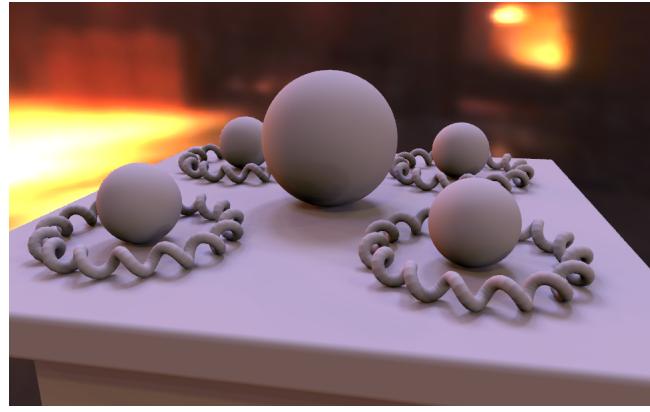
(a) 4 lights



(b) 16 lights

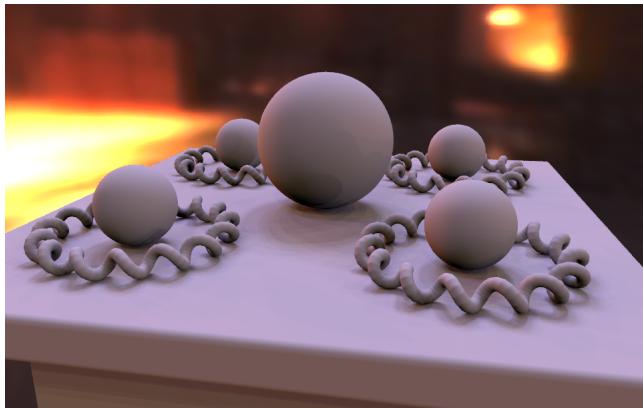


(c) 64 lights

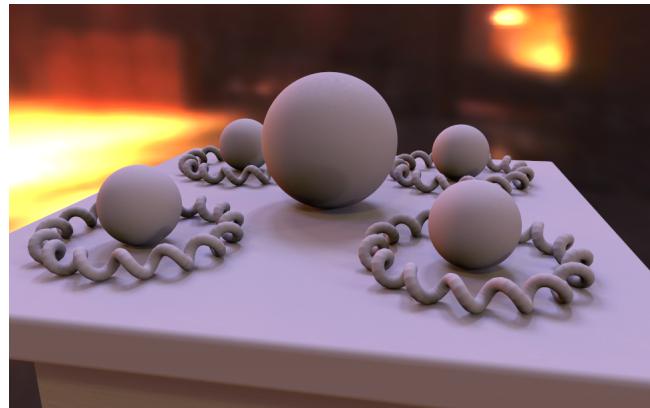


(d) 256 lights

Figure 3: Noise-free renderings in the Grace Cathedral environment approximated by 4, 16, 64, and 256 light sources.



(a) 64 lights



(b) 4096 ray samples

Figure 4: (a) Noise-free rendering in the Grace Cathedral approximated by 64 light sources, compared to (b) a not quite noise-free Monte Carlo rendering using 4096 randomly chosen rays per pixel.