

Computational Photography

SIGGRAPH 2006 Course 15 Notes

Organizers

Ramesh Raskar

Senior Research Scientist

MERL - Mitsubishi Electric Research Labs

201 Broadway

Cambridge, MA 02139

Email: raskar@merl.com

<http://www.merl.com/people/raskar/>

Jack Tumblin

Assistant Professor

Computer Science Dept., Northwestern University

1890 Maple Avenue, Evanston IL 60201,

Email: jet@cs.northwestern.edu

<http://www.cs.northwestern.edu/~jet>

Course Web Page

<http://www.merl.com/people/raskar/photo/>

Contents

- **Cover**
 - **Course Description**
 - **Prerequisites**
 - **Course Schedule**
 - **Lecturer Biographies**
 - **Course Syllabus**
- **Course Slides**
- **Papers**
- **Code**
- **Bibliography**

Acknowledgements

We would like to thank **Amit Agrawal** for his help on several aspects of this course and for providing the pseudo-code included in these notes.

We would also like to thank the following for providing us their slides and videos which have been extensively used throughout this course.

Stanford University: **Bennett Wilburn, Vaibhav Vaish, Ren Ng**

Washington University: **Aseem Agrawala**

MERL/Brown University: **Morgan McGuire**

University of Pennsylvania: **Jianbo Shi**

Course Description

Digital photography is evolving rapidly with advances in electronic sensing, processing and storage. The emerging field of computational photography attempts to exploit the cheaper and faster computing to overcome the physical limitations of a camera, such as dynamic range, resolution or depth of field, and extend the possible range of applications. The computational techniques encompass methods from modification of imaging parameters during capture to modern image reconstruction methods from the captured samples.

Many ideas in computational photography are still relatively new to digital artists and programmers although they are familiar with photography and image manipulation techniques. A larger problem is that a multi-disciplinary field that combines ideas from computational methods and modern digital photography involves a steep learning curve. For example photographers are not always familiar with advanced algorithms now emerging to capture high dynamic range images, but image processing researchers face difficulty in understanding the capture and noise issues in digital cameras. These topics, however, can be easily learned without extensive background. The goal of this course is to present both aspects in a compact form.

The new capture methods include sophisticated sensors, electromechanical actuators and on-board processing. Examples include adaptation to sensed scene depth and illumination, taking multiple pictures by varying camera parameters or actively modifying the flash illumination parameters. A class of modern reconstruction methods is emerging. The methods can achieve a 'photomontage' by optimally fusing information from multiple images, improve signal to noise ratio and extract scene features such as depth edges. The course briefly reviews fundamental topics in digital imaging and then provides a practical guide to underlying techniques beyond image processing such as gradient domain operations, graph cuts, bilateral filters and optimizations.

The participants learn about topics in image capture and manipulation methods for generating compelling pictures for computer graphics and for extracting scene properties for computer vision, with several examples. We hope to provide enough fundamentals to satisfy the technical specialist without intimidating the curious graphics researcher interested in photography.

Prerequisites

A basic understanding of camera operation and image processing is required. Familiarity with concepts of linear systems, convolution, and machine vision will be useful.

Photographers, digital artists, image processing programmers and vision researchers using or building applications for digital cameras or images will learn about camera fundamentals and powerful computational tools, along with many real world examples.

Course Schedule

A.1	Introduction	(Raskar, 5 mins)
A.2	Concepts in Computational Photography	(Tumblin, 15 mins)
A.3	Understanding Film-like Photography	(Tumblin, 10 mins)
A.4	Image Processing Tools	(Raskar, 10 mins)
	Q&A (5 minutes)	
B.1	Image Reconstruction Techniques	(Tumblin, 20 mins)
B.2	Computational Camera : Convergence of Optics and Software	(Nayar, 35 minutes)
	Q&A	(5 minutes)
	Break (15 mins)	
C.1	Computational Imaging in the Sciences	(Levoy, 35 minutes)
	Q&A	(5 minutes)
D.1	Computational Illumination	(Raskar, 15 mins)
D.2	Smart Optics, Modern Sensors and Future Cameras	(Raskar, 20 mins)
E.1	Panel Discussion and Q&A	(Nayar, Levoy, Raskar, Tumblin 30 mins)

Topics not covered: film cameras, advanced optics, traditional image processing, image based rendering (IBR) and novel view synthesis, lighting and illumination hardware technologies, camera assisted projector display systems, geometric calibration and photometric calibration techniques, compression, storage, photo-editing software packages, file formats and standards.

Course Web Page <http://www.merl.com/people/raskar/photo/>

Lecturer Biographies

Marc Levoy

Marc Levoy is an Associate Professor of Computer Science and Electrical Engineering at Stanford University. He received his PhD in Computer Science from the University of North Carolina in 1989. In the 1970's Levoy worked on computer animation, developing an early computer-assisted cartoon animation system. In the 1980's Levoy worked on volume rendering, a family of techniques for displaying sampled three-dimensional functions, such as CT and MR data. In the 1990's he worked on technology and algorithms for 3D scanning. This led to the Digital Michelangelo Project, in which he and a team of researchers spent a year in Italy digitizing the statues of Michelangelo using laser rangefinders. His current interests include light field sensing and display, computational imaging, and digital photography. Levoy received the NSF Presidential Young Investigator Award in 1991 and the SIGGRAPH Computer Graphics Achievement Award in 1996 for his work in volume rendering.



<http://graphics.stanford.edu/~levoy/>

Shree Nayar

Shree K. Nayar received his PhD degree in Electrical and Computer Engineering from the Robotics Institute at Carnegie Mellon University in 1990. He heads the Columbia Automated Vision Environment (CAVE), which is dedicated to the development of advanced computer vision systems. His research is focused on three areas; the creation of novel vision sensors, the design of physics based models for vision, and the development of algorithms for scene understanding. His work is motivated by applications in the fields of digital imaging, computer graphics, and robotics. Professor Nayar has received best paper awards at ICCV 1990, ICPR 1994, CVPR 1994, ICCV 1995, CVPR 2000 and CVPR 2004. He is the recipient of the David and Lucile Packard Fellowship (1992), the National Young Investigator Award (1993), the NTT Distinguished Scientific Achievement Award (1994), and the Keck Foundation Award for Excellence in Teaching (1995). He has published over 100 scientific papers and has several patents on inventions related to vision and robotics.



<http://www.cs.columbia.edu/~nayar/>

Ramesh Raskar

Ramesh Raskar is a Senior Research Scientist at MERL. His research interests include projector-based graphics, computational photography and non-photorealistic rendering. He has published several articles on imaging and photography including multi-flash photography for depth edge detection, image fusion, gradient-domain imaging and projector-camera systems. His papers have appeared in SIGGRAPH, EuroGraphics, IEEE Visualization, CVPR and many other graphics and vision conferences. He was a course organizer at Siggraph 2002 through 2005. He was the panel organizer at the Symposium on Computational Photography and Video in Cambridge, MA in May 2005 and taught a graduate level class on Computational Photography at Northeastern University, Fall 2005. He is a member of the ACM and IEEE.



<http://www.merl.com/people/raskar/>

Jack Tumblin

Jack Tumblin is an assistant professor of computer science at Northwestern University. His interests include novel photographic sensors and lighting devices to assist museum curators in historical preservation, computer graphics and visual appearance, and image-based modeling and rendering. During his doctoral studies at Georgia Tech and post-doc at Cornell, he investigated tone-mapping methods to depict high-contrast scenes. His MS in Electrical Engineering (December 1990) and BSEE (1978), also from Georgia Tech, bracketed his work as co-founder of IVEX Corp., (>45 people as of 1990) where his flight simulator design work was granted 5 US Patents. He was co-organizer of a course on Computational Photography at Siggraph 2005. He is an Associate Editor of ACM Transactions on Graphics.



<http://www.cs.northwestern.edu/~jet>

Copyright Information

The following copyright information appears on the first page of each corresponding paper.

© 2006 IEEE. Reprinted, with permission, from (...).

© 2006 ACM Reprinted, with permission, from (...).

All trademarks are the property of their respective owners.
Papers and Photos are copyrighted and shall not be reproduced without permission.

Course 15: Computational Photography



Organisers

Ramesh Raskar

Mitsubishi Electric Research Labs

Jack Tumblin

Northwestern University

Course WebPage :

<http://www.merl.com/people/raskar/photo>

Course 15: Computational Photography

Course WebPage

<http://www.merl.com/people/raskar/photo>

Course Evaluation

http://www.siggraph.org/courses_evaluation

Welcome

- Understanding Film-like Photography
 - Parameters, Nonlinearities, Ray-based concepts
- Image Processing and Reconstruction Tools
 - Multi-image Fusion, Gradient domain, Graph Cuts
- Improving Camera Performance
 - Better dynamic range, focus, frame rate, resolution
- Future Directions
 - HDR cameras, Gradient sensing, Smart optics/lighting

Speaker: Marc Levoy

Marc Levoy is an Associate Professor of Computer Science and Electrical Engineering at Stanford University.

He received his PhD in Computer Science from the University of North Carolina in 1989. In the 1970's Levoy worked on computer animation, developing an early computer-assisted cartoon animation system. In the 1980's Levoy worked on volume rendering, a family of techniques for displaying sampled three-dimensional functions, such as CT and MR data. In the 1990's he worked on technology and algorithms for 3D scanning. This led to the Digital Michelangelo Project, in which he and a team of researchers spent a year in Italy digitizing the statues of Michelangelo using laser rangefinders. His current interests include light field sensing and display, computational imaging, and digital photography. Levoy received the NSF Presidential Young Investigator Award in 1991 and the SIGGRAPH Computer Graphics Achievement Award in 1996 for his work in volume rendering.



<http://graphics.stanford.edu/~levoy/>

Speaker: Shree Nayar

Shree K. Nayar is a Professor at Columbia University.

He received his PhD degree in Electrical and Computer Engineering from the Robotics Institute at Carnegie Mellon University in 1990. He heads the Columbia Automated Vision Environment (CAVE), which is dedicated to the development of advanced computer vision systems. His research is focused on three areas; the creation of novel vision sensors, the design of physics based models for vision, and the development of algorithms for scene understanding. His work is motivated by applications in the fields of digital imaging, computer graphics, and robotics. Professor Nayar has received best paper awards at ICCV 1990, ICPR 1994, CVPR 1994, ICCV 1995, CVPR 2000 and CVPR 2004. He is the recipient of the David and Lucile Packard Fellowship (1992), the National Young Investigator Award (1993), the NTT Distinguished Scientific Achievement Award (1994), and the Keck Foundation Award for Excellence in Teaching (1995). He has published over 100 scientific papers and has several patents on inventions related to vision and robotics.



<http://www.cs.columbia.edu/~nayar/>

Speaker: Ramesh Raskar

Ramesh Raskar is a Senior Research Scientist at MERL.

His research interests include projector-based graphics, computational photography and non-photorealistic rendering. He has published several articles on imaging and photography including multi-flash photography for depth edge detection, image fusion, gradient-domain imaging and projector-camera systems. His papers have appeared in SIGGRAPH, EuroGraphics, IEEE Visualization, CVPR and many other graphics and vision conferences. He was a course organizer at Siggraph 2002 through 2005. He was the panel organizer at the Symposium on Computational Photography and Video in Cambridge, MA in May 2005 and taught a graduate level class on Computational Photography at Northeastern University, Fall 2005. He is a member of the ACM and IEEE.



<http://www.merl.com/people/raskar/raskar.html>

Speaker: Jack Tumblin

Jack Tumblin is an Assistant Professor of Computer Science at Northwestern University.

His interests include novel photographic sensors to assist museum curators in historical preservation, computer graphics and visual appearance, and image-based modeling and rendering. During his doctoral studies at Georgia Tech and post-doc at Cornell, he investigated tone-mapping methods to depict high-contrast scenes. His MS in Electrical Engineering (December 1990) and BSEE (1978), also from Georgia Tech, bracketed his work as co-founder of IVEX Corp., (~45 people as of 1990) where his flight simulator design work was granted 5 US Patents. He is an Associate Editor of ACM Transactions on Graphics, was a member of the SIGGRAPH Papers Committee (2003, 2004), and in 2001 was a Guest Editor of IEEE Computer Graphics and Applications.

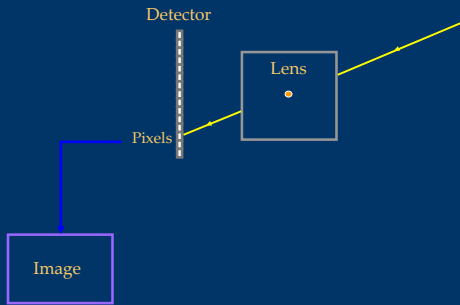


<http://www.cs.northwestern.edu/~jet>

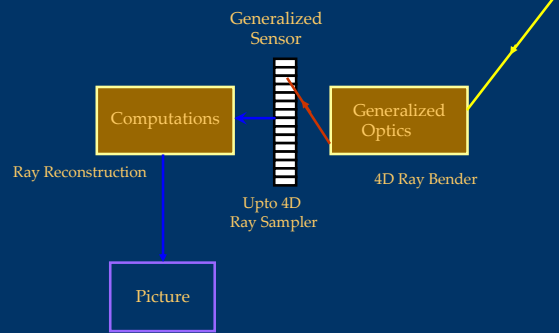
Opportunities

- **Unlocking Photography**
 - How to expand camera capabilities
 - Digital photography that goes beyond film-like photography
- **Opportunities**
 - Computing corrects for lens, sensor and lighting limitations
 - Computing merges results from multiple images
 - Computing reconstructs from coded image samples
 - Cameras benefit from computerized light sources
- **Think beyond post-capture image processing**
 - Computation well before image processing and editing
- **Learn how to build your own camera-toys**

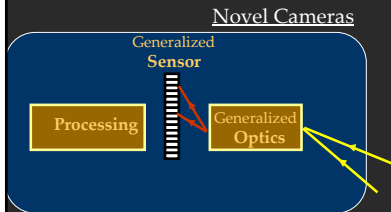
Traditional Photography



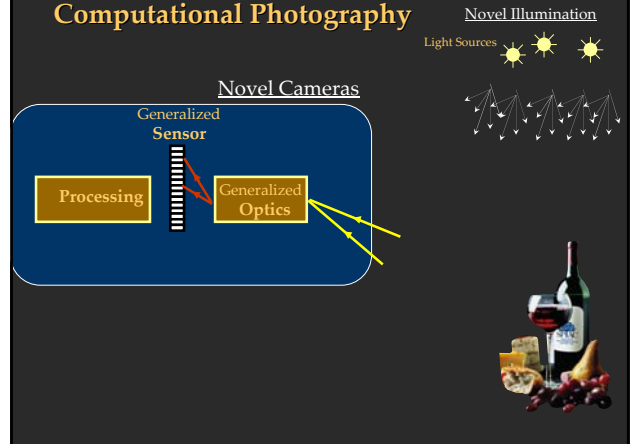
Computational Photography: Optics, Sensors and Computations

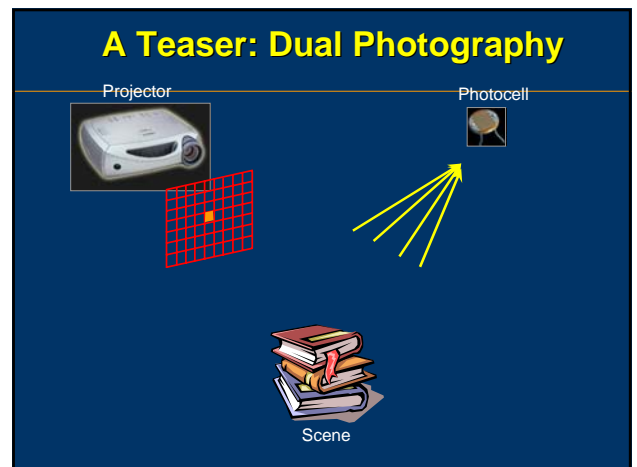
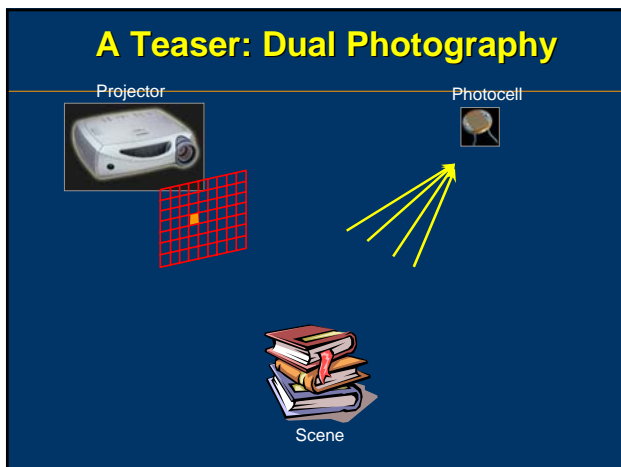
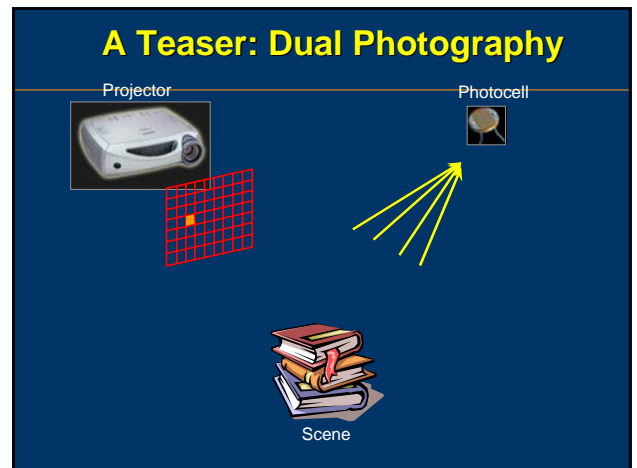
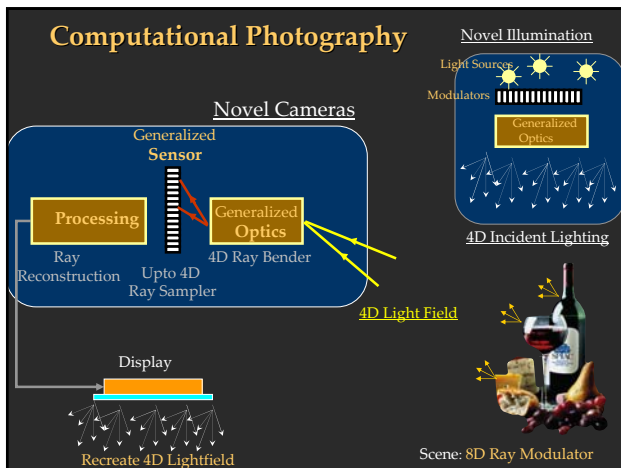
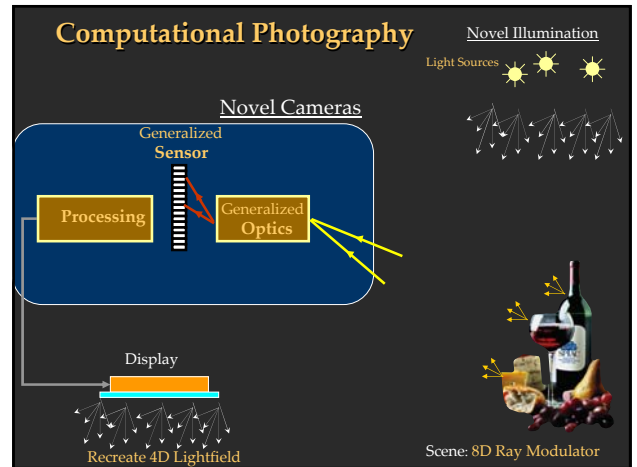
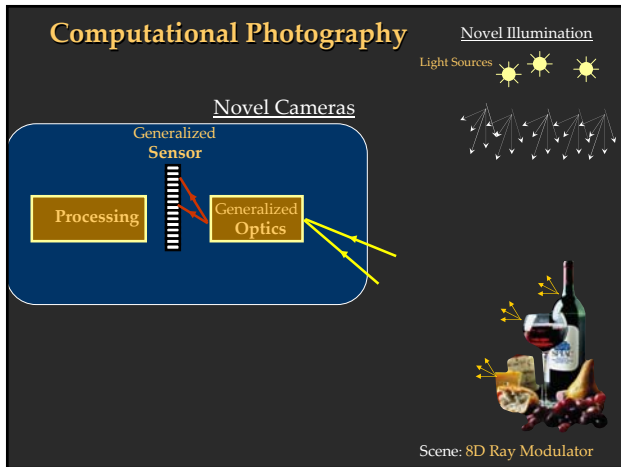


Computational Photography



Computational Photography

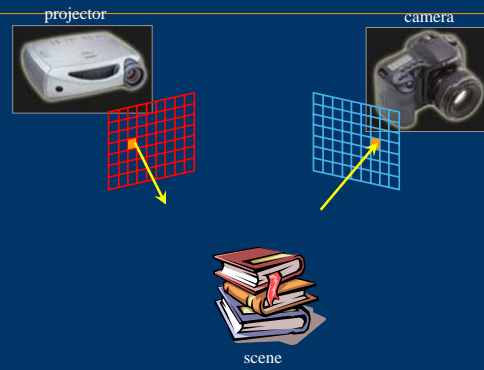




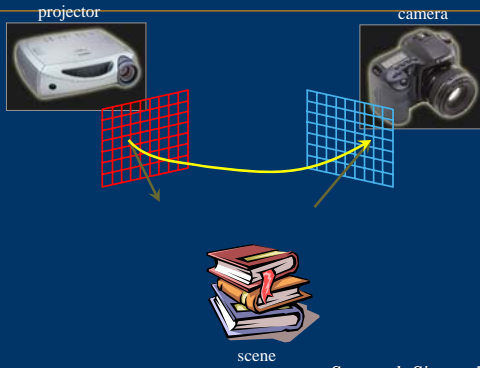
A Teaser: Dual Photography



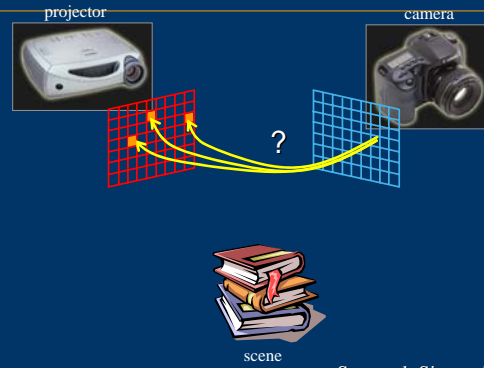
The 4D transport matrix: Contribution of each projector pixel to each camera pixel



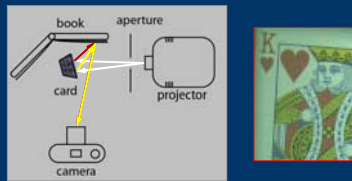
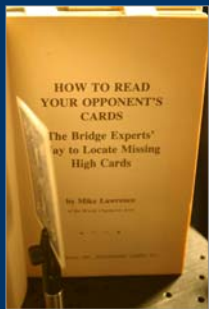
The 4D transport matrix: Contribution of each projector pixel to each camera pixel



The 4D transport matrix: Which projector pixel contribute to each camera pixel



Dual photography from diffuse reflections



Sen et al, Siggraph 2005

A Brief History of Images

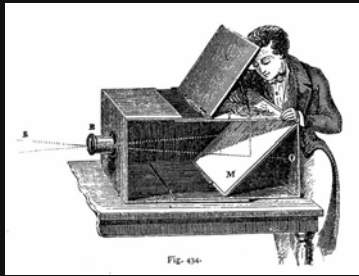
1558



Camera Obscura, Gemma Frisius, 1558

A Brief History of Images

● 1558
● 1568



Lens Based Camera Obscura, 1568

A Brief History of Images

● 1558
● 1568

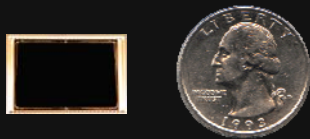


Still Life, Louis Jaques Mande Daguerre, 1837

● 1837

A Brief History of Images

● 1558
● 1568



Silicon Image Detector, 1970

● 1837

● 1970

A Brief History of Images

● 1558
● 1568



Digital Cameras

● 1837

● 1970
● 1994

Dream of A New Photography

	Old	New
• People and Time	~Cheap	Precious
• Each photo	Precious	Free
• Lighting	Critical	Automated*
• External Sensors	No	Yes
• 'Stills / Video'	Disjoint	Merged
• Exposure Settings	Pre-select	Post-Process
• Exposure Time	Pre-select	Post-Process
• Resolution/noise	Pre-select	Post-Process
• 'HDR' range	Pre-select	Post-Process

Course: Computational Photography

A2: What is the Ideal Photographic Signal?

Jack Tumblin
Northwestern University

What *is* Photography?

Safe answer:

A wholly new,
expressive medium
(ca. 1830s)



- Manipulated display of what we think, feel, want, ...
 - Capture a memory, a visual experience in tangible form
 - ‘painting with light’; express the subject’s visual essence
 - “*Exactitude is not the truth.*” –Henri Matisse

What *is* Photography?

- A 'bucket' word: a neat container for messy notions (e.g. aviation, music, comprehension)

- A record of what we see, or would like to see, in tangible form.

- **Does photography always capture it?** no.

- **So, what do we see?**



Harold 'Doc' Edgerton 1936

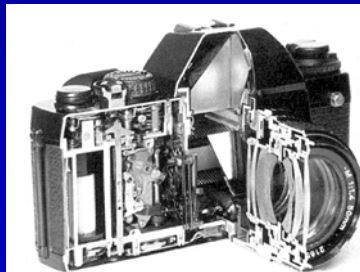
'Film-Like' Photography

Film Camera designs still dominate:

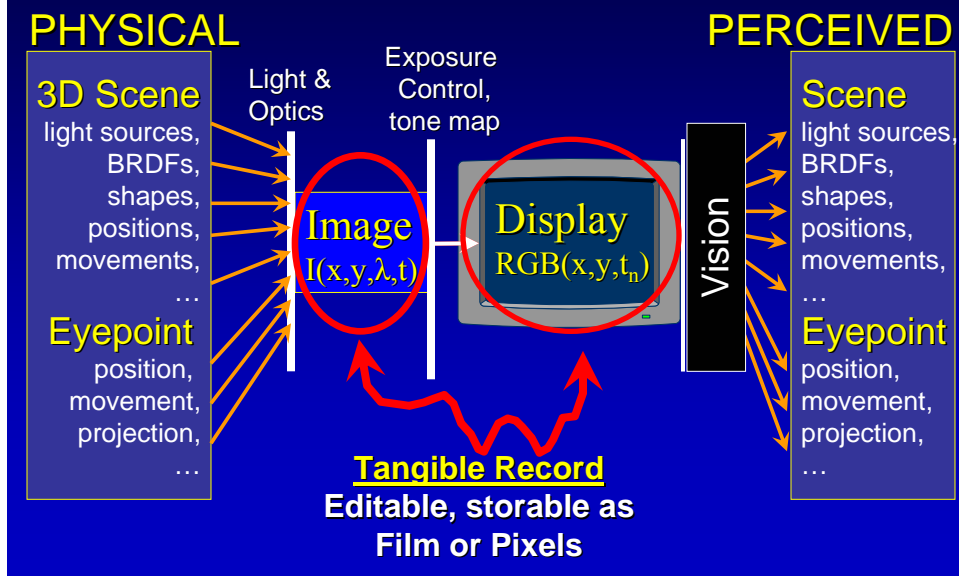
- 'Instantaneous' light measurement...
- Of focal plane image behind a lens.
- Reproduce those amounts of light;
- EXACT MATCH!

Implied:

"What we see is \cong
focal-plane intensities."
well, no...



Why we like Photography



What's Beyond Film-Like Photography?

Thought Experiment:

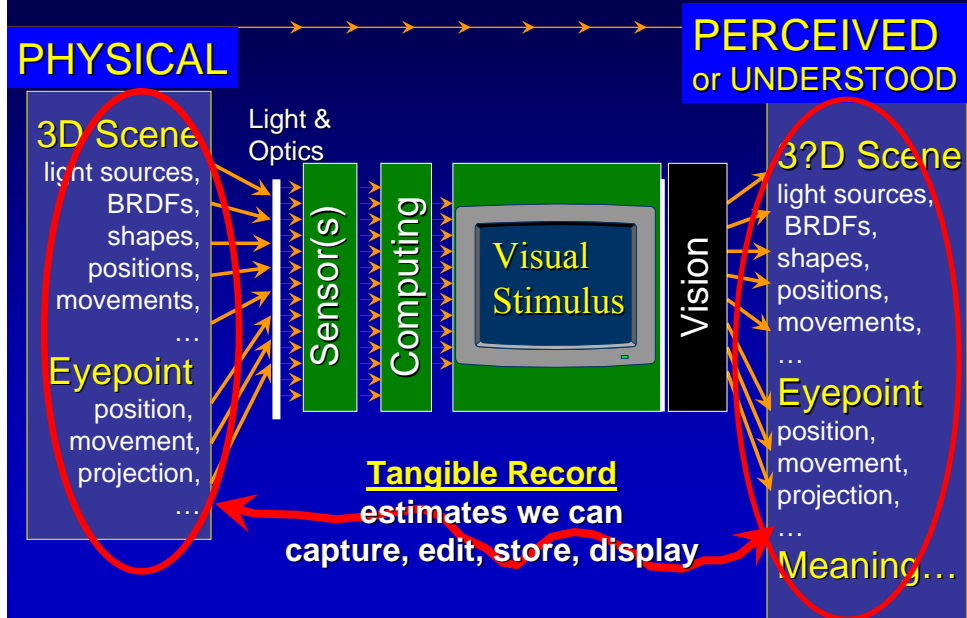
Side-by-side digital camera & film camera.

- COMPARE:
 - Digital Camera result.
 - Digitized (Scanned) Film result.

? Can we See more, Do more, Feel more?

? Has photography really *changed* yet?

Goals for a New Photography



Missing: **Dynamic Display, Interactive...**

What other ways
better reveal
appearance to
human viewers?
(Without direct shape
measurement?)

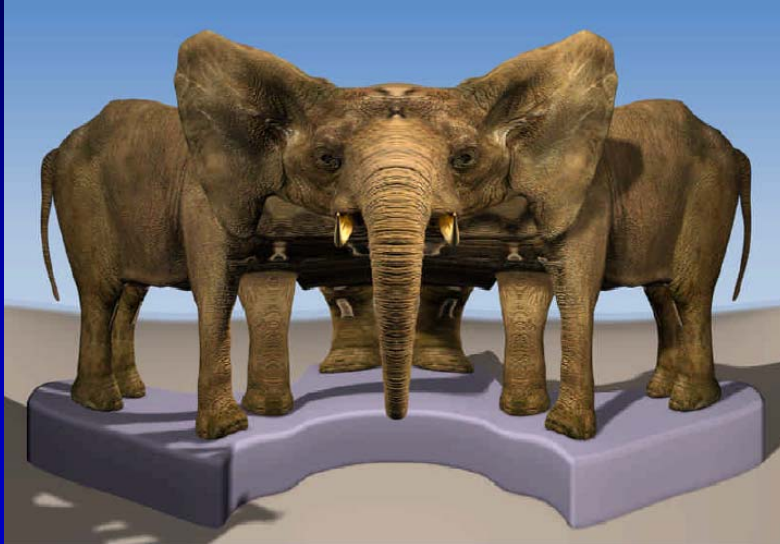
Can you understand
this shape better?



Time for space wiggle, Gasparini, 1998.

Missing: **More Revealing Sets of Rays**

"Multiple-Center-of-Projection Images" Rademacher, P, Bishop, G., SIGGRAPH '98



Taking Images versus Taking Pictures

Image: A copy of light intensities.

(Just one kind of picture, made by copying a scaled map of scene light intensities, as a lens might)

Visual Appearance: What we think we see.

(Consciously-available estimates of our surroundings, made from the light reaching our eyes)

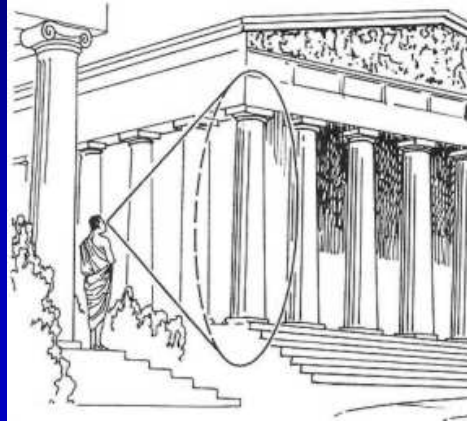
Picture: A 'container' for visual appearance.

(something we make to hold what we see, or what would like to see. Includes non-photorealistic drawings)

Photographic Signal: ~~Pixels~~ Rays

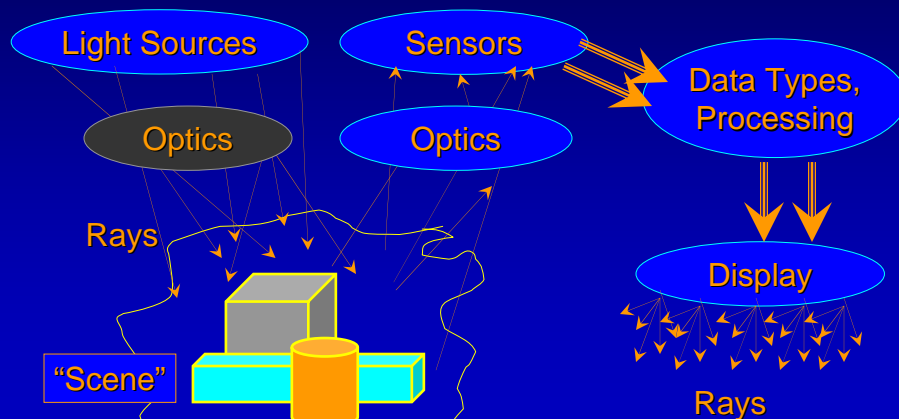
- Core ideas are ancient, simple, seem obvious:
 - Lighting: ray sources
 - Optics: ray bending/folding devices
 - Sensor: measure light
 - Processing: assess it
 - Display: reproduce it
- Ancient Greeks:
'eye rays' wipe the world to feel its contents...

<http://www.mlahanas.de/Greeks/Optics.htm>



The Photographic Signal Path

Computing can improve every component:



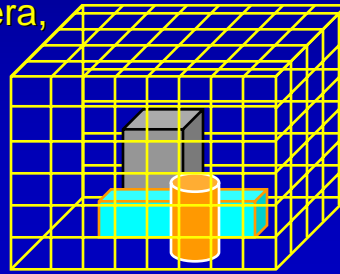
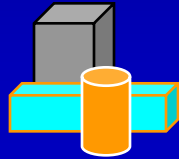
Review: How many Rays are there?

(Gortler et al. '96)

(Levoy et al. SIGG'96)

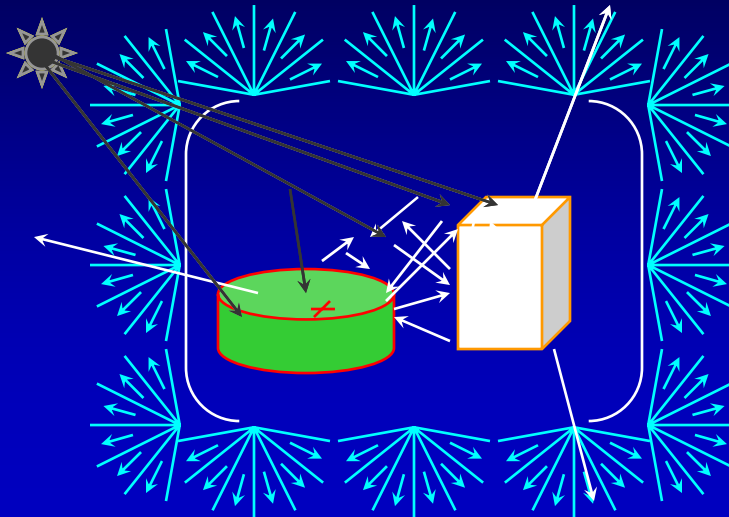
4-D set; infinitesimal members. Imagine:

- Convex Enclosure of a 3D scene
- Inward-facing ray camera at every surface point
- Pick the rays you need for ANY camera outside.
- 2D surface of cameras,
- + 2D ray set for each camera,
- 4D set of rays.



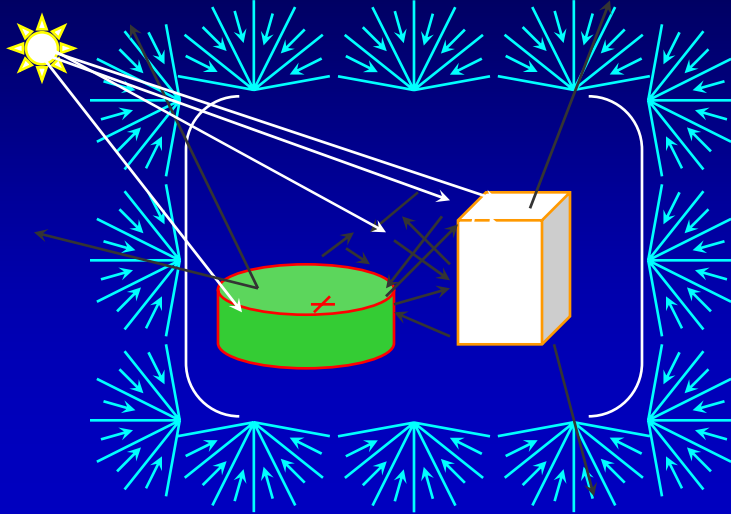
4-D Light Field / Lumigraph

Measure all the outgoing light rays.



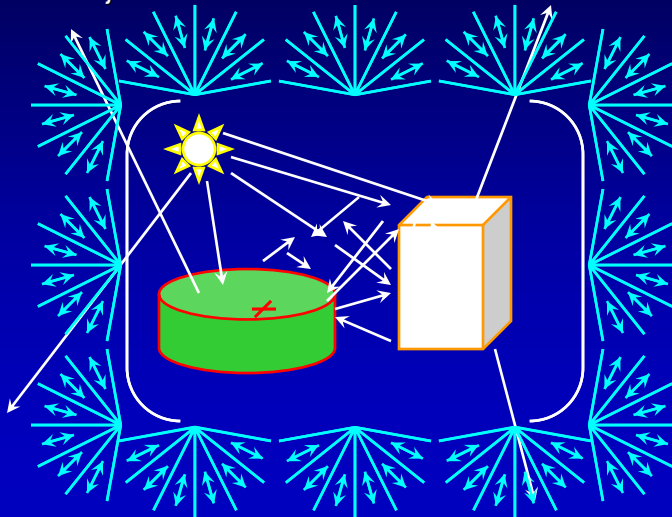
4-D Illumination Field

Same Idea: Measure all the incoming light rays



4D x 4D = 8-D Reflectance Field

Ratio: $R_{ij} = (\text{outgoing ray}_i) / (\text{incoming ray}_j)$



Is a 4-D Light Source Required?



[Debevec et al. 2000]



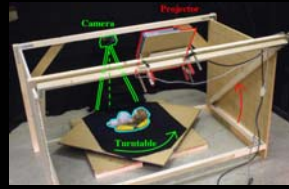
[Masselus et al. 2002]



[Matusik et al. 2002]



[Debevec et al. 2002]



[Masselus et al. 2003]

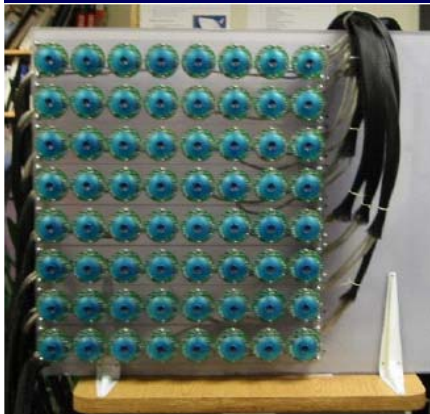


[Malzbender et al. 2002]

Is A 4D Camera Required?

e.g. MIT Dynamic Light Field Camera 2002

Is this the whole answer?

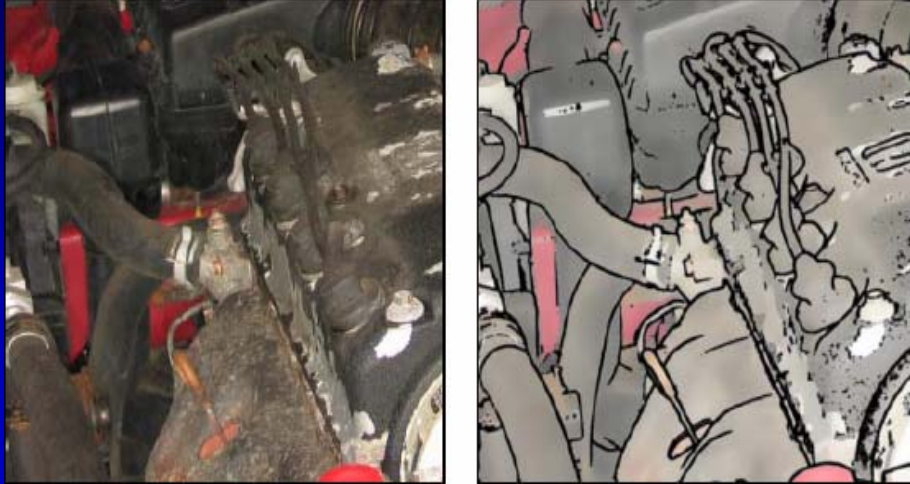


- Multiple dynamic Virtual Viewpoints
- Efficient Bandwidth usage: 'send only what you see'
- Yang, et al 2002
- 64 tightly packed commodity CMOS webcams
- 30 Hz, Scalable, Real-time:

or is it just "more film-like cameras, but now with computers!" ?

Or do Ray Changes Convey Appearance?

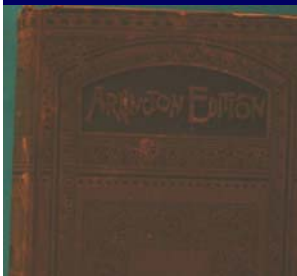
5 ray sets → explicit geometric occlusion boundaries



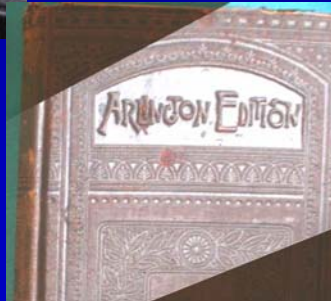
Ramesh Raskar, MERL, 2004

Or do Ray Changes Convey Appearance?

- These rays + all these rays give me...



- MANY more useful details I can examine...



Mild Viewing & Lighting Changes; Are these Enough?

Convincing visual appearance:
Is Accurate Depth really necessary?

a few good 2-D images may be enough...



"Image jets, Level Sets,
and Silhouettes"
Lance Williams,
talk at Stanford, 1998.

'The Ideal Photographic Signal'

I CLAIM IT IS:

~~All Rays?~~ ~~Some Rays?~~ **Changes in Rays**

Photographic ray space is vast and boring.
>8 dimensions: 4D view, 4D light, time, λ ,

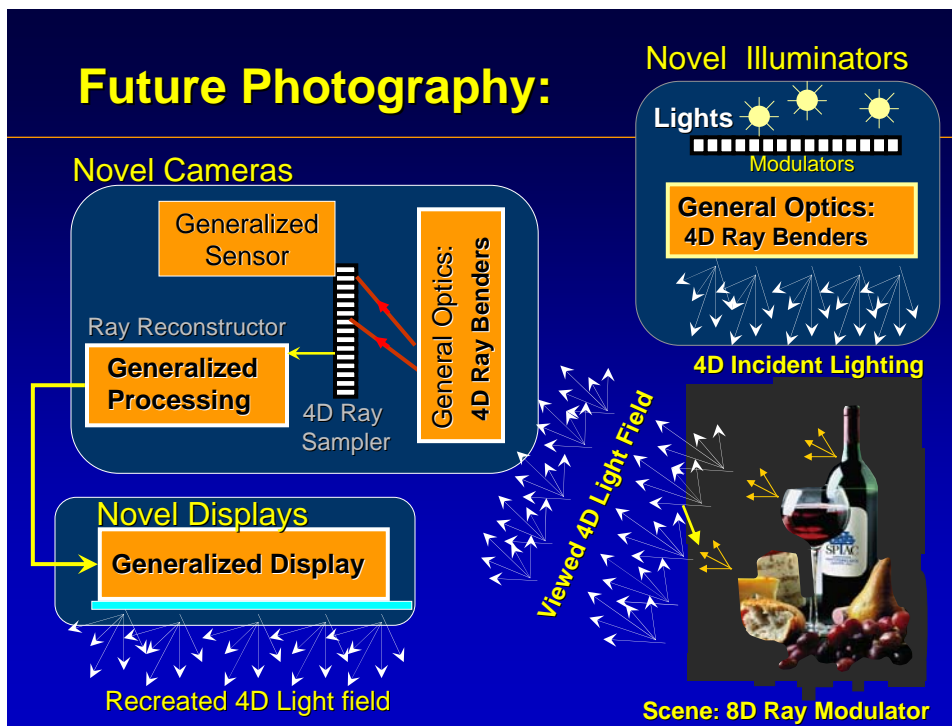
Gather only 'visually significant' ray changes

? What rays should we measure ?

? How should we combine them ?

? How should we display them ?

Future Photography:



Beyond 'Film-Like' Photography

'Computational Photography';

To make 'meaningful ray changes' tangible,

- Sensors can do more...
- Displays can do more...
- Light Sources can do more...
- Optics can do more...
- Ray Descriptors can do more...

**by applying low-cost storage,
computation, and control.**

Course : Computational Photography

A3: Film-like Photography: The Ray-Optics Model

Jack Tumblin
Northwestern University

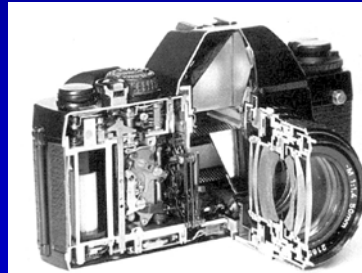
'Film-Like' Photography

Film Camera designs still dominate:

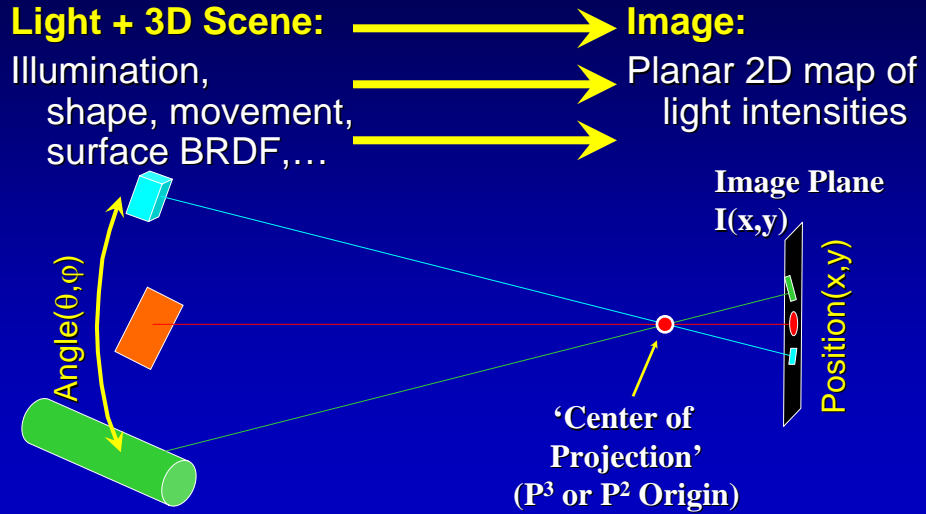
- 'Instantaneous' light measurement...
- Of focal plane image behind a lens.
- Reproduce those amounts of light;
- Display 'exactly matches' the scene

Implied:

"What we see is \cong
focal-plane intensities."
well, no...



'Film-Like Photography': Ray Model



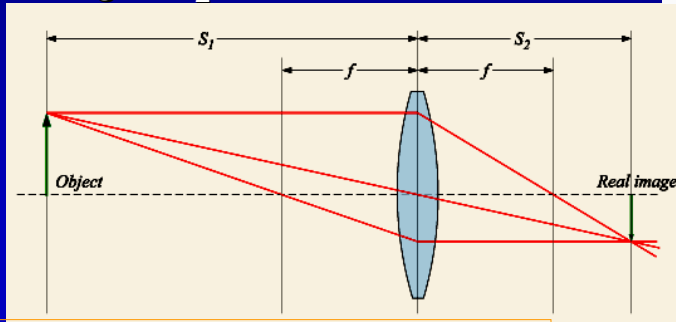
Film-Like Photography

• Lighting:	Ray Sources	(external)
• Scene:	Ray Modulator	(external)
• Optics:	Ray Benders	Thin Lens Approx.
• Sensors:	Ray Bundle Measurement	Sensor Irradiance $E(x,y)$
• Processing	Ray	Normalized $E(x,y)$
• Display	Recreate Rays	Normalized $E(x,y)$

Film-like Optics: 'Thin Lens Law'

- Focal length f : where parallel rays converge
- Object at distance S_1 forms image at S_2
- Focus at infinity: Adjust for $S_2=f$
Larger S_2 for closer focus

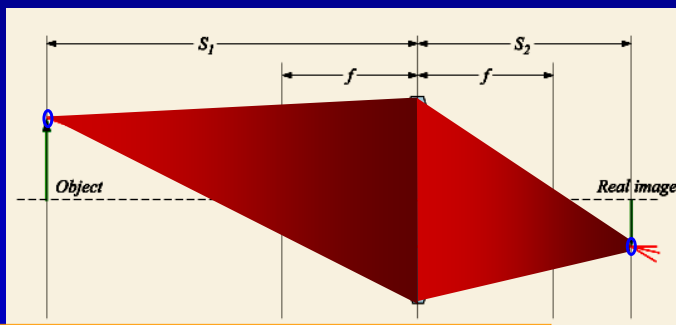
$$\frac{1}{S_1} + \frac{1}{S_2} = \frac{1}{f}$$



[http://www.nationmaster.com/encyclopedia/Lens-\(optics\)](http://www.nationmaster.com/encyclopedia/Lens-(optics))

Rays Are *Doubly* Differential

- Lens Systems: *approximate* rays with bundles
- Finite angle, not rays (*lens aperture*)
- Finite area, not points (*circle of confusion*)



[http://www.nationmaster.com/encyclopedia/Lens-\(optics\)](http://www.nationmaster.com/encyclopedia/Lens-(optics))

Ray BUNDLES approximate Rays

- Rays are doubly infinitesimal!
 - A 'ray' leaves a span of infinitesimal area 0^+
 - And covers a span of infinitesimal directions 0^+
- Ray Bundles:
 - Finite, measurable power from combined rays
 - A finite span of SOLID ANGLE, and
 - A finite span of SURFACE AREA

Ray BUNDLES approximate Rays

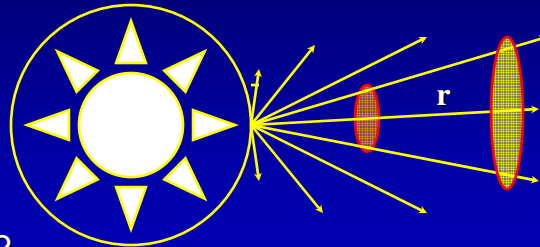
- Rays are doubly infinitesimal!
 - A 'ray' Leaves 0^+ area in 0^+ directions

EXAMPLE:

- Power from 1 point on a spherical lamp?

$$0^+ = 60\text{Watts} / (\infty \text{ points})$$

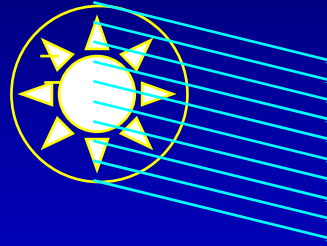
- BUT has a finite, measurable ratio: (flux/area)
(60Watts / 30 cm² area) = 2 W / cm²



Ray BUNDLES approximate Rays

- Rays are doubly infinitesimal!

A 'ray' Leaves 0^+ area in 0^+ directions



EXAMPLE:

- Power from spherical lamp in just 1 direction?

$$0^+ = (60\text{Watts} / (\infty \text{ directions}))$$

– BUT has finite ratio:

$$(60\text{Watts} / 4\pi \text{ steradians}) = 4.77 \text{ W} / \text{cm}^2$$

Ray Measurement: Radiance L

- Incoming light directions form hemisphere Ω ;
Ray == one point on the hemisphere

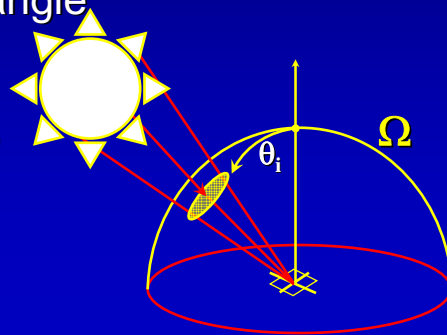
THUS

'Incident Rays' measured in Radiance Units L:

Irradiance per unit solid angle

$$L = (\text{watts} / \text{area}) / \text{sr}$$

(sr = steradians; solid angle;
= surface area on unit sphere)



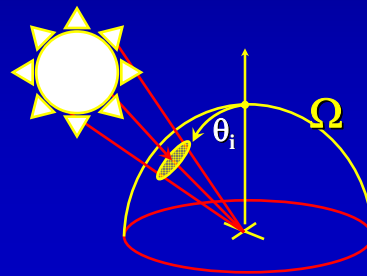
Ray 'Bundles'

- Rays have no surface area (just a point)
- Rays have no solid angle (just a point)

THUS:

- A Ray carries *infinitesimal* power (0^+ Watts).
- Only BUNDLES of rays are measurable!

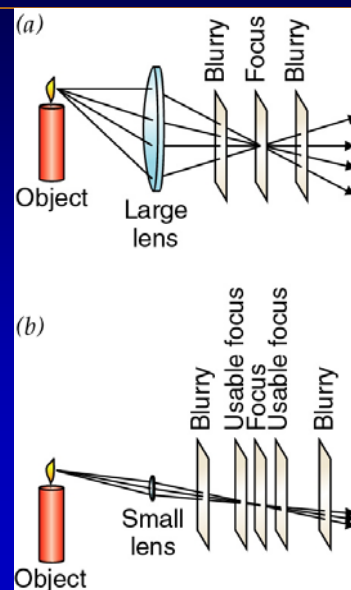
?How can estimate the 'Photographic Signal' when we can't directly measure it?



Lens Flaws: Depth of Focus

For the *same* focal length:

- **Larger lens**
 - Gathers a wider ray bundle:
 - More light: brighter image
 - Shallower depth-of-focus
- **Smaller lens**
 - Gathers a narrower ray bundle:
 - Less light: dimmer image
 - Deeper depth-of-focus

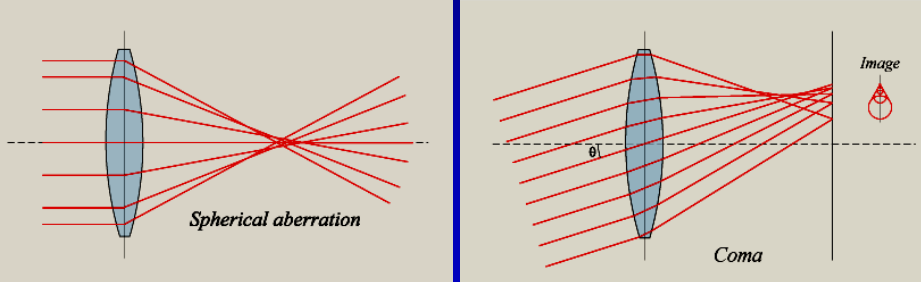


Lens Flaws: Geometric Aberration

- **Aberrations:**

Real lenses don't converge rays perfectly

- **Spherical:** edge rays \neq center rays
- **Coma:** diagonal rays focus deeper at edge



[http://www.nationmaster.com/encyclopedia/Lens-\(optics\)](http://www.nationmaster.com/encyclopedia/Lens-(optics))

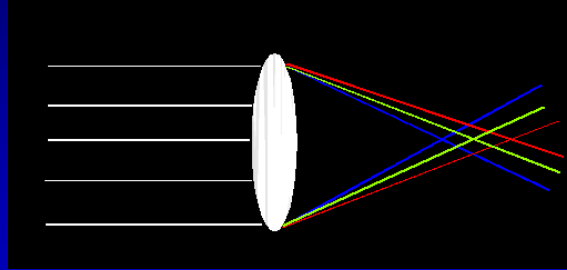
Radial Distortion (e.g. 'Barrel' and 'pin-cushion')

straight lines curve around the image center



Lens Flaws: Chromatic Aberration

- Dispersion: wavelength-dependent refractive index
 - (enables prism to spread white light beam into rainbow)
- Modifies ray-bending and lens focal length: $f(\lambda)$



- color fringes near edges of image
- Corrections: add 'doublet' lens of flint glass, etc.

<http://www.swgc.mun.ca/physics/physics/opticalbench.html>

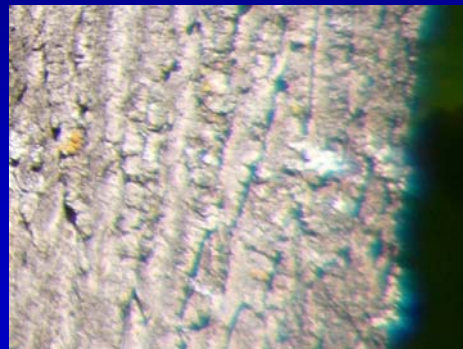
Lens Flaws: Chromatic Aberration

- **Lens Design Fix:** Multi-element lenses
Complex, expensive, many tradeoffs!
- **Computed Fix:** Geometric warp for R,G,B.

Near Lens Center



Near Lens Outer Edge

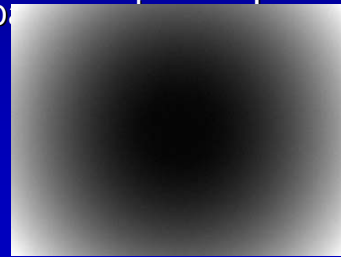


Lens Flaws: Intensity Aberrations

Image 'Vignette': bright at center, dark at edges.

Several compounded causes:

- Internal shadowing—angle-dependent Ray bundles
- Longer paths for off-axis Rays; Dark Glass
- Planar detector: outer pixels spread over larger area
- Compensation:
 - Use anti-vignetting filters, (darkest at center)
 - OR Position-dependent pixel-detector sensitivity.



<http://homepage.ntlworld.com/j.houghton/vignette.htm>

Polarization

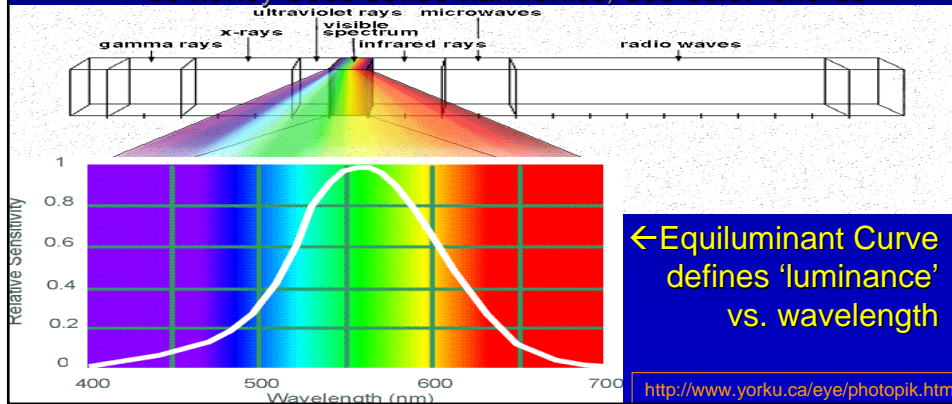


Sunlit haze is often strongly polarized. Polarization filter yields much richer sky colors



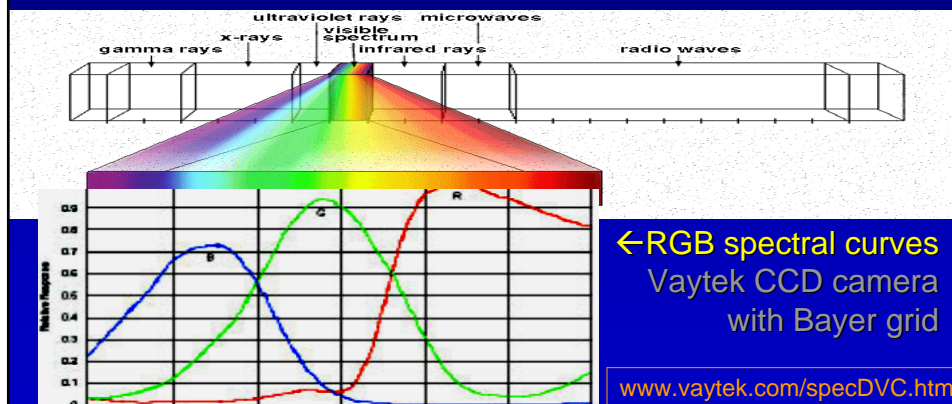
Film-like Color Sensing

- Visible Light: narrow band of emag spectrum
- $\lambda \approx 400-700 \text{ nm}$ (nm = 10⁻⁹ meter wavelength)
- (humans:<1 octave \leftrightarrow honey bees: 3-4 'octaves'
do honey bees sense harmonics, see color 'chords' ?



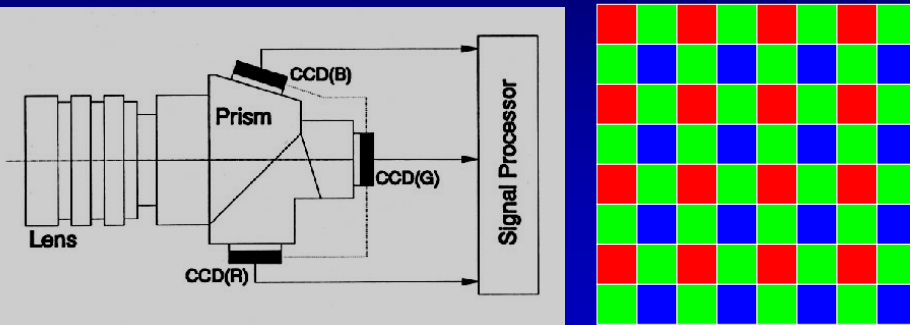
Film-like Color Sensing

- Visible Light: narrow band of emag spectrum
- $\lambda \approx 400-700 \text{ nm}$ (nm = 10⁻⁹ meter wavelength)
- At least 3 spectral bands required (e.g. R,G,B)



Color Sensing

- 3-chip vs. 1-chip: quality vs. cost



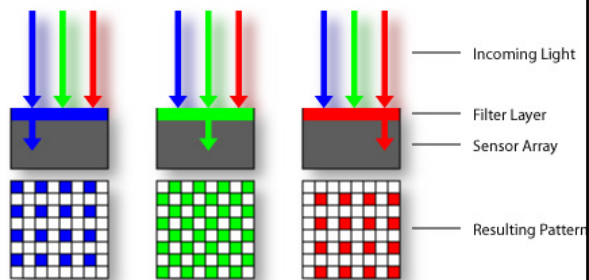
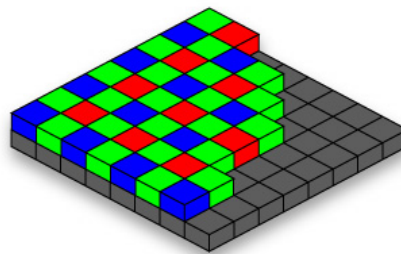
<http://www.cooldictionary.com/words/Bayer-filter.wikipedia>

Practical Color Sensing: Bayer Grid



- Estimate RGB at 'G' cells from neighboring values

<http://www.cooldictionary.com/words/Bayer-filter.wikipedia>



Conclusions

- Film-like photography methods limit digital photography to film-like results or less.
- Broaden, unlock our views of photography:
- 4D, 8D, even 10D Ray Space holds the photographic signal. Look for new solutions by creating, gathering, processing RAYS, not focal-plane intensities.
- Choose the best, most expressive sets of rays, THEN find the best way to measure them.

Useful links:

- Interactive Thin Lens Demo
(or search 'physlet optical bench')
www.swgc.mun.ca/physics/physlets/opticalbench.html
- For more about color:
- Prev. SIGGRAPH courses (Stone et al.)
 - Good: www.cs.rit.edu/~ncs/color/a_spectr.html
 - Good: www.colourware.co.uk/cpfaq.htm
 - Good: www.yorku.ca/eye/toc.htm

Image Processing and Reconstructions Tools



Ramesh Raskar
Mitsubishi Electric Research Labs
Cambridge, MA

Image Tools

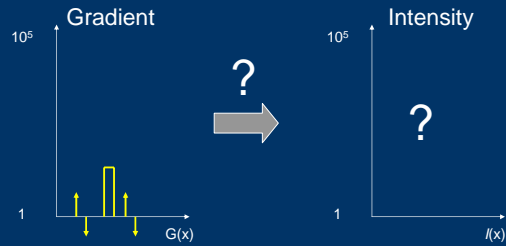
- Gradient domain operations,
 - Applications in tone mapping, fusion and matting
- Graph cuts,
 - Applications in segmentation and mosaicing
- Bilateral and Trilateral filters,
 - Applications in image enhancement

Intensity Gradient in 1D



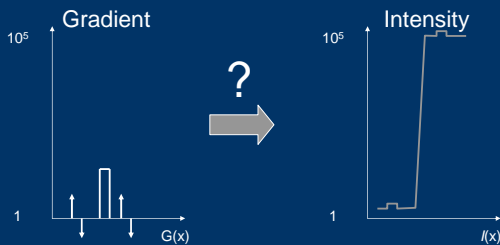
Gradient at x ,
 $G(x) = I(x+1) - I(x)$
 Forward Difference

Reconstruction from Gradients



For n intensity values, about n gradients

Reconstruction from Gradients



1D Integration

$I(x) = I(x-1) + G(x)$
 Cumulative sum

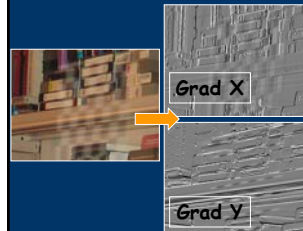
Intensity Gradient in 2D

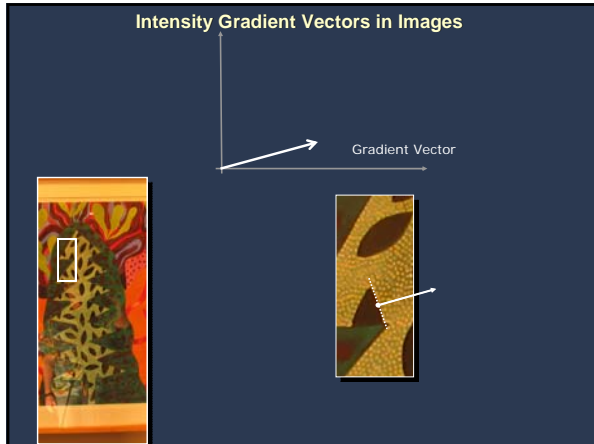
Gradient at x,y as Forward Differences

$$G_x(x,y) = I(x+1, y) - I(x,y)$$

$$G_y(x,y) = I(x, y+1) - I(x,y)$$

$$G(x,y) = (G_x, G_y)$$





Reconstruction from Gradients

Given $G(x,y) = (G_x, G_y)$
 How to compute $I(x,y)$ for the image ?
 For n^2 image pixels, $2n^2$ gradients !

Intensity Gradient in 2D

Recovering Original Image

Intensity Gradient Manipulation

Recovering Manipulated Image

Intensity Gradient Manipulation

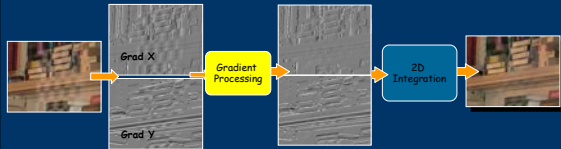
Recovering Manipulated Image

Intensity Gradient Manipulation

Recovering Manipulated Image

Intensity Gradient Manipulation

A Common Pipeline



Reconstruction from Gradients

- Look for image I with gradient closest to G in the least squares sense.
- I minimizes the integral: $\iint F(\nabla I, G) dx dy$

$$F(\nabla I, G) = \|\nabla I - G\|^2 = \left(\frac{\partial I}{\partial x} - G_x\right)^2 + \left(\frac{\partial I}{\partial y} - G_y\right)^2$$

Euler-Lagrange Equation

- I must satisfy: $\frac{\partial F}{\partial I} - \frac{d}{dx} \frac{\partial F}{\partial I_x} - \frac{d}{dy} \frac{\partial F}{\partial I_y} = 0$

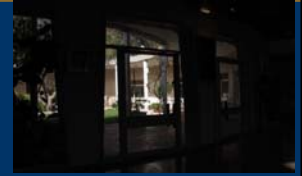
- Substituting F we get:

$$2\left(\frac{\partial^2 I}{\partial x^2} - \frac{\partial G_x}{\partial x}\right) + 2\left(\frac{\partial^2 I}{\partial y^2} - \frac{\partial G_y}{\partial y}\right)$$

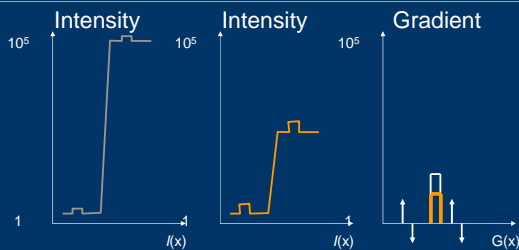
$$\nabla^2 I = \text{div } G$$

Application: Compressing Dynamic Range

How could you put all this information into one Image ?



Attenuate High Gradients

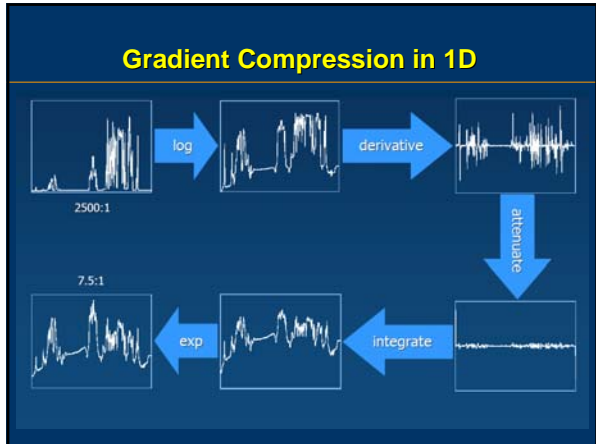


Maintain local detail at the cost of global range

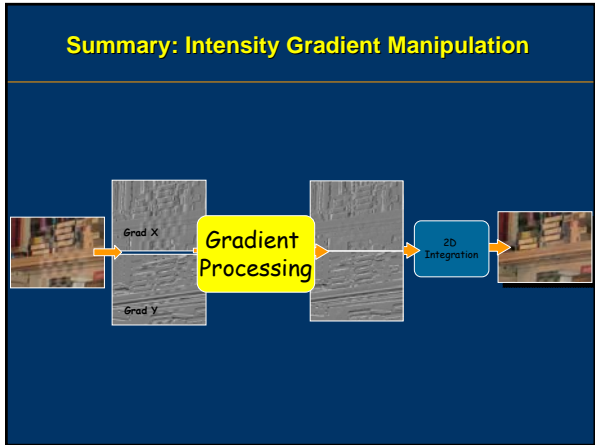
Fattal et al Siggraph 2002

Basic Assumptions

- The eye responds more to local intensity differences (ratios) than global illumination
- A HDR image must have some large magnitude gradients
- Fine details consist only of smaller magnitude gradients



- ### Basic Method
- Take the log of the luminances
 - Calculate the gradient at each point
 - Scale the magnitudes of the gradients with a progressive scaling function (Large magnitudes are scaled down more than small magnitudes)
 - Re-integrate the gradients and invert the log to get the final image



Graph and Images

Credits: Jianbo Shi

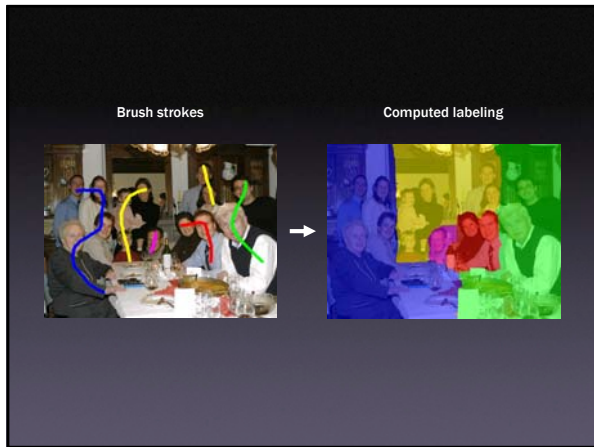
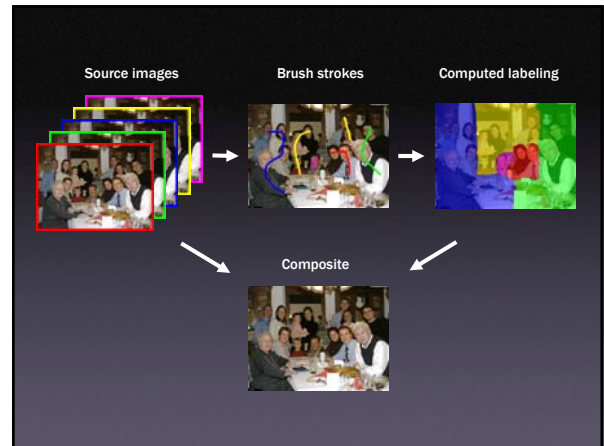




Agrawala et al, Digital Photomontage, Siggraph 2004



Agrawala et al, Digital Photomontage, Siggraph 2004



Graph Based Image Segmentation

Segmentation = Graph partition

$G = \{V, E\}$

- V: graph node \longleftrightarrow Image pixel
- E: edges connection nodes \longleftrightarrow Link to neighboring pixels
- W_{ij} : Edge weight \longleftrightarrow Pixel similarity

Minimum Cost Cuts in a graph

Cut: Set of edges whose removal makes a graph disconnected

$S_{i,j}$: Similarity between pixel i and pixel j

Cost of a cut, $cut(A, \bar{A}) = \sum_{i \in A, j \in \bar{A}} S_{i,j}$

A \bar{A}

Problem with min cuts



Min. cuts favors isolated clusters

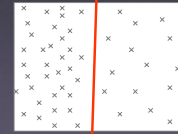
Normalize cuts in a graph

Ncut = balanced cut



$$Ncut(A, B) = cut(A, B) \left(\frac{1}{vol(A)} + \frac{1}{vol(B)} \right)$$

NP-Hard!



Graph Cuts for Segmentation and Mosaicing

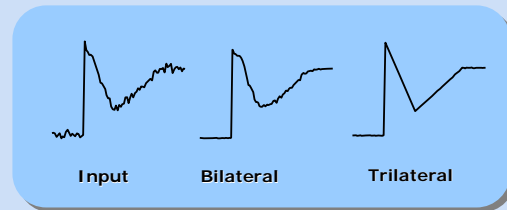
Cut ~ String on a height field

Brush strokes

Computed labeling



Bilateral and Trilateral Filter



Bilateral and Trilateral Filtering

Outline

- Unilateral filtering
 - Smoothing using filtering
- Bilateral filtering
 - Strength and 3 weaknesses
- Trilateral filtering
 - Key ideas
- Application in tone mapping
 - Detail preserving contrast reduction

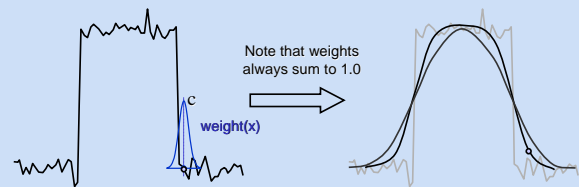
'Unilateral' Filter

Traditional, linear, FIR filters

Key Idea: Convolution

- Output(x) = local weighted avg. of inputs.
- Weights vary within a 'window' of nearby x

Smooths away details, **BUT** blurs result



'Unilateral' Filter

Forces a Tradeoff:

- Broad window: better detail removal
- OR --
- Narrow window: better large structure

But we want BOTH...

Bilateral Filter

A 2-D filter window: weights vary with intensity
[Tomasi&Manduchi1998]
 Further Analysis: [Black99] [Elad02] [Durand&Dorsey02], ...

Range
 $f(x)$
 Domain
 x

c: distance from input (**domain** of input)
 s: difference from input (**range** of input)

Bilateral Filter

Range
 $f(x)$
 Domain
 x

2 Gaussian Weights:
 product =
 ellisoidal footprint

Bilateral Filter

Why it works: graceful segmentation

- Filtering in one region ignores filtering in another
- Gaussian **s** acts as a 'filtered region' finder

Range
 $f(x)$
 Domain
 x

Bilateral Filter: Strengths

Piecewise smooth result

- averages local small details, ignores outliers
- preserves steps, large-scale ramps, and curves,...

• Equivalent to anisotropic diffusion and robust statistics
[Black98, Elad02, Durand02]

• Simple & Fast (esp. w/ [Durand02] FFT-based speedup)

Bilateral Filter: 3 Difficulties

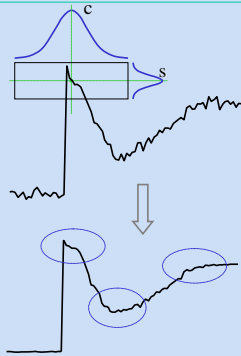
- Poor Smoothing in High Gradient Regions
- Smooths and blunts cliffs, valleys & ridges
- Can combine disjoint signal regions

Range
 $f(x)$
 Domain
 x

Output at \oplus is average of a tiny region

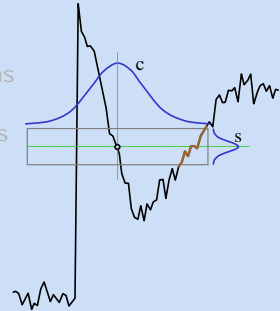
Bilateral Filter: 3 Difficulties

- Poor Smoothing in High Gradient Regions
- Smooths and blunts cliffs, valleys & ridges
- Can combine disjoint signal regions

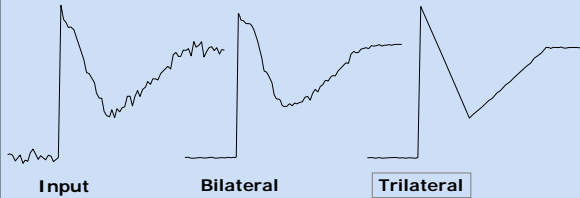


Bilateral Filter: 3 Difficulties

- Poor Smoothing in High Gradient Regions
- Smooths and blunts cliffs, valleys & ridges
- Disjoint regions can blend together



New Solution: "Trilateral" Filter

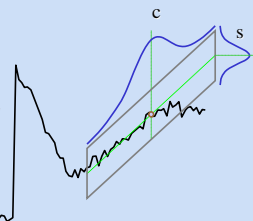


- Keep best features of bilateral, adds more
- Corner sharpening resembles PDE shocks
- User sets 1 parameter (good defaults for 7 internals)

Bilateral → Trilateral Filter

Three Key Ideas:

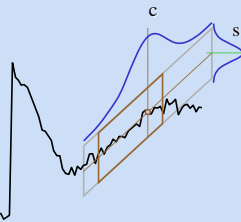
- **Tilt** the filter window according to bilaterally-smoothed gradients
- **Limit** the filter window to connected regions of similar smoothed gradient.
- **Adjust** Parameters from measurements of the windowed signal



Bilateral → Trilateral Filter

Key Ideas:

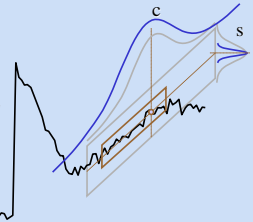
- **Tilt** the filter window according to bilaterally-smoothed gradients
- **Limit** the filter window to connected regions of similar smoothed gradient.
- **Adjust** Parameters from measurements of the windowed signal



Bilateral → Trilateral Filter

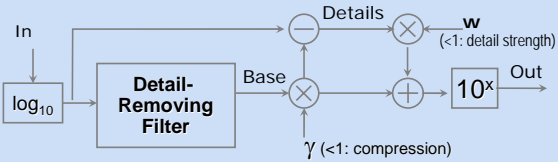
Key Ideas:

- **Tilt** the filter window according to bilaterally-smoothed gradients
- **Limit** the filter window to connected regions of similar smoothed gradient.
- **Adjust** Parameters from measurements of the windowed signal



Application: Tone Mapping

- Filter **removes** details.
- Goal: Detail-Preserving Contrast Reduction
 - in log domain, difference == contrast
 - remove details, compress contrast, replace details

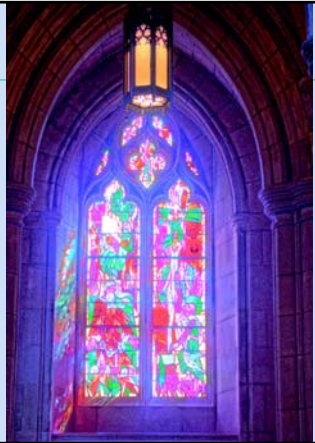


More Trilateral Results

Comparable to
Gradient Attenuation
[Fattal et al 2002]

Similar to LCIS
[Tumblin&Turk' 99]
[Bertozzi'03]

Simple, Robust



B1: Reconstruction



Ramesh Raskar

Mitsubishi Electric Research Labs

Jack Tumblin

Northwestern University

Course WebPage :

<http://www.merl.com/people/raskar/photo>

Image Fusion and Reconstruction

- Epsilon Photography
 - Vary time, view
 - Vary focus, exposure polarization, illumination
 - Better than any one photo
- Achieve effects via multi-image fusion
- Understand computer vision methods
- Exploit lighting

Time-Lapse

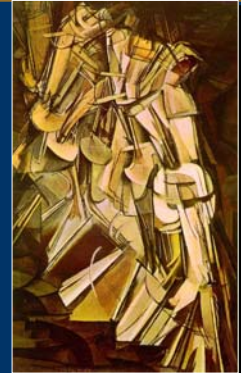
- Duchamp
 - Nude Descending a Staircase



Time-Lapse



Richard Hundley 2001



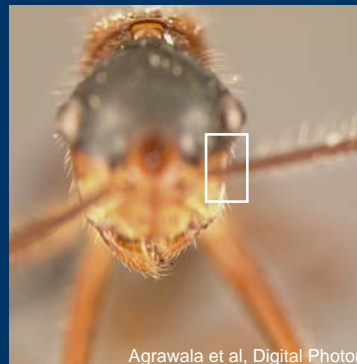
Shape Time Photography



Freeman et al 2003

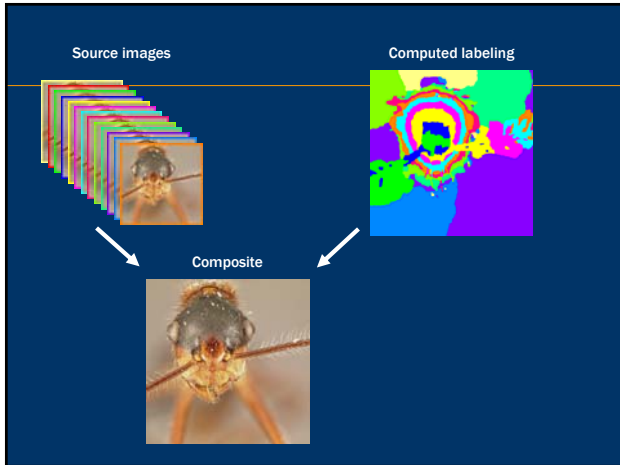


Varying Focus: Extended depth-of-field

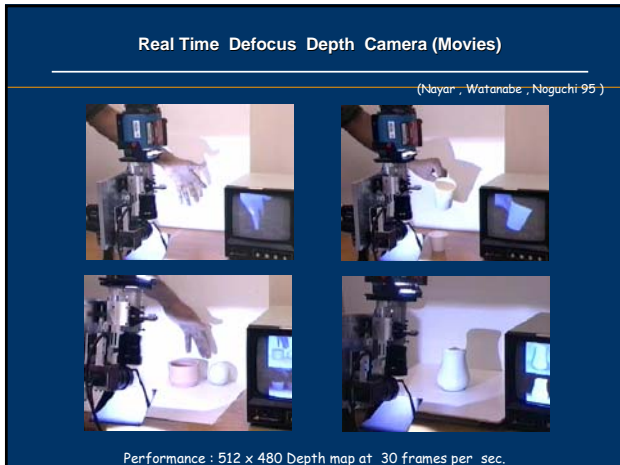
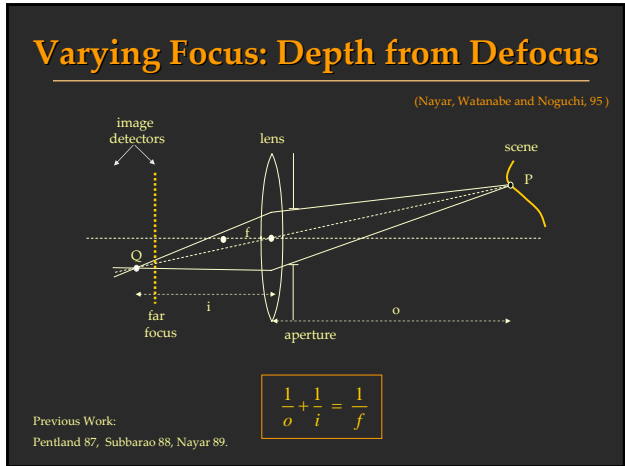
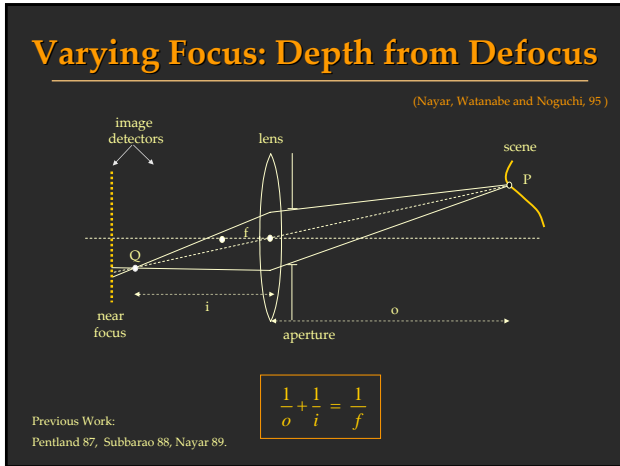


Agrawala et al, Digital Photomontage, Siggraph 2004





- ## Computer Vision Techniques
- Photometric Stereo
 - Varying light source positions
 - Estimate surface normal from shading
 - Diffuse objects: minimum 3 lights
 - Depth from Defocus
 - Varying focus
 - Defogging
 - Varying time and polarization



Varying Polarization

Yoav Y. Schechner, Nir Karpel 2005

Best polarization state

Worst polarization state

Best polarization state Recovered image

[Left] The raw images taken through a polarizer. [Right] White-balanced results: The recovered image is much clearer, especially at distant objects, than the raw image

Varying Polarization

- Schechner, Narasimhan, Nayar
- Instant dehazing of images using polarization

Best polarization state

Worst polarization state

Dehazed image

Varying Wavelength: Multispectral Fusion

Vegetation Mapping of the Forest

SAR + Optical Landsat =

- Mountain forest
- Mixed forest
- Okoume forest
- Grassland Savanna
- Fern Savanna
- Burnt Savanna

Varying IR Wavelength Image Fusion

NIR SWIR LWIR

Uniform fusion across image

Adaptive fusion by sub region

Non-photorealistic Camera: Depth Edge Detection and Stylized Rendering using Multi-Flash Imaging

Ramesh Raskar, Karhan Tan, Rogerio Feris, Jingyi Yu, Matthew Turk

Mitsubishi Electric Research Labs (MERL), Cambridge, MA

U of California at Santa Barbara

U of North Carolina at Chapel Hill

Left Flash Right Flash Max

Top Flash Bottom Flash Depth Edge Confidence Map

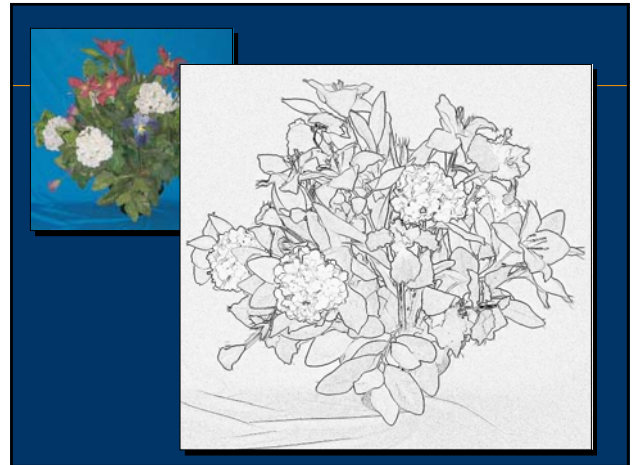


Image Fusion and Reconstruction

- Epsilon Photography
 - Vary focus, exposure polarization, illumination
 - Vary time, view
 - Better than any one photo
- Achieve effects via multi-image fusion
- Understand computer vision methods
- Exploit lighting

Improving FILM-LIKE Camera Performance

What would make it 'perfect' ?

- Dynamic Range

Film-Style Camera: Dynamic Range Limits

Under-Exposure

- Highlight details: Captured
- Shadow details: Lost

Over-Exposure

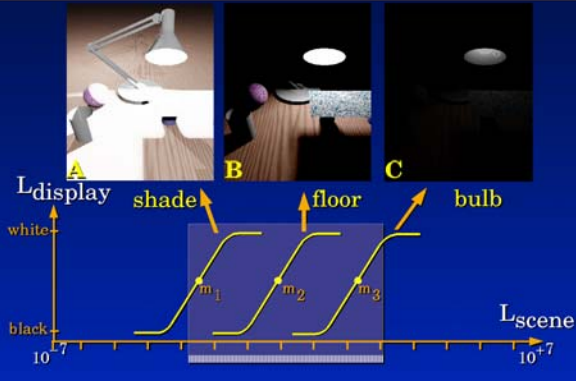
- Highlight details: Lost
- Shadow details: Captured

Problem: Map Scene to Display

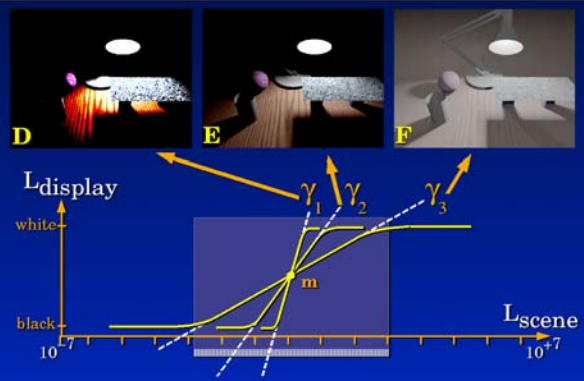
Domain of Human Vision:
from $\sim 10^{-6}$ to $\sim 10^{+8}$ cd/m²

Range of Typical Displays:
from ~ 1 to ~ 100 cd/m²

Backgnd: Global Scale: m

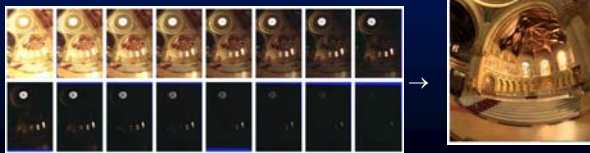


Backgnd: Global Contrast: γ

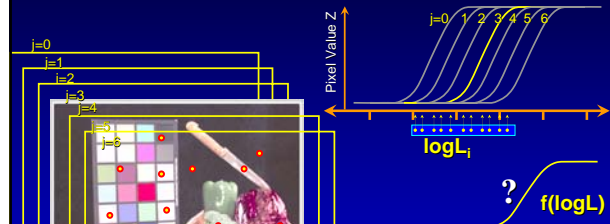


High dynamic range capture (HDR)

- overcomes one of photography's key limitations
 - negative film = 250:1 (8 stops)
 - paper prints = 50:1
 - [Debevec97] = 250,000:1 (18 stops)
 - hot topic at recent SIGGRAPHs

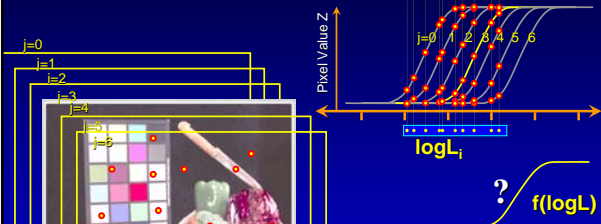


Debevec'97 (see www.HDRshop.com)



STEP 1:
 --number the images 'i',
 --pick fixed spots (x_i, y_i)
 that sample scene's
 radiance values $\log L_i$ well:

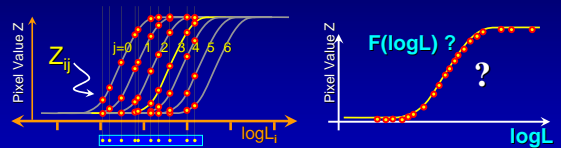
Debevec'97 (see www.HDRshop.com)



STEP 2:
 --Collect pixel values Z_{ij}
 (from image i , location j)
 --(All of them sample the
 response curve $f(\log L)$...)

Debevec'97 (see www.HDRshop.com)

- Use the multiple samples to reconstruct the response curve;



- Then use the inverse response curve to reconstruct the intensities that caused the responses

HDR Direct Sensing?

- An open problem! (esp. for video...)
- A direct (and expensive) solution:
 - Flying Spot Radiometer: brute force instrument, costly, slow, delicate
- Some Other Novel Image Sensors:
 - line-scan cameras (e.g. Spheron: multi-detector)
 - logarithmic CMOS circuits (e.g. Fraunhofer Inst)
 - Self-resetting pixels (e.g. sMaL /Cypress Semi)
 - Gradient detectors (CVPR 2005 Tumblin, Raskar et al)

HDR From Multiple Measurements



Captured Images



Computed Image

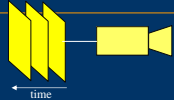
(Courtesy Shree Nayar, Tomoo Mitsunaga 99)

Ginosar et al 92, Burt & Kolczynski 93, Madden 93, Tsai 94, Saito 95, Mann & Picard 95, Debevec & Malik 97, Mitsunaga & Nayar 99, Robertson et al. 99, Kang et al. 03

MANY ways to make multiple exposure measurements

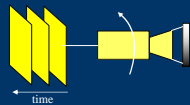
Sequential Exposure Change:

Ginosar et al 92, Burt & Kolczynski 93, Madden 93, Tsai 94, Saito 95, Mann 95, Debevec & Malik 97, Mitsunaga & Nayar 99, Robertson et al. 99, Kang et al. 03



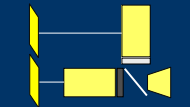
Mosaicing with Spatially Varying Filter:

Schechner and Nayar 01, Aggarwal and Ahuja 01



Multiple Image Detectors:

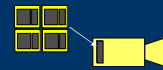
Doi et al. 86, Saito 95, Saito 96, Kimura 98, Ikeda 98, Aggarwal & Ahuja 01, ...



MANY ways to make multiple exposure measurements

Multiple Sensor Elements in a Pixel:

Handy 86, Wen 89, Murakoshi 94, Komishi et al. 95, Hamazaki 96, Street 98



Assorted Pixels:

Generalized Bayer Grid:
Trade resolution for multiple exposure,color

Nayar and Mitsunaga 00, Nayar and Narasimhan 02



Assorted-Pixel Camera Prototype

(Courtesy : Sony Kihara Research Lab)

Digital Still Camera

Camera with Assorted Pixels

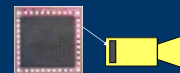


Another Approach: Locally Adjusted Sensor Sensitivity

Computational Pixels:

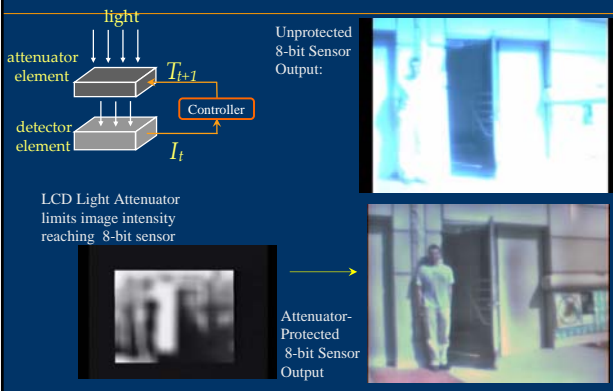
(pixel sensitivity set by its illumination)

Brijovic & Kanade 96, Ginosar & Gussin 97, Serafini & Sodini 00



NO GRADIENT CAMERA: RAMESH HAS IT

Sensor: LCD Adaptive Light Attenuator



Improving FILM-LIKE Camera Performance

- Vary Focus Point-by-Point

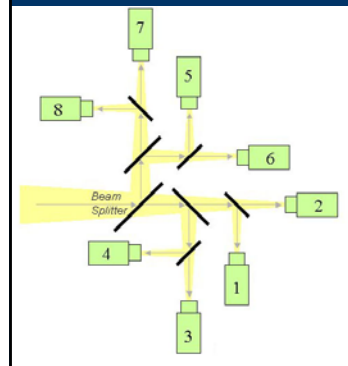
High depth-of-field

Levoy et al., SIGG2005

- adjacent views use different focus settings
- for each pixel, select sharpest view



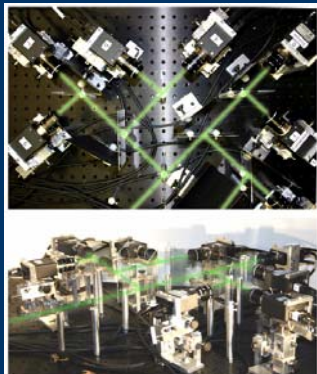
Single-Axis Multi-Parameter Camera (SAMP)



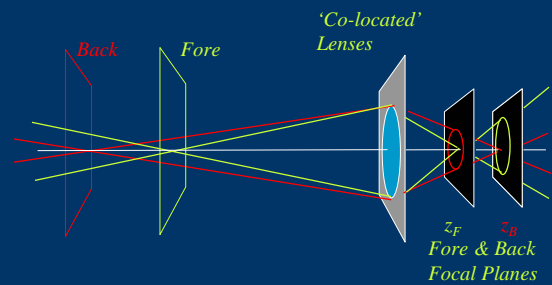
Idea:
Cameras + Beamsplitters
Place MANY (8) cameras
at same virtual location

2005: Morgan McGuire (Brown),
Wojciech Matusik (MERL),
Hanspeter Pfister (MERL),
Fredo Durand (MIT),
John Hughes (Brown),
Shree Nayar (Columbia)

SAMP Prototype System (Layout)

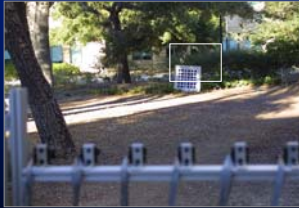


Multiple Simultaneous Focus Depths



Strongly desired in microscopy, too: see
<http://www.micrographia.com/articlz/artmicgr/mcspec/mcs0100.htm>

Long-range synthetic aperture photography



Focus Adjustment: Sum of Bundles



Improving FILM-LIKE Camera Performance

- Field of view vs. Resolution?

Are we done?

- Almost EVERY digital camera has panoramic stitching.

No; Much more is possible:

A tiled camera array



- 12×8 array of VGA cameras
- abutted: 7680×3840 pixels
- overlapped 50%: half of this
- total field of view = 29° wide
- seamless mosaicing isn't hard
- cameras individually metered
- Approx same center-of-proj.

Tiled panoramic image (before geometric or color calibration)



Tiled panoramic image (after calibration and blending)



same exposure in all cameras

1/60	1/60
1/60	1/60

individually metered

1/120	1/60
1/60	1/30

same and overlapped 50%

Improving FILM-LIKE Camera Performance

- Exposure time and Frame rate

High Speed Video

Say you want 120 frame per second (fps) video.

- You could get one camera that runs at 120 fps
- Or...

High Speed Video

Say you want 120 frame per second (fps) video.

- You could get one camera that runs at 120 fps
- Or... get 4 cameras running at 30 fps.

52 Camera Cluster, 1560 FPS

Levoy et al., SIGG2005

Conclusions

- Multiple measurements:
 - Multi-camera, multi-sensor, multi-optics, multi-lighting
- Intrinsic limits seem to require it
 - lens diffraction limits, noise, available light power.
- Are we eligible for Moore's law?
Or will lens making, mechanics limit us?

Computational Cameras: Convergence of Optics and Software

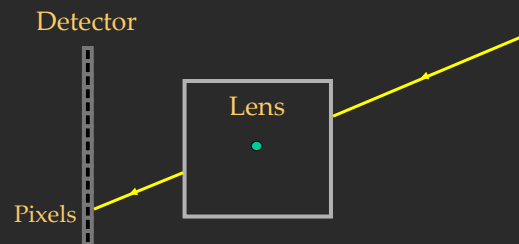
Shree K. Nayar

Computer Science
Columbia University

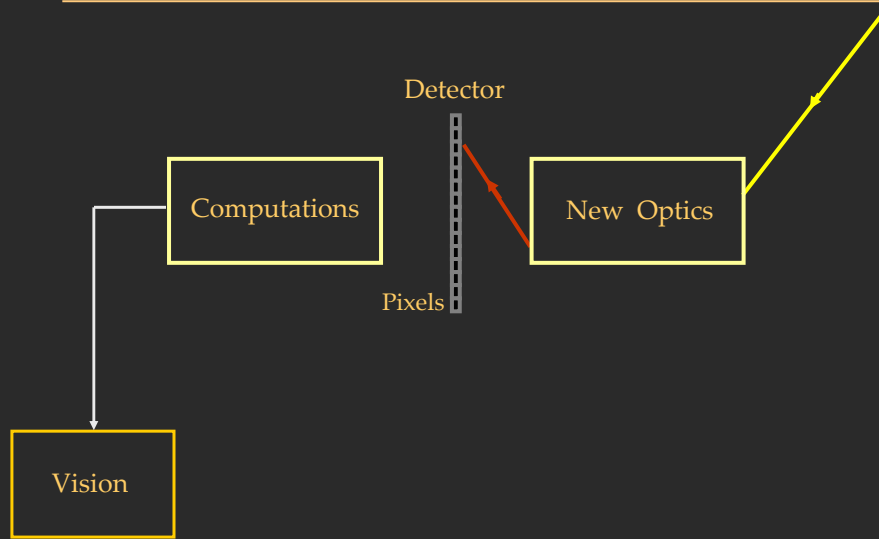
<http://www.cs.columbia.edu/CAVE/>

Support:
NSF, ONR, Packard Foundation
T. C. Chang Endowed Chair

Traditional Camera



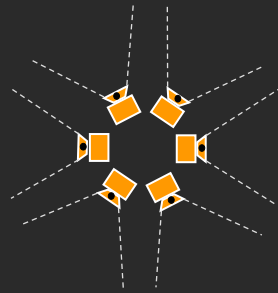
Computational Cameras



© Shree Nayar, Columbia University

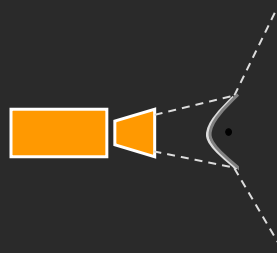
Wide Angle Imaging

Multiple Cameras



Examples: Disney 55, McCutchen 91, Nalwa 96, Swaminathan & Nayar 99, Cutler et al. 02

Catadioptric Imaging

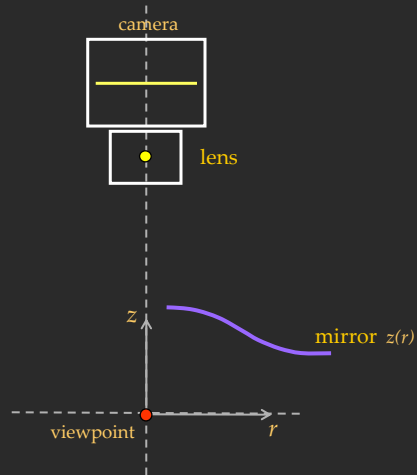


Examples: Rees 70, Charles 87, Nayar 88, Yagi 90, Hong 91, Yamazawa 95, Bogner 95, Nalwa 96, Nayar 97, Chahl & Srinivasan 97

© Shree Nayar, Columbia University

What's the Mirror's Shape ?

(with Simon Baker, ICCV 98)



Complete Class of Mirrors

$$\left(z - \frac{c}{2}\right)^2 - r^2 \left(\frac{k-1}{2}\right) = \frac{c^2}{4} \left(\frac{k-2}{k}\right) \quad (k > 0)$$

$$\left(z - \frac{c}{2}\right)^2 - r^2 \left(1 + \frac{c^2}{2k}\right) = \left(\frac{2k - c^2}{4}\right) \quad (k \geq 2)$$

© Shree Nayar, Columbia University

OneShot 360 by RemoteReality

(Nayar 97)



4 Megapixel (2000 × 2000)
360 degree still camera

© Shree Nayar, Columbia University



© Shree Nayar, Columbia University

Video Conferencing

(with Venkat Peri 96)

© Shree Nayar, Columbia University

Omnidirectional Periscope

Wide Area Surveillance

Vehicle Navigation

Perimeter Monitoring

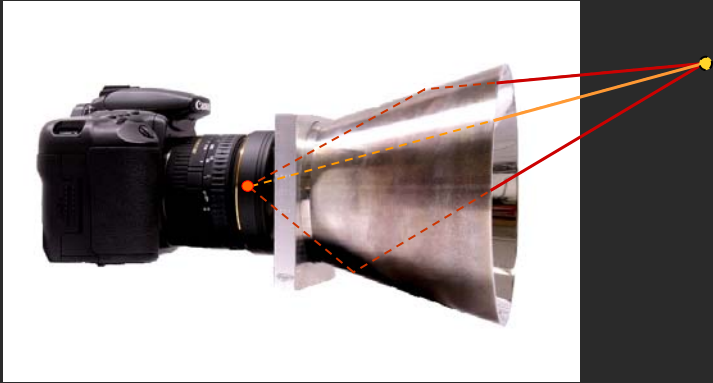
Commercial Security

Virtual Tours

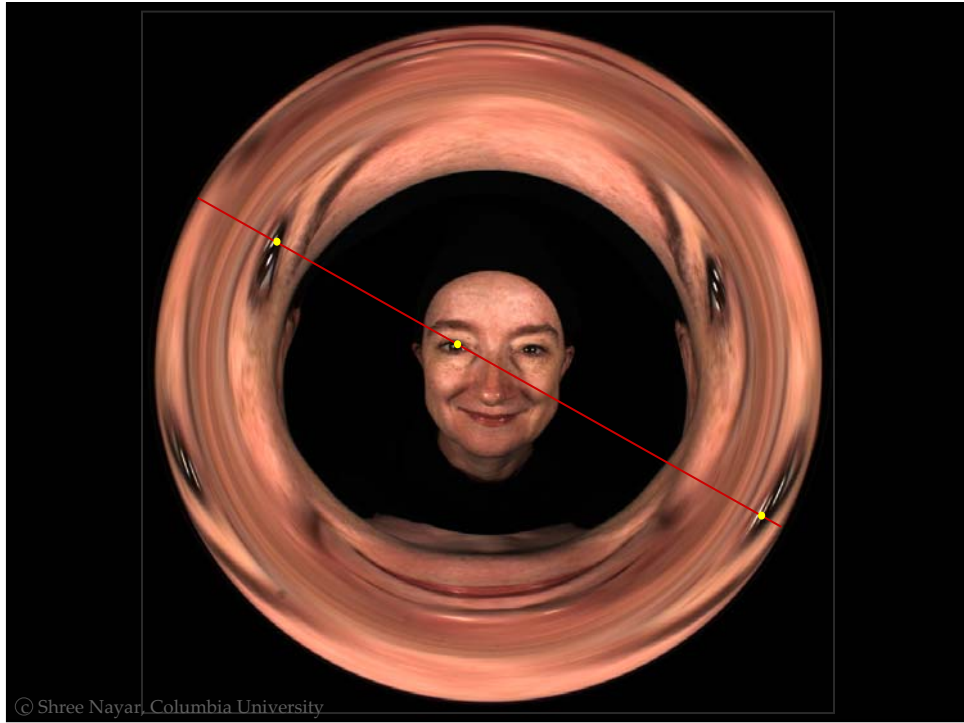
Near Vehicle Awareness

© Shree Nayar, Columbia University (Courtesy : RemoteReality Inc.)

Radial Stereoscopic Imaging



(with Sujit Kuthirummal, SIGGRAPH 06)



© Shree Nayar, Columbia University



© Shree Nayar, Columbia University

Mosaicing



© Shree Nayar, Columbia University

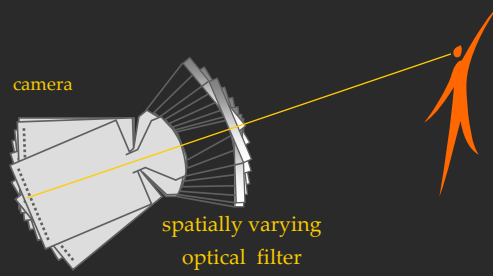
.....Redundant Measurements



© Shree Nayar, Columbia University

Generalized Mosaicing

(Schechner and Nayar, ICCV 2001)



Field of View

Dynamic Range

Spectrum

Depth of Field

Polarization

© Shree Nayar, Columbia University

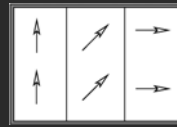
Exposure



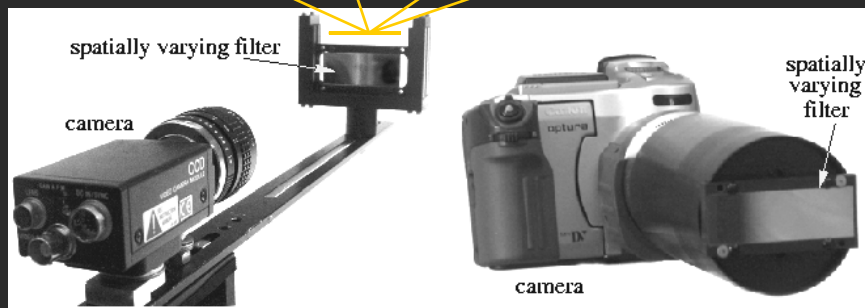
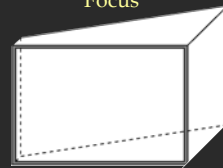
Spectrum



Polarization



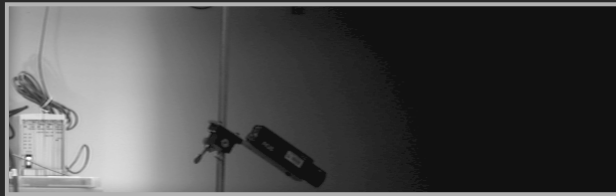
Focus



© Shree Nayar, Columbia University

High Dynamic Range Mosaicing

Attenuation

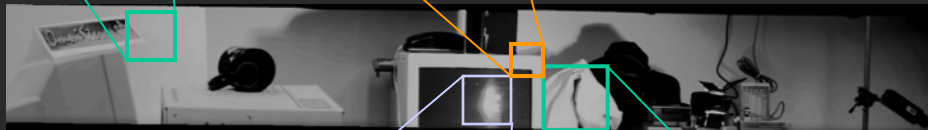


© Shree Nayar, Columbia University

High Dynamic Range Mosaic

137 - 610

34 - 1601



88 - 18,794

4 - 1,418

© Shree Nayar, Columbia University

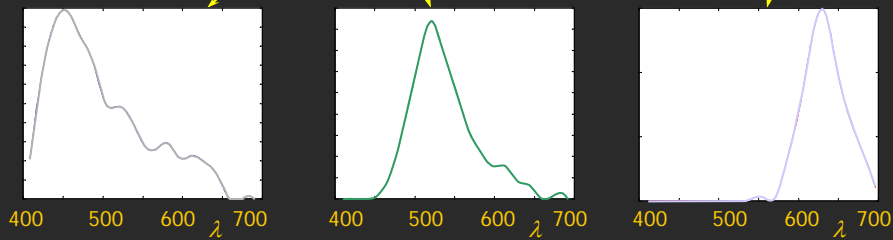
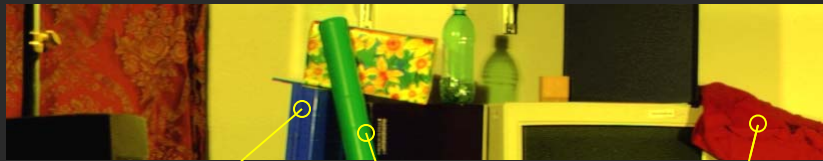
Multispectral Mosaicing

Spectral



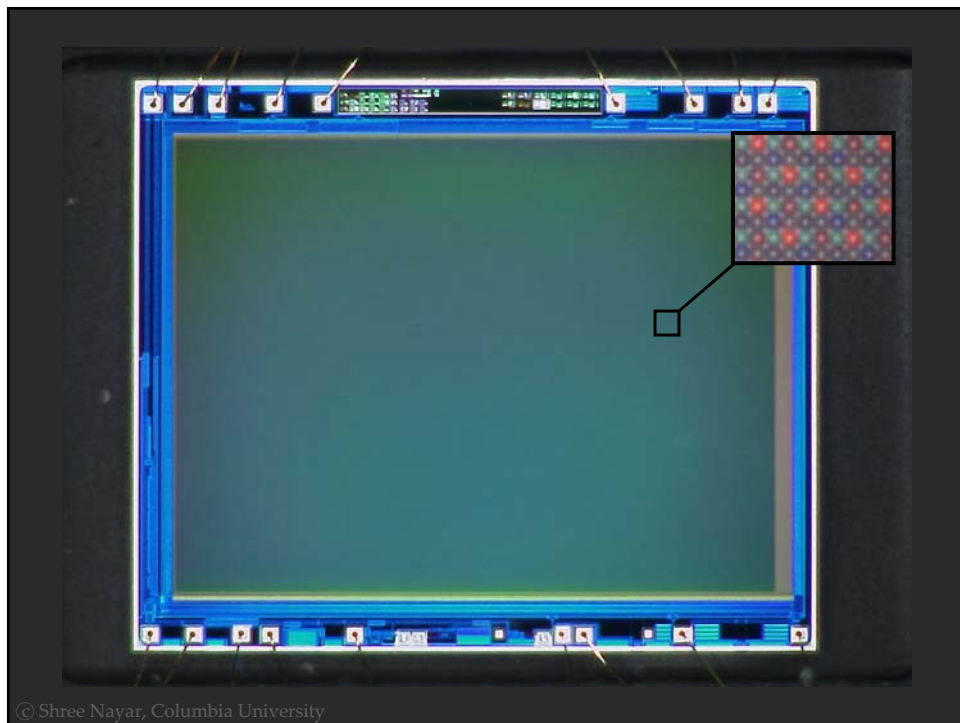
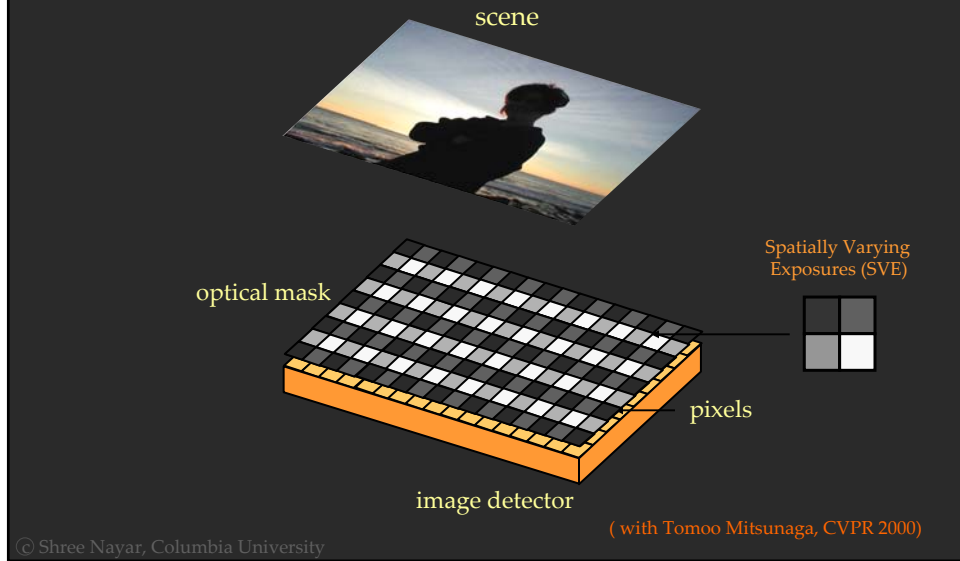
© Shree Nayar, Columbia University

Multispectral Mosaic



© Shree Nayar, Columbia University

High Dynamic Range Imaging: Assorted Pixels



Sony Cybershot



Sony Cybershot with Assorted Pixels



© Shree Nayar, Columbia University

Motion Blur



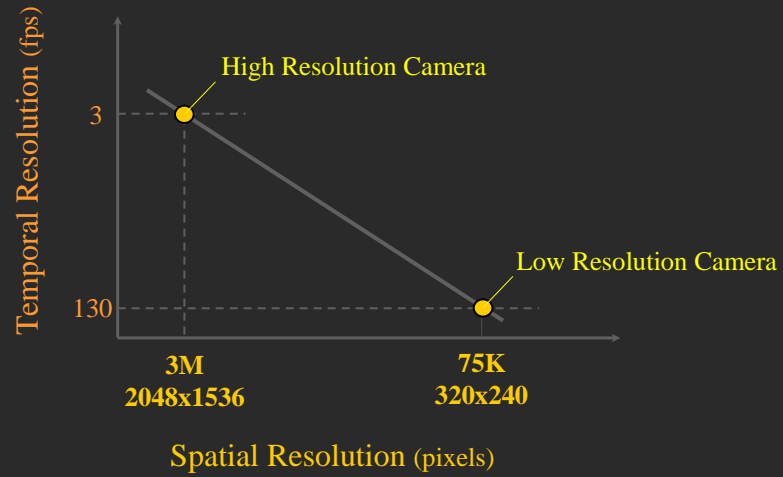
Object Motion



Camera Motion

© Shree Nayar, Columbia University

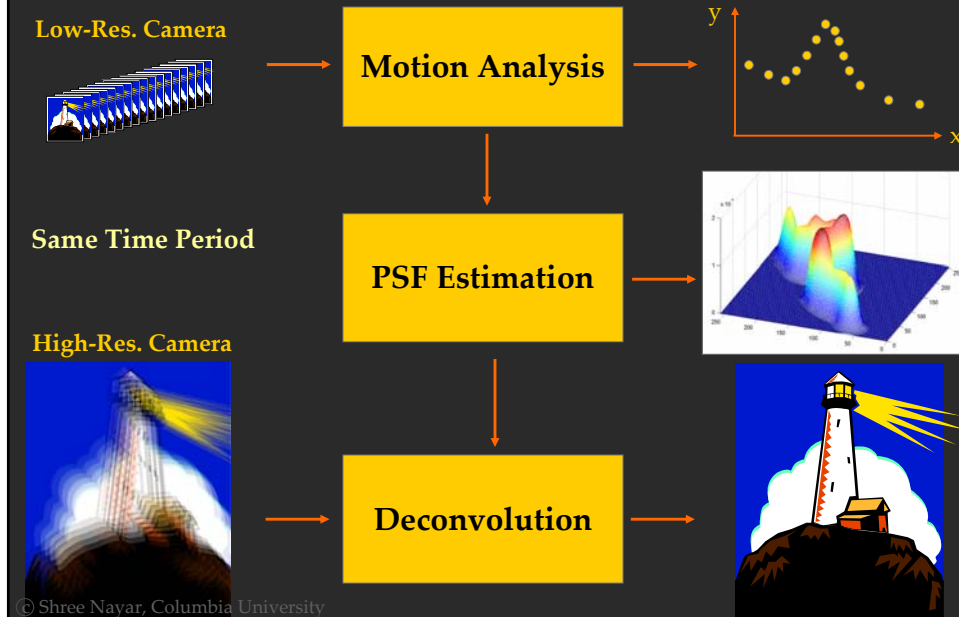
Fundamental Trade-Off in Imaging



© Shree Nayar, Columbia University

(with Moshe Ben-Ezra, CVPR 2003)

Deblurring Approach: Hybrid Imaging



© Shree Nayar, Columbia University

Hybrid Imaging System: Prototype

Primary Detector
(2048x1536)

Secondary Detector
(360x240)



Resolution Ratio of 1 : 36

© Shree Nayar, Columbia University

Example: Blurred High Resolution Image

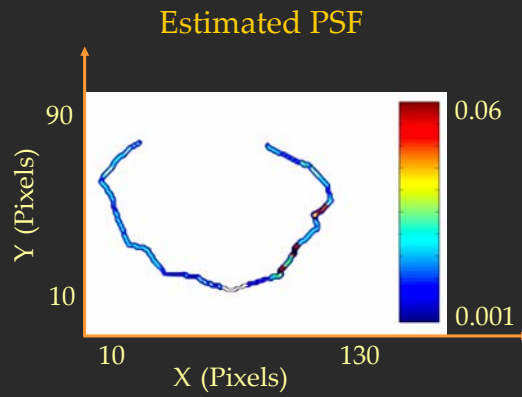


$f = 633\text{mm}$, Exp. Time 1 Sec (> -9 stops)

© Shree Nayar, Columbia University

Example: PSF Estimation from Motion

Low Resolution Video



© Shree Nayar, Columbia University

Example: PSF Estimation from Motion



Blurred Image



Deblurred image



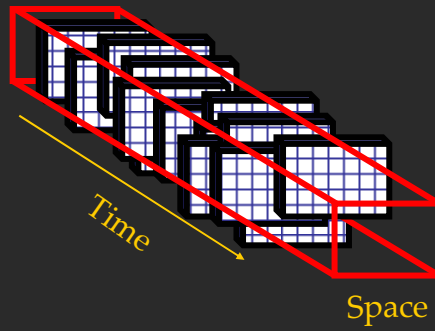
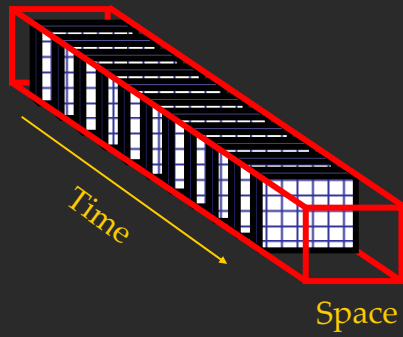
Tripod image (Ground Truth)

© Shree Nayar, Columbia University

Super-Resolution using Jitter Video

Conventional Video

Jitter Video



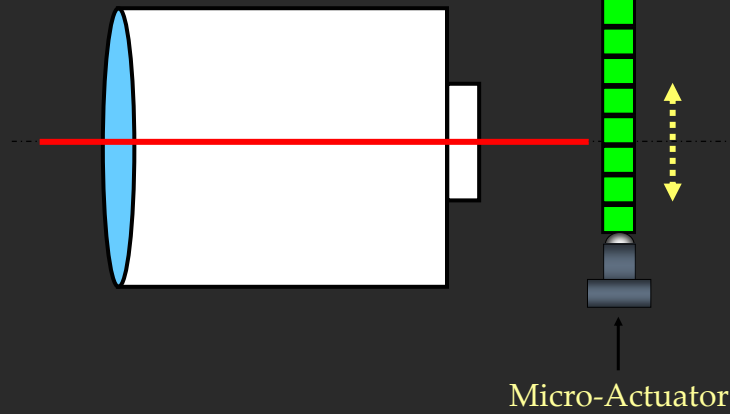
(with Moshe Ben-Ezra and Assaf Zomet, CVPR 2004)

© Shree Nayar, Columbia University

Jitter Camera

Lens

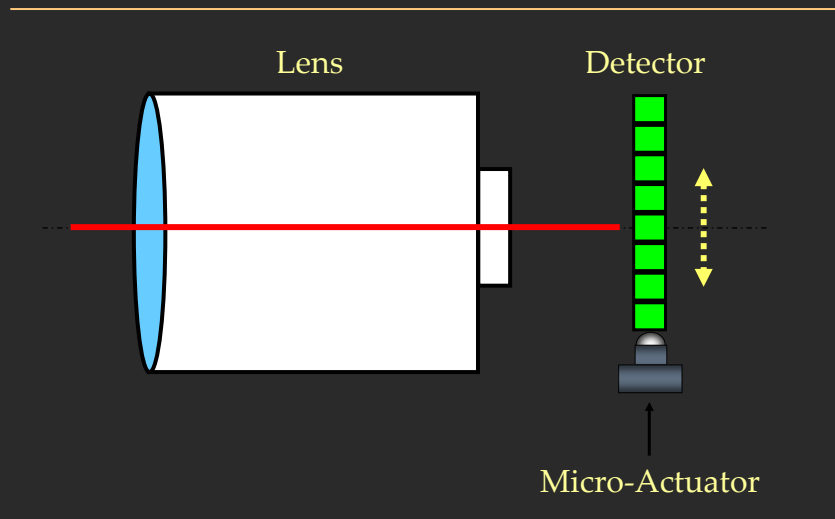
Detector



Jitter is Instantaneous and Synchronized

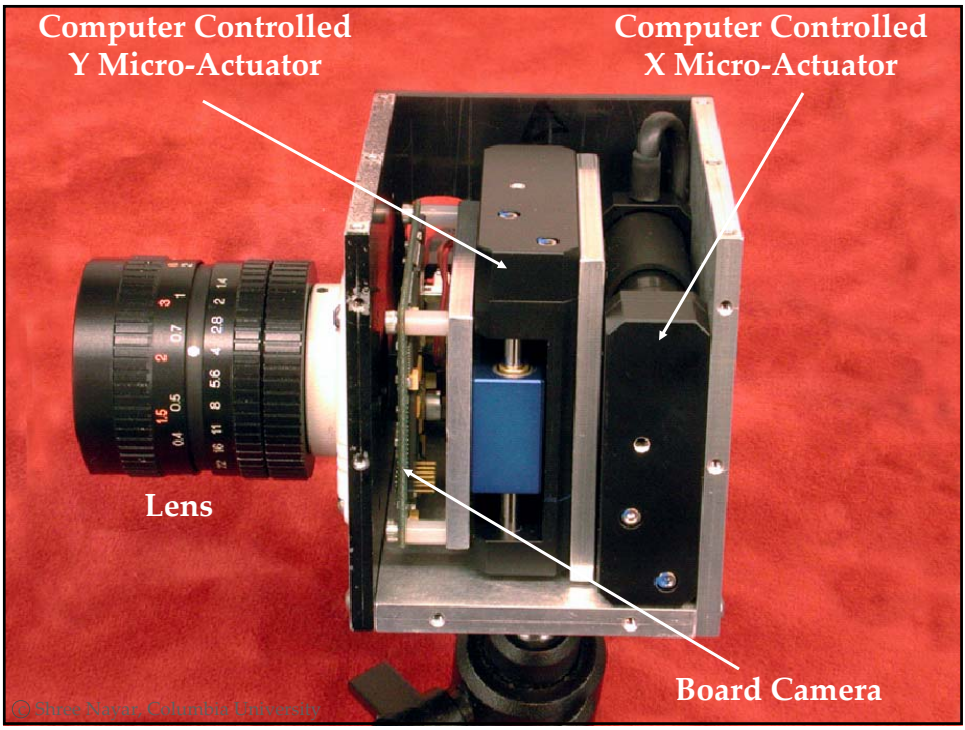
© Shree Nayar, Columbia University

Jitter Camera



Jitter is Instantaneous and Synchronized

© Shree Nayar, Columbia University



© Shree Nayar, Columbia University

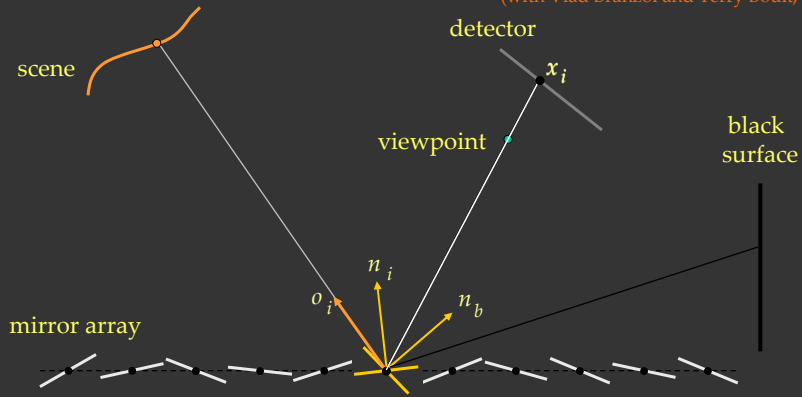
Super-Resolution using Jitter Video



**Video Super-Resolution
Using Controlled
Sub-Pixel Detector Shifts**

Imaging Through Micro-Mirrors

(with Vlad Branzoi and Terry Boult, 2004)



Geometry: Ray Orientation

Photometry: Ray Attenuation

$$o_i(x_i) = G(n_i)$$

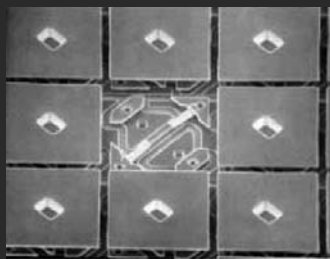
$$a_i(x_i) = \frac{t(n_i)}{t(n_i) + t(n_b)}$$

© Shree Nayar, Columbia University

Digital Micromirror Device (DMD)

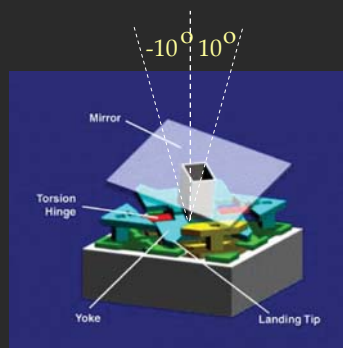
(by Texas Instruments)

DMD Array:



14 μm

Micromirror Architecture:

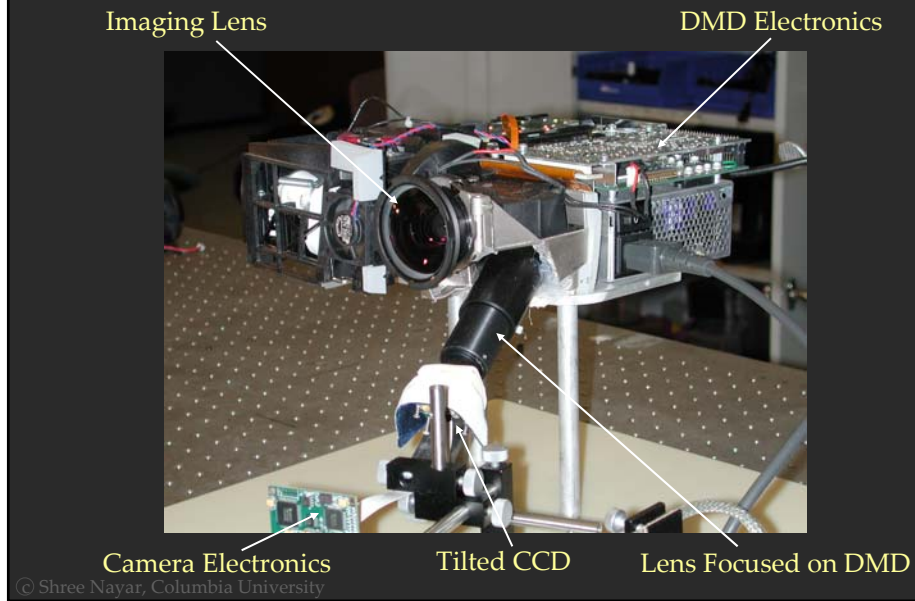


DMD for Imaging:

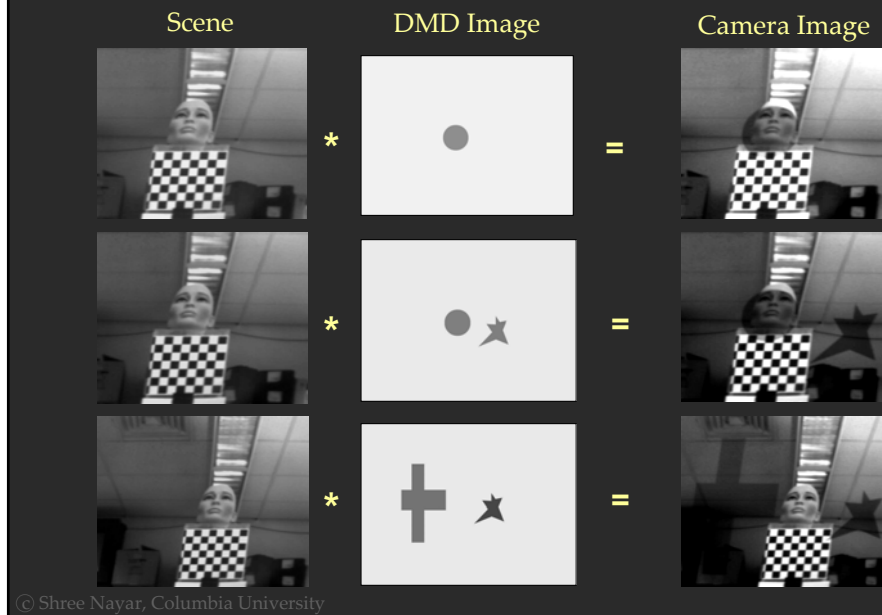
(Malbet et al. 95, Kearney et al. 98, Castracane et al. 99, Christensen et al. 02)

© Shree Nayar, Columbia University

Programmable Imaging System



Modulation: Examples



Adaptive Dynamic Range Imaging (ADR)

Normal (Constant Exposure) Video



Pixel-wise Dynamic Range Control

(Nayar & Branzoi 03)
(Christensen et al. 02)

ADR Video

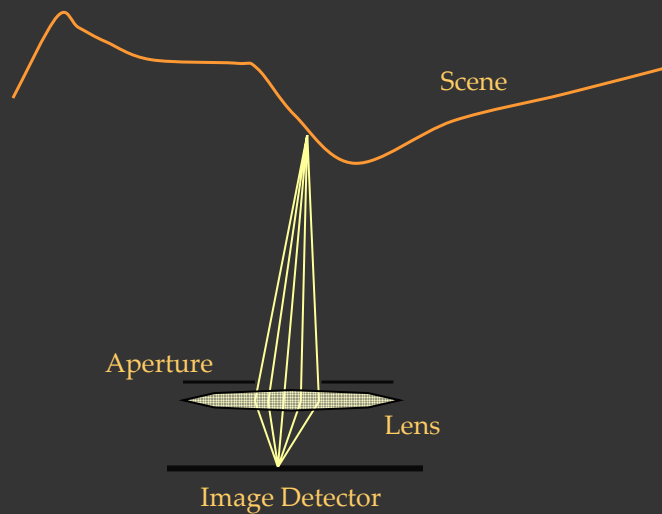


DMD Control Video



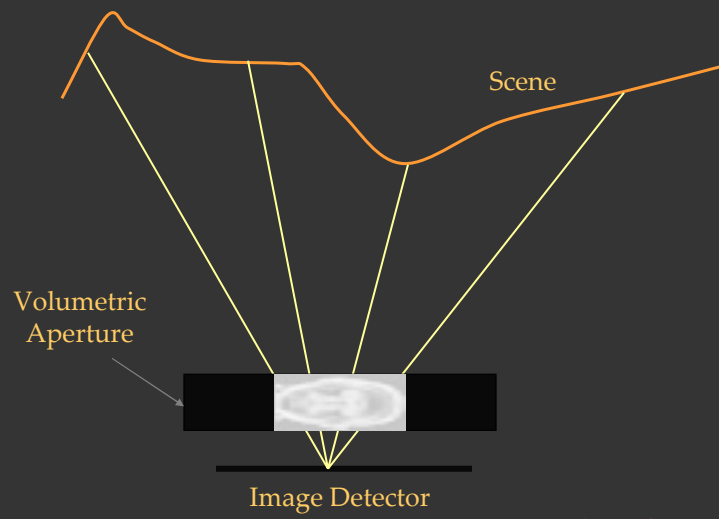
© Shree Nayar, Columbia University

Camera with a Lens



© Shree Nayar, Columbia University

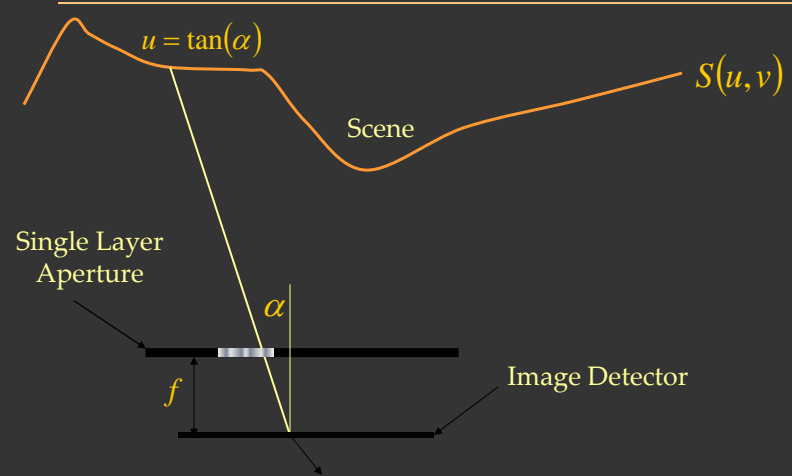
Lensless Camera with Volumetric Aperture



© Shree Nayar, Columbia University

(with Assaf Zomet, 2005)

Single Aperture Layer

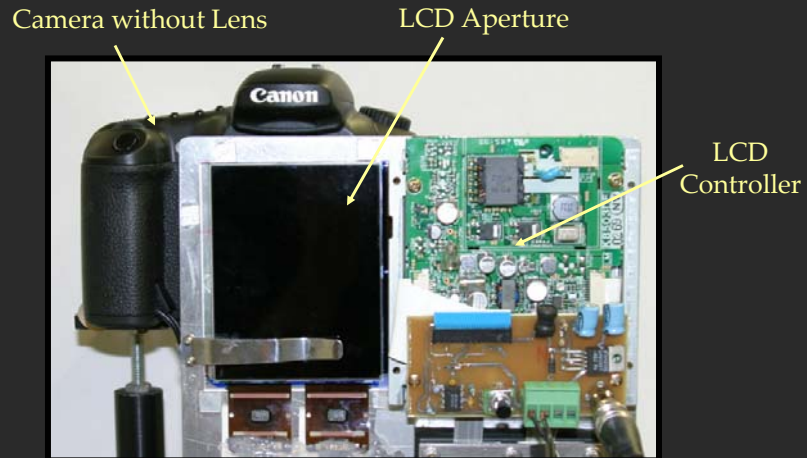


Pixel Brightness:
$$I(x, y) = \iint S(u, v) T(x - fu, y - fv) dudv$$

© Shree Nayar, Columbia University

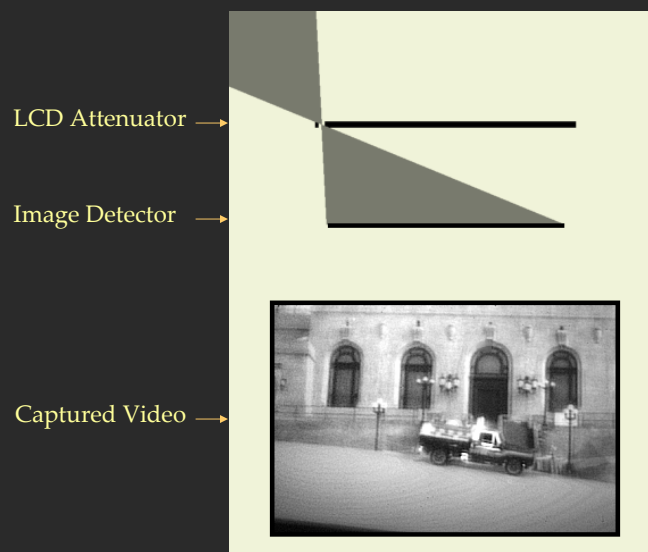
Scene Transmittance Function

Initial Implementation: LCD Attenuator



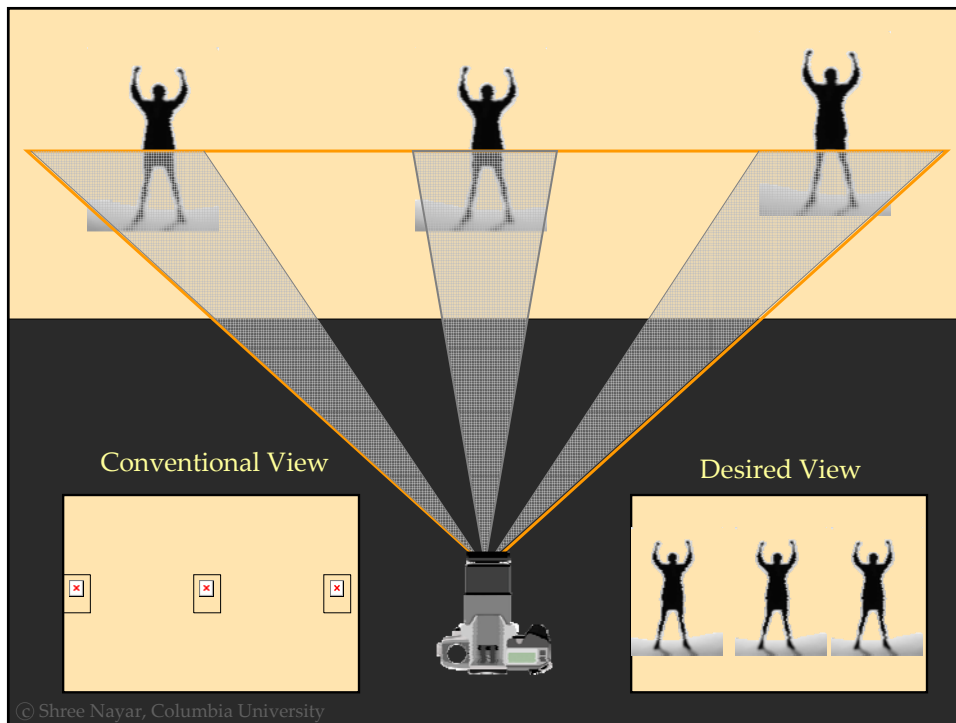
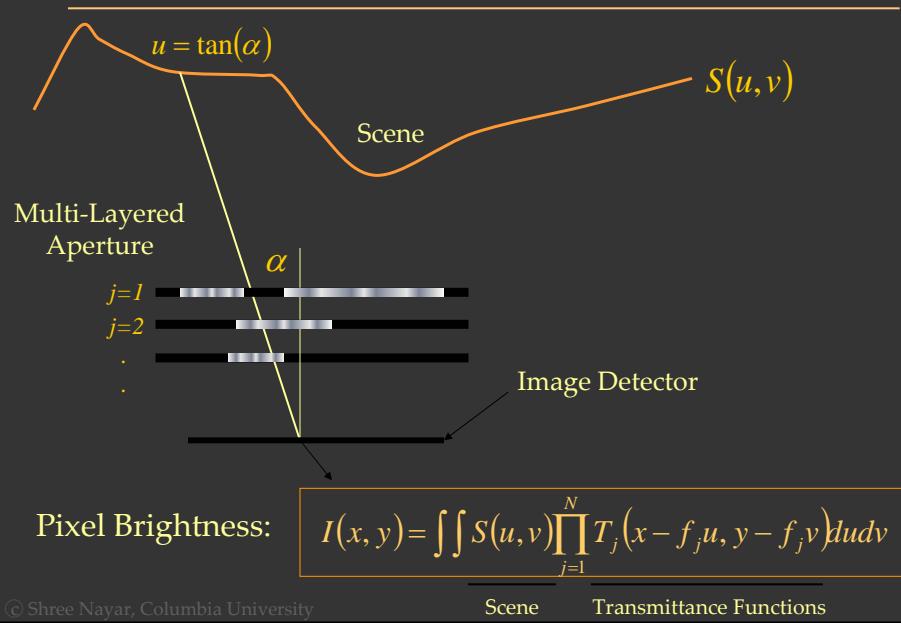
© Shree Nayar, Columbia University

Panning without Moving Parts

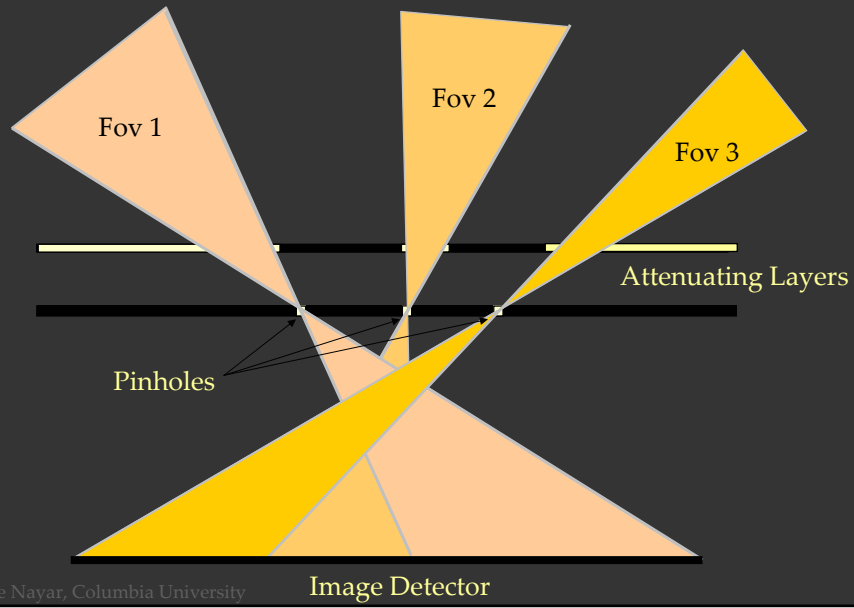


© Shree Nayar, Columbia University

Multiple Aperture Layers



Split Field of View using Multiple Layers



Split Field of View



Lens Camera

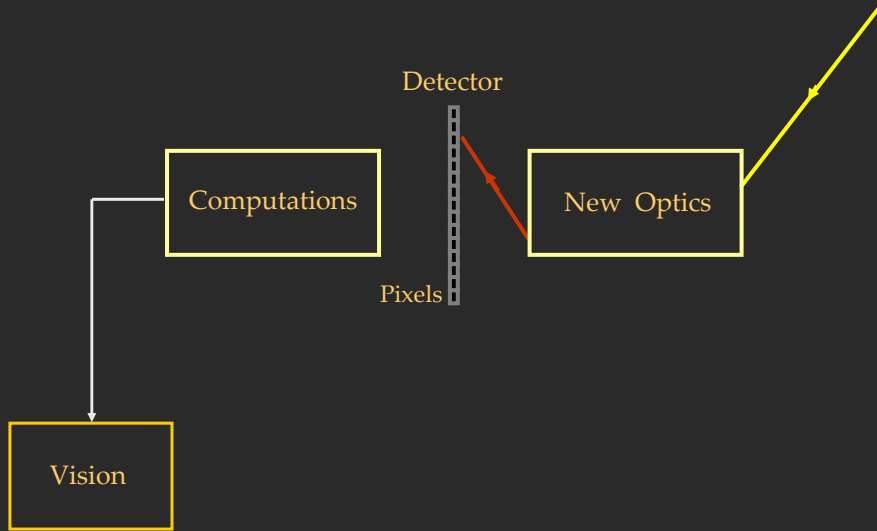


Lensless Camera



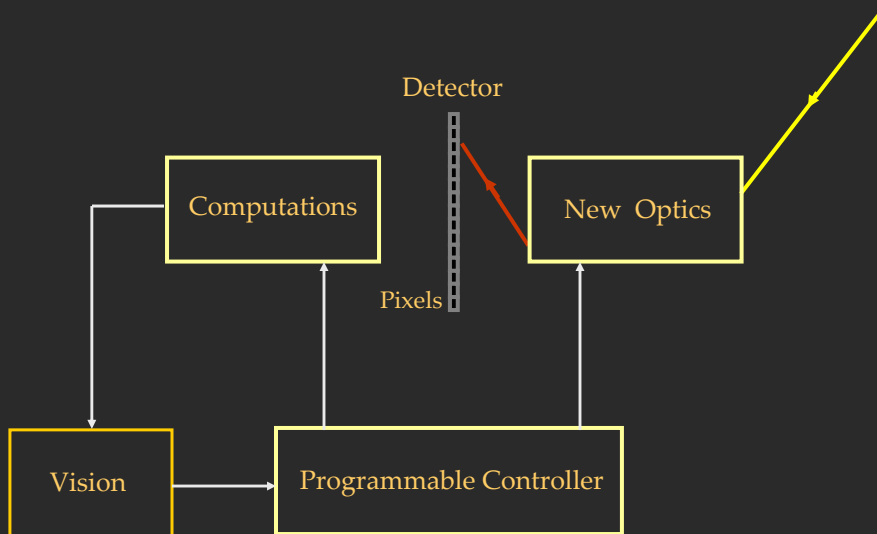
© Shree Nayar, Columbia University

Computational Cameras



© Shree Nayar, Columbia University

Programmable Imaging



© Shree Nayar, Columbia University

Light field photography and videography

Marc Levoy



Computer Science Department
Stanford University

34:15 total + 30% = ~45 minutes

Time =

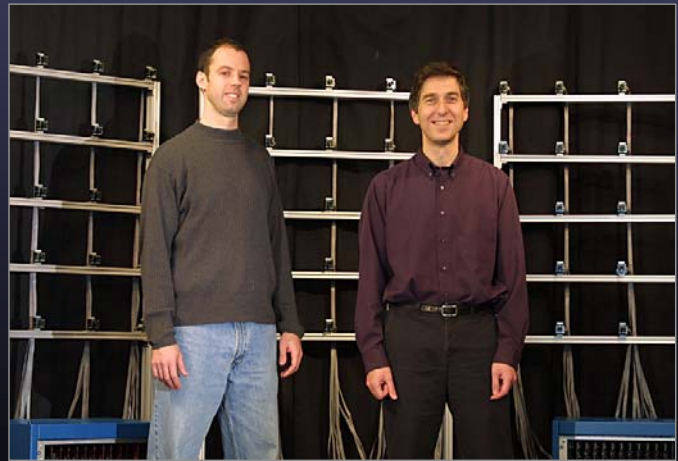
List of projects

- high performance imaging using large camera arrays
- light field photography using a handheld plenoptic camera
- dual photography

High performance imaging using large camera arrays

*Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Eino-Ville Talvala, Emilio Antunez,
Adam Barth, Andrew Adams, Mark Horowitz, Marc Levoy*

(Proc. SIGGRAPH 2005)



Stanford multi-camera array



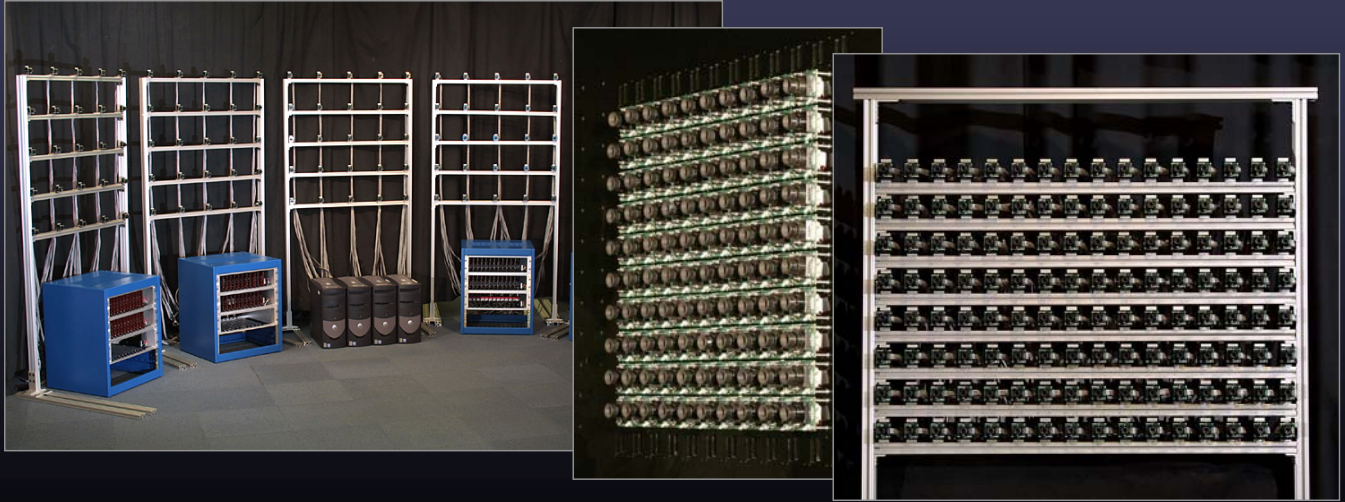
- 640×480 pixels \times
30 fps \times 128 cameras
- synchronized timing
- continuous streaming
- flexible arrangement



© 2005 Marc Levoy

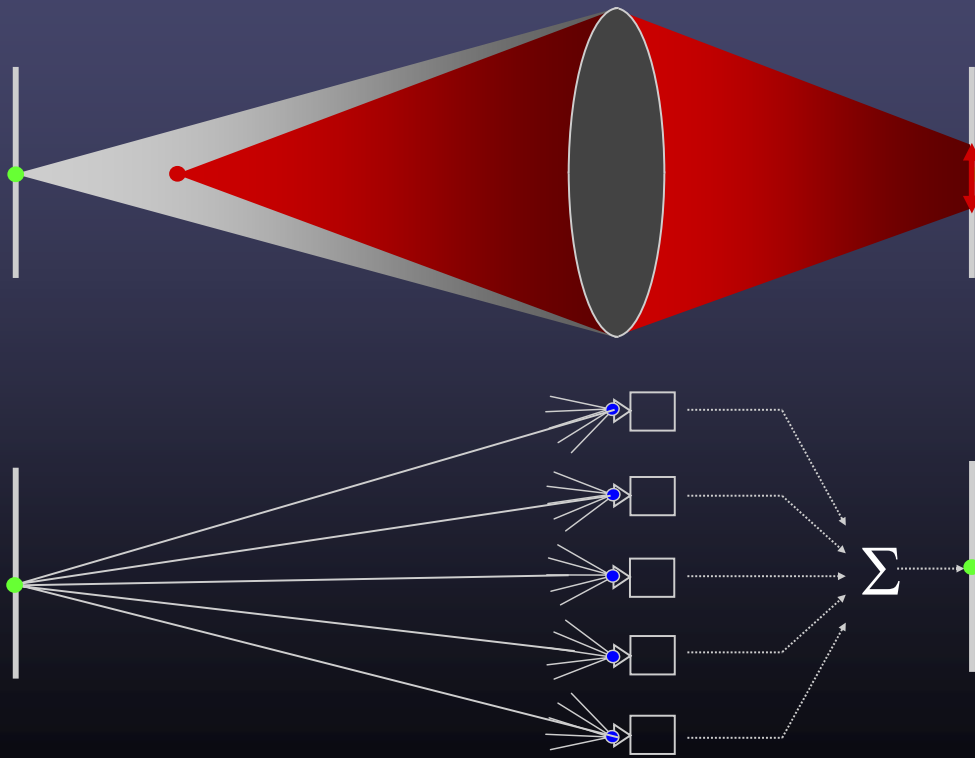
Ways to use large camera arrays

- widely spaced → light field capture
- tightly packed → high-performance imaging
- intermediate spacing → synthetic aperture photography



© 2005 Marc Levoy

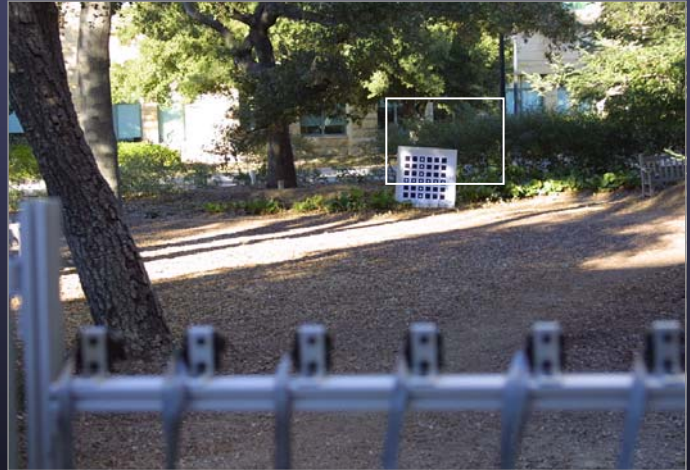
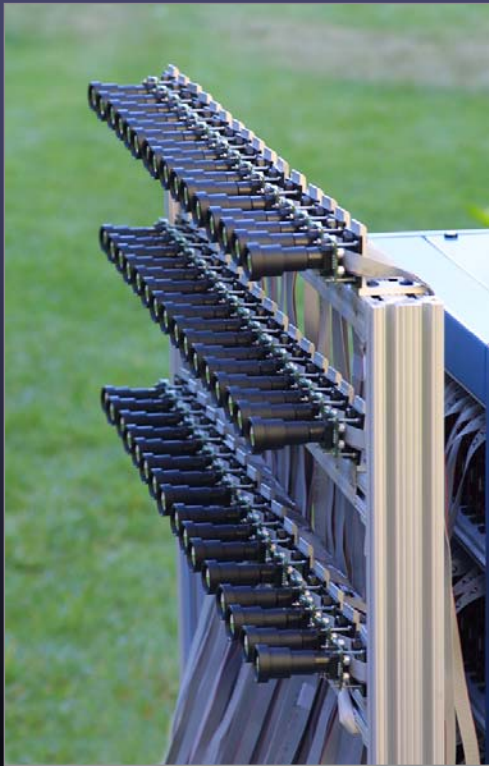
Intermediate camera spacing: synthetic aperture photography



© 2005 Marc Levoy

Example using 45 cameras

[Vaish CVPR 2004]



© 2005 Marc Levoy

- Reference:

- Vaish, V., Wilburn, B., Joshi, N., Levoy, M., Using Plane + Parallax for Calibrating Dense Camera Arrays,
Proc. CVPR 2004.



- At left is a single view, at right is a sum of all views, appropriately shifted.
- For the movie, see 2nd bullet of “Slides and videos” on <http://graphics.stanford.edu/projects/array/>

Video

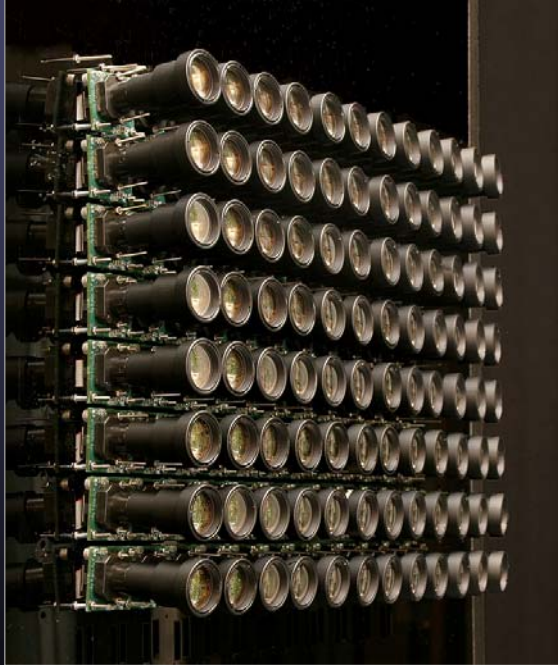


© 2005 Marc Levoy

- Video available at <http://graphics.stanford.edu/papers/CameraArray/>

Tiled camera array

Can we match the image quality of a cinema camera?



- world's largest video camera
- no parallax for distant objects
- poor lenses limit image quality
- seamless mosaicing isn't hard

•poor lenses limit image quality

–we set out to answer the question, “Can we **match the image quality** of an SLR?”

Tiled panoramic image (before geometric or color calibration)



Tiled panoramic image (after calibration and blending)



Tiled camera array

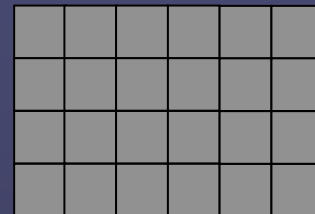
Can we match the image quality of a cinema camera?



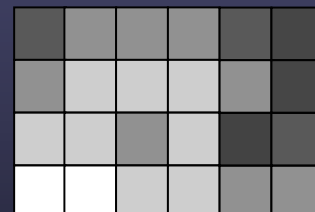
- world's largest video camera
- no parallax for distant objects
- poor lenses limit image quality
- seamless mosaicing isn't hard
- per-camera exposure metering
- HDR within and between tiles



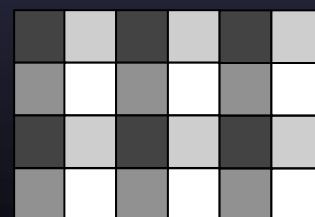
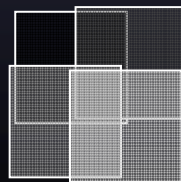
same exposure
in all cameras



individually
metered



checkerboard
of exposures

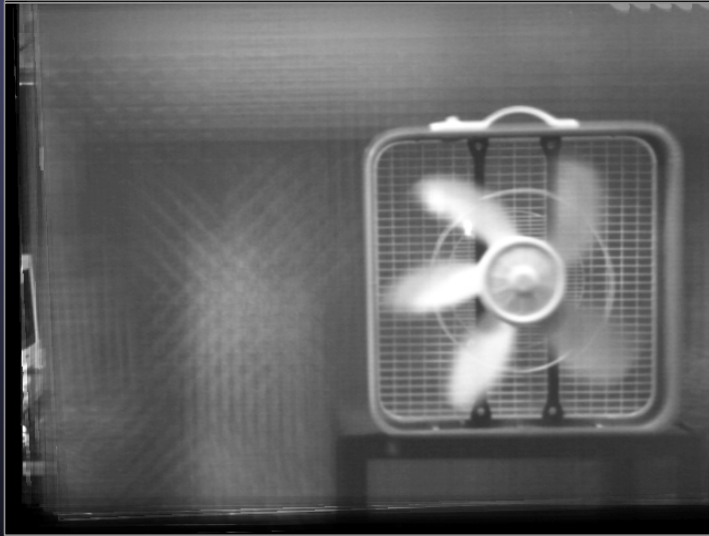


High-performance photography as multi-dimensional sampling

- spatial resolution
- field of view
- frame rate
- dynamic range
- bits of precision
- depth of field
- focus setting
- color sensitivity

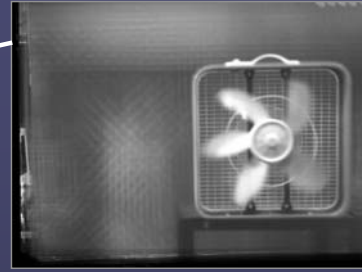
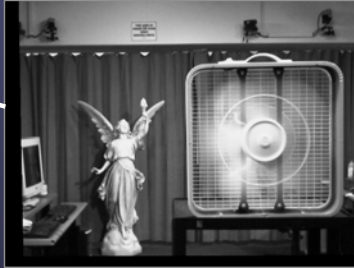
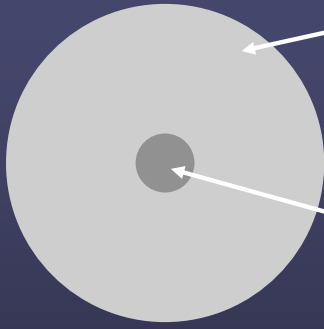
© 2005 Marc Levoy

Spacetime aperture shaping



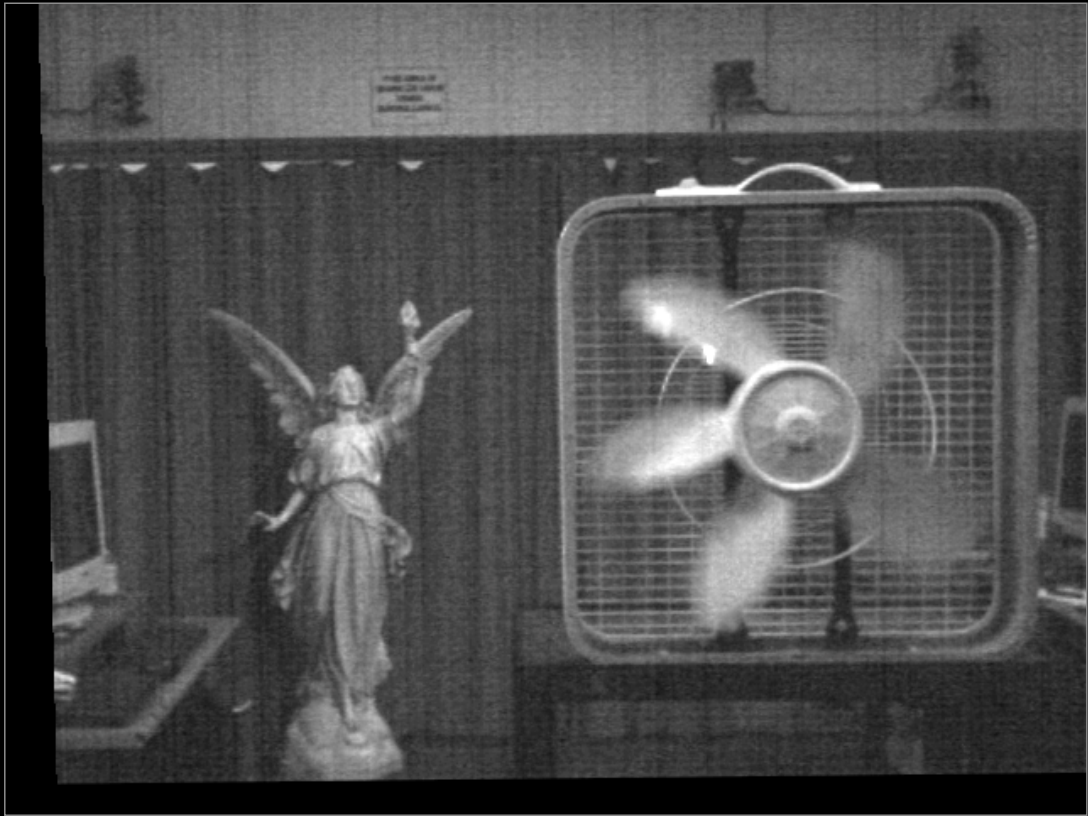
- shorten exposure time to freeze motion → dark
- stretch contrast to restore level → noisy
- increase (synthetic) aperture to capture more light → decreases depth of field

© 2005 Marc Levoy



- center of aperture: few cameras, long exposure → high depth of field, low noise, but action is blurred
- periphery of aperture: many cameras, short exposure → freezes action, low noise, but low depth of field





Light field photography using a handheld plenoptic camera

*Ren Ng, Marc Levoy, Mathieu Brédif,
Gene Duval, Mark Horowitz and Pat Hanrahan*

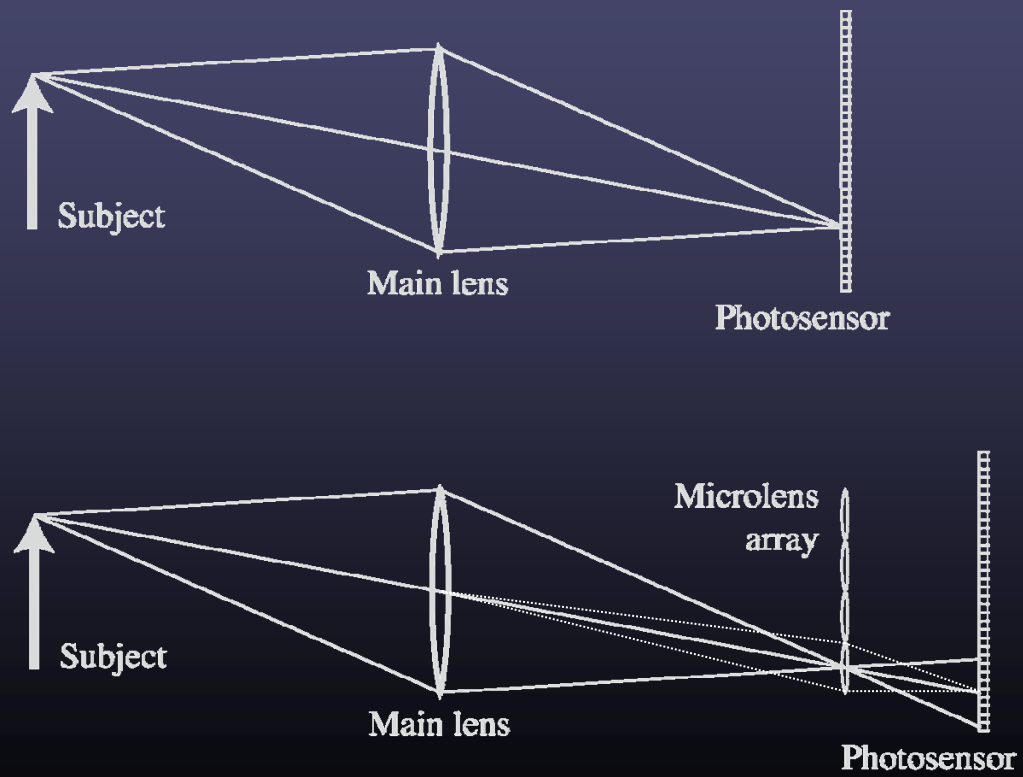
*(Proc. SIGGRAPH 2005
and TR 2005-02)*



© 2005 Marc Levoy

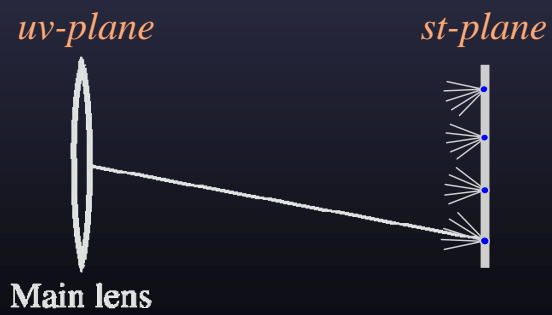
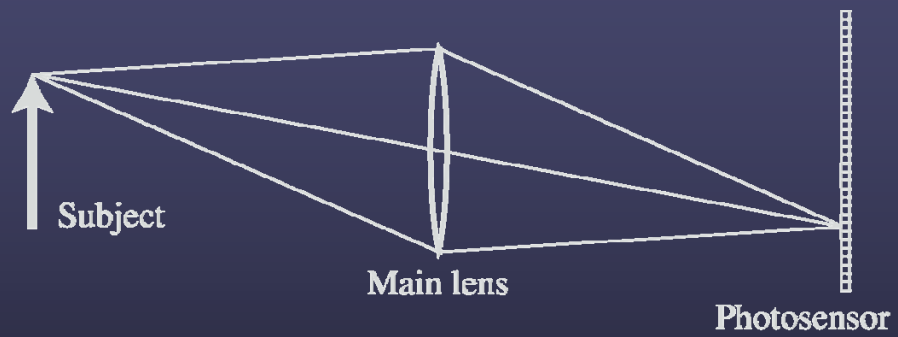
- Light field capture not using an array of cameras, but
 - using a single, handheld camera
- What we'll do with these light fields is not seeing through crowds, but
 - **refocusing** a picture after we take it, and
 - **moving the observer** (slightly) after we take the picture

Conventional versus light field camera



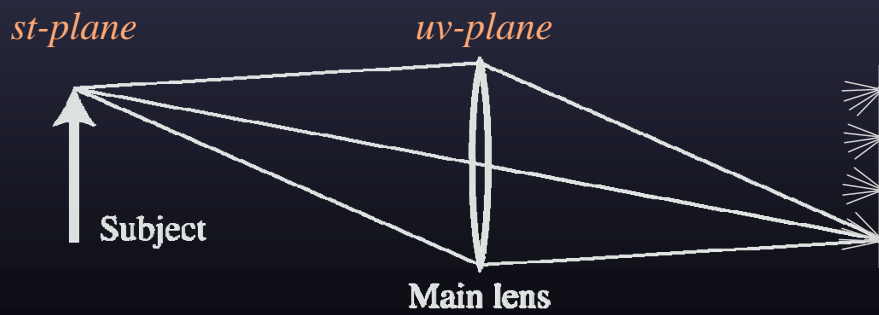
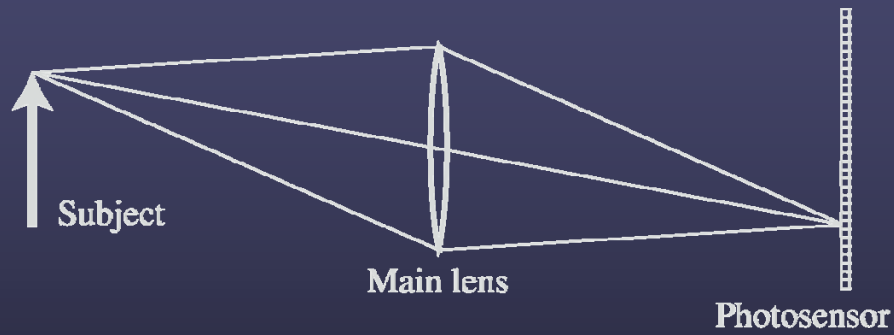
© 2006 Marc Levoy

Conventional versus light field camera



© 2006 Marc Levoy

Conventional versus light field camera



© 2006 Marc Levoy

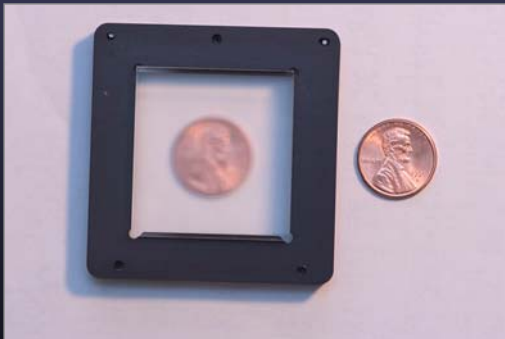
Prototype camera



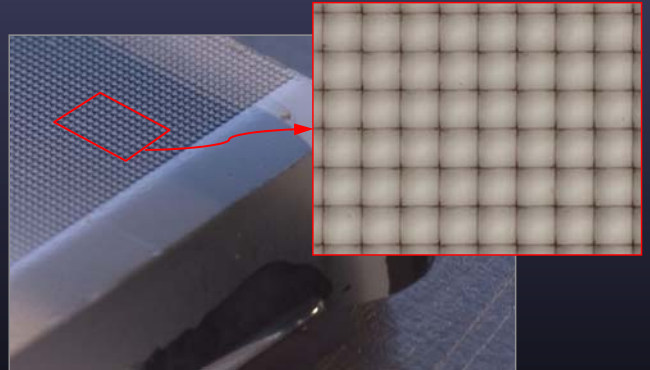
Contax medium format camera



Kodak 16-megapixel sensor



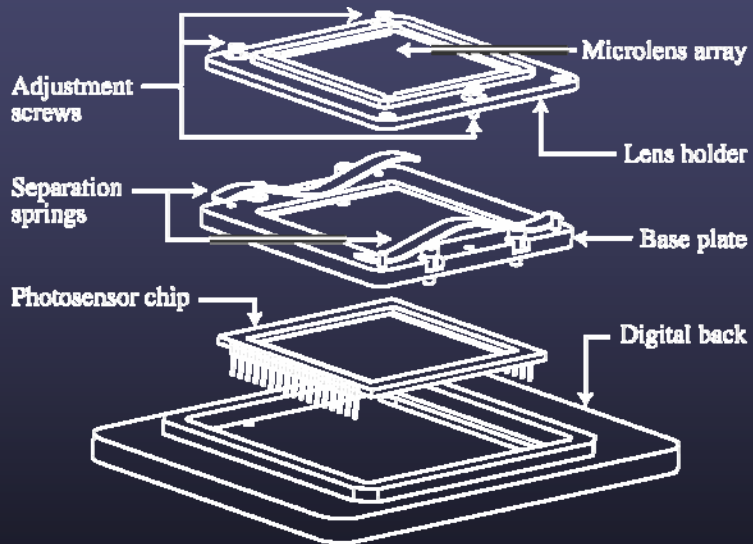
Adaptive Optics microlens array



125 μ square-sided microlenses

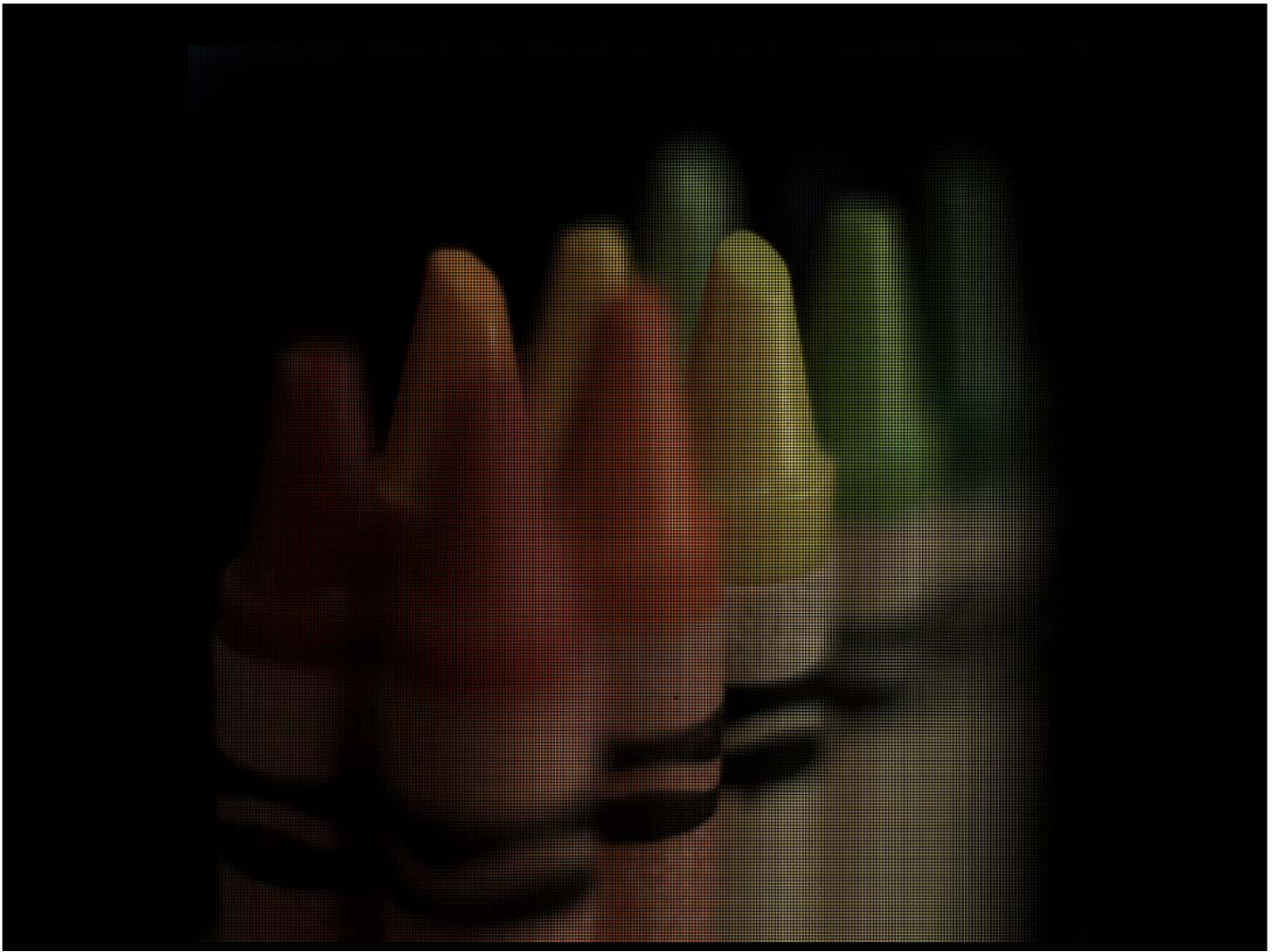
$$4000 \times 4000 \text{ pixels} \div 292 \times 292 \text{ lenses} = 14 \times 14 \text{ pixels per lens}$$

Mechanical design



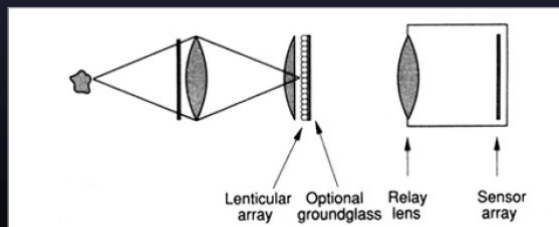
- microlenses float 500μ above sensor
- focused using 3 precision screws

© 2005 Marc Levoy



Prior work

- integral photography
 - microlens array + film
 - application is autostereoscopic effect
- [Adelson 1992]
 - proposed this camera
 - built an optical bench prototype using relay lenses
 - application was stereo vision, not photography



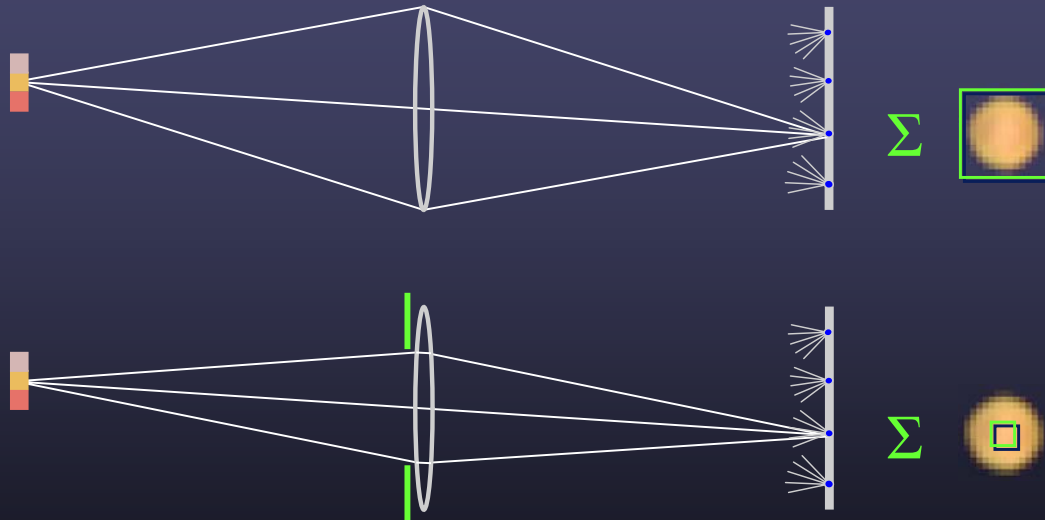
© 2006 Marc Levoy

- Reference:

- Adelson, E.H., Wang, J.Y.A., Single Lens Stereo with a Plenoptic Camera ,

- IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*,
Vol. 14, No. 2, 1992, pp. 99-106.

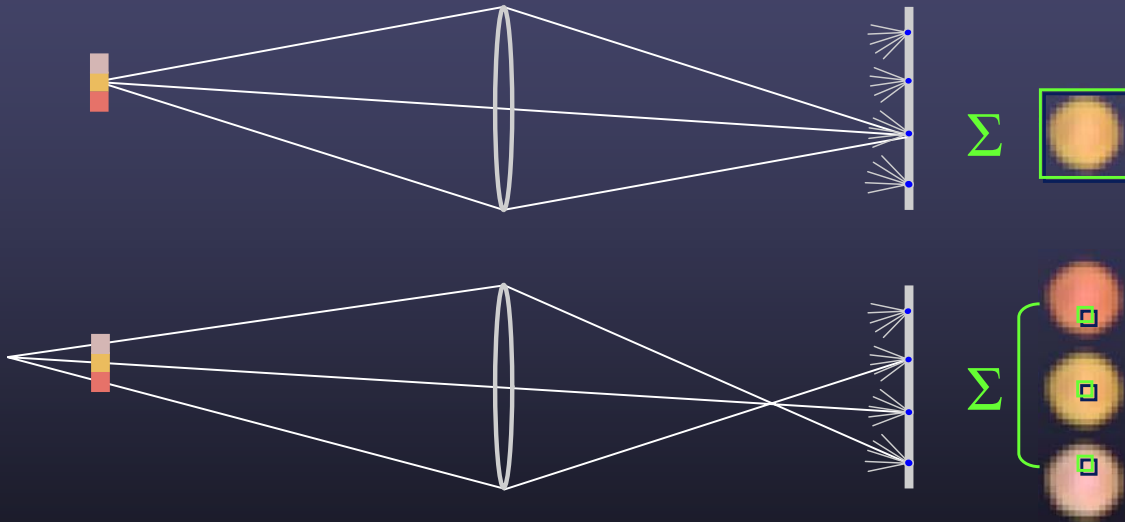
Digitally stopping-down



- stopping down = summing only the central portion of each microlens

© 2006 Marc Levoy

Digital refocusing



- refocusing = summing windows extracted from several microlenses

© 2006 Marc Levoy

A digital refocusing theorem

- an f/N light field camera, with $P \times P$ pixels under each microlens, can produce views as sharp as an $f/(N \times P)$ conventional camera

– or –

- it can produce views with a shallow depth of field (f/N) focused anywhere within the depth of field of an $f/(N \times P)$ camera

© 2006 Marc Levoy

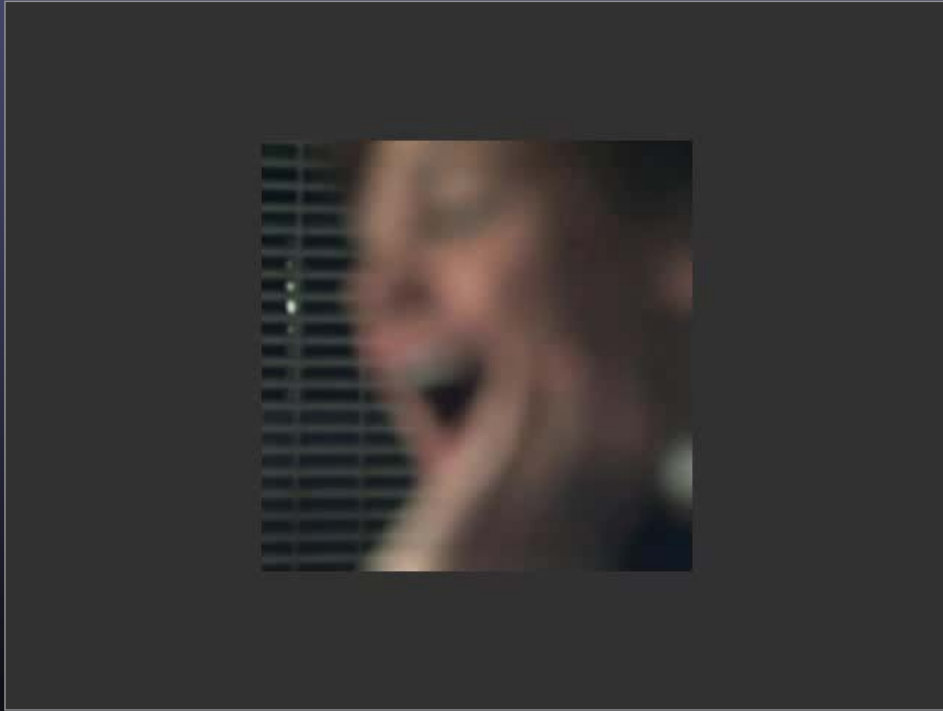
Example of digital refocusing



© 2006 Marc Levoy

•For full sequence, see
<http://graphics.stanford.edu/papers/lfcamera/>

Refocusing portraits



© 2006 Marc Levoy

•For full sequence, see
<http://graphics.stanford.edu/papers/lfcamera/>

Action photography



Focusing through a splash of water

© 2006 Marc Levoy

•For full sequence, see
<http://graphics.stanford.edu/papers/lfcamera/>

Extending the depth of field



conventional photograph,
main lens at $f/4$



conventional photograph,
main lens at $f/22$



light field, main lens at $f/4$,
after all-focus algorithm
[Agarwala 2004]

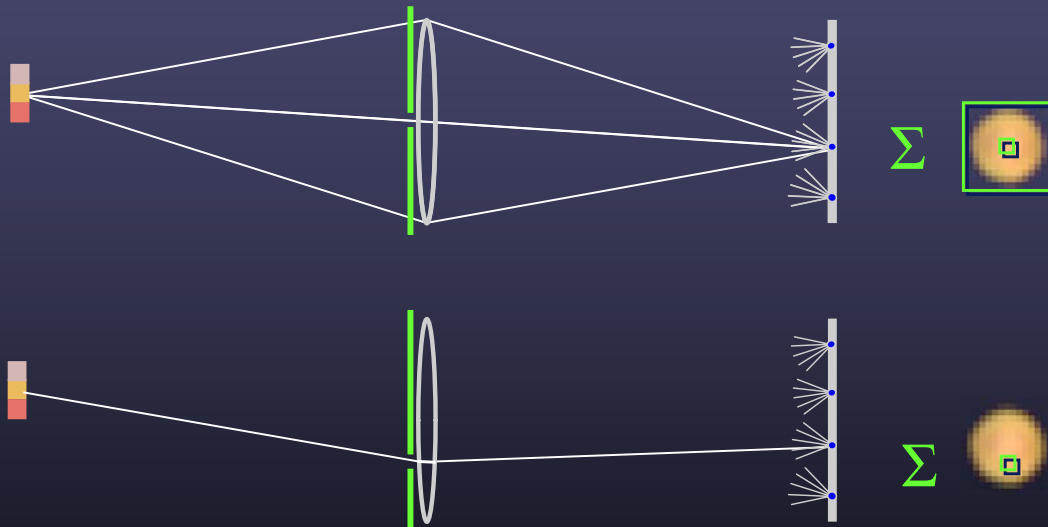
- main lens at $f/22$
 - captured with light field camera and $f/4$ lens,
 - computed by extracting only the middle pixel of that image
 - would be the same image if no microlenses and larger pixels
- Reference:
 - Agarwala, A., Dontcheva, M., Agrawala, M., Drucker, S., Colburn, A., Curless, B., Salesin, D., Cohen, M., Interactive digital photomontage, *Proc. SIGGRAPH 2004*.

Macrophotography



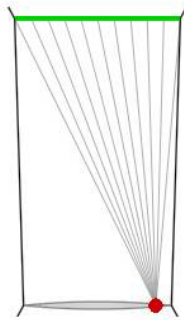
© 2005 Marc Levoy

Digitally moving the observer



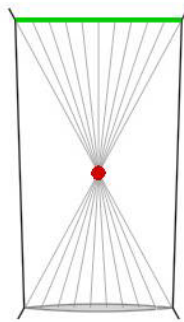
- moving the observer = moving the window we extract from the microlenses

Example of moving the observer



© 2006 Marc Levoy

Moving backward and forward



© 2006 Marc Levoy

Implications

- cuts the unwanted link between exposure (due to the aperture) and depth of field
- trades off (excess) spatial resolution for ability to refocus and adjust the perspective
- sensor pixels should be made even smaller, subject to the diffraction limit

$$36\text{mm} \times 24\text{mm} \div 2\mu \text{ pixels} = 216 \text{ megapixels}$$

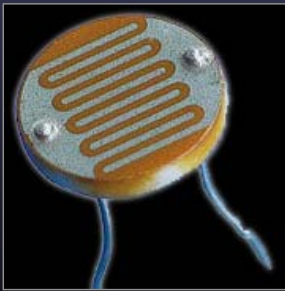
$$18\text{K} \times 12\text{K} \text{ pixels}$$

$$1800 \times 1200 \text{ pixels} \times 10 \times 10 \text{ rays per pixel}$$

Dual Photography

*Pradeep Sen, Billy Chen, Gaurav Garg, Steve Marschner,
Mark Horowitz, Marc Levoy, Hendrik Lensch*

(Proc. SIGGRAPH 2005)

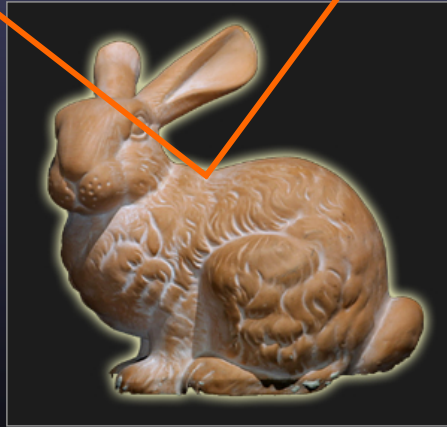


Helmholtz reciprocity

camera

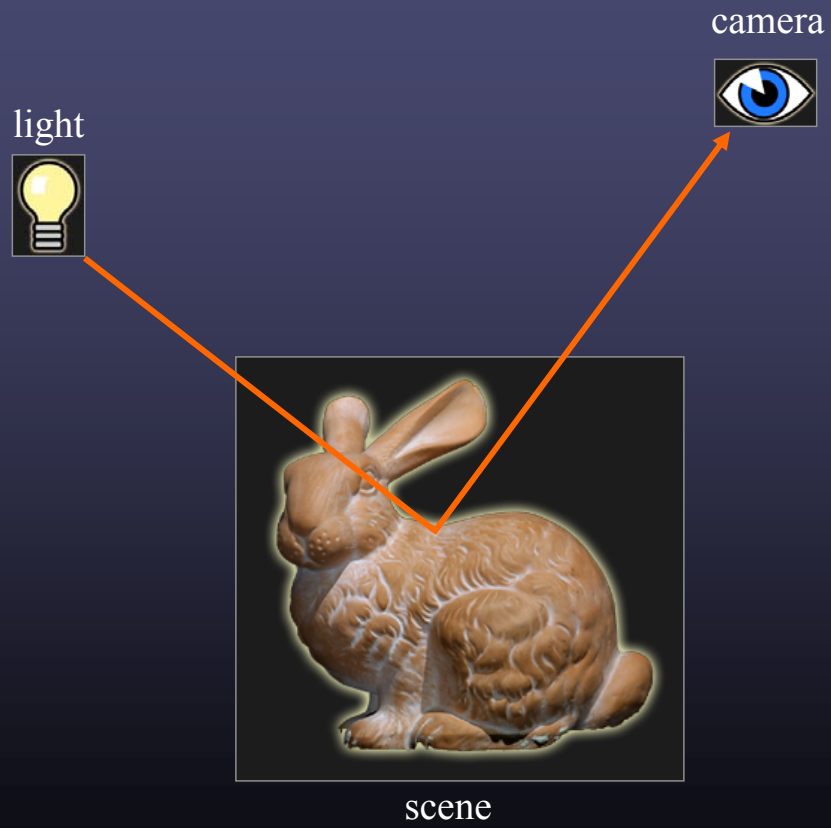


light

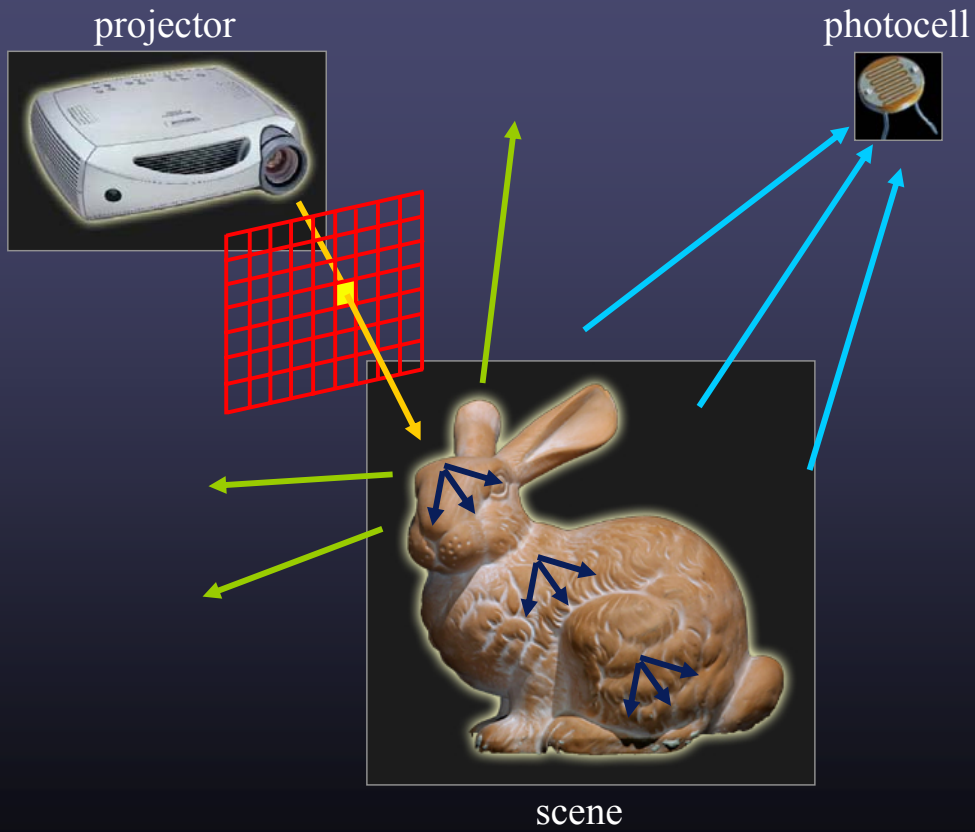


scene

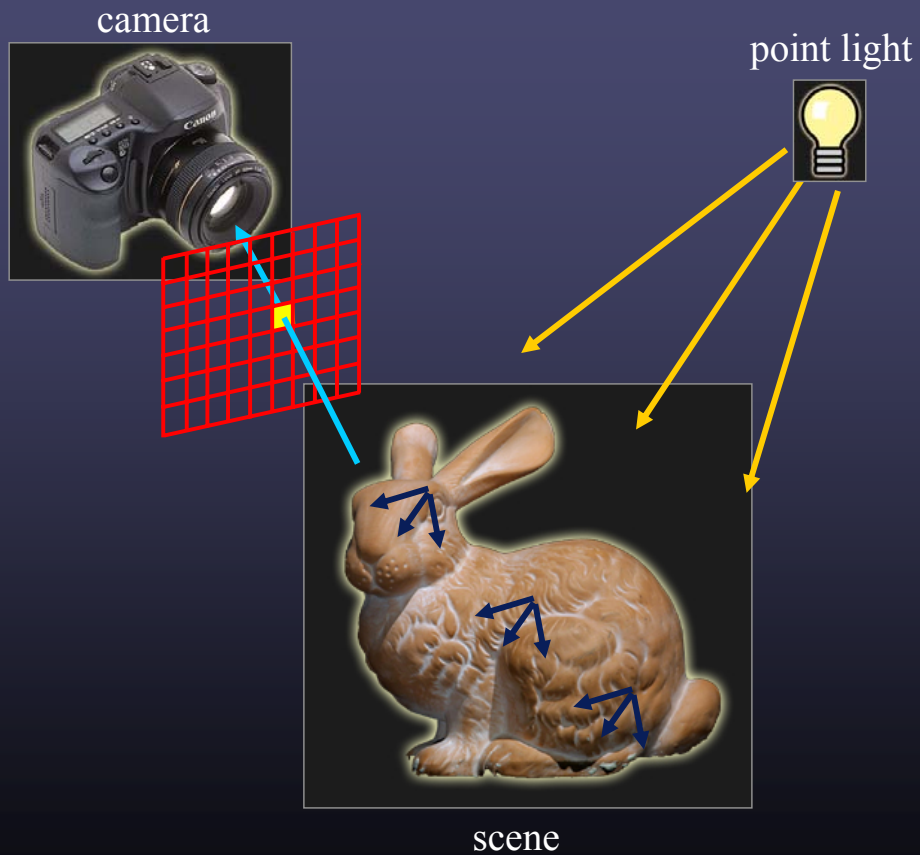
Helmholtz reciprocity



Measuring transport along a set of paths

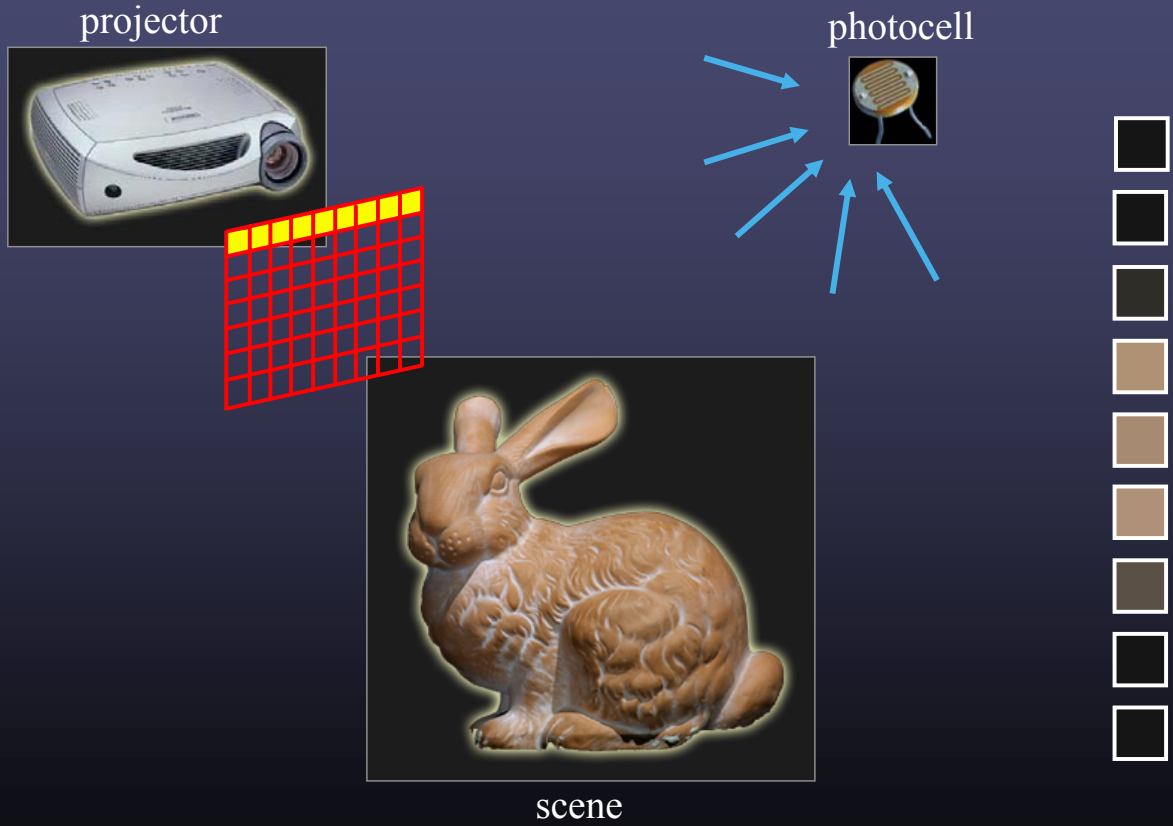


Reversing the paths



- The transport will be the same
 - up to a global scaling factor
 - because we replaced a projector by a different kind of light

Forming a dual photograph



Forming a dual photograph

“dual” camera

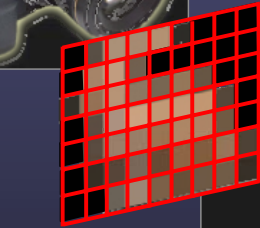


image of scene

“dual” light



scene

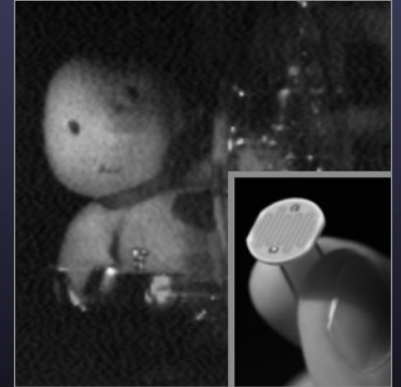


Physical demonstration

- light replaced with projector
- camera replaced with photocell
- projector scanned across the scene



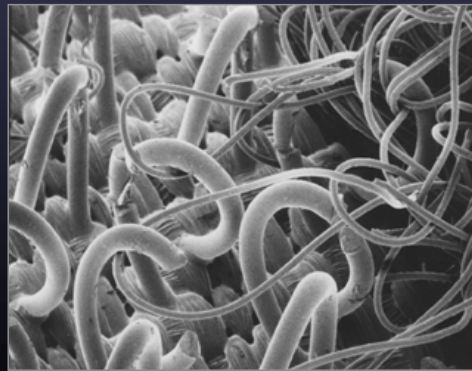
conventional photograph,
with light coming from right



dual photograph,
as seen from projector's position
and as illuminated from photocell's position

Related imaging methods

- time-of-flight scanner
 - if they return reflectance as well as range
 - but their light source and sensor are typically coaxial
- scanning electron microscope



Velcro® at 35x magnification,
Museum of Science, Boston

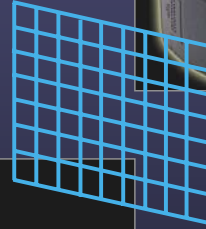
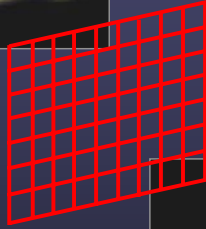
© 2006 Marc Levoy

The 4D transport matrix

projector

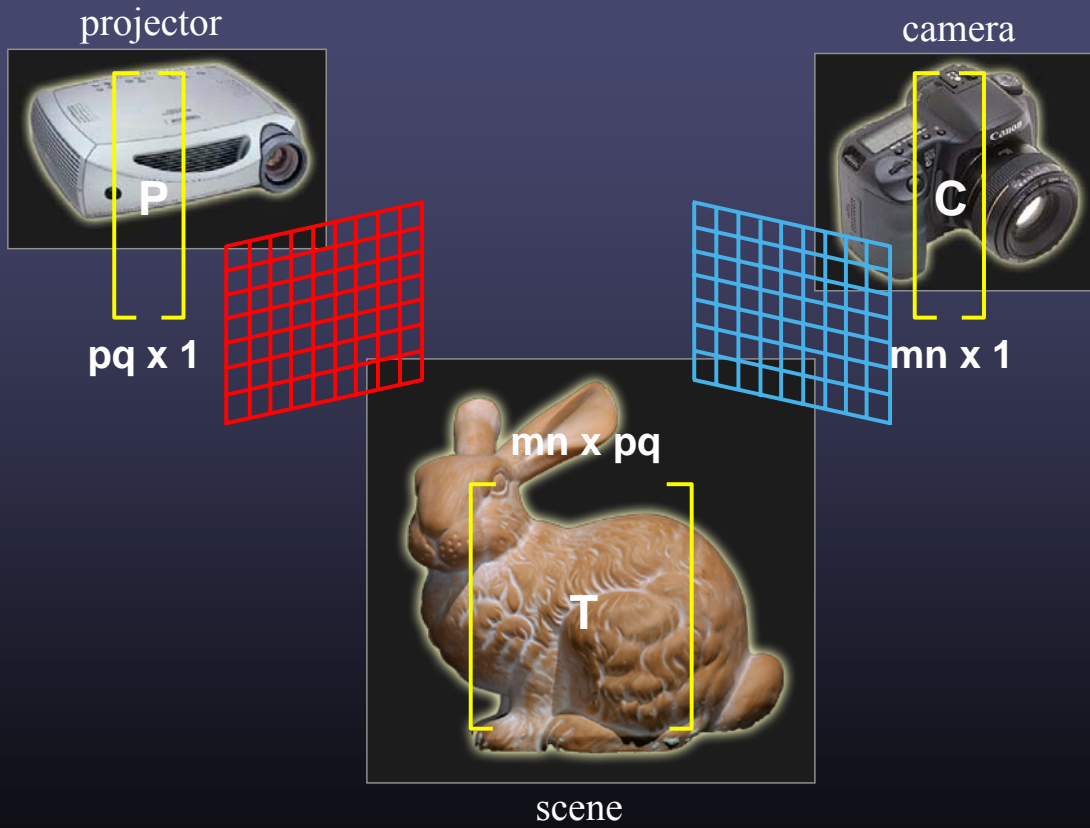


camera



scene

The 4D transport matrix



The 4D transport matrix

$$\begin{array}{c} \left[\begin{array}{c} \mathbf{C} \\ \mathbf{C} \end{array} \right] \\ mn \times 1 \end{array} = \begin{array}{c} mn \times pq \\ \left[\begin{array}{c} \mathbf{T} \\ \mathbf{T} \end{array} \right] \end{array} \begin{array}{c} \left[\begin{array}{c} \mathbf{P} \\ \mathbf{P} \end{array} \right] \\ pq \times 1 \end{array}$$

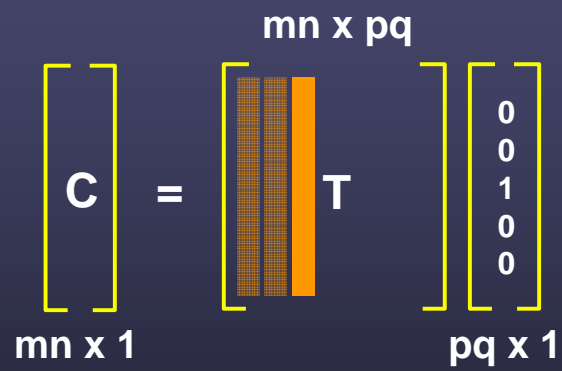
The 4D transport matrix

$$\begin{array}{c} \left[\begin{array}{c} \mathbf{C} \end{array} \right] \\ mn \times 1 \end{array} = \begin{array}{c} mn \times pq \\ \left[\begin{array}{c} \text{orange bar} \end{array} \right] \end{array} \mathbf{T} \begin{array}{c} \left[\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{array} \right] \\ pq \times 1 \end{array}$$

The 4D transport matrix

$$\begin{array}{c} \left[\begin{array}{c} \mathbf{C} \end{array} \right] \\ mn \times 1 \end{array} = \begin{array}{c} mn \times pq \\ \left[\begin{array}{c} \text{matrix} \end{array} \right] \\ \mathbf{T} \end{array} \begin{array}{c} \left[\begin{array}{c} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{array} \right] \\ pq \times 1 \end{array}$$

The 4D transport matrix

$$\begin{array}{c} \left[\begin{array}{c} \mathbf{C} \end{array} \right] \\ mn \times 1 \end{array} = \begin{array}{c} mn \times pq \\ \left[\begin{array}{c} \mathbf{T} \end{array} \right] \end{array} \begin{array}{c} \left[\begin{array}{c} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{array} \right] \\ pq \times 1 \end{array}$$


The 4D transport matrix

$$\begin{array}{c} \left[\begin{array}{c} \mathbf{C} \\ \mathbf{C} \end{array} \right] \\ mn \times 1 \end{array} = \begin{array}{c} mn \times pq \\ \left[\begin{array}{c} \mathbf{T} \\ \mathbf{T} \end{array} \right] \end{array} \begin{array}{c} \left[\begin{array}{c} \mathbf{P} \\ \mathbf{P} \end{array} \right] \\ pq \times 1 \end{array}$$

The 4D transport matrix

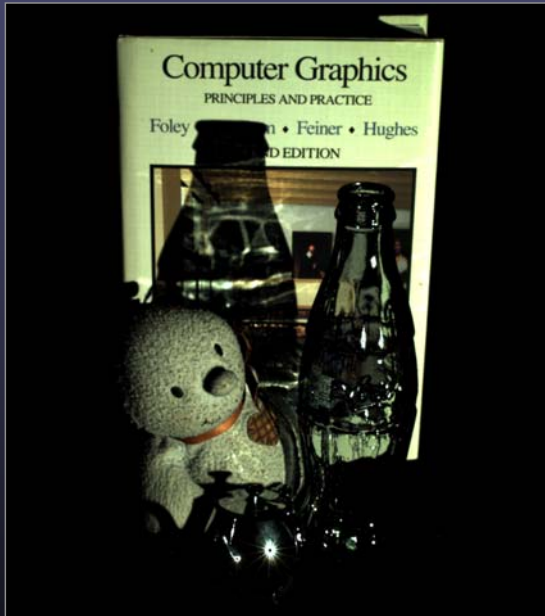
$$\begin{array}{c} \left[\begin{array}{c} \mathbf{C} \\ \mathbf{C} \end{array} \right] \\ mn \times 1 \end{array} = \begin{array}{c} mn \times pq \\ \left[\begin{array}{c} \mathbf{T} \\ \mathbf{T} \end{array} \right] \end{array} \begin{array}{c} \left[\begin{array}{c} \mathbf{P} \\ \mathbf{P} \end{array} \right] \\ pq \times 1 \end{array}$$

applying Helmholtz reciprocity...

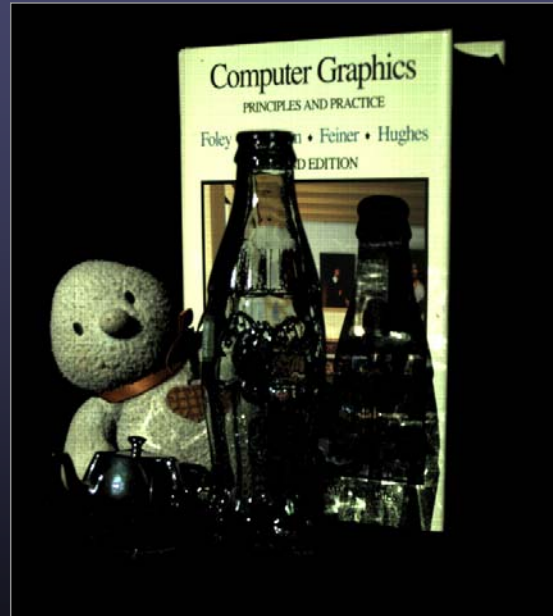
$$\begin{array}{c} \left[\begin{array}{c} \mathbf{C}' \\ \mathbf{C}' \end{array} \right] \\ pq \times 1 \end{array} = \begin{array}{c} pq \times mn \\ \left[\begin{array}{c} \mathbf{T}^T \\ \mathbf{T}^T \end{array} \right] \end{array} \begin{array}{c} \left[\begin{array}{c} \mathbf{P}' \\ \mathbf{P}' \end{array} \right] \\ mn \times 1 \end{array}$$

- This lets us relight the scene
 - as viewed from the projector's position, and
 - not just as illuminated by a uniform point light, but
 - as illuminated by a point source with arbitrary directional control,
 - i.e. as illuminated by a programmable video projector

Example



conventional photograph
with light coming from right



dual photograph
as seen from projector's position

© 2006 Marc Levoy

Video available at
http://graphics.stanford.edu/papers/dual_photography/

Properties of the transport matrix

- little interreflection → sparse matrix
- many interreflections → dense matrix
- convex object → diagonal matrix
- concave object → full matrix

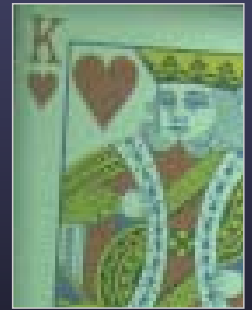
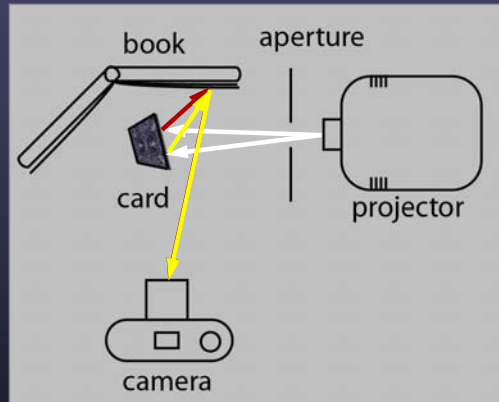
Can we create a dual photograph entirely from diffuse reflections?

© 2006 Marc Levoy

Dual photography from diffuse reflections



the camera's view



The relighting problem

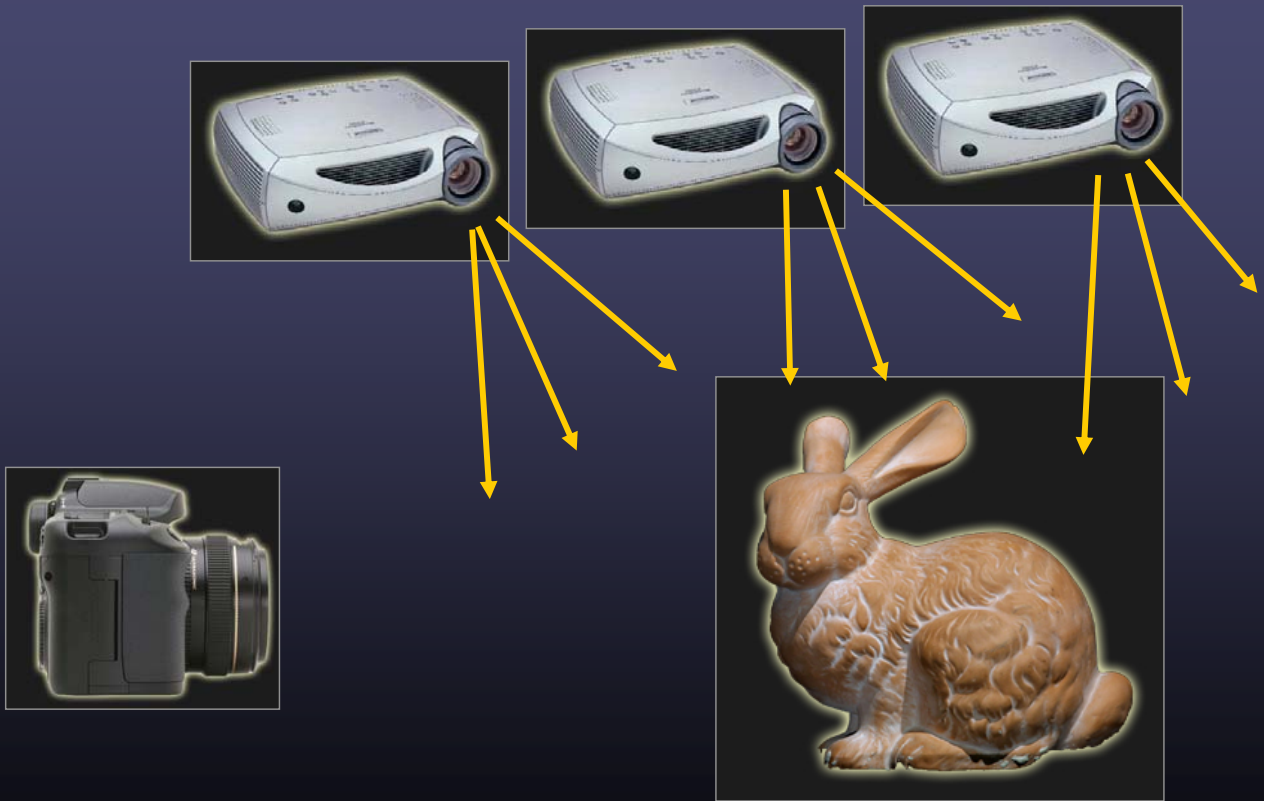


Paul Debevec's
Light Stage 3

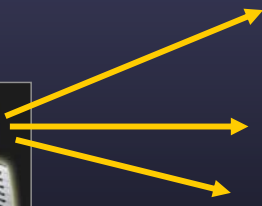
- subject captured under multiple lights
- one light at a time, so subject must hold still
- point lights are used, so can't relight with cast shadows

© 2006 Marc Levoy

The 6D transport matrix



The 6D transport matrix



The advantage of dual photography

- capture of a scene as illuminated by different lights cannot be parallelized
- capture of a scene as viewed by different cameras can be parallelized

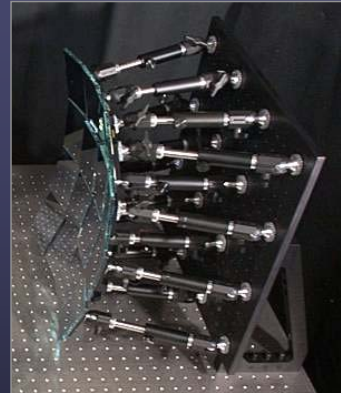
© 2006 Marc Levoy

Measuring the 6D transport matrix

projector



mirror array



scene

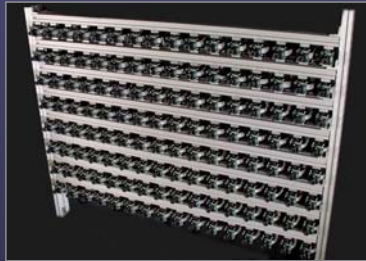


Relighting with complex illumination

projector



camera array



scene

$$\begin{matrix} \left[\begin{matrix} C' \end{matrix} \right] \\ pq \times 1 \end{matrix} = \begin{matrix} pq \times mn \times uv \\ \left[\begin{matrix} T^T \end{matrix} \right] \end{matrix} \begin{matrix} \left[\begin{matrix} P' \end{matrix} \right] \\ mn \times uv \times 1 \end{matrix}$$



- step 1: measure 6D transport matrix T
- step 2: capture a 4D light field
- step 3: relight scene using captured light field

© 2006 Marc Levoy

Running time

- the different rays within a projector can in fact be parallelized to some extent
- this parallelism can be discovered using a coarse-to-fine adaptive scan
- can measure a 6D transport matrix in 5 minutes

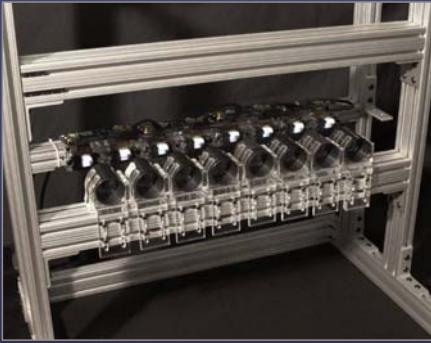
© 2006 Marc Levoy

- 5 minutes

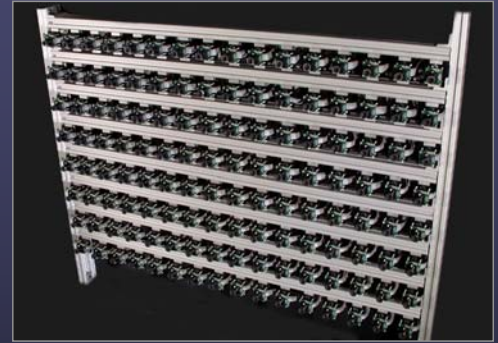
- using a video-rate camera
- and (effectively) measuring 1M x 1M transport entries
- for scenes having average amounts of diffuse interreflection
- everything depends on the **density of the T matrix**

Can we measure an 8D transport matrix?

projector array



camera array



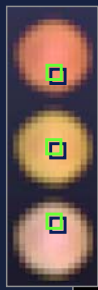
scene

- 8D transport matrix

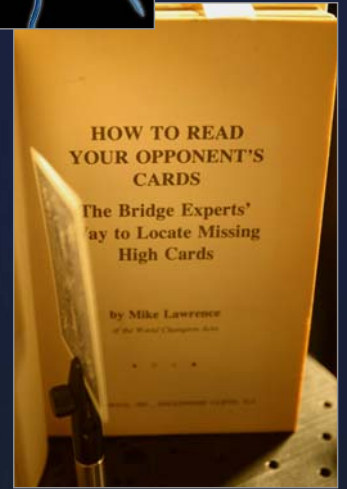
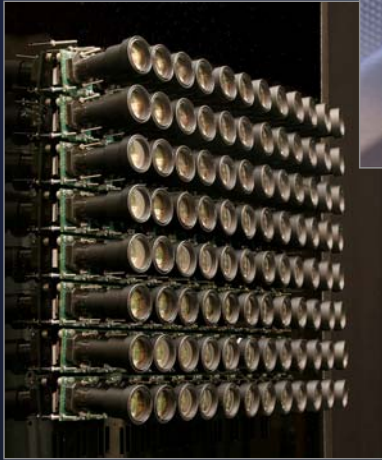
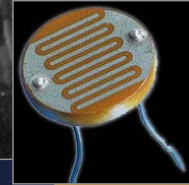
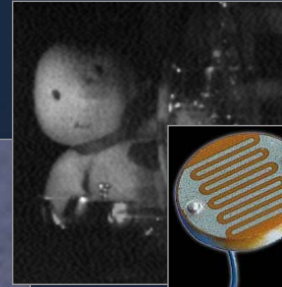
- the full scattering function

- if this were a surface, we'd call it the BSSRDF

- what should we call this? the **bidirectional light field transport distribution function (BLFTDF)?**



$$\begin{matrix} \left[\begin{matrix} C \\ \end{matrix} \right] \\ mn \times 1 \end{matrix} = \begin{matrix} \left[\begin{matrix} T \\ \end{matrix} \right] \left[\begin{matrix} P \\ \end{matrix} \right] \\ mn \times pq \quad pq \times 1 \end{matrix}$$



<http://graphics.stanford.edu>

Computational Imaging in the Sciences (and Medicine)

Marc Levoy

Due to copyright restrictions, some images have been removed from this version of the slides.
To see presentation with these images intact, go to:

<http://graphics.stanford.edu/courses/cs448a-06-winter/>

and look under the heading “Lectures for SIGGRAPH 2006 course on Computational Photography”.



Computer Science Department
Stanford University

- Based on lectures given in:
 - Stanford CS 448A (Computational Photography), Winter quarter 2006

34:15 total + 30% = ~45 minutes

Time =

Some examples

- medical imaging

- rebinning ← *inspiration for light field rendering*
- ✓ – transmission tomography
- ✗ – reflection tomography (for ultrasound)

- geophysics

- ✓ – borehole tomography
- ✗ – seismic reflection surveying

- applied physics

- ✓ – diffuse optical tomography
 - ✓ ✗ – diffraction tomography
 - inverse scattering
- ✓ in this lecture
✗ time-of-flight or wave-based

©2006 Marc Levoy

- biology

- ✓ – confocal microscopy ← *applicable at macro scale too*

- ✓ – deconvolution microscopy ← *related to tomography*

- astronomy

- ✓ – coded-aperture imaging

- ✗ – interferometric imaging

- airborne sensing

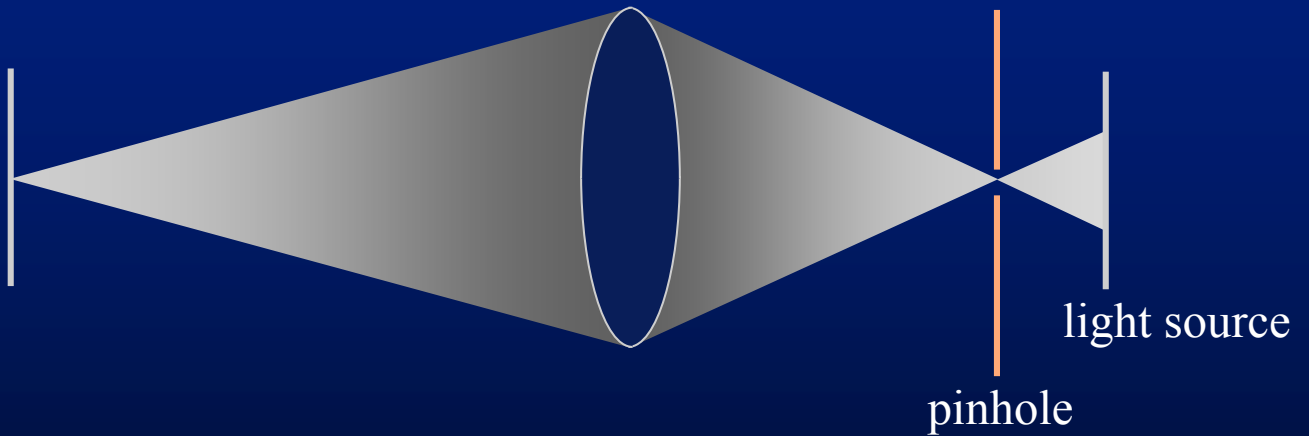
- multi-perspective panoramas

- ✓ ✗ – synthetic aperture radar

- optics

- ✓ ✗ – holography
 - wavefront coding

Confocal scanning microscopy



© 2004 Marc Levoy

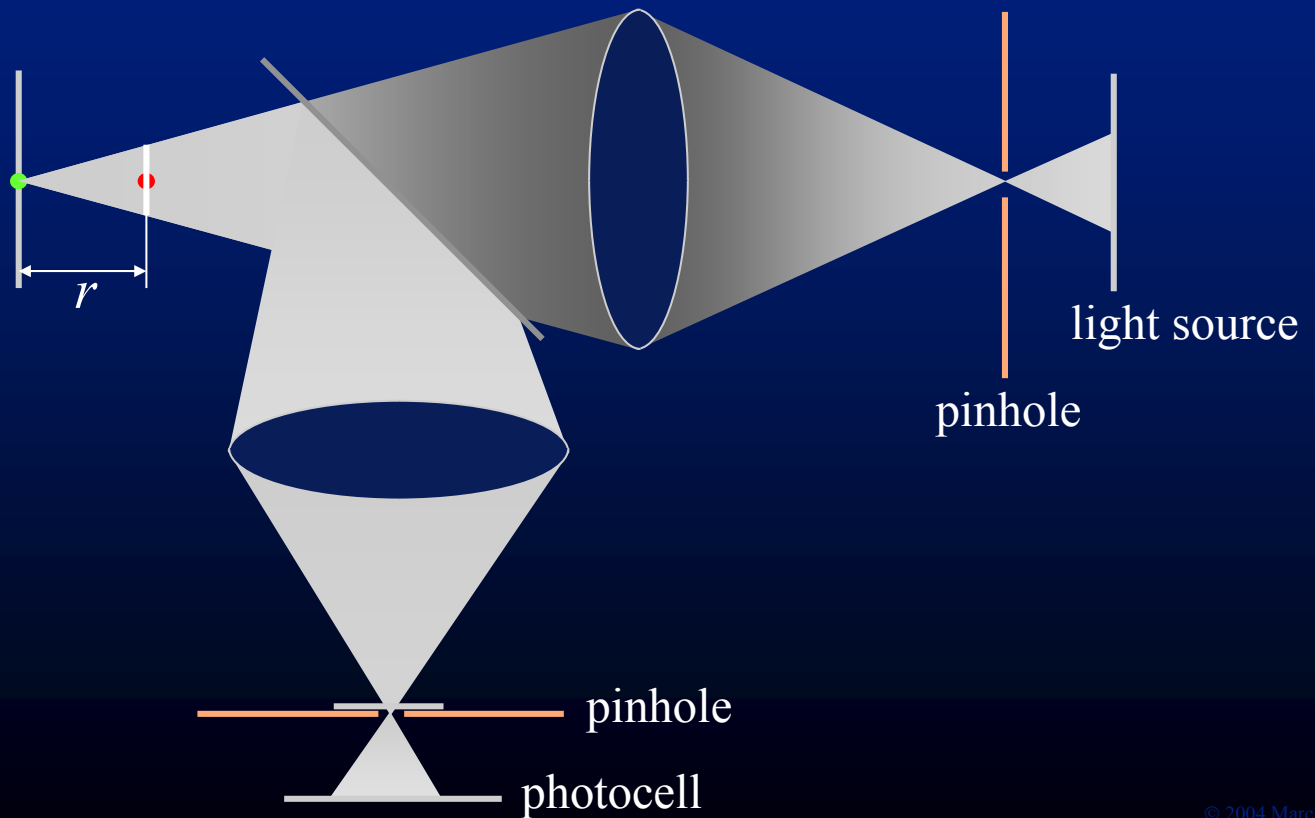
- typical reference:

- Corle, T.R.. Kino, G.S. *Confocal Scanning Optical Microscopy and Related Imaging Systems*, Academic Press, 1996.

- if you introduce a **pinhole**

- only one point on the focal plane will be illuminated

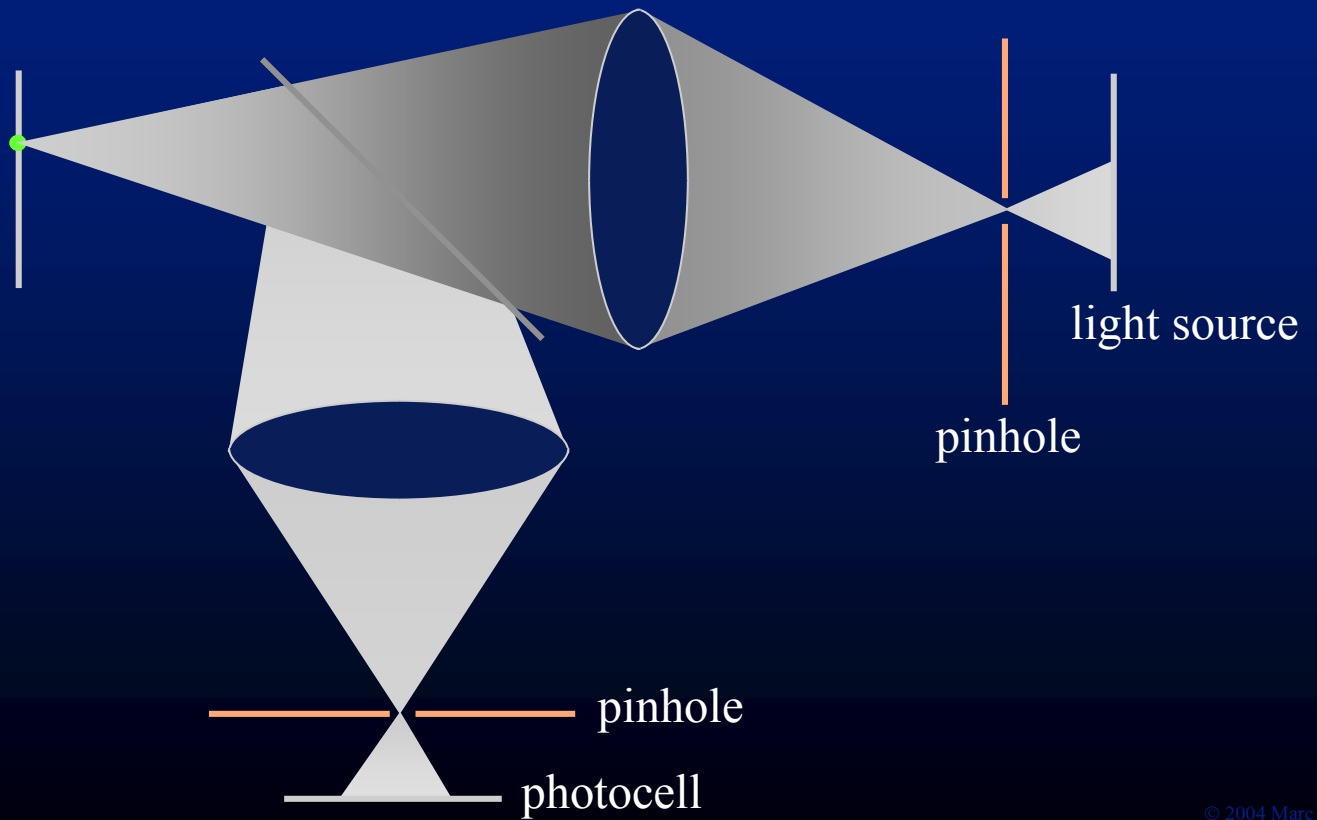
Confocal scanning microscopy



© 2004 Marc Levoy

- ...and a **matching optical system**,
 - hence the word **confocal**
- this **green dot**
 - will be both **strongly illuminated** and **sharply imaged**
- while this **red dot**
 - will have **less light** falling on it by the **square of distance r** ,
 - because the light is spread over a disk
 - and it will also be **more weakly imaged** by the **square of distance r** ,
 - because its image is blurred out over an disk on the pinhole mask, and only a little bit is permitted through
- so the extent to which the red dot contributes to the final image
 - falls off as the **fourth power of r** , the distance from the focal plane

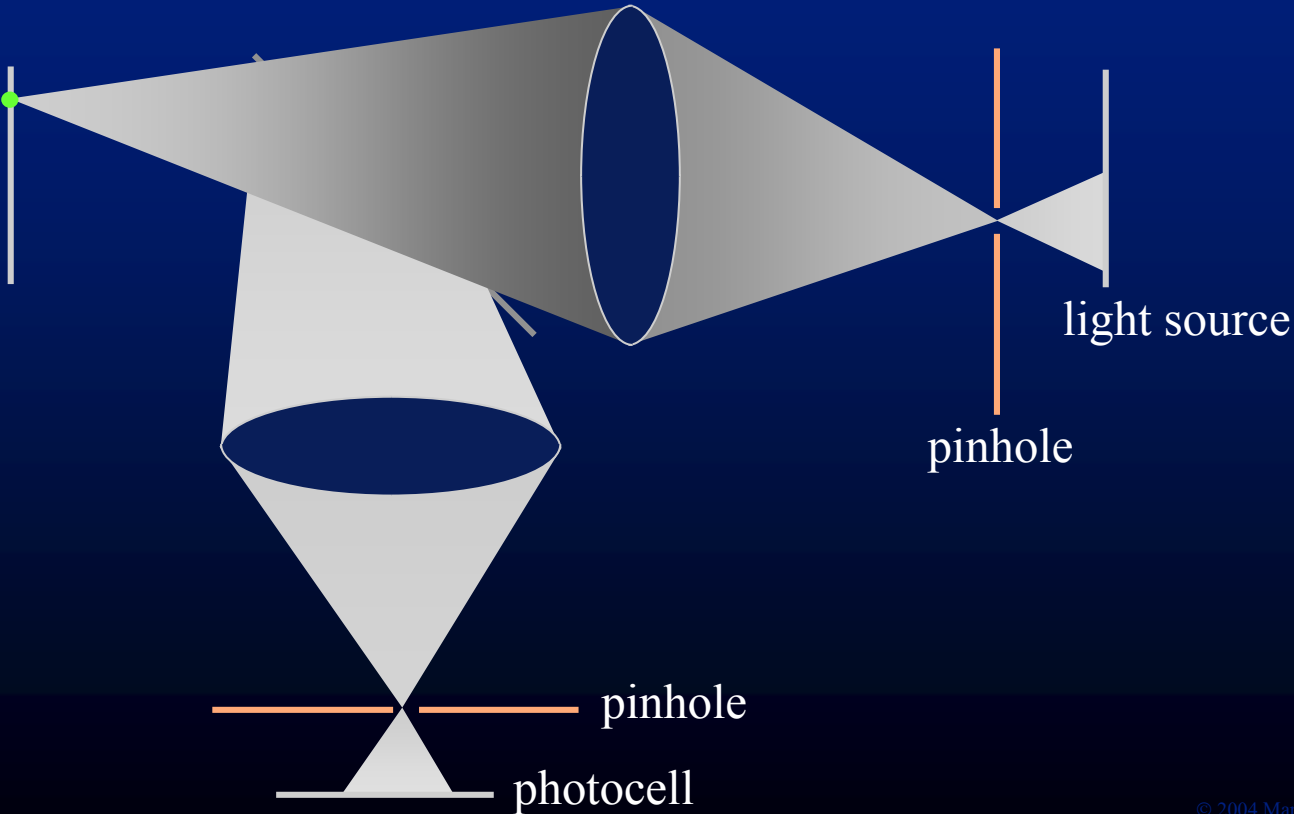
Confocal scanning microscopy



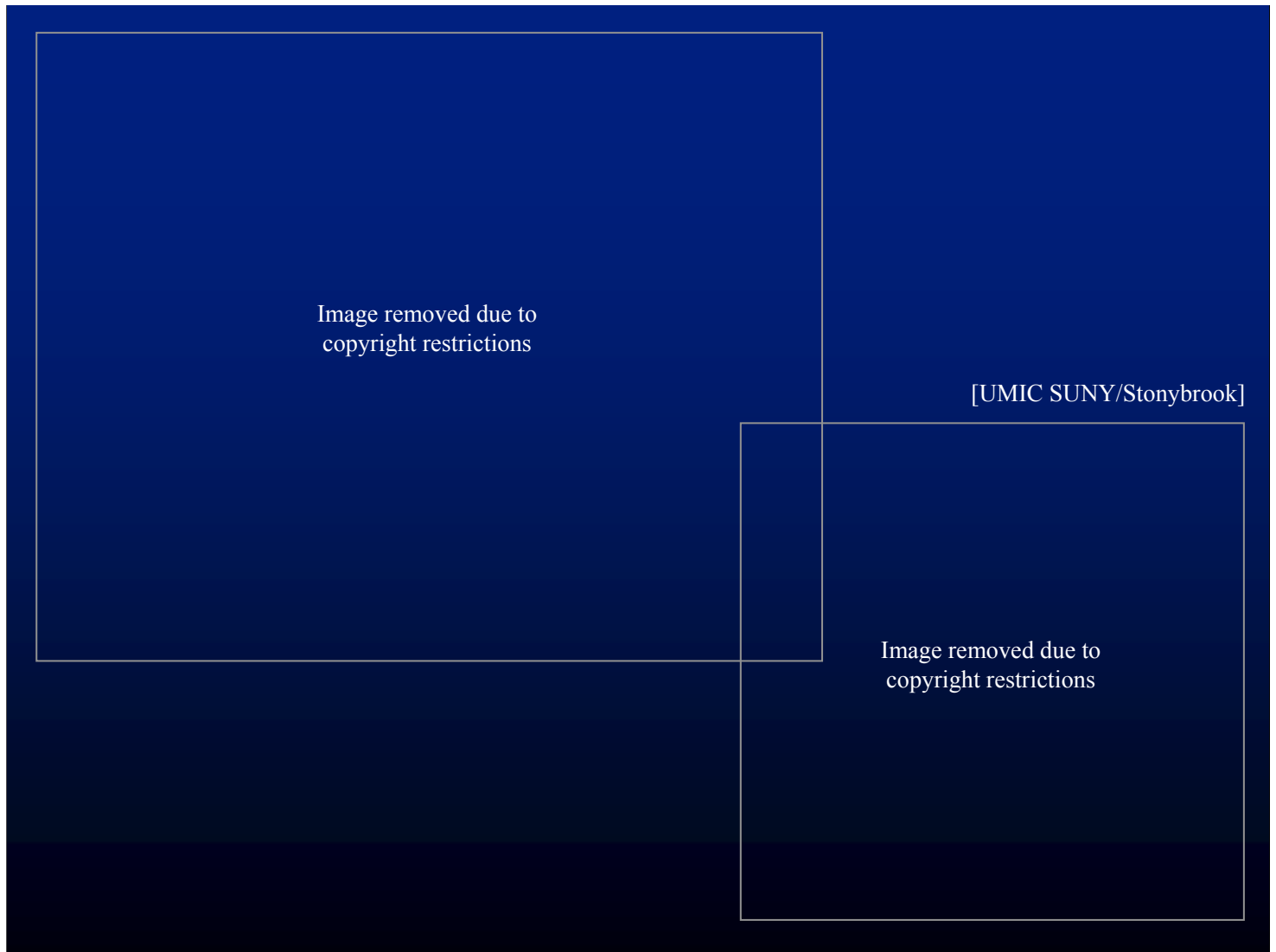
© 2004 Marc Levoy

- of course, you've only imaged one point
 - so you need to **move the pinholes**
 - and scan across the focal plane

Confocal scanning microscopy



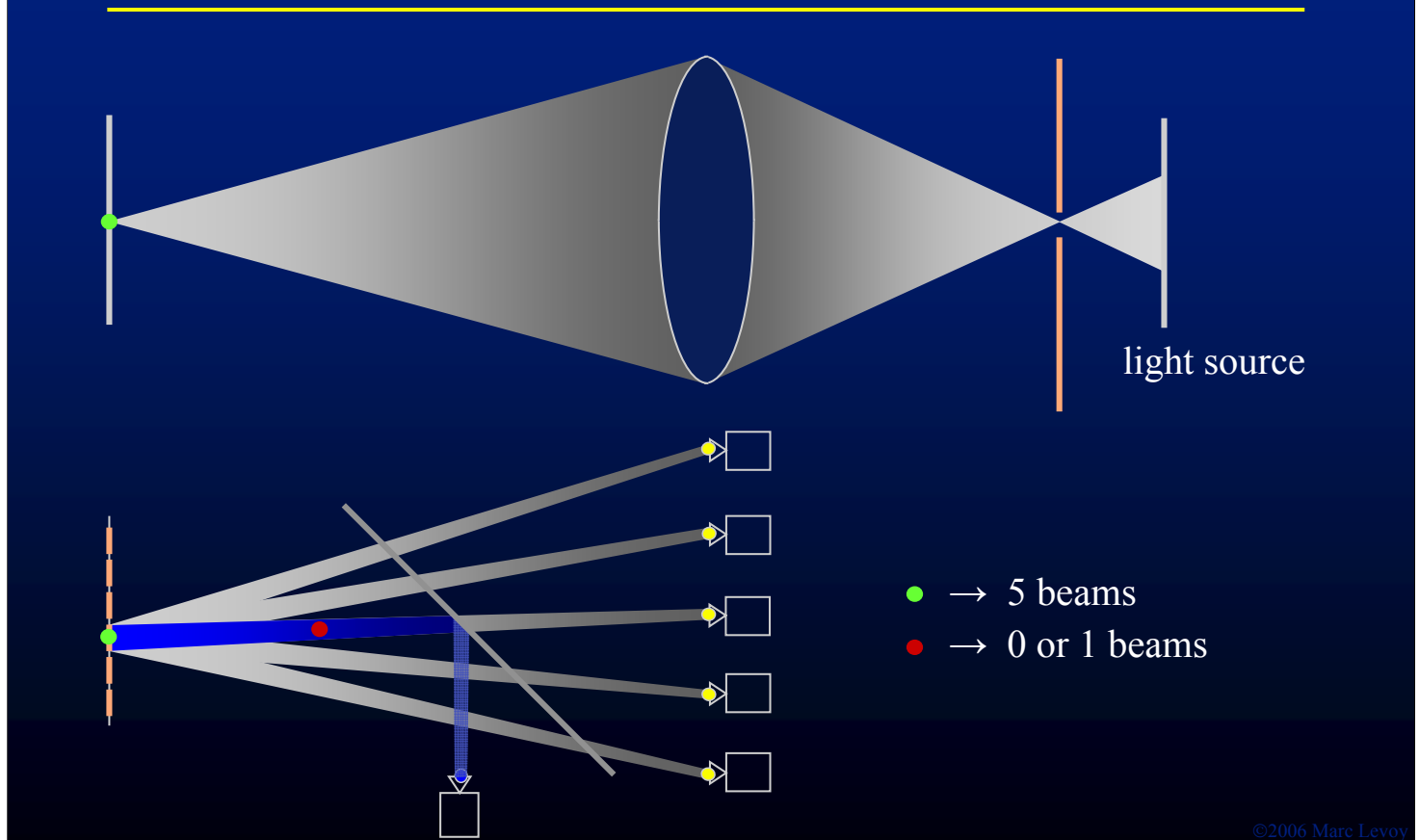
© 2004 Marc Levoy



- the object in the lower-right image is actually **spherical**,
 - but portions of it that are off the focal plane are both **blurry** and **dark**,
 - effectively disappearing

Synthetic confocal scanning

[Levoy 2004]



- our goal

- is to approximate this effect **at the large scale**

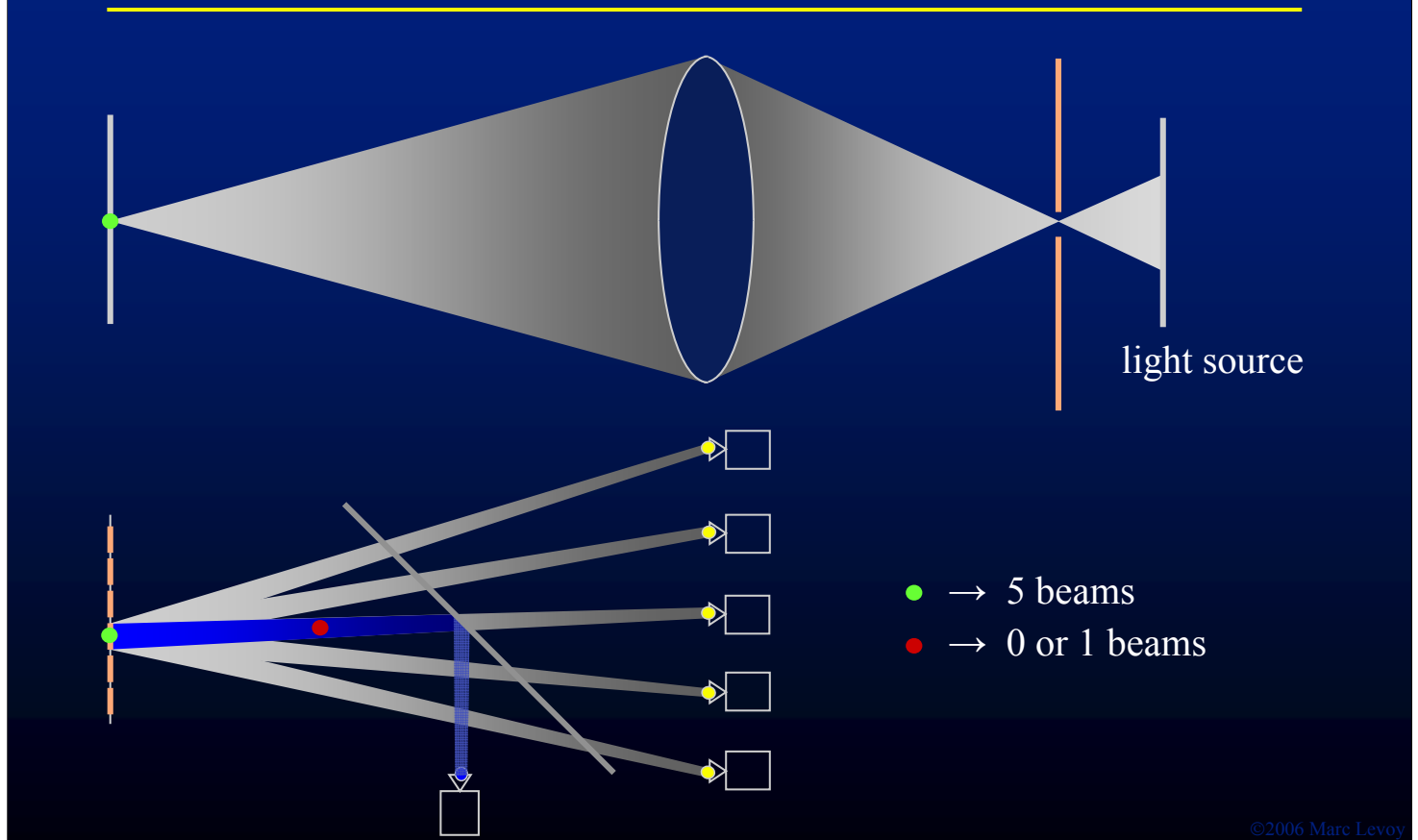
- we can understand the photometry of this setup

- using a **simplified counting argument**

- Reference:

- Levoy, M., Chen, B., Vaish, V., Horowitz, M., McDowall, I., Bolas, M., Synthetic aperture confocal imaging, *Proc. SIGGRAPH 2004*.

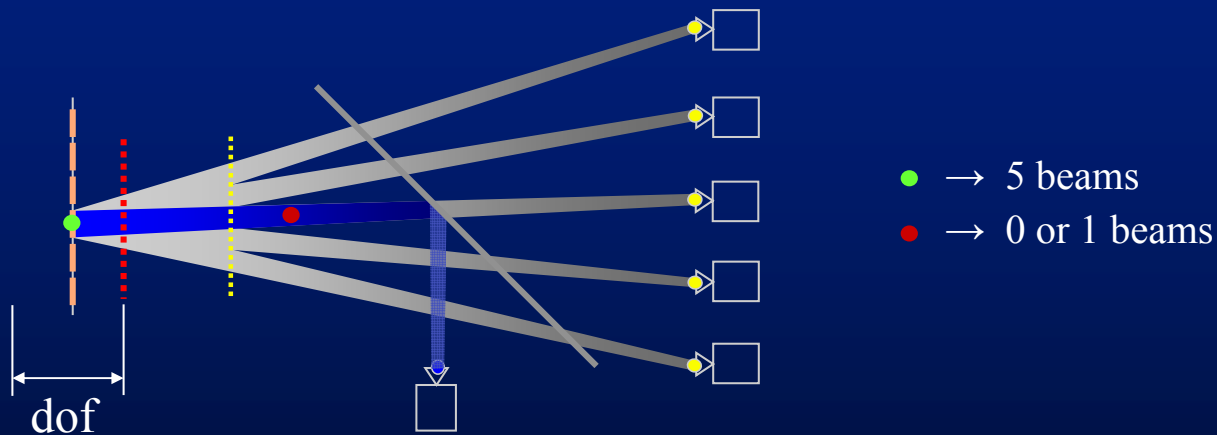
Synthetic confocal scanning



•5:0 or 5:1

–if we had 5 cameras as well as 5 projectors,
then the ratio would be 25:0 or **25:1**

Synthetic confocal scanning



- works with any number of projectors ≥ 2
- discrimination degrades if • point to left of
- no discrimination for • points to left of
- slow!
- poor light efficiency

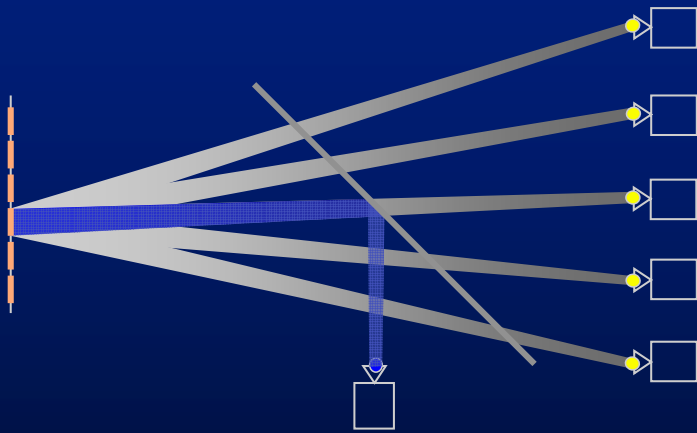
©2006 Marc Levoy

•depth of field

–a microscoper would call it the **axial resolution**

–to make the depth of field **shallower**, **spread out** the projectors, i.e. a larger synthetic aperture

Synthetic coded-aperture confocal imaging



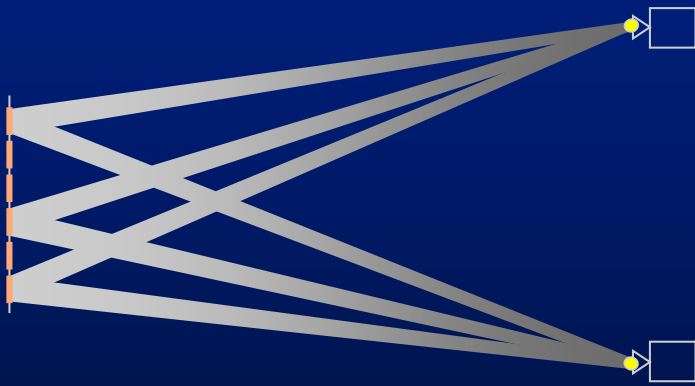
- different from coded aperture imaging in astronomy
- [Wilson, Confocal Microscopy by Aperture Correlation, 1996]

©2006 Marc Levoy

•Reference:

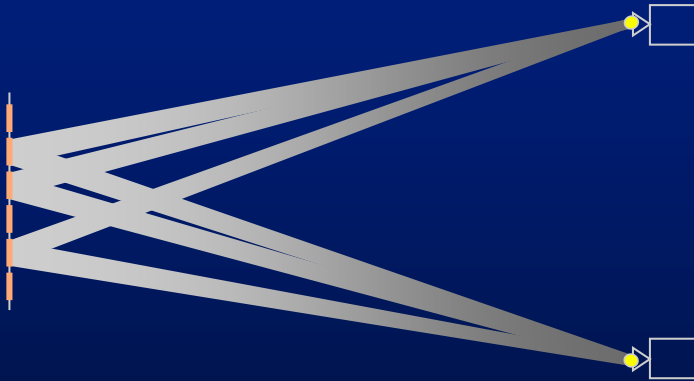
–Wilson, T., Juskaitis, R., Neil, M.A.A.,
Kozubek, M., Confocal microscopy by
aperture correlation,
Optics Letters, Vol. 21, No. 23, December 1,
1996.

Synthetic coded-aperture confocal imaging



©2006 Marc Levoy

Synthetic coded-aperture confocal imaging



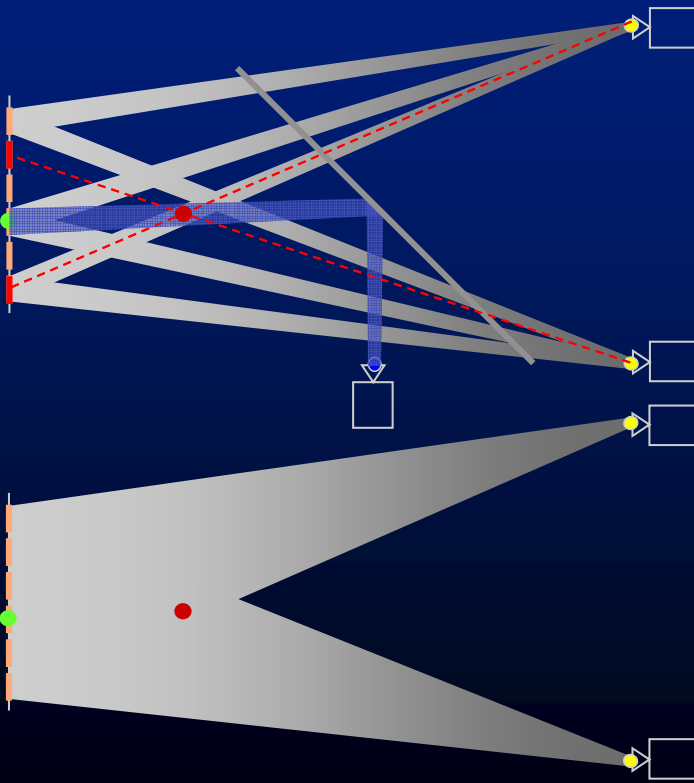
©2006 Marc Levoy

Synthetic coded-aperture confocal imaging



©2006 Marc Levoy

Synthetic coded-aperture confocal imaging



100 trials

- → 2 beams \times 50/100 trials = 1
- → \sim 1 beam \times 50/100 trials = 0.5

floodlit

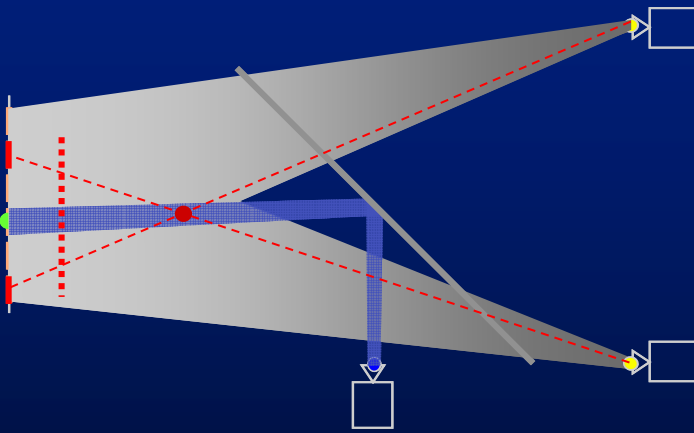
- → 2 beams
- → 2 beams

trials - $\frac{1}{4} \times$ floodlit

- → $1 - \frac{1}{4}(2) = 0.5$
- → $0.5 - \frac{1}{4}(2) = 0$

©2006 Marc Levoy

Synthetic coded-aperture confocal imaging



- works with relatively few trials (~16)
- 50% light efficiency
- works with any number of projectors ≥ 2
- discrimination degrades if \bullet point vignettted for some projectors
- no discrimination for \bullet points to left of
- needs patterns in which illumination of tiles are uncorrelated

100 trials

- $\rightarrow 2 \text{ beams} \times 50/100 \text{ trials} = 1$
- $\rightarrow \sim 1 \text{ beam} \times 50/100 \text{ trials} = 0.5$

floodlit

- $\rightarrow 2 \text{ beams}$
- $\rightarrow 2 \text{ beams}$

trials $= \frac{1}{4} \times \text{floodlit}$

- $\rightarrow 1 - \frac{1}{4}(2) = 0.5$
- $\rightarrow 0.5 - \frac{1}{4}(2) = 0$

©2006 Marc Levoy

•note all the **tildas** in the formulas

- this algorithm is **statistical** in nature
- for example, if we **flip a coin** to decide whether to illuminate a particular tile on a particular trial
- the **binomial theorem** tells us how much **variability** we'll get over a given number of trials

•the effect of this variability

- the image of our focal plane will be slightly **non-uniform**, and
- objects off the focal plane won't be entirely **dark** after the confocal subtraction

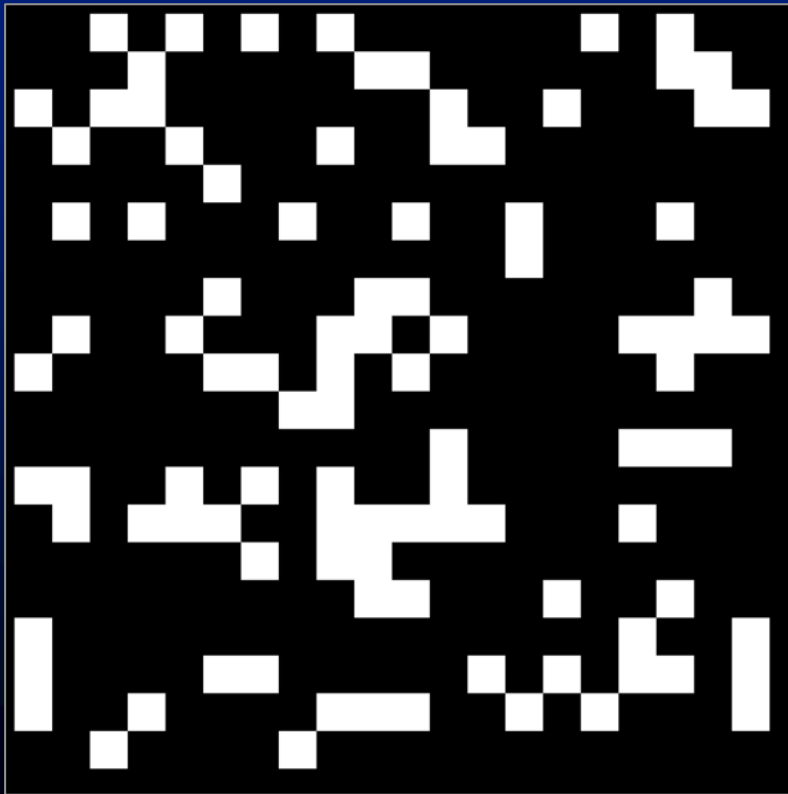
•but for visual purposes

- our technique works well with a **modest number of trials**, like 16
- far fewer than would be required to **scan** out the focal plane, as in the usual confocal scanning algorithm
- we need patterns in which the **illumination of different tiles are spatially uncorrelated**

•patterns

- Hadamard patterns don't work well because
 - »they need to be square
 - »they are too structured, creating aliasing on foreground objects

Example pattern



©2006 Marc Levoy

What are good patterns?

pattern

one trial

16 trials



(a) pseudo-random tiling



(b) randomly permuted tiling

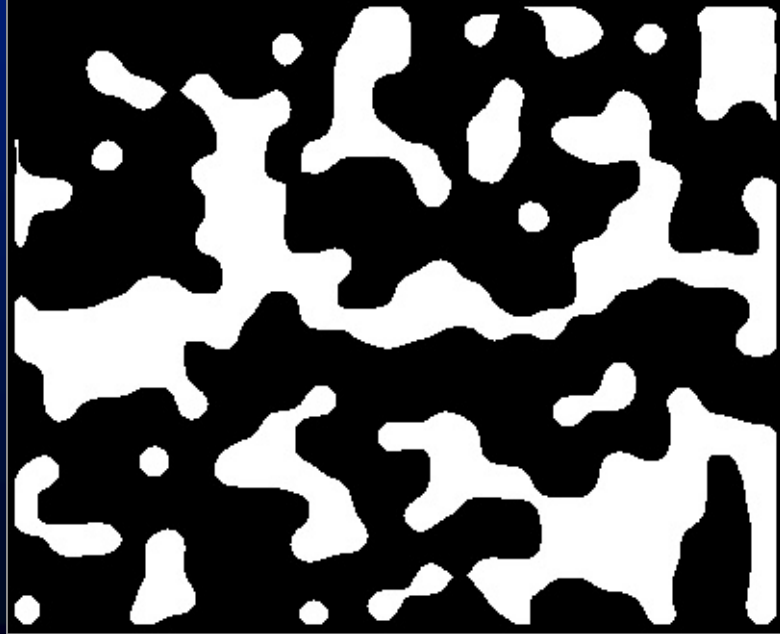


(c) randomly placed tiles



(d) sinuous patterns

Patterns with less aliasing



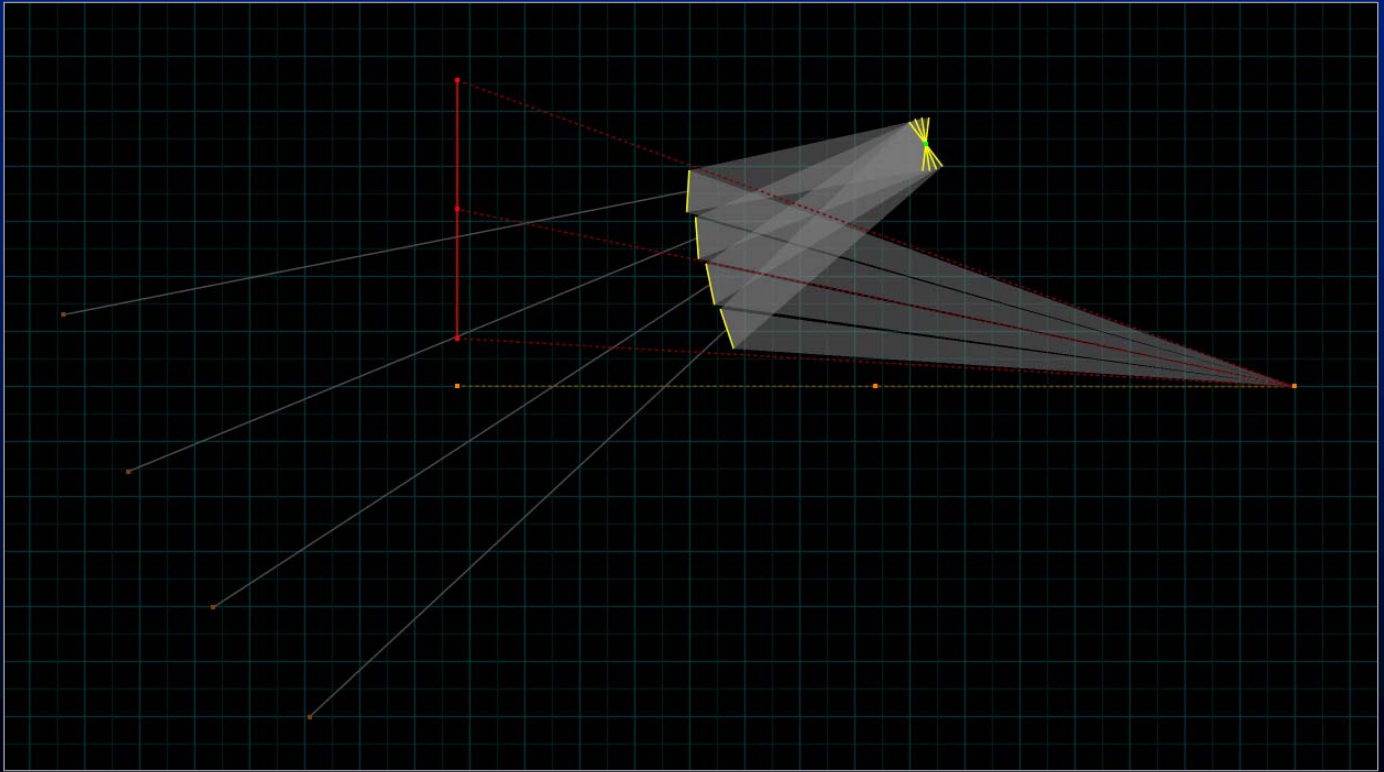
multi-phase
sinusoids?
[Neil 1997]

©2006 Marc Levoy

•reference:

–Neil, M.A.A., Juskaitis, R., Wilson, T.,
Method of obtaining optical sectioning by
using structured light in a conventional
microscope,
Optics Letters, Vol. 22, No. 24, December
15, 1997.

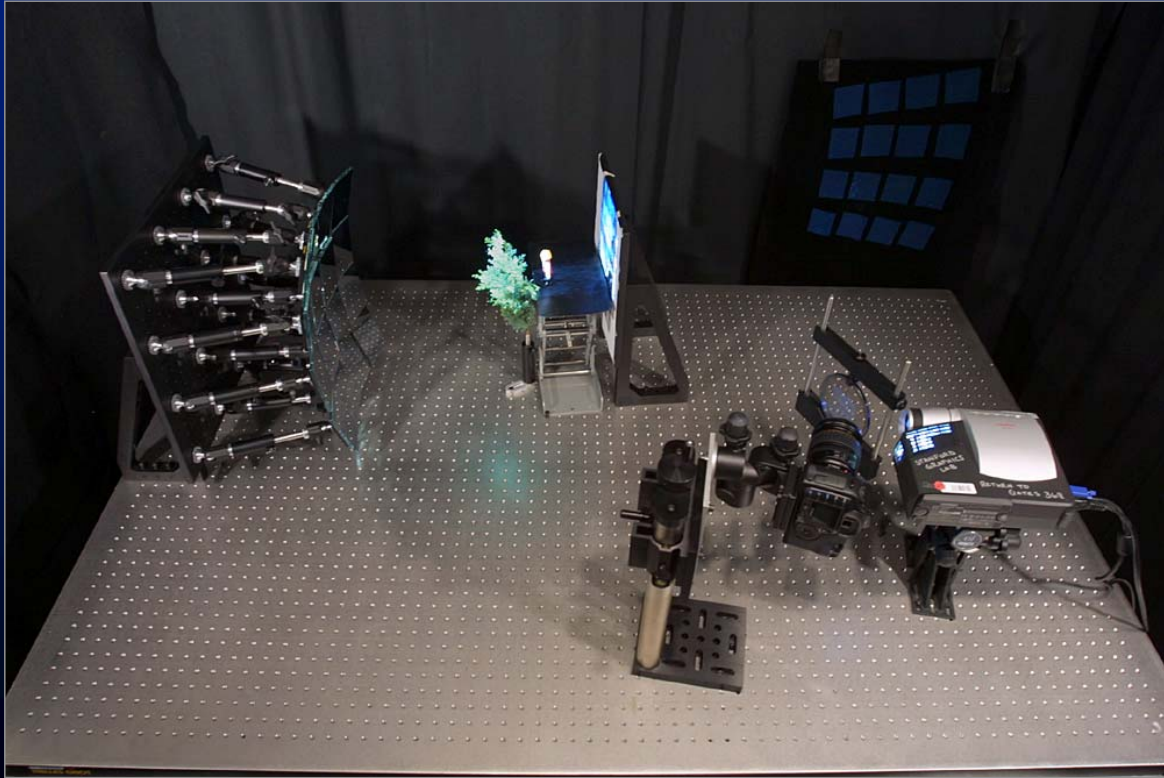
Implementation using an array of mirrors



©2006 Marc Levoy

- video projector is orange dot at right, off-axis screen (not used) is red line, mirrors are yellow lines, scene is green dot, focal planes are superimposed yellow lines at green dot
- virtual projectors are orange dots at left
- this is an interactive program that let's you adjust all the parameters
- main tradeoff is between angular spread of rays arriving at scene versus depth of field of patterns arriving there

Implementation using an array of mirrors



©2006 Marc Levoy

- URL of SIGGRAPH 2004 paper and movie is:
–<http://graphics.stanford.edu/papers/confocal/>

Confocal imaging in scattering media



- small tank
 - too short for attenuation
 - lit by internal reflections

© 2006 Marc Levoy

- I observed a confocal effect
 - but it was **modest**
 - theory said the effect **should be stronger**

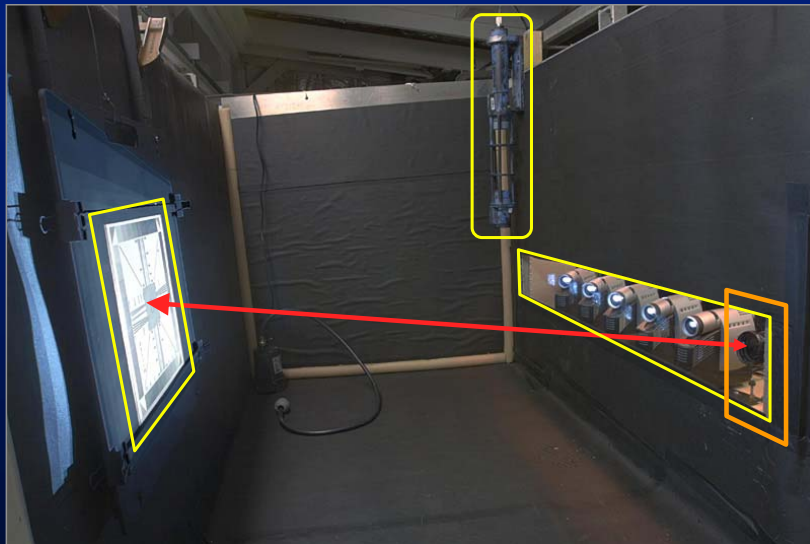
Experiments in a large water tank



50-foot flume at Wood's Hole Oceanographic Institution (WHOI)

© 2004 Marc Levoy

Experiments in a large water tank



- 4-foot viewing distance to target
- surfaces blackened to kill reflections
- titanium dioxide in filtered water
- transmissometer to measure turbidity

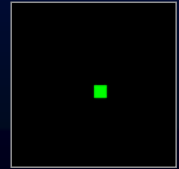
© 2004 Marc Levoy

- titanium dioxide
 - the stuff in **white paint**

Experiments in a large water tank

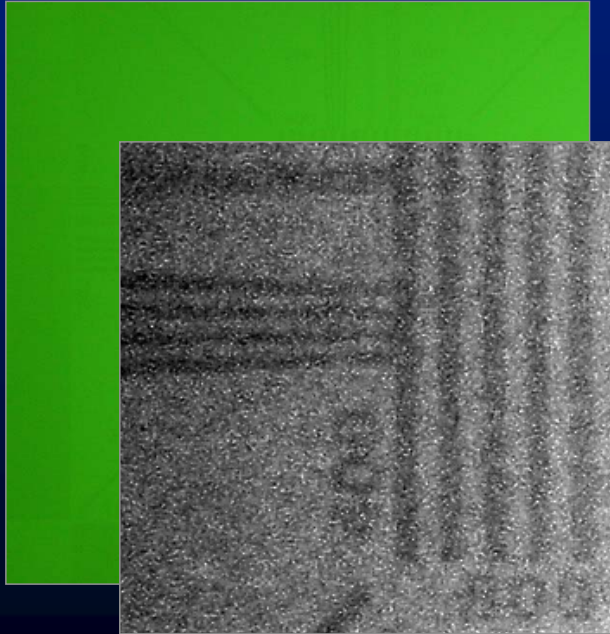


- stray light limits performance
- one projector suffices if no occluders

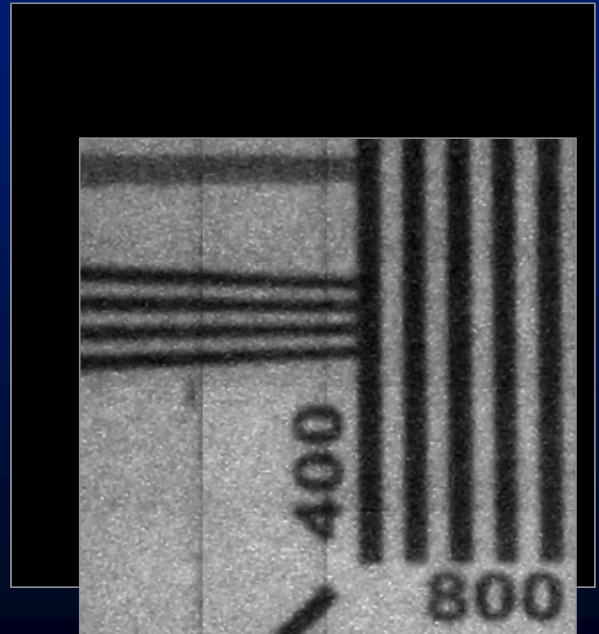


© 2004 Marc Levoy

Seeing through turbid water



floodlit

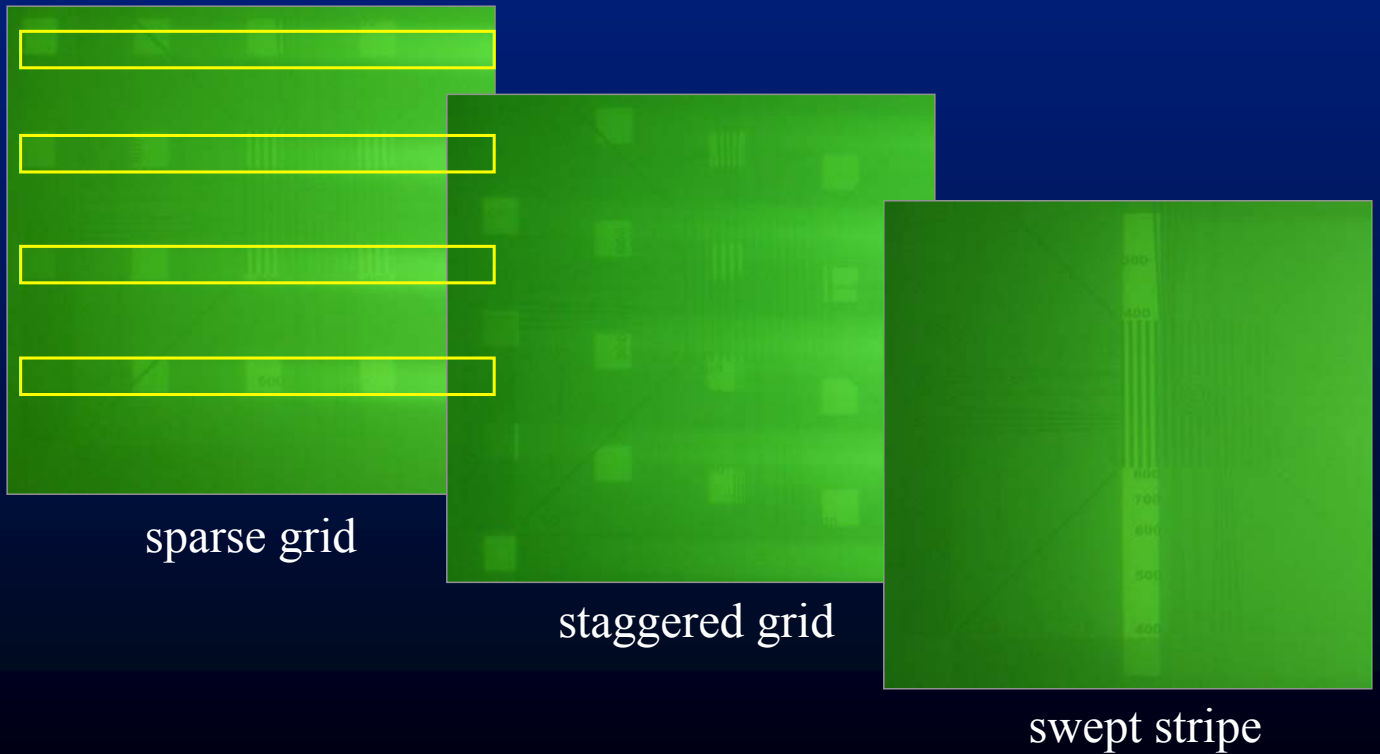


scanned tile

© 2004 Marc Levoy

- this is very turbid water
 - the “**attenuation length**” (a technical term that roughly translates to “how far you can see clearly”) is about **8 inches**
 - and I’m trying to see through **4 feet**
- if you contrast enhance these images,
 - you can see the improvement in **signal-to-noise ratio**
- Reference:
 - M. Levoy, Improving underwater vision using confocal imaging. In preparation.

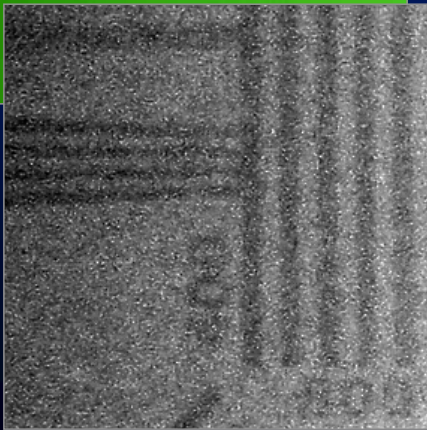
Other patterns



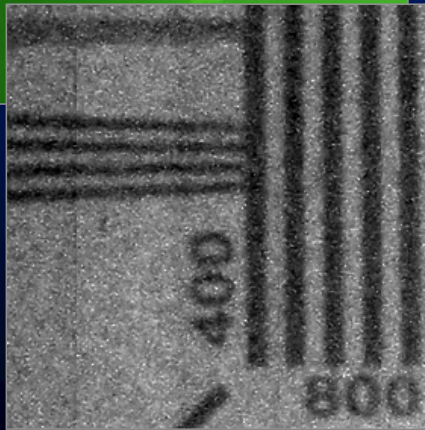
© 2004 Marc Levoy

- to speed things up,
 - one can use patterns that illuminate **several tiles at once**
 - similar strategies have been used in confocal microscopy
- the problem with this one
 - the illumination beams (coming in from the right side) **intersect the lines of sight** to other tiles, **degrading contrast**
- here's a pattern in which they don't
 - a **staggered grid**
- here's another pattern in which they don't
 - a simple **swept stripe**
 - with the light coming in from the right side

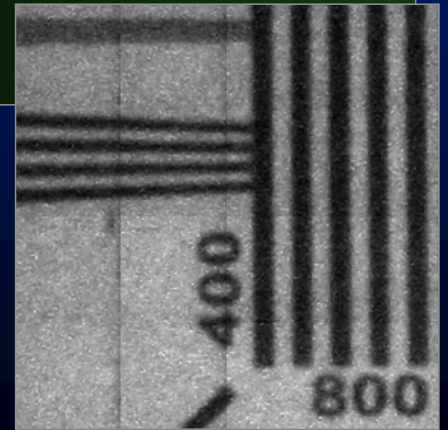
Other patterns



floodlit



swept stripe



scanned tile

© 2004 Marc Levoy

- here's how a swept stripe stacks up against the other patterns
 - somewhere **in the middle**, in terms of quality
- a swept stripe has a number of **advantages**, though...

Stripe-based illumination

- if vehicle is moving, no sweeping is needed!
- can triangulate from leading (or trailing) edge of stripe, getting range (depth) for free

Image removed due to
copyright restrictions

© 2004 Marc Levoy

- no sweeping is needed
 - the **forward motion of the vehicle** will sweep out the stripe!
- in addition, since the **stripe is coming from the side...**
 - essentially constitutes a **structured-light rangefinder**
- this is not a new idea
 - Jules Jaffe** proposed it 14 years ago
 - »Jaffe, J.S., Computer modeling and the design of optimal underwater imaging systems, *IEEE J. Oceanic Eng.* **15**(2), 101-111 (1990).
 - but he had no technology to implement it
 - compact video projectors** provide that technology
- so we can easily envision
 - video projectors being mounted on **future underwater vehicles**
 - this is the **Hercules** remotely operated vehicle
 - exploring the wreck of the **Titanic** two months ago in the North Atlantic

Application to underwater exploration

Image removed due to
copyright restrictions

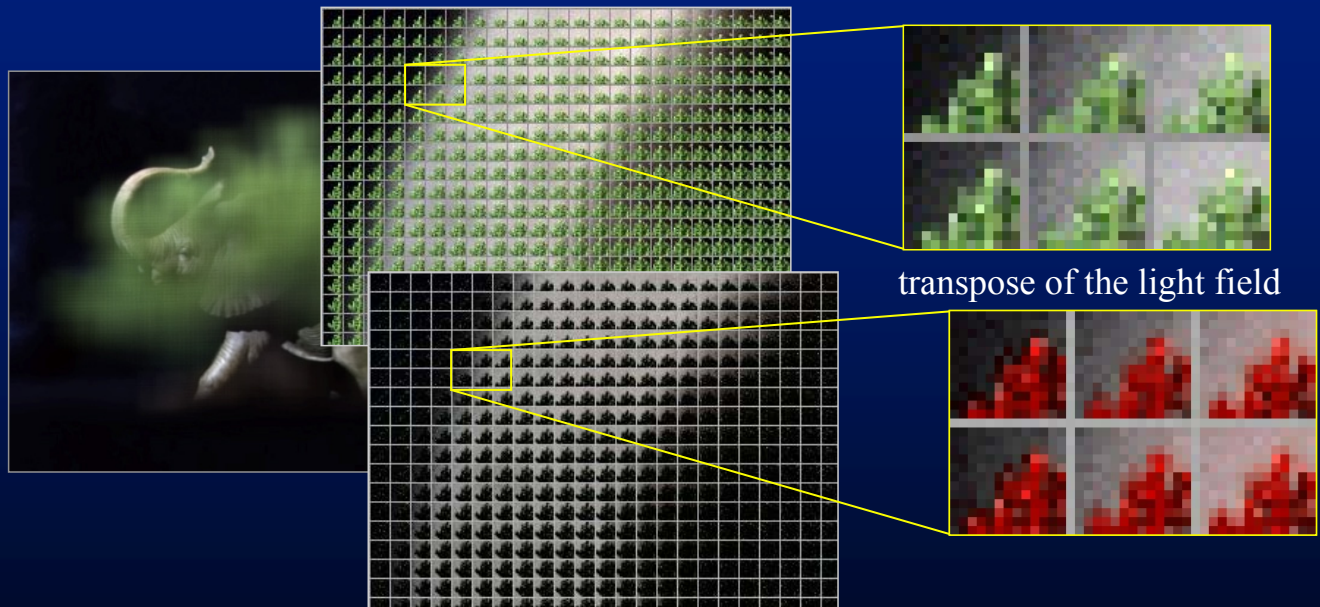
[Ballard/IFE 2004]

Image removed due to
copyright restrictions

© 2004 Marc Levoy

- so I think we can expect
 - video projectors being mounted on **future underwater vehicles**
 - this is the **Hercules** remotely operated vehicle
 - exploring the wreck of the **Titanic** two months ago in the North Atlantic
 - Pictures from Robert Ballard, Institute for Exploration, 2004.
- the question is
 - can you produce an **overhead view** like this, of the Titanic
 - in a **single shot** taken from **far away** using **shaped illumination**
 - rather than by **mowing the lawn** with the underwater vehicle
 - which is **difficult, dangerous, time consuming**, and produces a **mosaic with parallax errors**

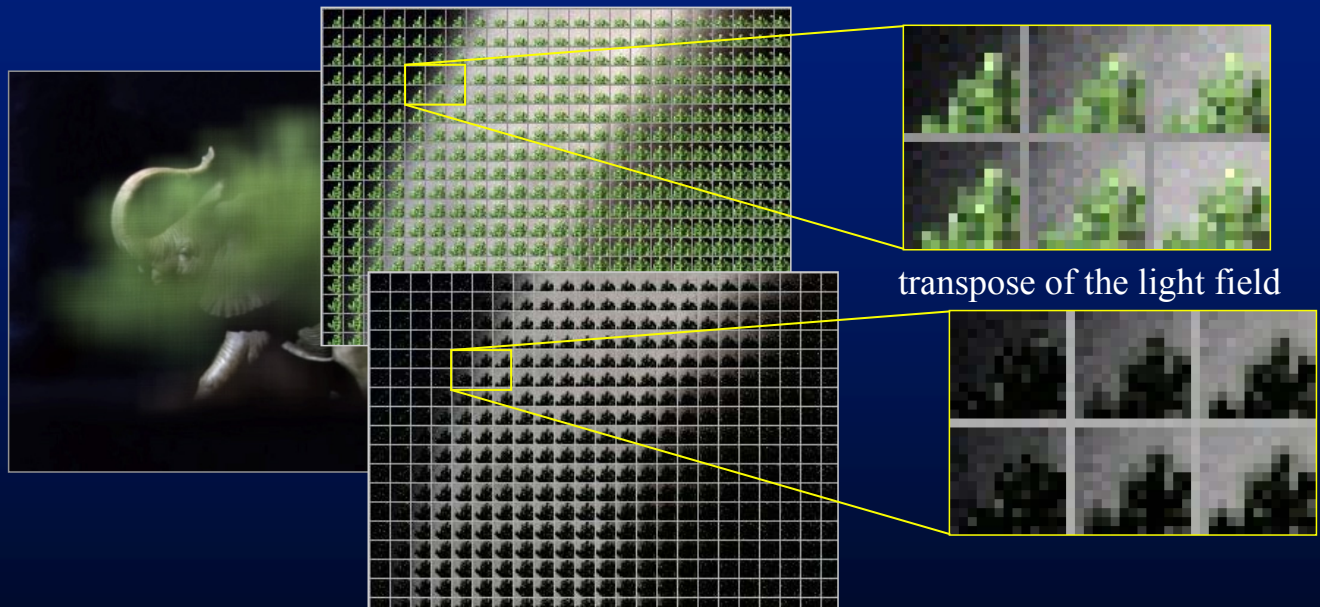
Shaped illumination in a computer vision algorithm



- low variance within one block = stereo constraint
- sharp differences between adjacent blocks = focus constraint
- both algorithms are confused by occluding objects

© 2006 Marc Levoy

Shaped illumination in a computer vision algorithm



transpose of the light field

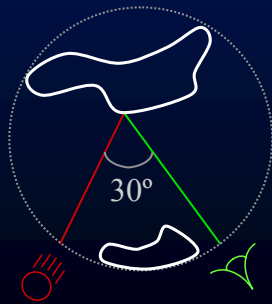
- confocal estimate of projector mattes → re-shape projector beams
- re-capture light field → run vision algorithm on new light field
- re-estimate projector mattes from model and iterate

© 2006 Marc Levoy

Confocal imaging versus triangulation rangefinding

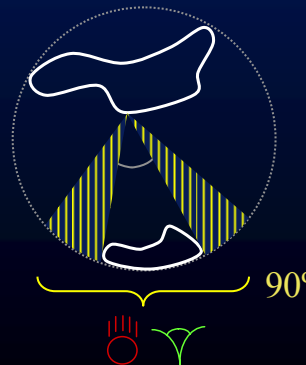
- triangulation

- line sweep of W pixels or $\log(W)$ time sequence of stripes, $W \approx 1024$
- projector and camera lines of sight must be unoccluded, so requires S scans, $10 \leq S \leq 100$
- one projector and camera
- $S \log(W) \approx 100-1000$



- confocal

- point scan over W^2 pixels or time sequence of T trials, $T \approx 32-64$
- works if some fraction of aperture is unoccluded, but gets noisier, max aperture $\approx 90^\circ$, so 6-12 sweeps?
- multiple projectors and cameras
- $6 T = 200-800$ ← *no moving parts*



© 2006 Marc Levoy

The Fourier projection-slice theorem

(a.k.a. the central section theorem) [Bracewell 1956]



- $P_{\square}(t)$ is the integral of $g(x,y)$ in the direction \square
- $G(u,v)$ is the 2D Fourier transform of $g(x,y)$
- $G_{\square}(\omega)$ is a 1D slice of this transform taken at \square
- $\mathcal{F}^{-1} \{ G_{\square}(\omega) \} = P_{\square}(t) !$

©2006 Marc Levoy

•References:

–Bracewell, R. N. Strip Integration in Radio Astronomy,
Australian Journal of Physics, Vol. 9, 1956,
No. 2, pp. 198-217.

–Kak, A, Slaney, M., *Principles of Computerized Tomographic Imaging*,
IEEE Press, 1988.

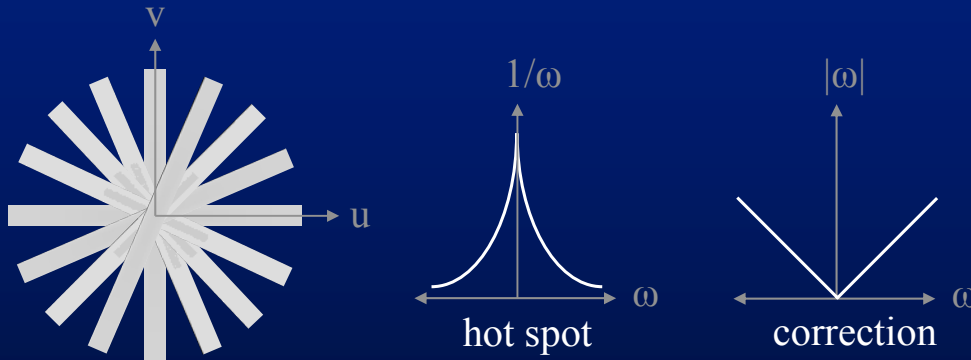
Reconstruction of $g(x,y)$ from its projections



- add slices $G_{\square}(\omega)$ into u,v at all angles \square and inverse transform to yield $g(x,y)$, or
- add 2D backprojections $P_{\square}(t, s)$ into x,y at all angles \square

©2006 Marc Levoy

The need for filtering before (or after) backprojection



- sum of slices would create $1/\omega$ hot spot at origin
- correct by multiplying each slice by $|\omega|$, or
- convolve $P_{\square}(t)$ by $\mathcal{F}^{-1} \{ |\omega| \}$ before backprojecting
- this is called filtered backprojection

©2006 Marc Levoy

$$\begin{aligned} \mathcal{F}^{-1} \{ |\omega| \} &= \text{Hilbert transform of } (\partial / \partial t) P_{\square}(t) \\ &= -1 / (\pi t) * (\partial / \partial t) P_{\square}(t) \\ &= \mathcal{F}^{-1} \left\{ \lim_{\epsilon \rightarrow 0} |\omega| e^{-\epsilon |\omega|} \right\} \end{aligned}$$

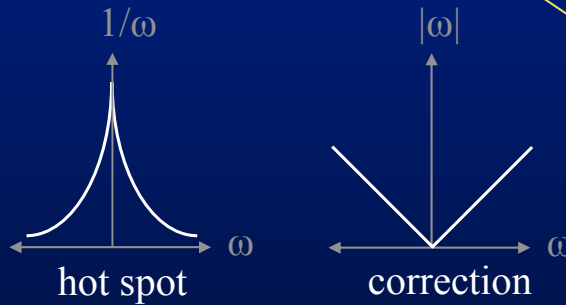
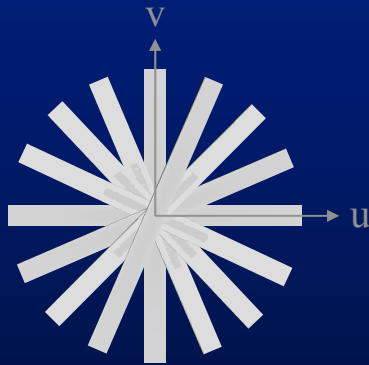
Image removed due to copyright restrictions

Image removed due to copyright restrictions

(from Bracewell)

Image removed due to copyright restrictions

$\sim 2^{\text{nd}}$ derivative



- sum of slices would create $1/\omega$ hot spot at origin
- correct by multiplying each slice by $|\omega|$, or
- convolve $P_{\square}(t)$ by $\mathcal{F}^{-1} \{ |\omega| \}$ before backprojecting
- this is called filtered backprojection

•Reference:

–Bracewell, R.N., *The Fourier Transform and its Applications*,
2nd ed., McGraw-Hill, 1985.

Summing filtered backprojections

Image removed due to
copyright restrictions

(from Kak)

©2006 Marc Levoy

Example of reconstruction by filtered backprojection



X-ray



sinogram



filtered sinogram



reconstruction

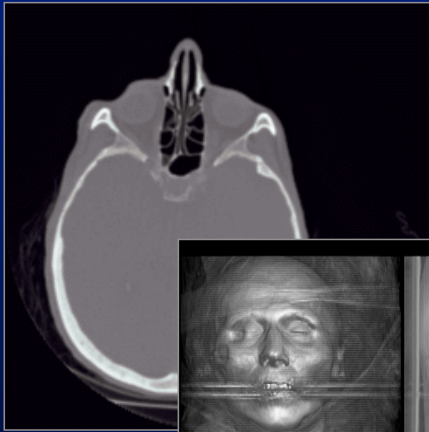
(from Herman)

©2006 Marc Levoy

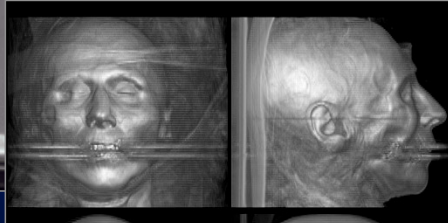
•Reference:

–Herman, G.T., *Image Reconstruction from Projections*, Academic Press, 1980.

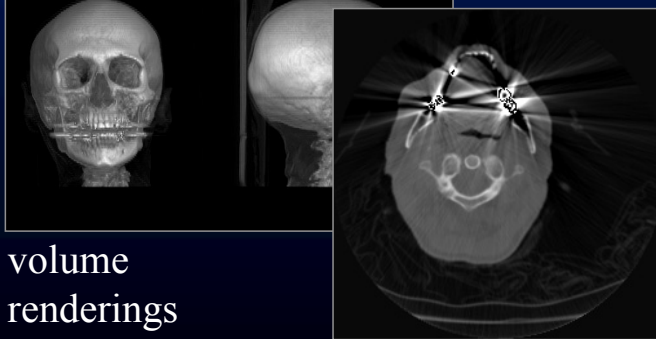
More examples



CT scan
of head



volume
renderings

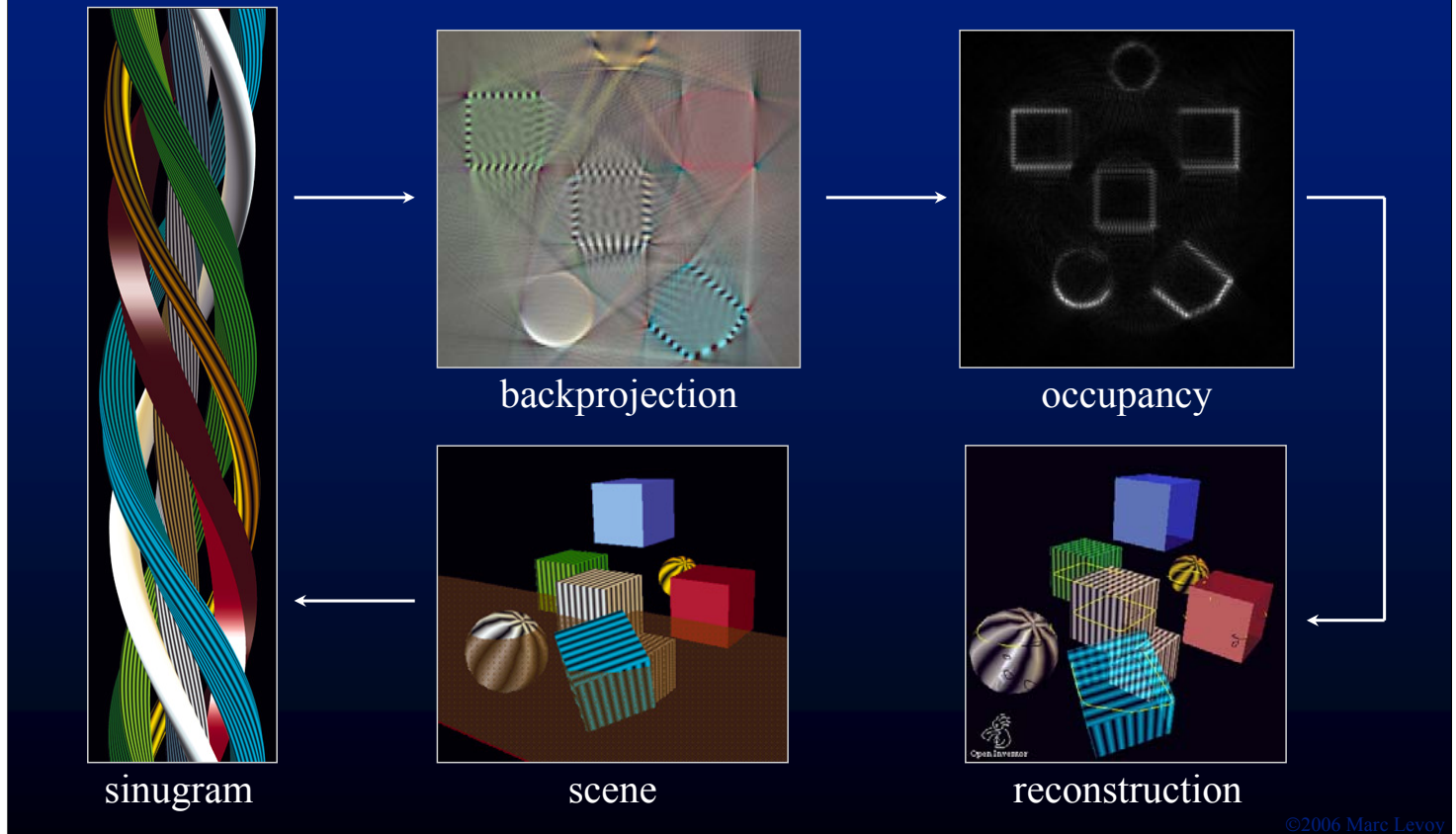


the effect
of occlusions



©2006 Marc Levoy

Shape from light fields using filtered backprojection



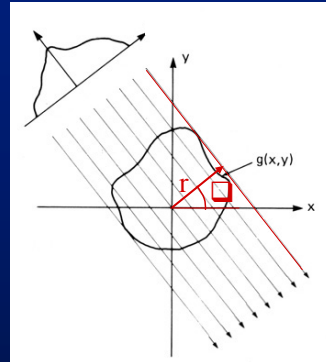
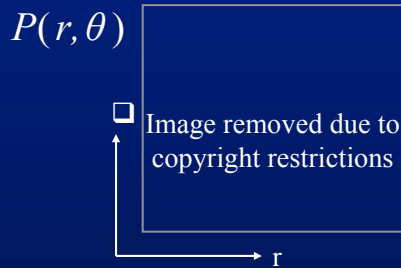
- Reference:

- M. Levoy, unpublished.

- but also see:

- Brady, D.J., Stack, R., Feller, S., Cull, E., Fernandez, L., Kammeyer, D., and Brady, R., Information Flow in Streaming 3D Video, *Proc. SPIE*, Vol. CR76-13, 2000.

Relation to Radon Transform



- Radon transform

$$P(r, \theta) = \int_{-\infty}^{+\infty} g(r \cos \theta - s \sin \theta, r \sin \theta + s \cos \theta) ds$$

- Inverse Radon transform

$$g(x, y) = -\frac{1}{2\pi^2} \int_0^{2\pi} d\theta \int_{-\infty}^{+\infty} \frac{1}{q} P_1(x \cos \theta + y \sin \theta + q, \theta) dq$$

where P_1 where is the partial derivative of P with respect to t

$= [B H_1 D_1 P](r, \theta)$ where

D_1 is partial differentiation wrt the 1st variable,

H_1 is the Hilbert transform wrt the 1st variable, and

B is backprojection

- Radon transform

$$P(r, \theta) = \int_{-\infty}^{+\infty} g(r \cos \theta - s \sin \theta, r \sin \theta + s \cos \theta) ds$$

- Inverse Radon transform

$$g(x, y) = -\frac{1}{2\pi^2} \int_0^{2\pi} d\theta \int_{-\infty}^{+\infty} \frac{1}{q} P_1(x \cos \theta + y \sin \theta + q, \theta) dq$$

where P_1 where is the partial derivative of P with respect to t

Higher dimensions

- Fourier projection-slice theorem in \mathbb{R}^n

- $G_\xi(\omega)$, where ξ is a unit vector in \mathbb{R}^n , ω is the basis for a hyperplane in \mathbb{R}^{n-1} , and G contains integrals over lines
- in 2D: a slice (of G) is a line through the origin at angle θ , each point on \mathcal{F}^{-1} of that slice is a line integral (of g) perpendicular to θ
- in 3D: each slice is a plane through the origin at angles (θ, ϕ) , each point on \mathcal{F}^{-1} of that slice is a line integral perpendicular to the plane

(from Deans)

- Radon transform in \mathbb{R}^n

- $P(r, \xi)$, where ξ is a unit vector in \mathbb{R}^n , r is a scalar, and P contains integrals over $(n-1)$ -D hyperplanes
- in 2D: each point (in P) is the integral along the line (in g) perpendicular to a ray connecting that point and the origin
- in 3D: each point is the integral across a plane normal to a ray connecting that point and the origin

Image removed due to copyright restrictions

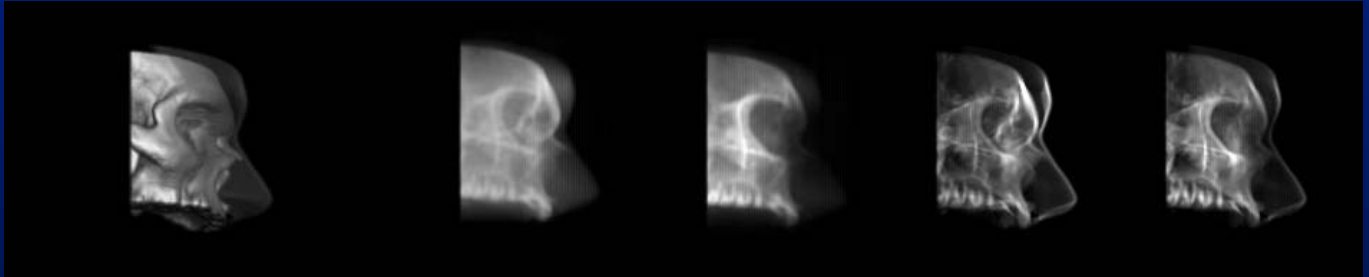
©2006 Marc Levoy

- Reference:

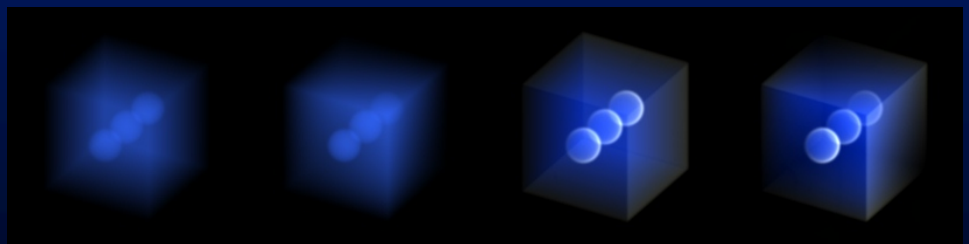
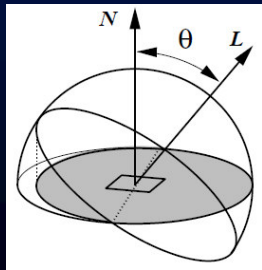
–Deans, S., *The Radon Transform and Some of its Applications*, Krieger, 1983.

Frequency domain volume rendering

[Totsuka and Levoy, SIGGRAPH 1993]



volume rendering



X-ray

with
depth cueing

with
directional
shading

with
depth cueing
and shading

©2006 Marc Levoy

•Reference:

–Totsuka, T. and Levoy, M., Frequency Domain Volume Rendering, *Proc. SIGGRAPH 1993*.

•Example depth cueing

–sinusoidal falloff = multiplication of volume by large sinusoid

–= convolution of spectrum by $F(\text{sinusoid})$ = convolution by spike = shifting the 3D spectrum before extracting slice!

•Example directional shading

–we can't compute $|N.L|$ or $\max(N.L, 0)$

–Lambertian under hemispherical shading = $\frac{1}{2} + \frac{1}{2} N.L$, which smoothly maps $N.L$ to $0..1$

– $N.L$ = first derivative of volume in direction x of pole = directional first moment ($x f(x)$) of spectrum, which again can be computed while extracting a slice

Other issues in tomography

- resample fan beams to parallel beams
- extendable (with difficulty) to cone beams in 3D
- modern scanners use helical capture paths
- scattering degrades reconstruction

Limited-angle projections



Image removed due to
copyright restrictions

(from Olson)

©2006 Marc Levoy

- Reference:

- Olson, T., Jaffe, J.S., An explanation of the effects of squashing in limited angle tomography,
IEEE Trans. Medical Imaging, Vol. 9, No. 3., September 1990.

Reconstruction using Algebraic Reconstruction Technique (ART)

Image removed due to
copyright restrictions

(from Kak)

$$p_i = \sum_{j=1}^N w_{ij} f_j, \quad i = 1, 2, \dots, M$$

M projection rays

N image cells along a ray

p_i = projection along ray i

f_j = value of image cell j (n^2 cells)

w_{ij} = contribution by cell j to ray i
(a.k.a. resampling filter)

- applicable when projection angles are limited or non-uniformly distributed around the object
- can be under- or over-constrained, depending on N and M

©2006 Marc Levoy

Image removed due to
copyright restrictions

Image removed due to
copyright restrictions

$$\bar{f}^{(k)} = \bar{f}^{(k-1)} - \frac{\bar{f}^{(k-1)} \bullet (\bar{w}_i - p_i)}{\bar{w}_i \bullet \bar{w}_i} \bar{w}_i$$

$\bar{f}^{(k)}$ = k^{th} estimate of all cells

\bar{w}_i = weights ($w_{i1}, w_{i2}, \dots, w_{iN}$) along ray i

Procedure

- make an initial guess, e.g. assign zeros to all cells
- project onto p_1 by increasing cells along ray 1 until $\Sigma = p_1$
- project onto p_2 by modifying cells along ray 2 until $\Sigma = p_2$, etc.
- to reduce noise, reduce by $\alpha \Delta \bar{f}^{(k)}$ for $\alpha < 1$

•Formula is derived in Kak, chapter 7, p. 278

- linear system, but big, sparse, and noisy
- ART is solution by *method of projections* [Kaczmarz 1937]
- to increase angle between successive hyperplanes, jump by 90°
- SART modifies all cells using $f^{(k-1)}$, then increments k
- overdetermined if $M > N$, underdetermined if missing rays
- optional additional constraints:
 - $f > 0$ everywhere (positivity)
 - $f = 0$ outside a certain area

Procedure

- make an initial guess, e.g. assign zeros to all cells
- project onto p_1 by increasing cells along ray 1 until $\Sigma = p_1$
- project onto p_2 by modifying cells along ray 2 until $\Sigma = p_2$, etc.
- to reduce noise, reduce by $\alpha \Delta \bar{f}^{(k)}$ for $\alpha < 1$

•SIRT = Simultaneous Iterative Reconstruction Technique

•SART = Simultaneous ART

- linear system, but big, sparse, and noisy
- ART is solution by *method of projections* [Kaczmarz 1937]
- to increase angle between successive hyperplanes, jump by 90°
- SART modifies all cells using $f^{(k-1)}$, then increments k
- overdetermined if $M > N$, underdetermined if missing rays
- optional additional constraints:
 - $f > 0$ everywhere (positivity)
 - $f = 0$ outside a certain area

Image removed due to
copyright restrictions

(Olson)

•Reference:

–Olson, T., A stabilized inversion for limited
angle tomography. Manuscript.

–35 degrees missing

Image removed due to
copyright restrictions

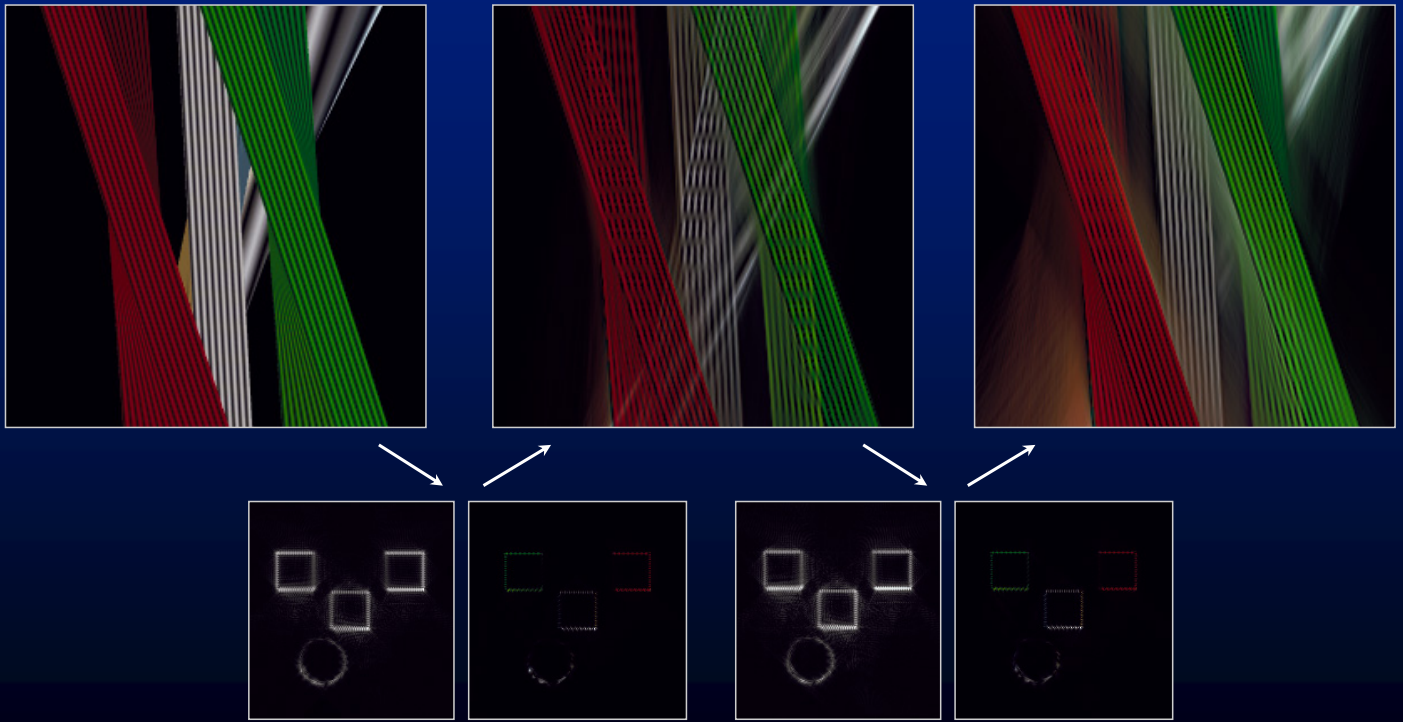
Image removed due to
copyright restrictions

(Olson)

•Nonlinear constraints

– $f = 0$ outside of circle (oval?)

Shape from light fields using iterative relaxation



©2006 Marc Levoy

- Reference:
 - M. Levoy, unpublished
- but also see:
 - DeBonet, J., Viola, P., Poxels: responsibility weighted 3D volume reconstruction, *Proc. ICCV 1999*.

Borehole tomography

Image removed due to
copyright restrictions

(from Reynolds)

- receivers measure end-to-end travel time
- reconstruct to find velocities in intervening cells
- must use limited-angle reconstruction method (like ART)

©2006 Marc Levoy

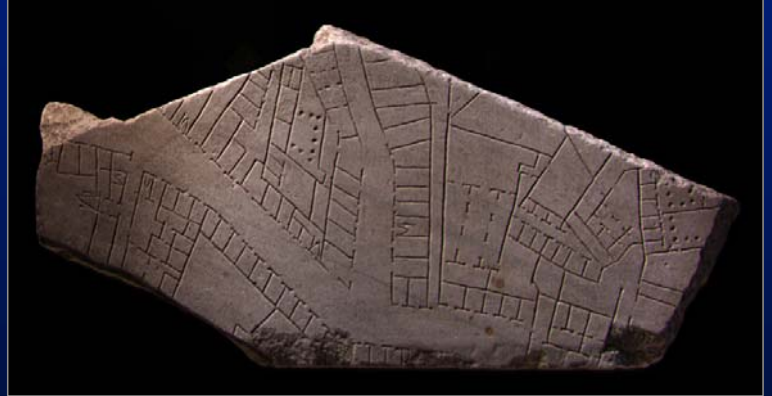
•Reference:

–Reynolds, J.M., *An Introduction to Applied and Environmental Geophysics*,
Wiley, 1997.

Applications

Image removed due to
copyright restrictions

mapping a *seismosaurus* in sandstone
using microphones in 4 boreholes and
explosions along radial lines



mapping ancient Rome using
explosions in the subways and
microphones along the streets?

©2006 Marc Levoy

•Left picture is from Reynolds, right picture is from
Stanford's Forma Urbis Romae project

From microscope light fields to volumes

- 4D light field → *digital refocusing* →
3D focal stack → *deconvolution microscopy* →
3D volume data

Image removed due to
copyright restrictions

Image removed due to
copyright restrictions

- 4D light field → *tomographic reconstruction* →
3D volume data

Image removed due to
copyright restrictions

© 2006 Marc Levoy

- Reference:

–[http://www.api.com/lifescience/DeltaVision
RT.html](http://www.api.com/lifescience/DeltaVisionRT.html)

3D deconvolution

[McNally 1999]

Image removed due to
copyright restrictions

focus stack of a point in 3-space is the 3D PSF of that imaging system

- object * PSF → focus stack
- $\{ \text{object} \} \times \{ \text{PSF} \} \rightarrow \{ \text{focus stack} \}$
- $\{ \text{focus stack} \} \otimes \{ \text{PSF} \} \rightarrow \{ \text{object} \}$
- spectrum contains zeros, due to missing rays
- imaging noise is amplified by division by ~zeros
- reduce by regularization (smoothing) or completion of spectrum
- improve convergence using constraints, e.g. object > 0

Image removed due to
copyright restrictions

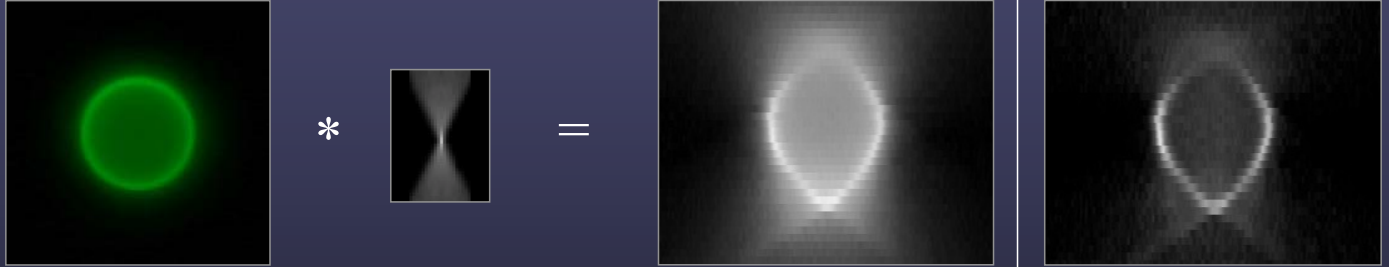
•Reference:

–McNally, J.G., Karpova, T., Cooper, J.,
Conchello, J.A., Three-Dimensional Imaging
by Deconvolution Microscopy,
Methods, Vol. 19, 1999.

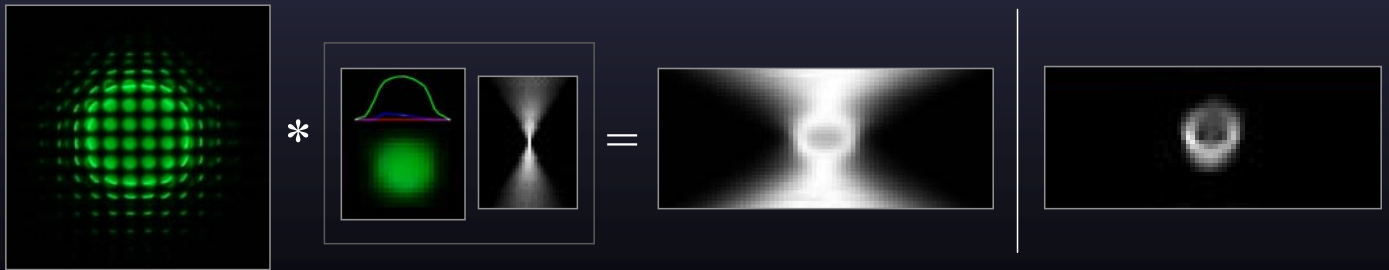
© 2006 Marc Levoy

Example: 15 μ hollow fluorescent bead

conventional microscope



light field microscope



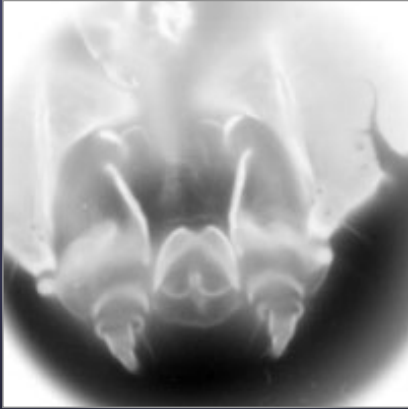
© 2006 Marc Levoy

•Images from:

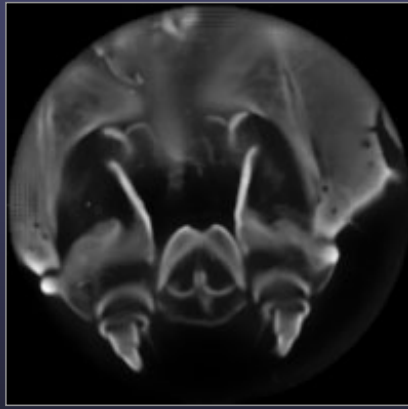
– Levoy, M., Ng, R., Adams, A., Footer, M., Horowitz, M., Light field microscopy, *ACM Transactions on Graphics (Proc. SIGGRAPH)*, **25**(3), 2006.

Silkworm mouth

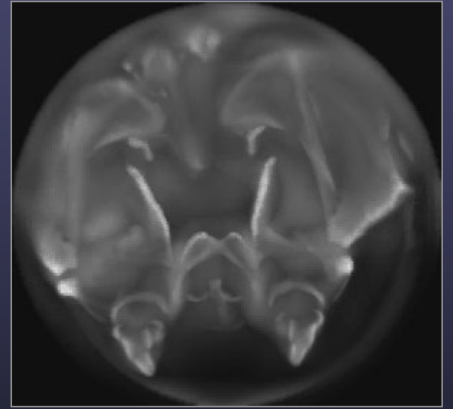
(collection of B.M. Levoy)



slice of focal stack



slice of volume

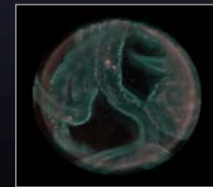
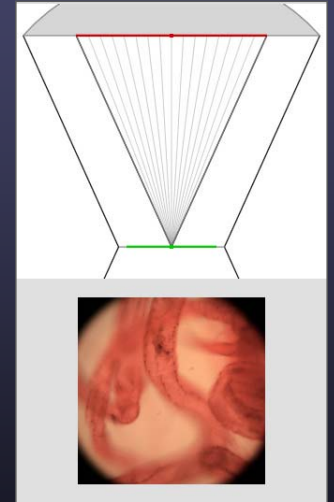
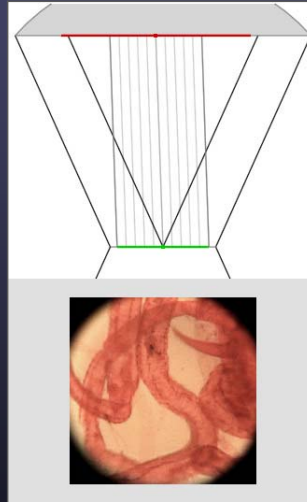
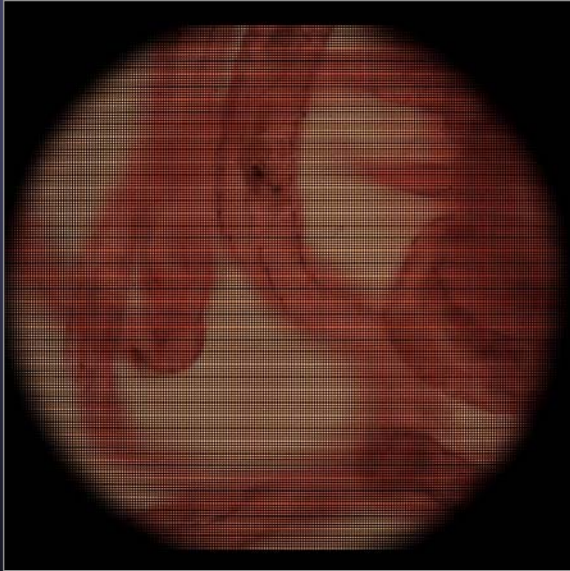


volume rendering

© 2006 Marc Levoy

Legs of unknown insect

(collection of B.M. Levoy)




© 2006 Marc Levoy

Tomography and 3D deconvolution: how different are they?

Fourier domain

- deconvolution

– 4D LF \rightarrow *refocusing* \rightarrow 3D spectrum \rightarrow   {PSF} \rightarrow  ⁻¹

- tomography

– 4D LF \rightarrow 2D slices in 3D spectrum \rightarrow   ⁻¹ \rightarrow volume
spatial domain

- deconvolution

– 4D LF \rightarrow *refocusing* \rightarrow 3D stack \rightarrow  *inverse filter* \rightarrow volume

- tomography



© 2006 Marc Levoy

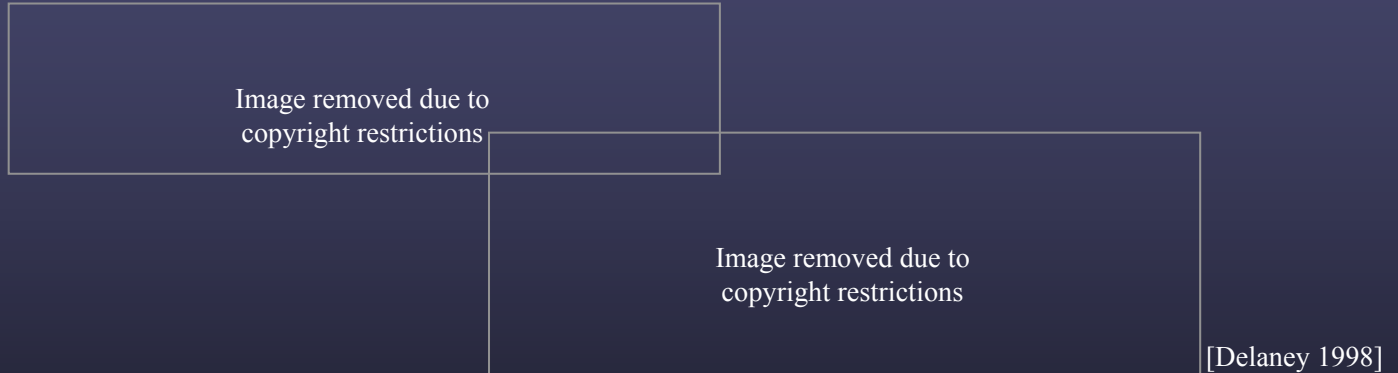
•Full proof appears in [Levoy 2006] (previously cited)

For finite apertures, they are still the same

- deconvolution
 - nonblind iterative deconvolution with positivity constraint on 3D reconstruction
- limited-angle tomography
 - Simultaneous Algebraic Reconstruction Technique (SART) with same constraint

Their artifacts are also the same

- tomography from limited-angle projections



- deconvolution from finite-aperture images



•References:

–Delaney, A.H., Bresler, Y., Globally Convergent Edge-Preserving Regularized Reconstruction: An Application to Limited-Angle Tomography, *IEEE Transactions on Image Processing*, Vol. 7, No. 2, February 1998.

–others are previously cited

Diffraction tomography

Image removed due to
copyright restrictions

(from Kak)

Image removed due to
copyright restrictions

limit as $\lambda \rightarrow 0$ (relative to
object size) is Fourier
projection-slice theorem

- Wolf (1969) showed that a broadband hologram gives the 3D structure of a semi-transparent object
- Fourier Diffraction Theorem says $\{ \text{scattered field} \} = \text{arc in } \{ \text{object} \}$ as shown above, can use to reconstruct object
- assumes weakly refractive media and coherent plane illumination, must record amplitude and phase of forward scattered field

©2006 Marc Levoy

•Wolf

- applicable only to semi-transparent objects
- Wolf E 1969, Three-dimensional structure determination of semi-transparent objects from holographic data, *Opt. Commun.* **1** 153–6

Image removed due to
copyright restrictions

Image removed due to
copyright restrictions

Image removed due to
copyright restrictions

limit as $\lambda \rightarrow 0$ (relative to
object size) is Fourier
projection-slice theorem

(from Kak)

- Wolf (1969) showed that a broadband hologram gives the 3D structure of a semi-transparent object
- Fourier Diffraction Theorem says $\{ \text{scattered field} \} = \text{arc in } \{ \text{object} \}$ as shown above, can use to reconstruct object
- assumes weakly refractive media and coherent plane illumination, must record amplitude and phase of forward scattered field

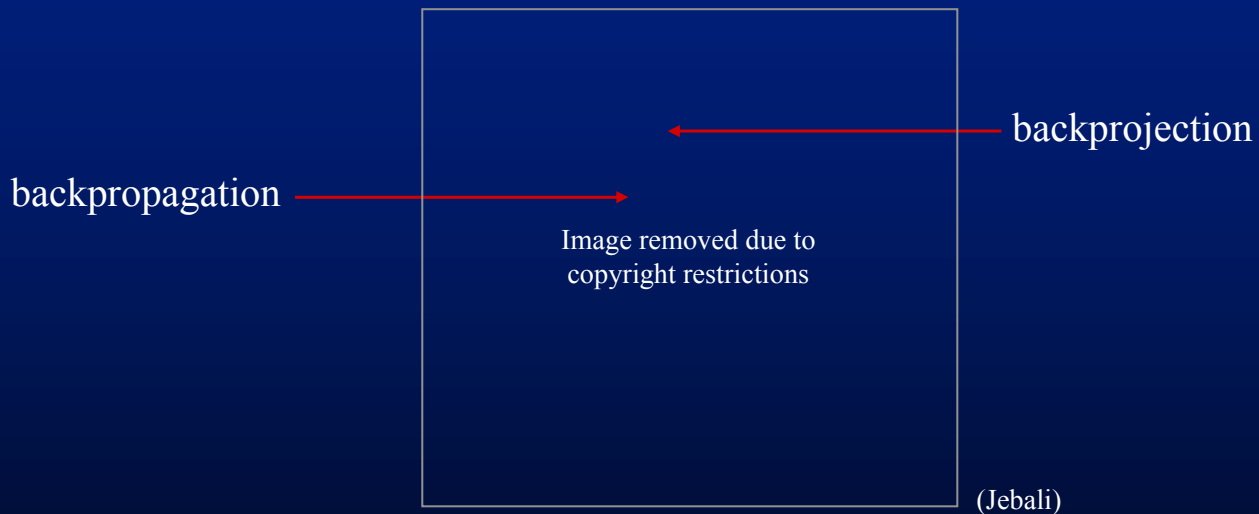
- measuring phase

- typically requires a reference beam and interference between it and the main beam, i.e. a holographic procedure

- Reference:

- Devaney, A., Inverse scattering and optical diffraction tomography, Powerpoint presentation.

Inversion by filtered backpropagation



- depth-variant filter, so more expensive than tomographic backprojection, also more expensive than Fourier method
- applications in medical imaging, geophysics, optics

©2006 Marc Levoy

•Reference:

–Jebali, A., *Numerical Reconstruction of semi-transparent objects in Optical Diffraction Tomography*,
Diploma Project, Ecole Polytechnique,
Lausanne, 2002.

Diffuse optical tomography

Image removed due to
copyright restrictions

(Arridge)

- assumes light propagation by multiple scattering
- model as diffusion process (similar to Jensen01)

©2006 Marc Levoy

•References:

–Arridge, S.R., Methods for the Inverse Problem in Optical Tomography, *Proc. Waves and Imaging Through Complex Media*, Kluwer, 307-329, 2001.

–Schweiger, M., Gibson, A., Arridge, S.R., “Computational Aspects of Diffuse Optical Tomography,” *IEEE Computing*, Vol. 5, No. 6, Nov./Dec., 2003. (for image)

–Jensen, H.W., Marschner, S., Levoy, M., Hanrahan, P., A Practical Model for Subsurface Light Transport, *Proc. SIGGRAPH 2001*.

The optical diffusion equation

$$D\Delta^2\phi(x) = \sigma_a\phi(x) - Q_0(x) + 3D\vec{\Delta} \cdot \vec{Q}_1(x) \quad (\text{from Jensen})$$

- $D = \text{diffusion constant} = 1/3\sigma'_t$
where σ'_t is a reduced extinction coefficient
- $\phi(x) = \text{scalar irradiance at point } x$
- $Q_n(x) = n^{\text{th}}\text{-order volume source distribution, i.e.}$

$$Q_0(x) = \int_{4\pi} Q(x, \vec{\omega}) d\omega \quad Q_1(x) = \int_{4\pi} Q(x, \vec{\omega}) \vec{\omega} d\omega$$

- in DOT, σ_a source and σ_t are unknown

Diffuse optical tomography

Image removed due to
copyright restrictions

Image removed due to
copyright restrictions

female breast with
sources (red) and
detectors (blue)

- assumes light propagation by multiple scattering
- model as diffusion process (similar to Jensen01)
- inversion is non-linear and ill-posed
- solve use optimization with regularization (smoothing)

©2006 Marc Levoy

- acquisition

- 81 source positions, 81 detector positions

- for each source position, measure light at all detector positions

- use time-of-flight measurement to estimate initial guess for absorption, to reduce cross-talk between absorption and scattering

Coded aperture imaging

Image removed due to
copyright restrictions

Image removed due to
copyright restrictions

(from Zand)

(source assumed infinitely distant)

- optics cannot bend X-rays, so they cannot be focused
- pinhole imaging needs no optics, but collects too little light
- use multiple pinholes and a single sensor
- produces superimposed shifted copies of source

©2006 Marc Levoy

•Reference:

–Zand, J., Coded aperture imaging in high
energy astronomy,
http://hea-www.gsfc.nasa.gov/docs/cai/coded_intr.html

Reconstruction by matrix inversion

$$\begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix} = \begin{bmatrix} C_1 & C_2 & C_3 & C_4 \\ C_4 & C_1 & C_2 & C_3 \\ C_3 & C_4 & C_1 & C_2 \\ C_2 & C_3 & C_4 & C_1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{bmatrix}$$

detector

mask
(0/1)

source

$$\mathbf{d} = \mathbf{C} \mathbf{s}$$

$$\mathbf{s} = \mathbf{C}^{-1} \mathbf{d}$$

- ill-conditioned unless auto-correlation of mask is a delta function



source larger than detector,
system underconstrained



(from Zand)

collimators restrict source directions to those from which projection of mask falls completely within the detector

©2006 Marc Levoy

Reconstruction by backprojection

Image removed due to
copyright restrictions

(from Zand)

- backproject each detected pixel through each hole in mask
- superimposition of projections reconstructs source
- essentially a cross correlation of detected image with mask
- also works for non-infinite sources; use voxel grid
- assumes non-occluding source

©2006 Marc Levoy

- Another example:

- Carlisle, P., Coding aperture imaging ("Mmm, gumballs..."),
<http://paulcarlisle.net/old/codedaperture.html>

- cross correlation is just convolution (of detected image by mask) without first reversing detected image in x and y

- conversion of blacks to -1's in "decoding matrix" just serves to avoid normalization of resulting reconstruction

- performing this on an image of gumballs, rather than a 3D gumball scene, is equivalent to assuming the gumballs cover the sky at infinity, i.e. they are an angular function

- assumes non-occluding source

- otherwise it's the voxel coloring problem

Interesting techniques I didn't have time to cover

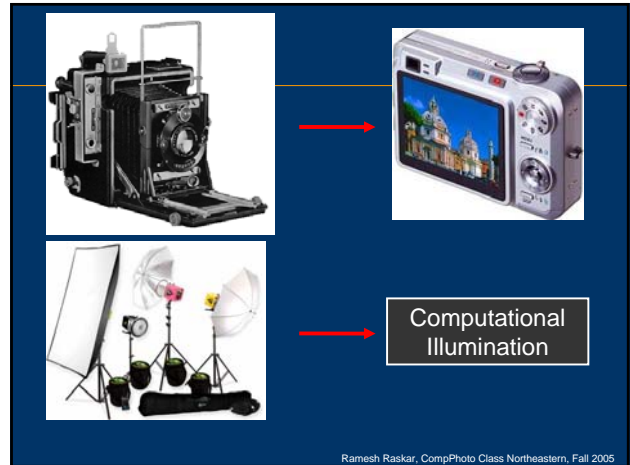
- reflection tomography
- synthetic aperture radar & sonar
- holography
- wavefront coding

Computational Illumination

Ramesh Raskar
Mitsubishi Electric Research Labs

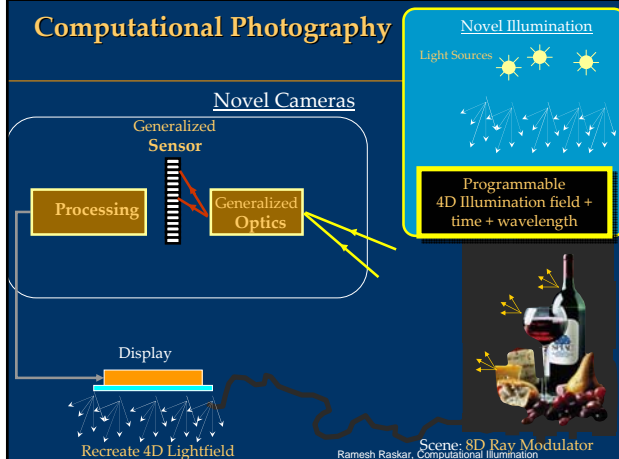
Course WebPage :
<http://www.merl.com/people/raskar/photo/course/>

Ramesh Raskar, Computational Illumination



Ramesh Raskar, CompPhoto Class Northeastern, Fall 2005

Computational Photography



Ramesh Raskar, Computational Illumination

Computational Illumination:

Programmable 4D Illumination Field + Time + Wavelength

- Presence or Absence
 - Flash/No-flash
- Light position
 - Multi-flash for depth edges
 - Programmable dome (image re-lighting and matting)
- Light color/wavelength
- Spatial Modulation
 - Synthetic Aperture Illumination
- Temporal Modulation
 - TV remote, Motion Tracking, Sony ID-cam, RFIG
- Exploiting (uncontrolled) natural lighting condition
 - Day/Night Fusion

Ramesh Raskar, Computational Illumination

Computational Illumination

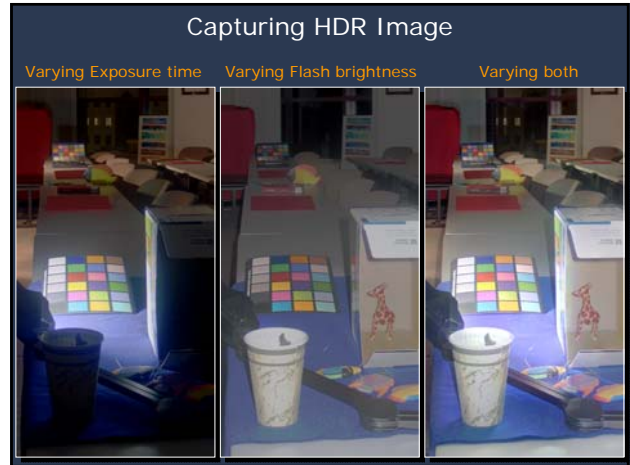
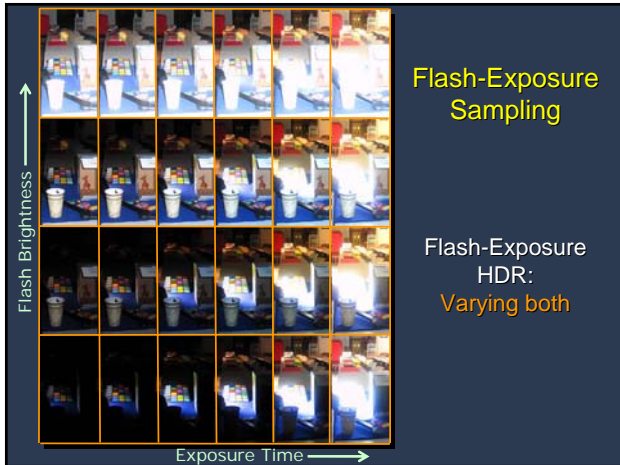
- Presence or Absence
 - Flash/No-flash
- Light position
 - Multi-flash for depth edges
 - Programmable dome (image re-lighting and matting)
- Light color/wavelength
- Spatial Modulation
 - Synthetic Aperture Illumination
- Temporal Modulation
 - TV remote, Motion Tracking, Sony ID-cam, RFIG
- General lighting condition
 - Day/Night

Ramesh Raskar, Computational Illumination

Flash and Ambient Images

[Agrawal, Raskar, Nayar, Li Siggraph05]





Denoising Challenging Images

Available light:
 + nice lighting
 - noise/blurriness
 - color

No-flash

Flash:
 + details
 + color
 - flat/artificial

Elmar Elsemann and Frédo Durand, Flash Photography Enhancement via Intrinsic Relighting
 Georg Petschnigg, Maneesh Agrawala, Hugues Hoppe, Richard Szeliski, Michael Cohen, Kentaro Toyama. [Digital Photography with Flash and No-Flash Image Pairs](#)

Use no-flash image relight flash image

No-flash

Flash

Result

Computational Illumination

- Presence or Absence
 - Flash/No-flash
- Light position
 - Multi-flash for depth edges
 - Programmable dome (image re-lighting and matting)
- Light color/wavelength
- Spatial Modulation
 - Synthetic Aperture Illumination
- Temporal Modulation
 - TV remote, Motion Tracking, Sony ID-cam, RFIG
- General lighting condition
 - Day/Night

Ramesh Raskar, Computational Illumination

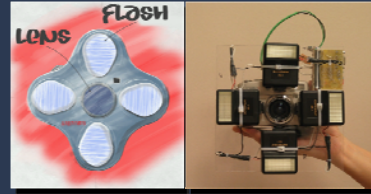
Synthetic Lighting

Paul Haeberli, Jan 1992



Ramesh Raskar, CompPhoto Class Northeastern, Fall 2005

Non-photorealistic Camera: Depth Edge Detection and Stylized Rendering using Multi-Flash Imaging

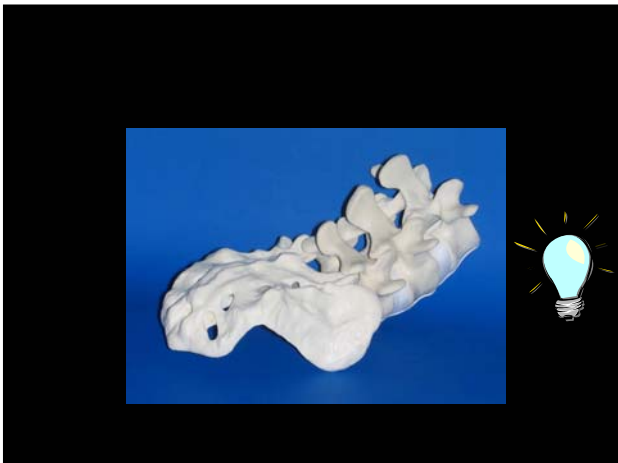
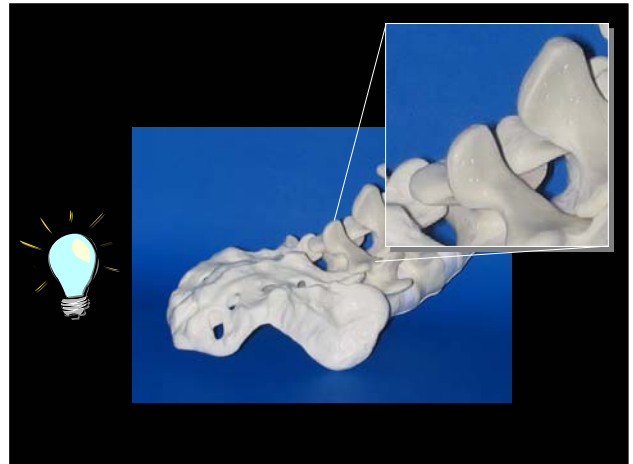


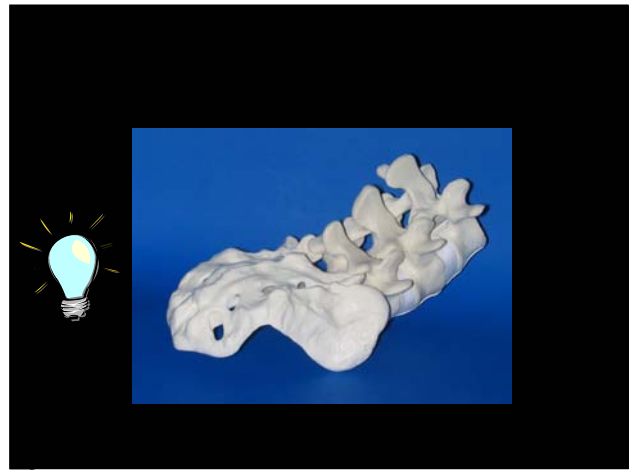
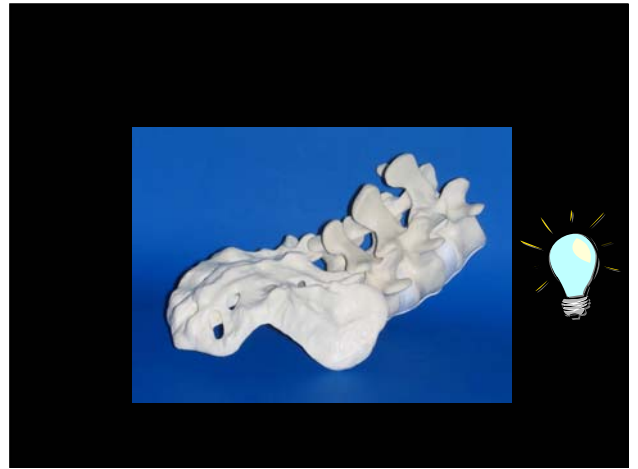
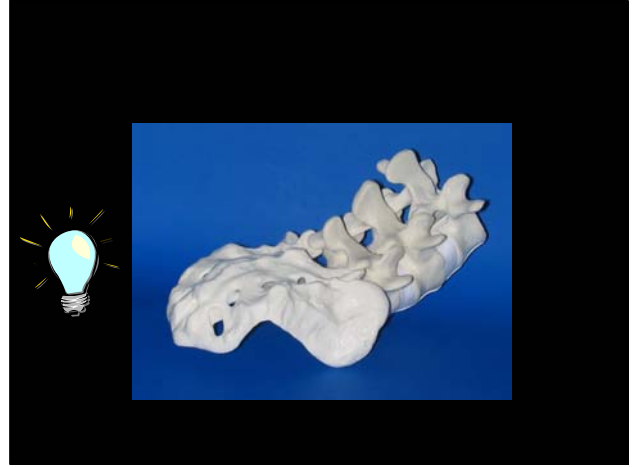
Ramesh Raskar, Karhan Tan, Rogerio Feris,
Jingyi Yu, Matthew Turk
Mitsubishi Electric Research Labs (MERL), Cambridge, MA
U of California at Santa Barbara
U of North Carolina at Chapel Hill

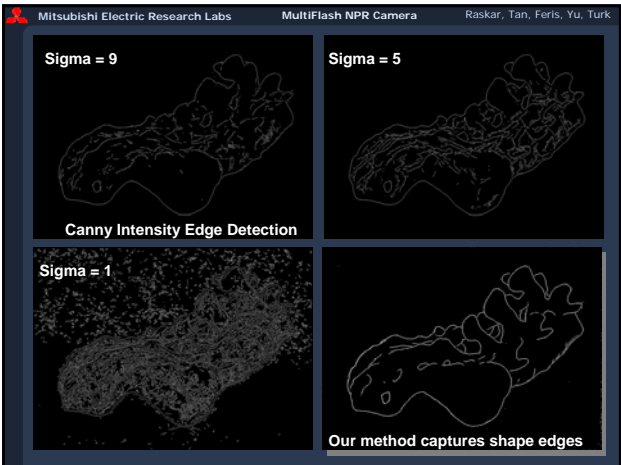
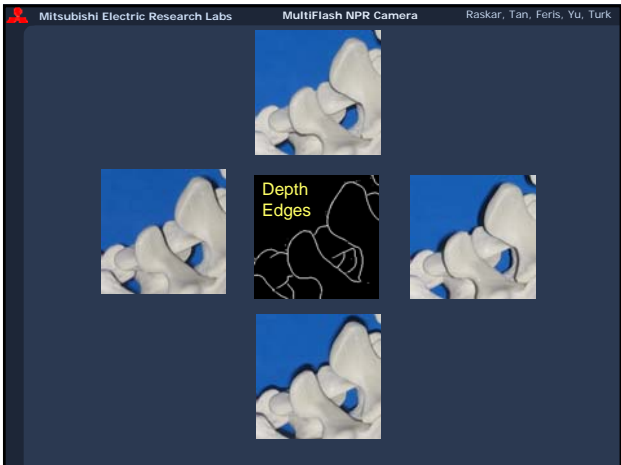
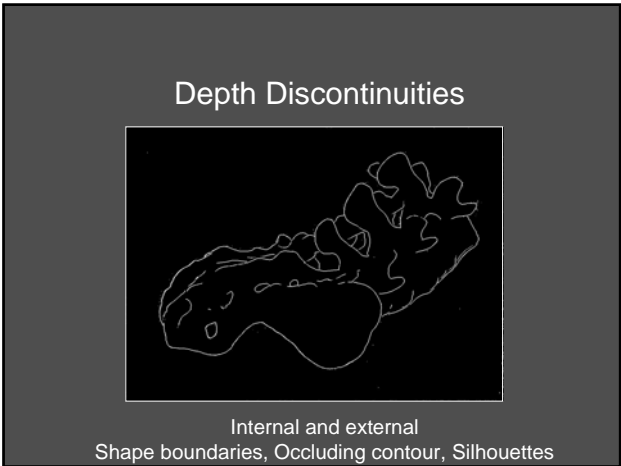
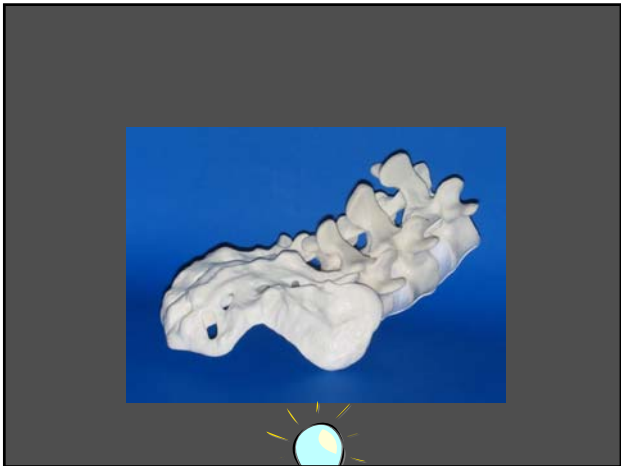
Mitsubishi Electric Research Labs MultiFlash NPR Camera Raskar, Tan, Feris, Yu, Turk

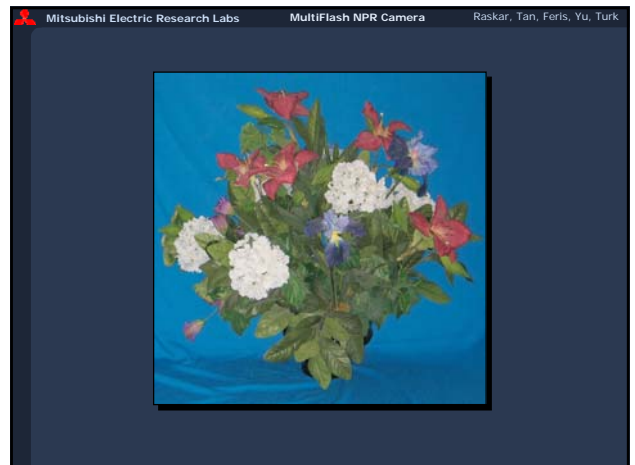
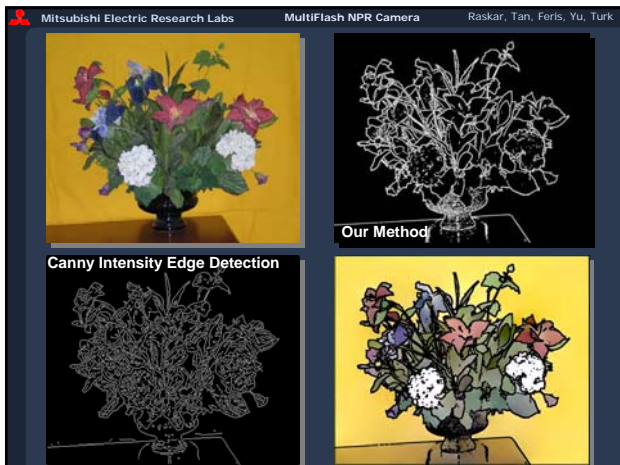
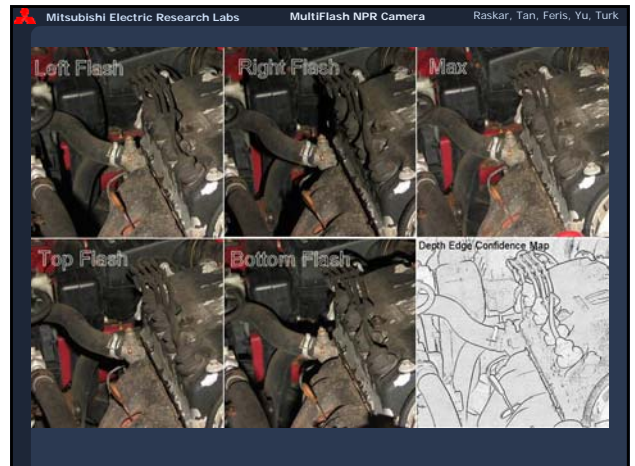
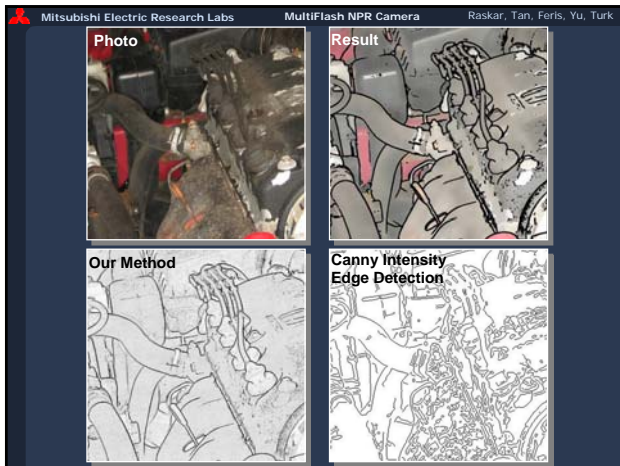
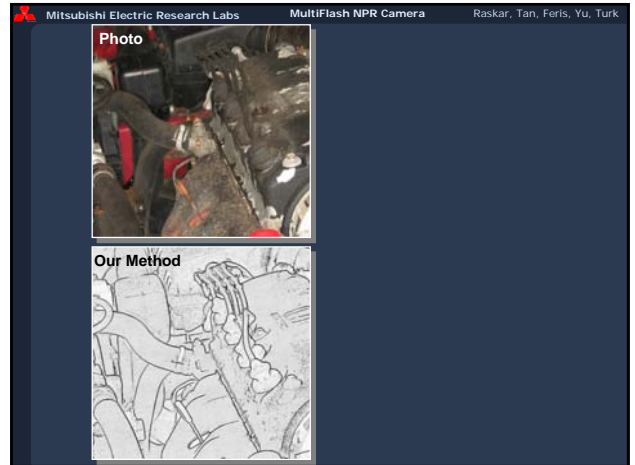
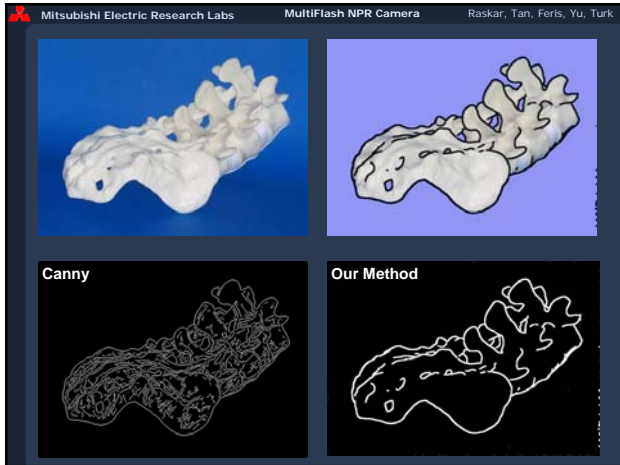
Depth Edge Camera

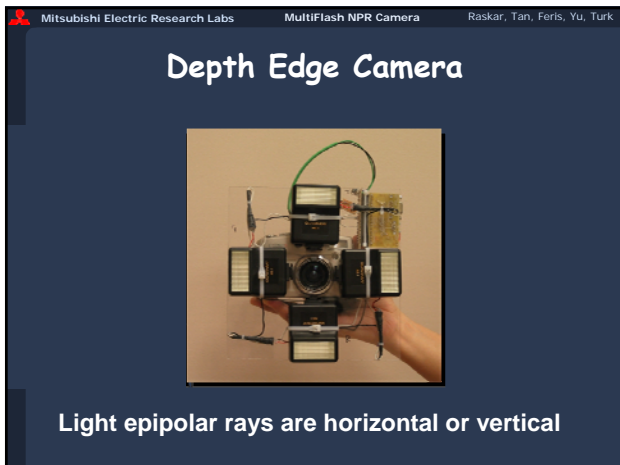
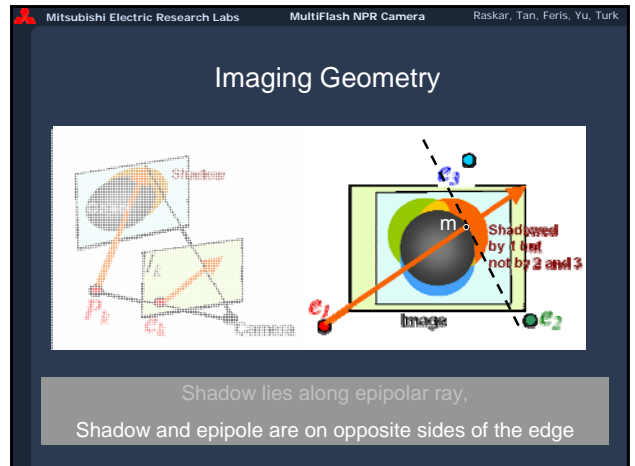
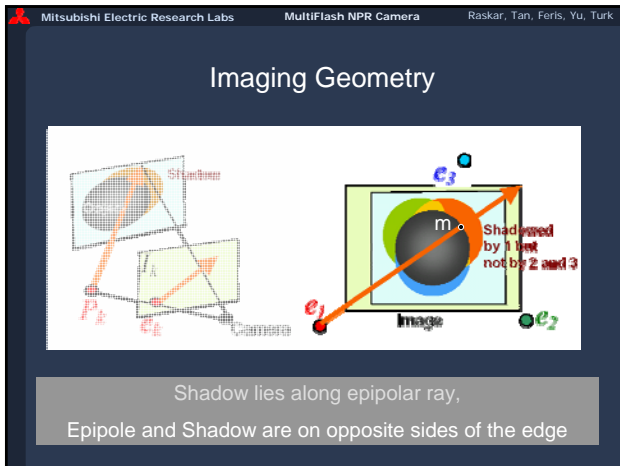
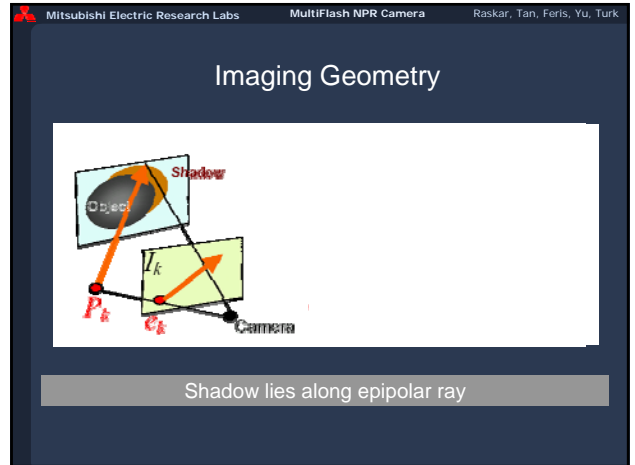
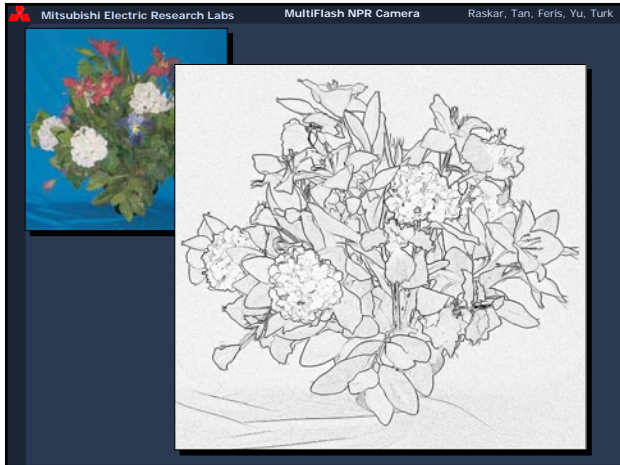
A diagram of a multi-flash camera lens with four flash units arranged in a cross pattern, labeled 'LENS' and 'FLASH'. To the right is a photograph of the physical camera hardware, showing a camera body with four external flash units attached.

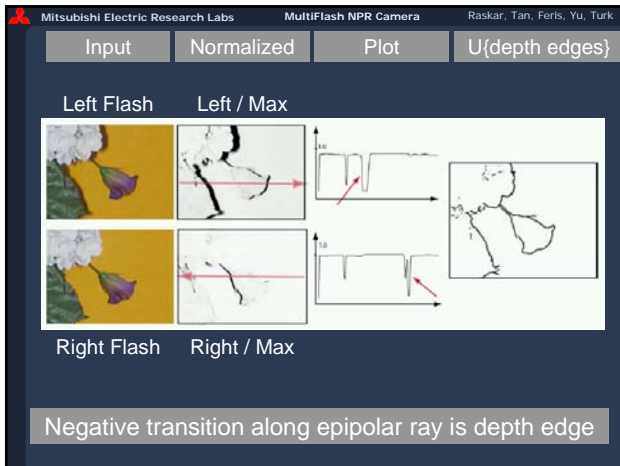
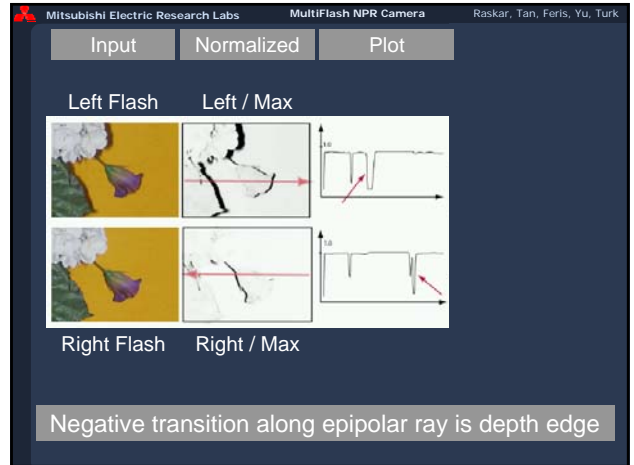
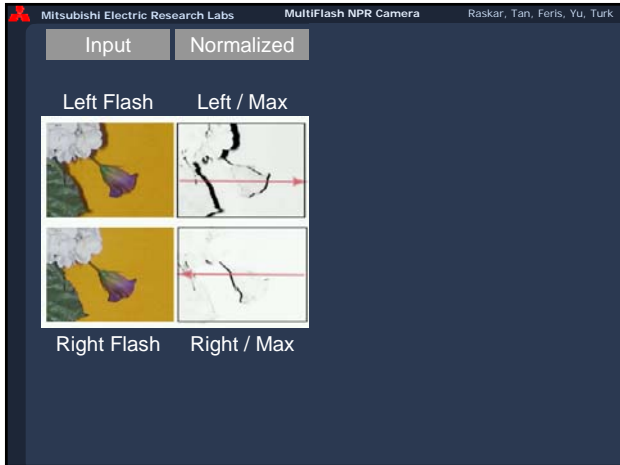












```

% Max composite
maximg = max( left, right, top, bottom);

% Normalize by computing ratio images
r1 = left ./ maximg;      r2 = top ./ maximg;
r3 = right ./ maximg;    r4 = bottom ./ maximg;

% Compute confidence map
v = fspecial( 'sobel' ); h = v';
d1 = imfilter( r1, v ); d3 = imfilter( r3, v ); % vertical sobel
d2 = imfilter( r2, h ); d4 = imfilter( r4, h ); % horizontal sobel

% Keep only negative transitions
silhouette1 = d1 .* (d1<0);
silhouette2 = abs( d2 .* (d2<0) );
silhouette3 = abs( d3 .* (d3<0) );
silhouette4 = d4 .* (d4<0);

% Pick max confidence in each
confidence = max( silhouette1, silhouette2, silhouette3, silhouette4 );
imwrite( confidence, 'confidence.bmp' );

```

No magic parameters !



Image-Based Actual Re-lighting

Debevec et al., SIGG2001

Light the actress in Los Angeles

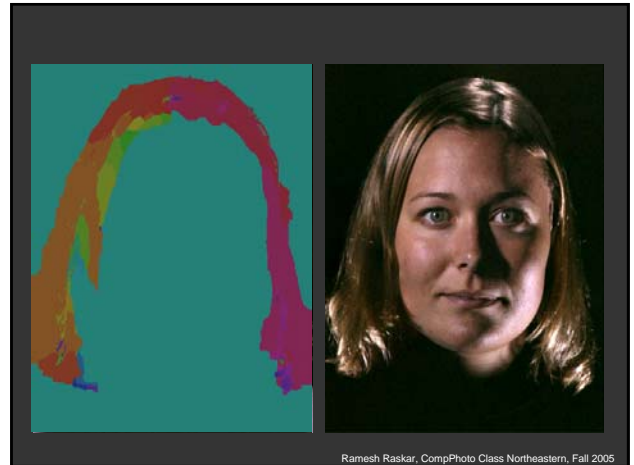
Film the background in Milan, Measure incoming light,

Matched LA and Milan lighting.

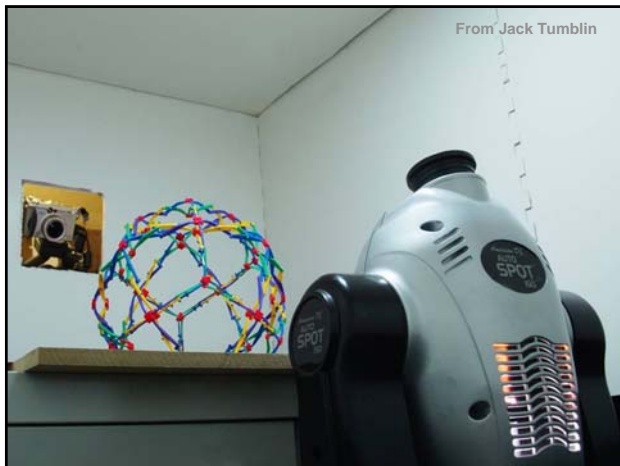
Matte the background



Ramesh Raskar, CompPhoto Class Northeastern, Fall 2005



Ramesh Raskar, CompPhoto Class Northeastern, Fall 2005





Computational Illumination

- Presence or Absence
 - Flash/No-flash
- Light position
 - Multi-flash for depth edges
 - Programmable dome (image re-lighting and matting)
- Light color/wavelength
- Spatial Modulation
 - **Synthetic Aperture Illumination**
- Temporal Modulation
 - TV remote, Motion Tracking, Sony ID-cam, RFIG
- General lighting condition
 - Day/Night

Ramesh Raskar, Computational Illumination

Synthetic Aperture Illumination: Comparison with Long-range synthetic aperture photography

- width of aperture 6'
- number of cameras 45
- spacing between cameras 5"
- camera's field of view 4.5°

© 2005 Intel

The scene



- distance to occluder 110'
- distance to targets 125'
- field of view at target 10'

Synthetic aperture photography using an array of mirrors

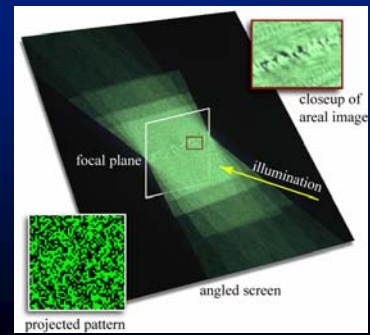


- 11-megapixel camera (4064 x 2047 pixels)
- 18 x 12 inch effective aperture, 9 feet to scene
- 22 mirrors, tilted inwards → 22 views, each 750 x 500 pixels

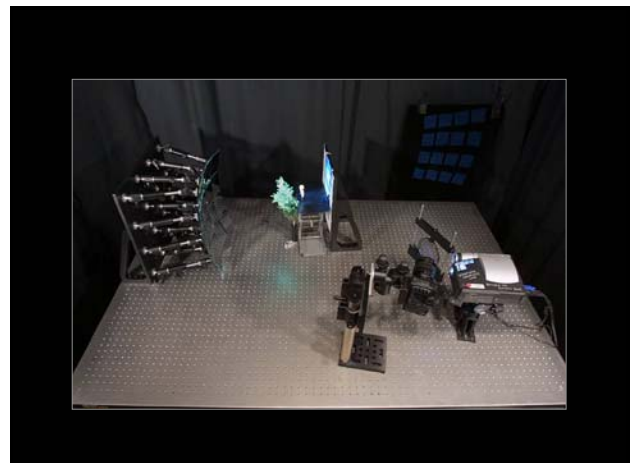
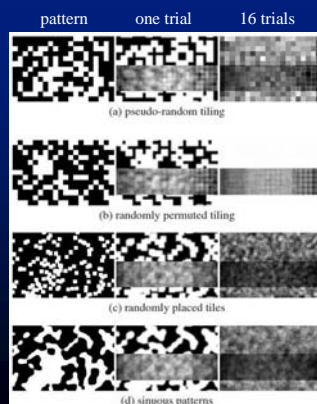
Synthetic aperture illumination

- technologies
 - array of projectors
 - array of microprojectors
 - single projector + array of mirrors

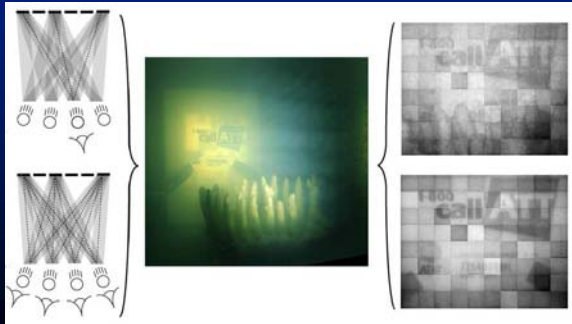
What does synthetic aperture illumination look like?



What are good patterns?

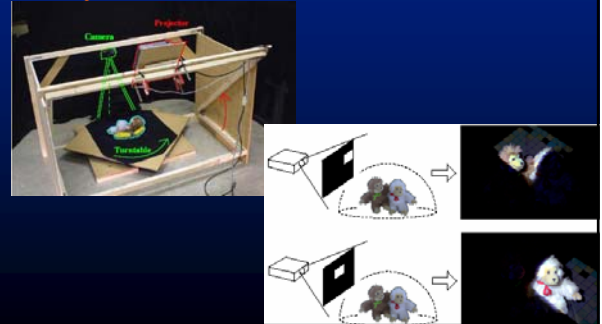


Underwater confocal imaging with and without SAP



6-D Methods and beyond...

Relighting with 4D Incident Light Fields Vincent Masselus, Pieter Peers, Philip Dutre and Yves D. Willems SIGG2003

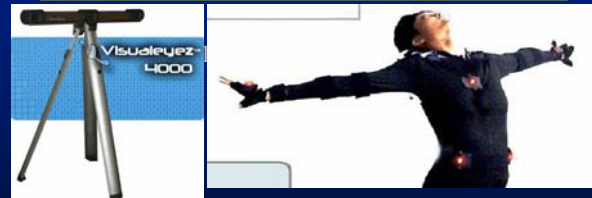


Computational Illumination

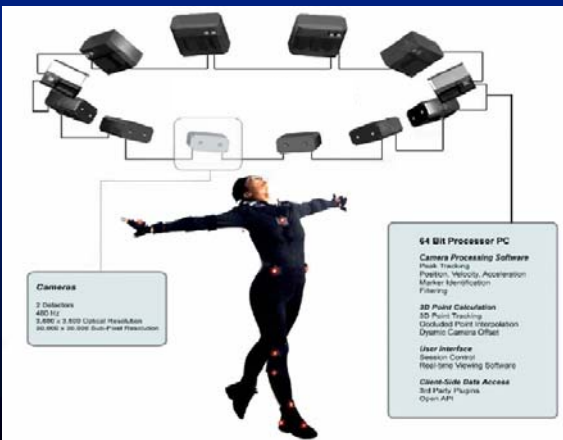
- Presence or Absence
 - Flash/No-flash
- Light position
 - Multi-flash for depth edges
 - Programmable dome (image re-lighting and matting)
- Light color/wavelength
- Spatial Modulation
 - Synthetic Aperture Illumination
- Temporal Modulation
 - TV remote, Motion Tracking, Sony ID-cam, RFIC
- General lighting condition
 - Day/Night

Ramesh Raskar, Computational Illumination

Demodulating Cameras

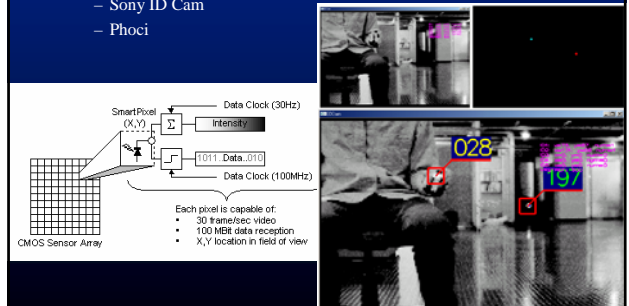


- Motion Capture Cameras
 - Visualeyez™ VZ4000 Tracking System
 - PhaseSpace motion digitizer




Demodulating Cameras


- Decode signals from blinking LEDs + image
 - Sony ID Cam
 - Phoci




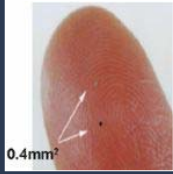
RFID Lamps :

Interacting with a Self-describing World via Photosensing Wireless Tags and Projectors





 Ramesh Raskar, Paul Beardsley, Jeroen van Baar, Yao Wang,
 Paul Dietz, Johnny Lee, Darren Leigh, Thomas Willwacher
 Mitsubishi Electric Research Labs (MERL), Cambridge, MA

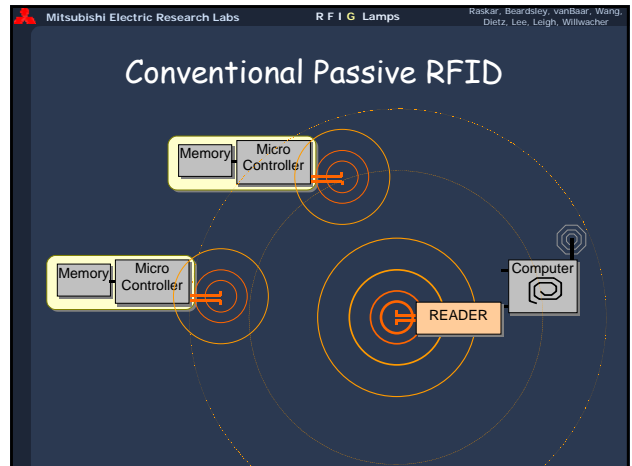
Radio Frequency Identification Tags (RFID)


No batteries,
 Small size,
 Cost few cents




Warehousing Routing Livestock tracking
 Library Baggage handling Currency



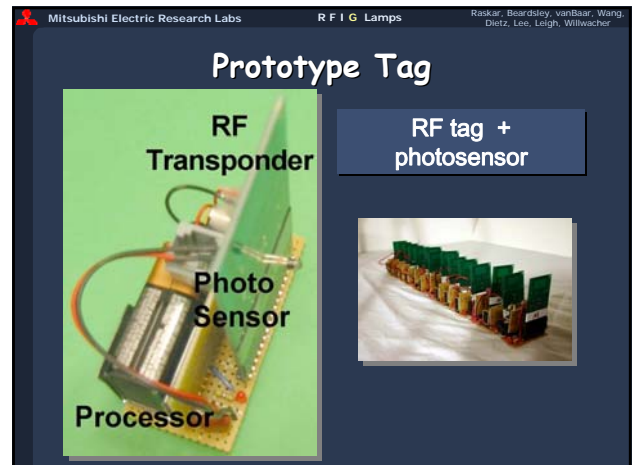
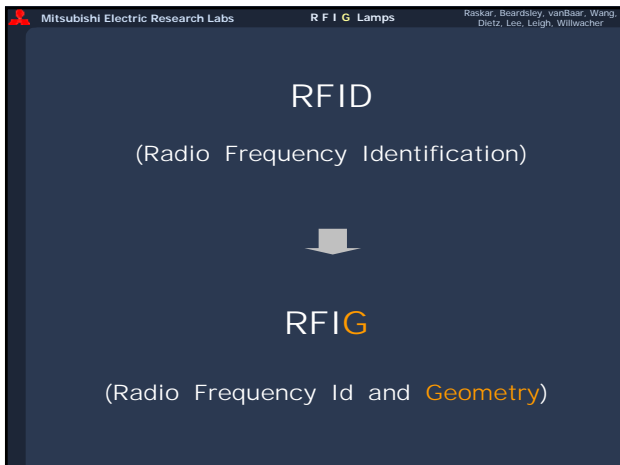
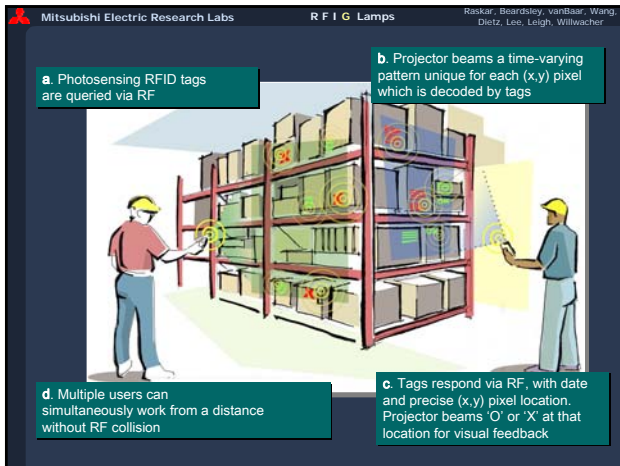
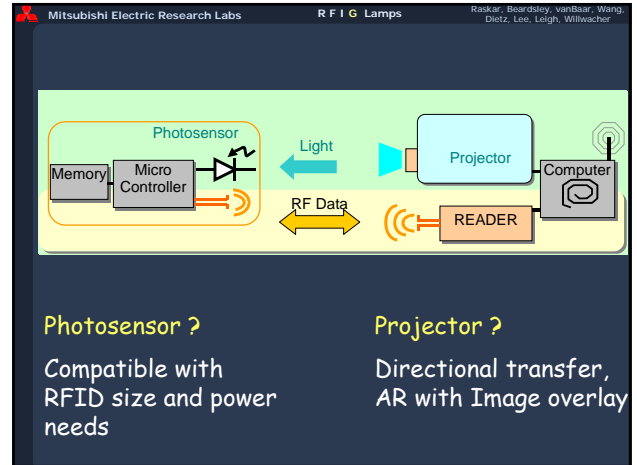
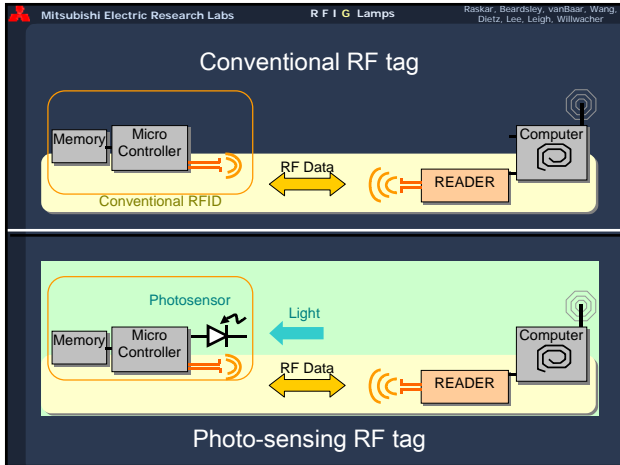
Tagged Books in a Library



- ✓ **Id**
Easy to get list of books in RF range
- ✗ **No Precise Location Data**
Difficult to find if the books in sorted order?
Which book is upside down?

Where are boxes with Products close to Expiry Date ?





Mitsubishi Electric Research Labs R F I G Lamps Raskar, Beardley, vanBaar, Wang, Dietz, Lee, Leigh, Willwacher

Projected Sequential Frames

- Handheld Projector beams binary coded stripes
- Tags decode temporal code

Mitsubishi Electric Research Labs R F I G Lamps Raskar, Beardley, vanBaar, Wang, Dietz, Lee, Leigh, Willwacher

Projected Sequential Frames

- Handheld Projector beams binary coded stripes
- Tags decode temporal code

Mitsubishi Electric Research Labs R F I G Lamps Raskar, Beardley, vanBaar, Wang, Dietz, Lee, Leigh, Willwacher

Projected Sequential Frames

- Handheld Projector beams binary coded stripes
- Tags decode temporal code

Mitsubishi Electric Research Labs R F I G Lamps Raskar, Beardley, vanBaar, Wang, Dietz, Lee, Leigh, Willwacher

Projected Sequential Frames

- Handheld Projector beams binary coded stripes
- Tags decode temporal code

Mitsubishi Electric Research Labs R F I G Lamps Raskar, Beardley, vanBaar, Wang, Dietz, Lee, Leigh, Willwacher

Projected Sequential Frames

- Handheld Projector beams binary coded stripes
- Tags decode temporal code

Mitsubishi Electric Research Labs R F I G Lamps Raskar, Beardley, vanBaar, Wang, Dietz, Lee, Leigh, Willwacher

Projected Sequential Frames

0 1 1 0 0 X=12

For each tag

- From light sequence, decode x and y coordinate
- Transmit back to RF reader (Id, x, y)

Mitsubishi Electric Research Labs R F I G Lamps Raskar, Beardsley, vanBaar, Wang, Dietz, Lee, Leigh, Willwacher

Visual feedback of 2D position

a. Receive via RF $\{(x_1, y_1), (x_2, y_2), \dots\}$ pixels

b. Illuminate those positions

Computational Illumination

- Presence or Absence
 - Flash/No-flash
- Light position
 - Multi-flash for depth edges
 - Programmable dome (image re-lighting and matting)
- Light color/wavelength
- Spatial Modulation
 - Synthetic Aperture Illumination
- Temporal Modulation
 - TV remote, Motion Tracking, Sony ID-cam, RFIG
- Natural lighting condition
 - Day/Night Fusion

Ramesh Raskar, Computational Illumination

A Night Time Scene:
Objects are Difficult to Understand due to Lack of Context

Dark Bldgs

Reflections on bldgs

Unknown shapes

Enhanced Context :
All features from night scene are preserved, but background in clear

Well-lit Bldgs

Reflections in bldgs windows

Tree, Street shapes

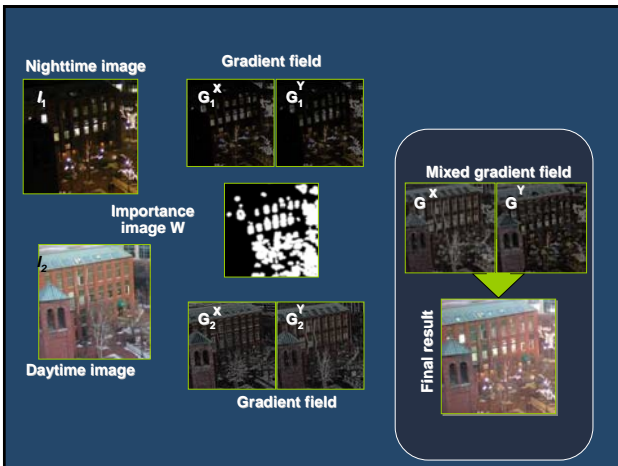
Night Image

Background is captured from day-time scene using the same fixed camera

Day Image

Result: Enhanced Image

Mask is automatically computed from scene contrast



Reconstruction from Gradient Field

- Problem: minimize error $|\nabla I' - G|$
- Estimate I' so that

$$G = \nabla I'$$

- Poisson equation
- $\nabla^2 I' = \text{div } G$
- Full multigrid solver

G^x

G^y

I'

Video Enhancement using Fusion

- Video from fixed cameras
 - Improve low quality InfraRed video using high-quality visible video
 - Fill in dark areas, enhance change in intensity
 - Output style: better context
- Current Demo
 - Fusion of Night video and Daytime image

Original Video Frame

Easy-to-understand Non-photorealistic (NPR) Image or Video

Ramesh Raskar, CompPhoto Class Northeastern, Fall 2005

Details

- Combine day and night time images
 - Night videos have low contrast, areas with no detail
 - Same camera during day can capture static information
 - Dark areas of night video are replaced to provide context
 - Moving object (from night) + Static scene (from day)

Modified Surveillance Camera

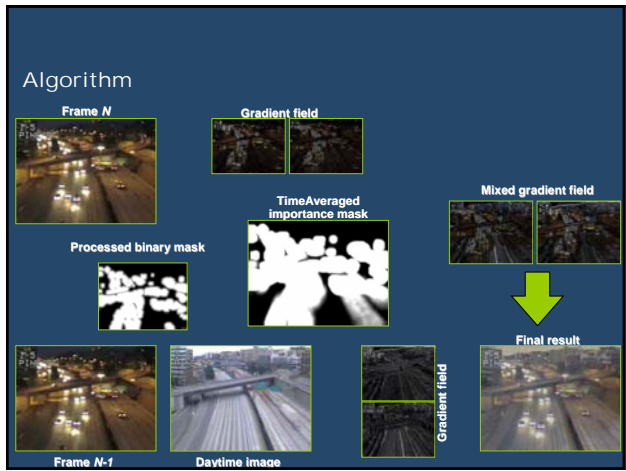
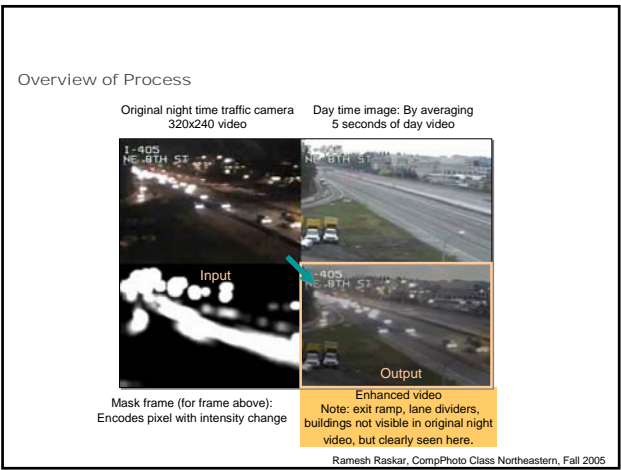
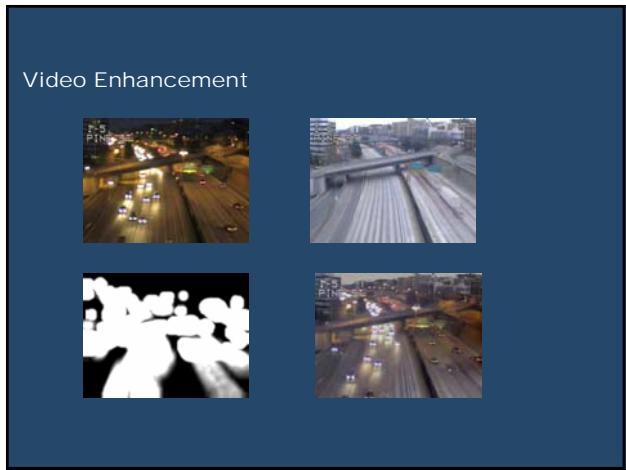
Day time Photograph

Night time Video (or Photo)

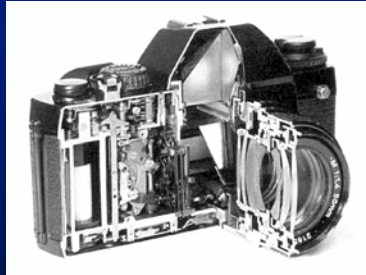
Combine pixels depending on context, image and temporal gradient

Enhanced Night Video (or Photo) with context

Ramesh Raskar, CompPhoto Class Northeastern, Fall 2005



D.2: Smart Optics, Modern Sensors and Future Cameras



Ramesh Raskar

Mitsubishi Electric Research Labs

Jack Tumblin

Northwestern University

Course WebPage :

<http://www.merl.com/people/raskar/photo>

Computational Photography

Novel Illumination

Light Sources



Novel Cameras

Generalized
Sensor

Processing

Generalized
Optics

Display

Recreate 4D Lightfield



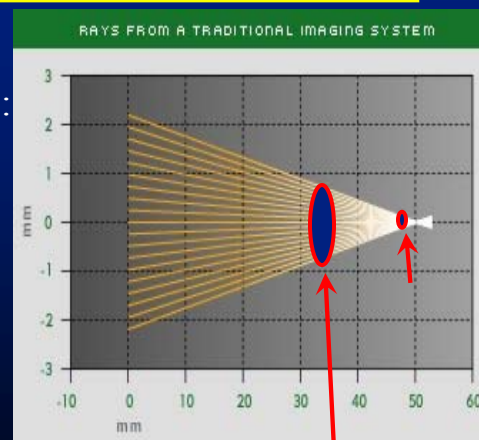
Scene: 8D Ray Modulator

Future Directions

- Smart Lighting
 - Light stages, Domes, Light waving, Towards 8D
- Computational Imaging outside Photography
 - Tomography, Coded Aperture Imaging
- Smart Optics
 - Handheld Light field camera, Programmable imaging/aperture
- Smart Sensors
 - HDR Cameras, Gradient Sensing, Line-scan Cameras, Demodulators
- Speculations

Wavefront Coding: 10X Depth of Field

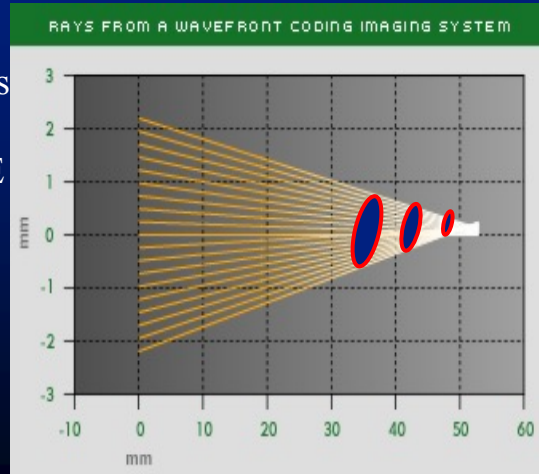
- In-focus: small 'Circle of Confusion':
- Out-of-focus: LARGE 'circle of confusion'
- Coma-like distortion: Make Circle MOVE as focus changes:



http://www.cdm-optics.com/site/extended_dof.php

Wavefront Coding: 10X Depth of Field

- In-focus: small 'Circle of Confusion'
- Out-of-focus: LARGE "circle of confusion"
- Coma-like distortion allows us to De-convolve, sharpen out-of-focus items



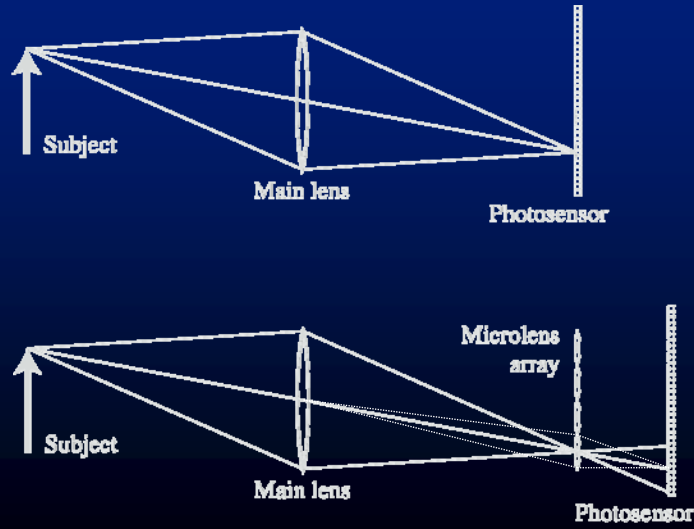
http://www.cdm-optics.com/site/extended_dof.php

Light field photography using a handheld plenoptic camera

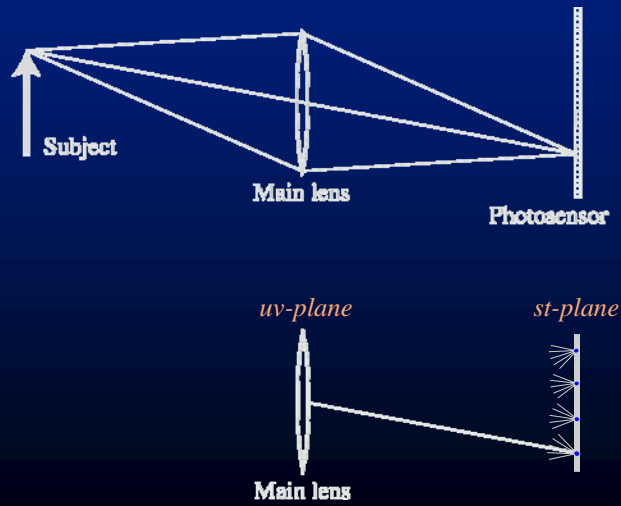
*Ren Ng, Marc Levoy, Mathieu Brédif,
Gene Duval, Mark Horowitz and Pat Hanrahan*



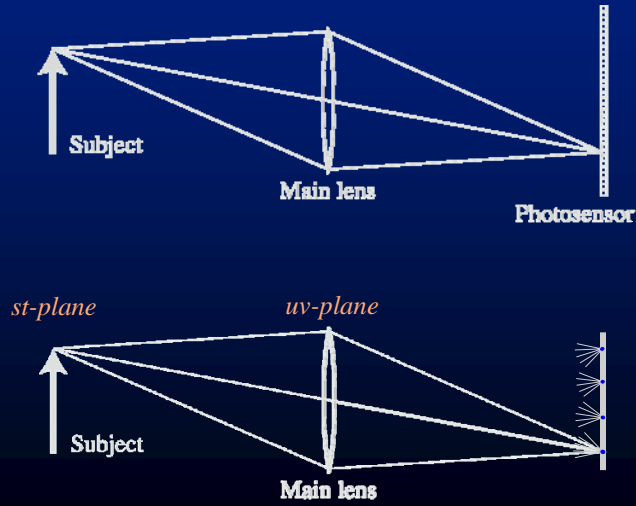
Conventional versus light field camera



Conventional versus light field camera



Conventional versus light field camera



Prototype camera



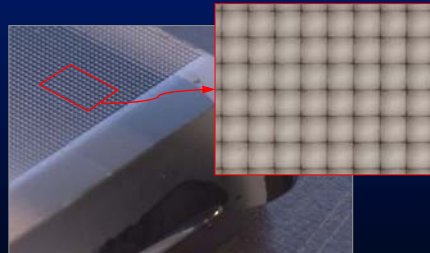
Contax medium format camera



Kodak 16-megapixel sensor

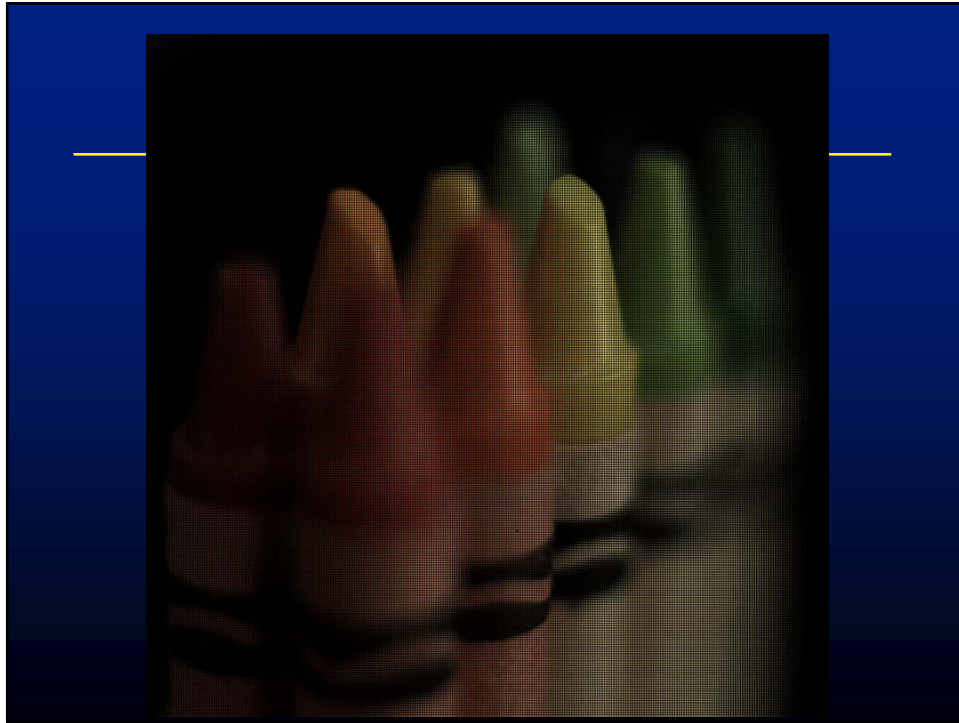


Adaptive Optics microlens array

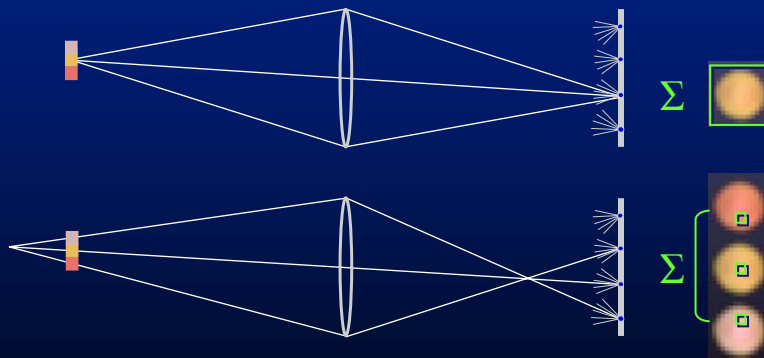


125 μ square-sided microlenses

$$4000 \times 4000 \text{ pixels} \div 292 \times 292 \text{ lenses} = 14 \times 14 \text{ pixels}$$



Digital refocusing



refocusing = summing windows
extracted from several microlenses

Example of digital refocusing



Extending the depth of field



conventional photograph,
main lens at $f/4$

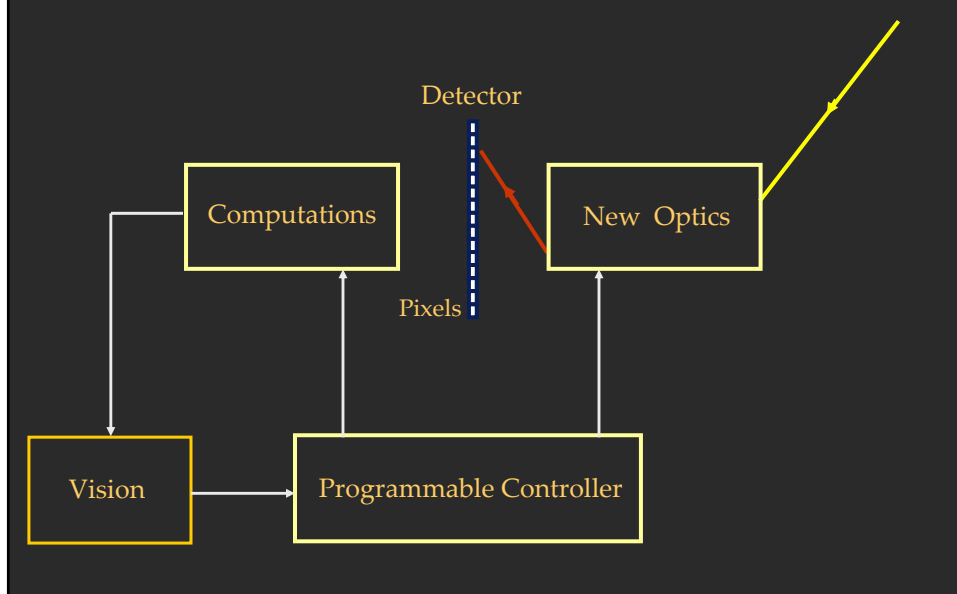


conventional photograph,
main lens at $f/22$



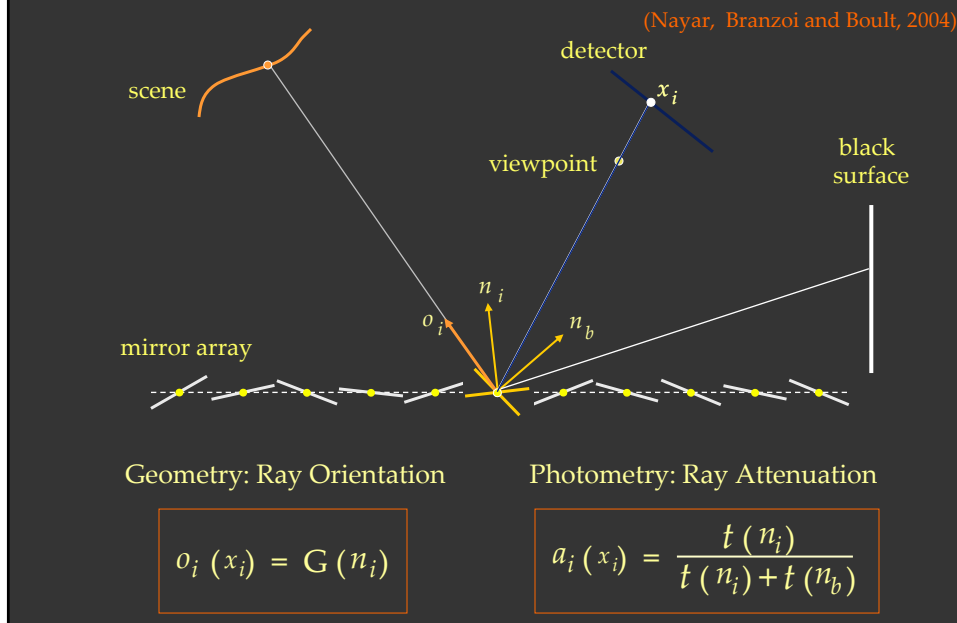
light field, main lens at $f/4$,
after all-focus algorithm
[Agarwala 2004]

Programmable Imaging



Imaging Through Micro-Mirrors

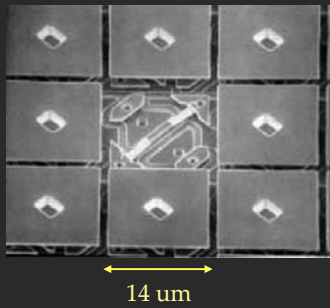
(Nayar, Branzoi and Boulton, 2004)



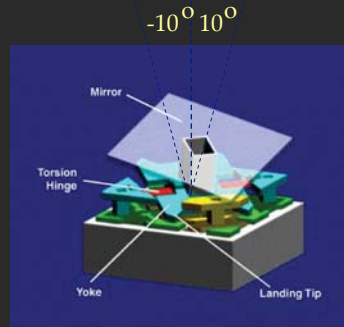
Digital Micromirror Device (DMD)

(by Texas Instruments)

DMD Array:



Micromirror Architecture:



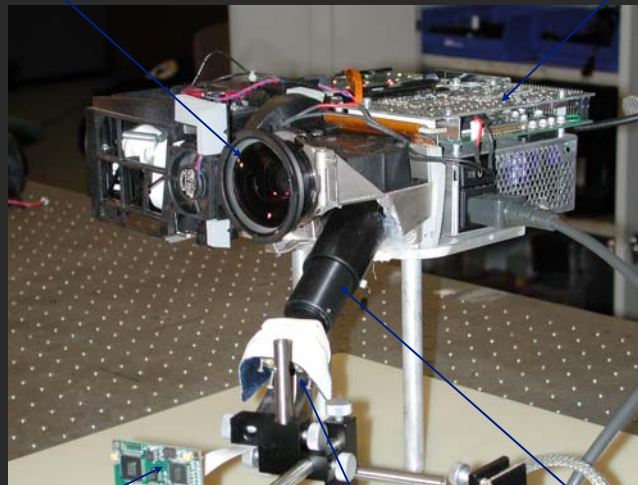
DMD for Imaging:

(Malbet et al. 95, Kearney et al. 98, Castracane et al. 99, Christensen et al. 02)

Programmable Imaging System

Imaging Lens

DMD Electronics

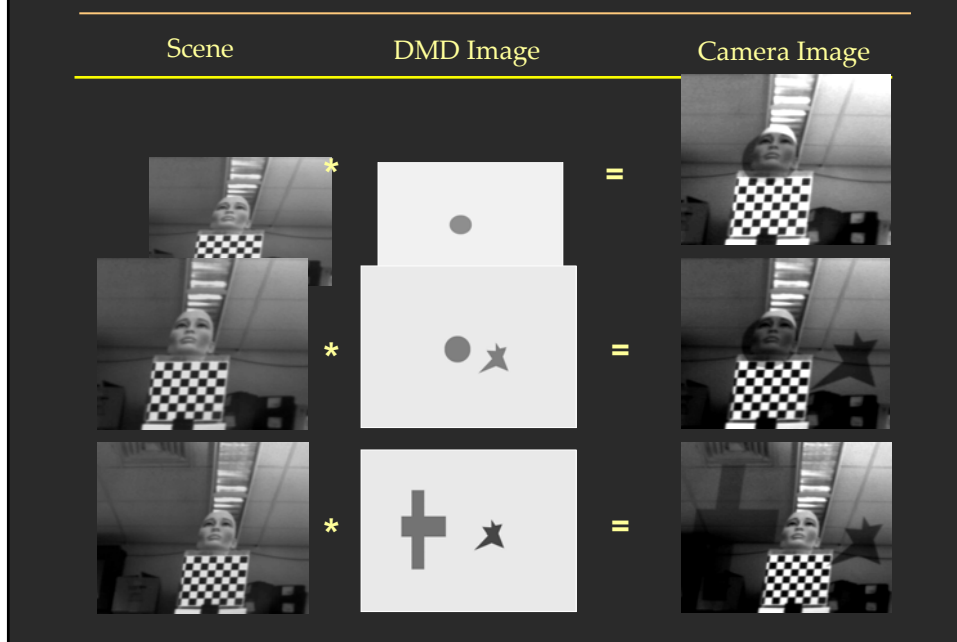


Camera Electronics

Tilted CCD

Lens Focused on DMD

Modulation: Examples



Optical Intra-Pixel Feature Detection

$$f * g = f * g^+ - f * g^-$$

CCD Pixels DMD Pixels

Laplacian

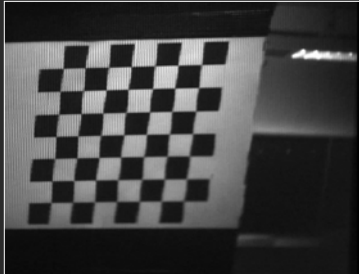
1	4	1	=	1	4	1	-	0	0	0
4	-20	4	=	4	0	4	-	0	20	0
1	4	1	=	1	4	1	-	0	0	0

Laplacian Image:

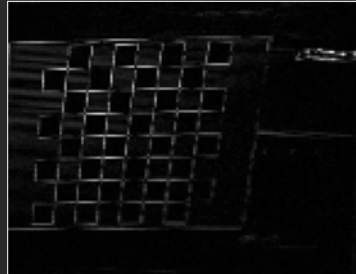
1	4	1	1	4	1	-	0	0	0	0	0	0
4	0	4	4	0	4	-	0	20	0	0	20	0
1	4	1	1	4	1	-	0	0	0	0	0	0
1	4	1	1	4	1	-	0	0	0	0	0	0
4	0	4	4	0	4	-	0	20	0	0	20	0
1	4	1	1	4	1	-	0	0	0	0	0	0

Optical Edge Detection

Scene Video



Edge Video



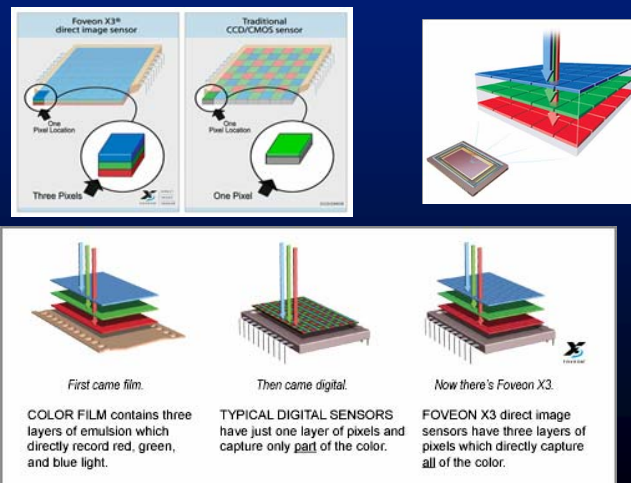
Generalized Optics and Sensors

- Smart Optics
 - Handheld Light field camera,
 - Programmable imaging/aperture
- Smart Sensors
 - HDR Cameras,
 - Gradient Sensing,
 - Line-scan Cameras,
 - Demodulators

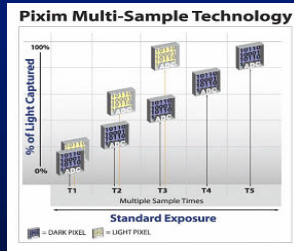
Future Directions

- Smart Lighting
 - Light stages, Domes, Light waving, Towards 8D
- Computational Imaging outside Photography
 - Tomography, Coded Aperture Imaging
- Smart Optics
 - Handheld Light field camera, Programmable imaging/aperture
- Smart Sensors
 - HDR Cameras, Gradient Sensing, Line-scan Cameras, Demodulators
- Speculations

Foveon: Thick Sensor

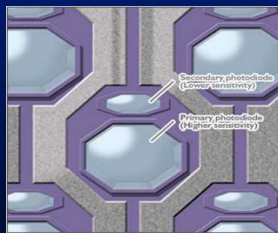


Pixim



High Dynamic Range

<http://www.cybergrain.com/tech/hdr/>



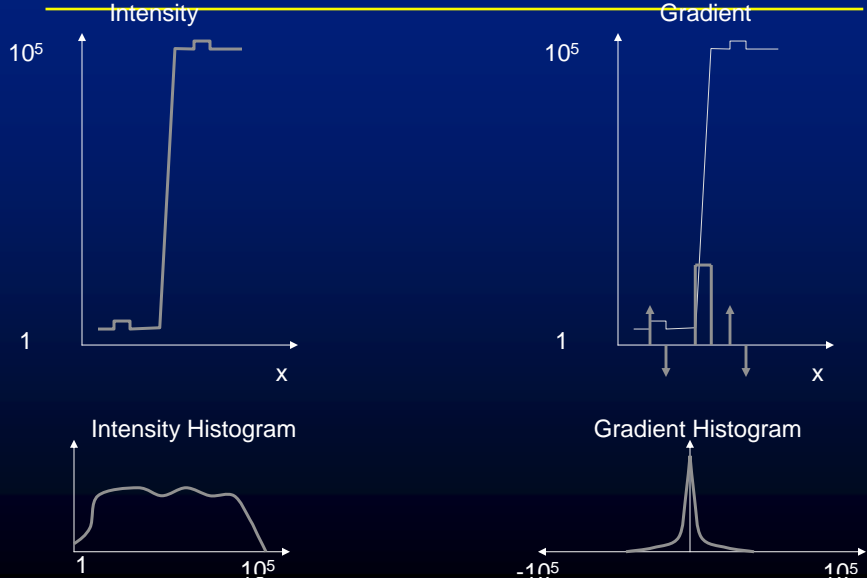
Fuji's SuperCCD S3 Pro camera has a chip with high and low sensitivity sensors per pixel location to increase dynamic range


Gradient Camera

Sensing Pixel Intensity Difference with
Locally Adaptive Gain


Ramesh Raskar, MERL
Work with Jack Tumblin, Northwestern U,
Amit Agrawal, U of Maryland

Natural Scene Properties







Original Image
Intensity values ranging from **0 to 1800**
Intensity ramp plus low contrast logo




Intensity Camera Image
8 bit camera for 1:1000 range
Problem: . **saturation** at high intensity regions



Log Camera Image
8 bit log for 1:10⁶ range
Problem: Visible **quantization effects** at high intensities



Locally Adaptive Gain
Pixel divided by the average of local neighborhood.
Thus the **low frequency contents are lost** and only detail remains.

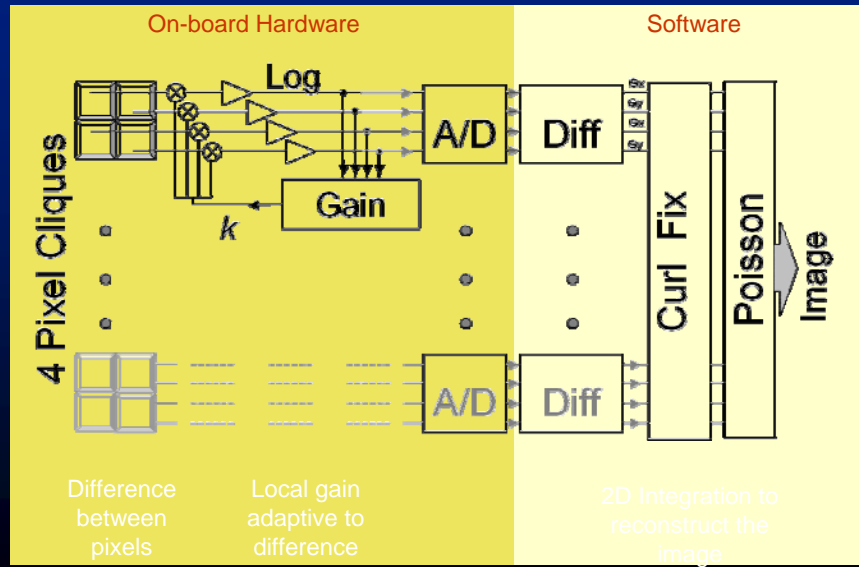


Gradient Camera Image
In proposed method, we sense intensity differences. We use a 8 bit A/D with resolution of $\log(1.02)$ to capture 2% contrast change between adjacent pixels. Notice that the details at both high and low intensities are captured.

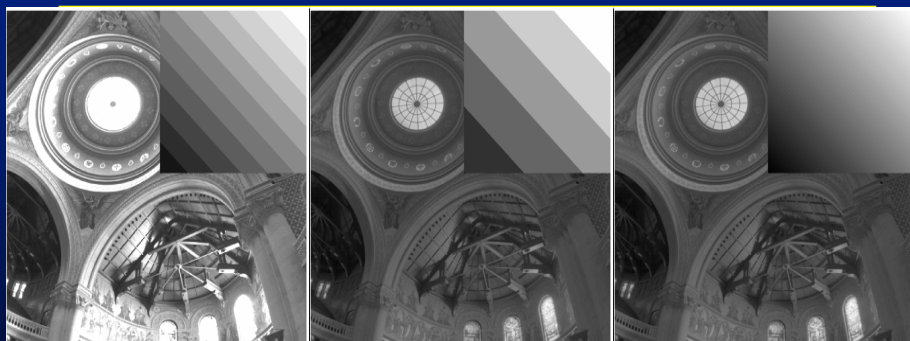
Gradient Camera

- Two main features
 1. Sense difference between neighboring pixel intensity
At each pixel, measure (∇_x, ∇_y) , $\nabla_x = I_{x+1,y} - I_{x,y}$, $\nabla_y = I_{x,y+1} - I_{x,y}$
 2. With locally adaptive gain
- Gradient camera is very similar to locally adaptive gain camera
- Locally Adaptive Gain Camera
 - Gain is **different** for each pixel
 - Problem: Loses low frequency detail and preserves only high frequency features (edges)
- Gradient Camera
 - The gain is **same** for four adjacent pixels
 - Difference between two pixels is measured with same gain on both pixels
 - Reconstruct original image in software from pixel differences by solving a linear system (solving Poisson Equation)

Camera Pipeline



Detail Preserving



Intensity Camera

Log Intensity Camera

Gradient Camera

Intensity cameras capture detail but lose range
Log cameras capture range but lose detail

Quantization



Original Image



Uniform quantization 3 bits



Log Uniform quantization 3 bits

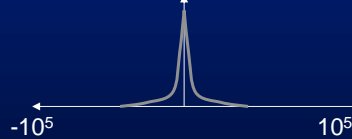


Log Uniform gradients quantization 3 bits

Intensity Histogram



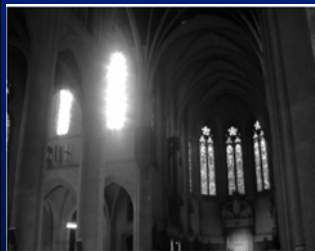
Gradient Histogram



GradCam requires fewer bits

In the reconstructed image, error is pushed to high gradient pixel positions which is visually imperceptible

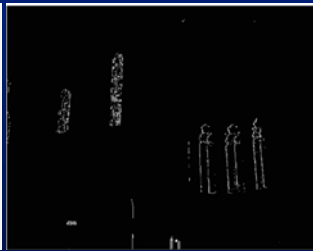
High Dynamic Range Images



Scene



Intensity camera saturation map



Gradient camera saturation map

*Intensity camera fail to capture range
Gradients saturate at very few isolated pixels*

Line Scan Camera: PhotoFinish 2000 Hz

GAMES: IAAF GRAND PRIX FINAL - PARIS 2002

100m MEN

WIND: +2.0

START: 16:28 AU.

LN	NUM.	TIME
1	79	9.78
2	19	9.87
3	72	9.97
4	60	9.98
5	55	10.03
6	85	10.05
7	25	10.20
8	78	10.21

*KEY RECORD

SEIKO

WIN

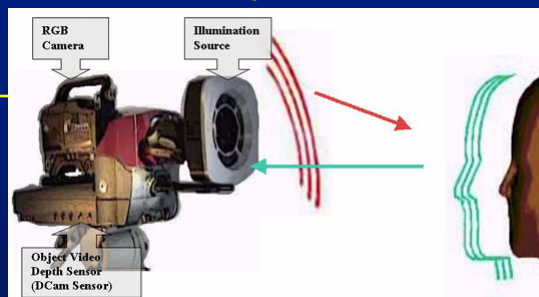
Dead Heat

10.1 10.0

3D Cameras

- Time of flight
 - ZCam (Shuttered Light Pulse)
- Phase Decoding of modulated illumination
 - Canesta (Phase comparison)
 - Phase difference = depth
 - Magnitude = reflectance
- Structured Light
 - Binary coded light and triangulation

ZCam (3Dvsystems), Shuttered Light Pulse



Resolution :
1cm for 2-7 meters

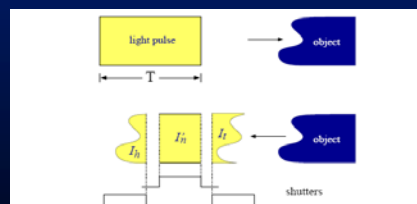


Figure 1: A light pulse of duration T radiates an object and is reflected back to the sensor. The signal is shuttered at the head, center and tail of the signal. The measured intensities I_h and I_t are functions of the distance travelled by the pulse, while the intensity I_c is a constant fraction of the unshuttered value I_n .



Objects that lay outside the designated range (both near the camera or further away in the background) are not included in the key.



Graphics can inserted behind and between characters

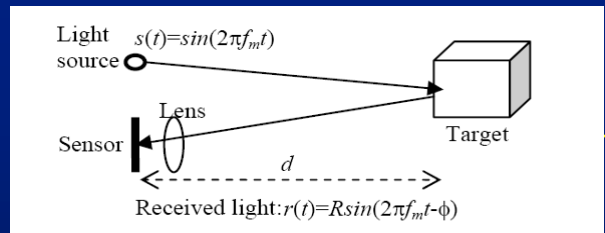


Figure 2. Time of flight measurement.

Canesta: Modulated Emitter

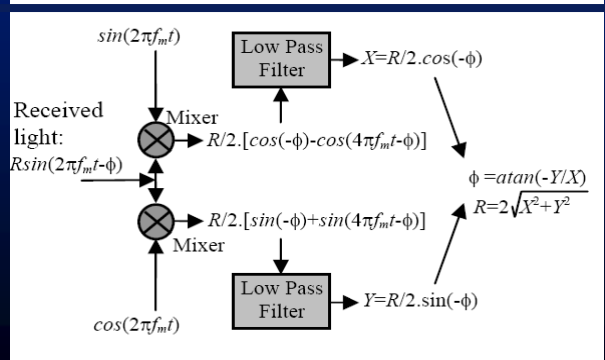


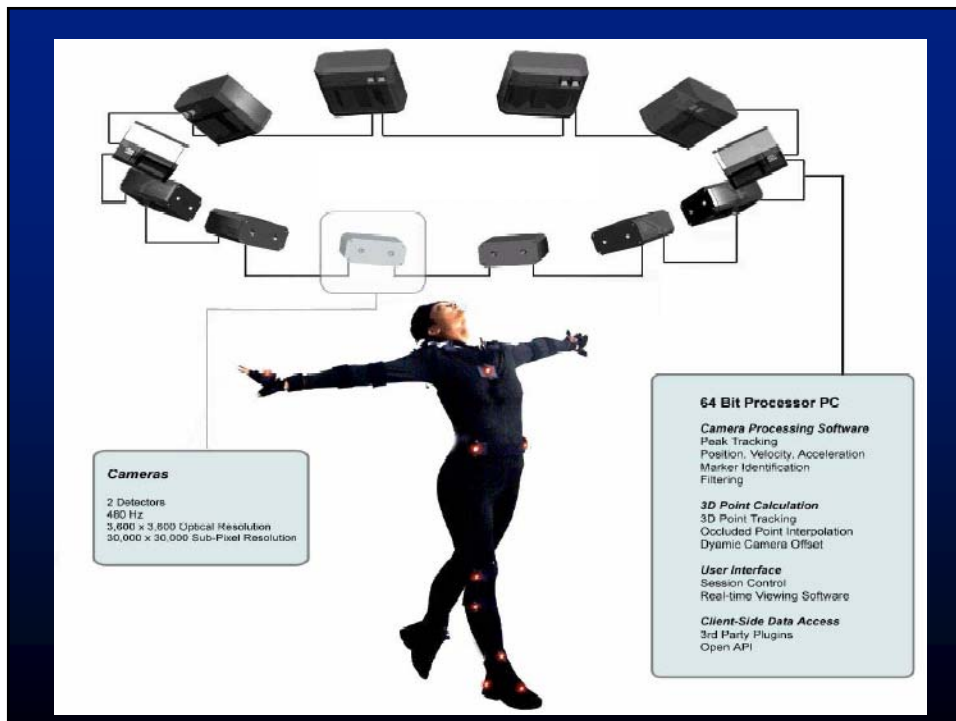
Figure 3. A method of phase/amplitude calculation.

Phase ~ distance
Amplitude ~ reflectance

Demodulating Cameras

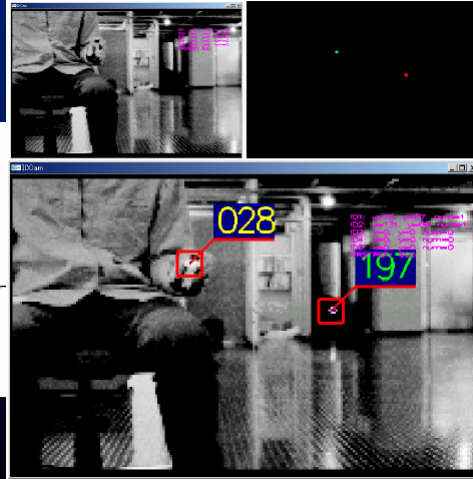
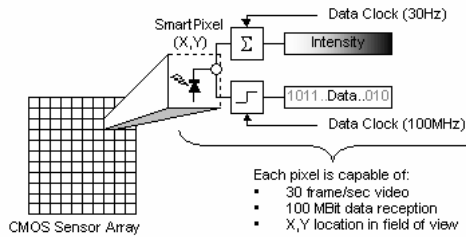


- Motion Capture Cameras
 - Visualeyez™ VZ4000 Tracking System
 - PhaseSpace motion digitizer



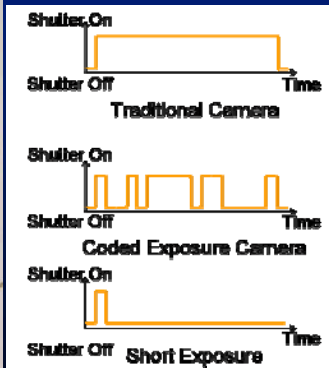
Demodulating Cameras

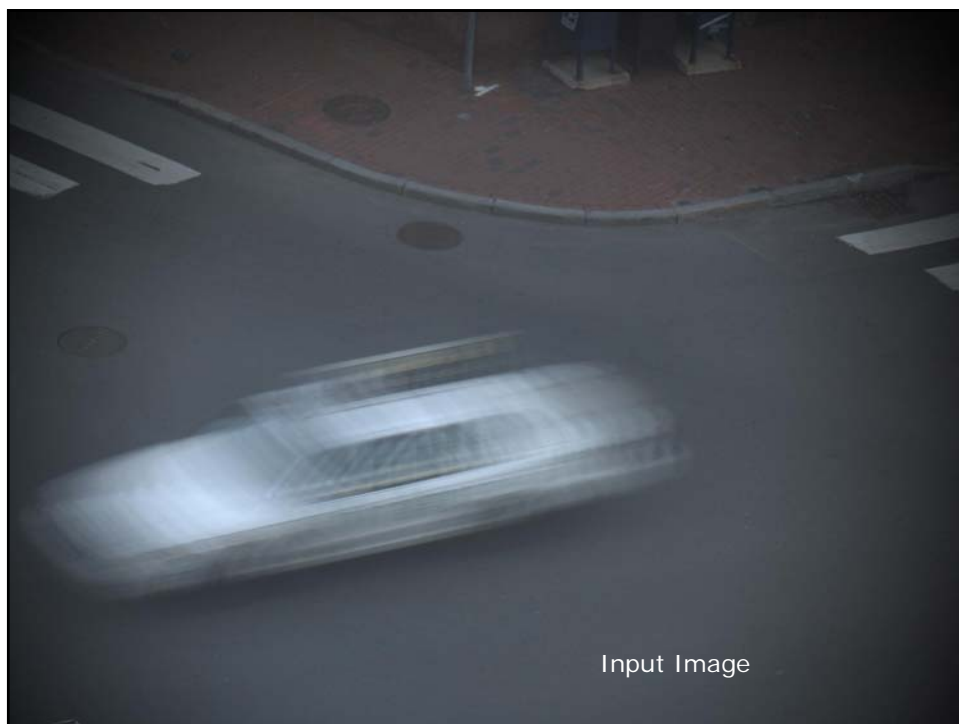
- Decode signals from blinking LEDs + image
 - Sony ID Cam
 - Phoci



Fluttered Shutter Camera

Raskar, Agrawal, Tumblin Siggraph2006





Input Image



Rectified Image to make motion lines parallel to scan lines.



Approximate cutout of the blurred image containing the taxi (vignetting on left edge). Exact alignment of cutout with taxi extent is not required.



Image Deblurred by solving a linear system. No post-processing

Coded Exposure Photography

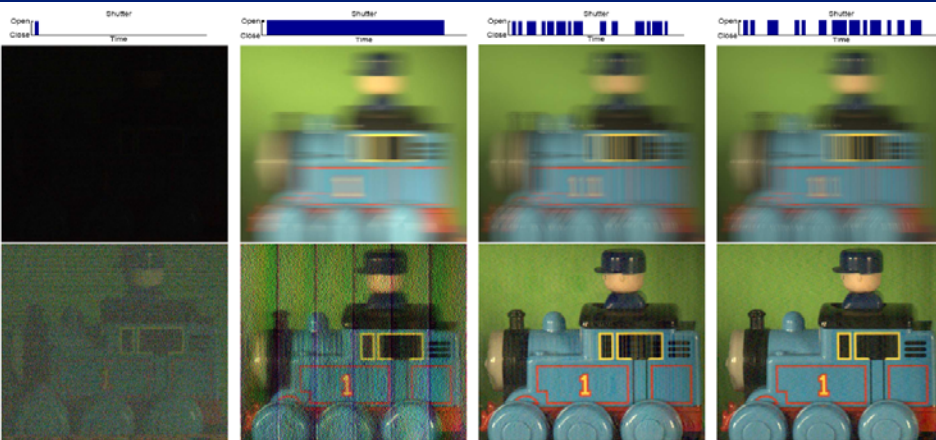
Raskar, Agrawal, Tumblin Siggraph2006

Short Exposure

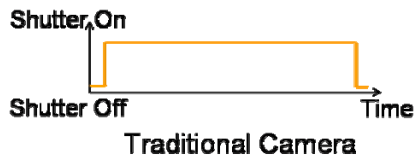
Traditional

MURA

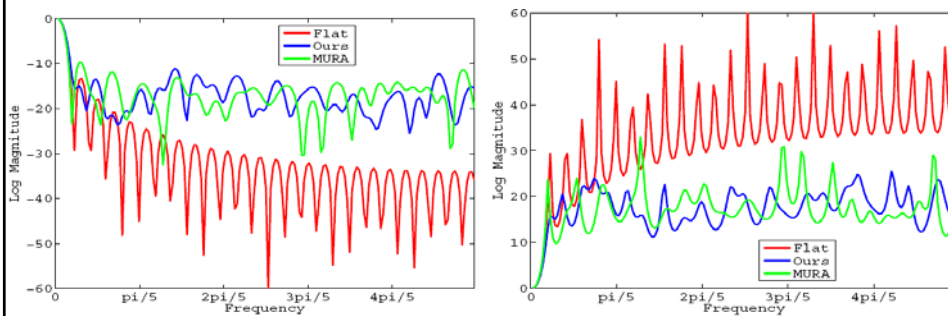
Coded



Coded Exposure Photography



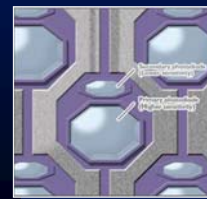
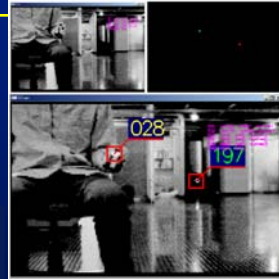
Discrete Fourier Transform of Convolution Filter



Converting Deblurring into a Well-posed Problem

Novel Sensors

- Gradient sensing
- HDR Camera, Log sensing
- Line-scan Camera
- Demodulating
- Motion Capture
- Fluttered Shutter
- 3D



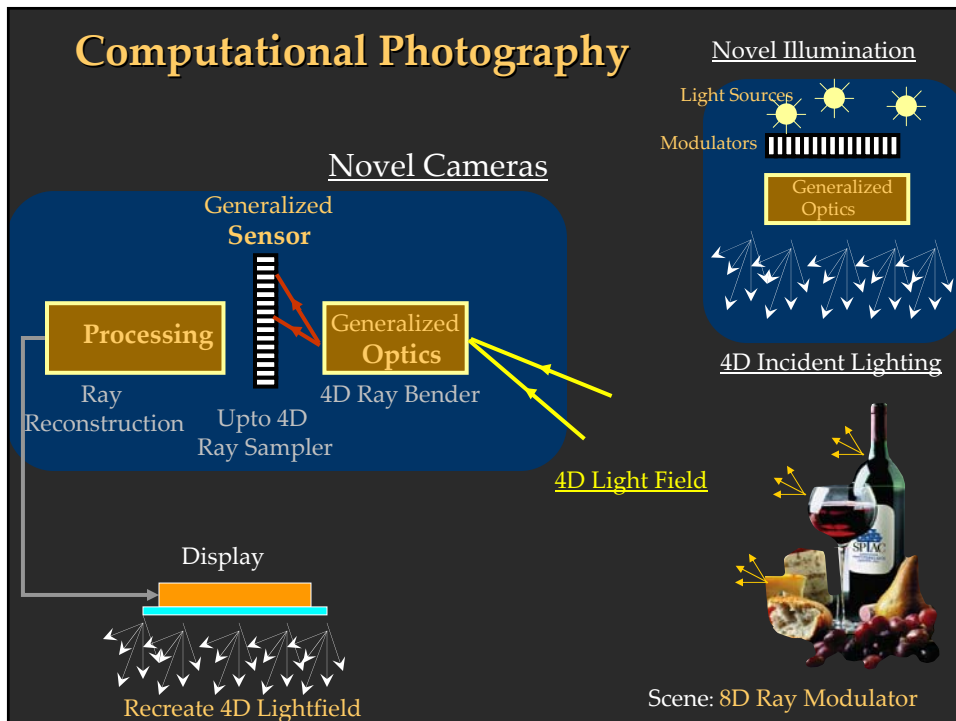
Fantasy Configurations

- 'Cloth-cam': 'Wallpaper-cam'
elements 4D light emission and 4D capture in the surface of a cloth...
- Floating Cam: ad-hoc wireless networks form camera arrays in environment...
- Other ray sets:
Multilinear cameras, canonical 'basis' cameras (linear combination of 8 types)
McMillan'04, '05

Dream of A New Photography

	Old	New
• People and Time	~Cheap	Precious
• Each photo	Precious	Free
• Lighting	Critical	Automated?
• External Sensors	No	Yes
• 'Stills / Video'	Disjoint	Merged
• Exposure Settings	Pre-select	Post-Process
• Exposure Time	Pre-select	Post-Process
• Resolution/noise	Pre-select	Post-Process
• 'HDR' range	Pre-select	Post-Process

Computational Photography



Film-like Photography with bits	Computational Photography			
	Computational Camera			Smart Light
Digital Photography	Computational Processing	Computational Imaging/Optics	Computational Sensor	Computational Illumination
Image processing applied to captured images to produce "better" images.	Processing of a set of captured images to create "new" images.	Capture of optically coded images and computational decoding to produce "new?" images.	Detectors that combine sensing and processing to create "smart" pixels.	Adapting and Controlling Illumination to Create 'revealing' image
Examples: Interpolation, Filtering, Enhancement, Dynamic Range Compression, Color Management, Morphing, Hole Filling, Artistic Image Effects, Image Compression, Watermarking.	Examples: Mosaicing, Matting, Super-Resolution, Multi-Exposure HDR, Light Field from Multiple View, Structure from Motion, Shape from X.	Examples: Coded Aperture, Optical Tomography, Diaphanography, SA Microscopy, Integral Imaging, Assorted Pixels, Catadioptric Imaging, Holographic Imaging.	Examples: Artificial Retina, Retinex Sensors, Adaptive Dynamic Range Sensors, Edge Detect Chips, Focus of Expansion Chips, Motion Sensors.	Examples: Flash/no flash, Lighting domes, Multi-flash for depth edges, Dual Photos, Polynomial texture Maps, 4D light source

Acknowledgements

- MERL, Northwestern Graphics Group
- Amit Agrawal
- Shree Nayar
- Marc Levoy
- Jinbo Shi
- Ankit Mohan, Holger Winnemoller
- Image Credits
 - Ren Ng, Vaibhav Vaish, William Bennet
 - Fredo Durand, Aseem Agrawala
 - Morgan McGuire, Paul Debevec
 - And more

Non-photorealistic Camera: Depth Edge Detection and Stylized Rendering using Multi-Flash Imaging

Ramesh Raskar*
Mitsubishi Electric Research Labs (MERL)

Kar-Han Tan

Rogério Feris
UC Santa Barbara

Jingyi Yu†
MIT

Matthew Turk
UC Santa Barbara



Figure 1: (a) A photo of a car engine (b) Stylized rendering highlighting boundaries between geometric shapes. Notice the four spark plugs and the dip-stick which are now clearly visible (c) Photo of a flower plant (d) Texture de-emphasized rendering.

Abstract

We present a non-photorealistic rendering approach to capture and convey shape features of real-world scenes. We use a camera with multiple flashes that are strategically positioned to cast shadows along depth discontinuities in the scene. The projective-geometric relationship of the camera-flash setup is then exploited to detect depth discontinuities and distinguish them from intensity edges due to material discontinuities.

We introduce depiction methods that utilize the detected edge features to generate stylized static and animated images. We can highlight the detected features, suppress unnecessary details or combine features from multiple images. The resulting images more clearly convey the 3D structure of the imaged scenes.

We take a very different approach to capturing geometric features of a scene than traditional approaches that require reconstructing a 3D model. This results in a method that is both surprisingly simple and computationally efficient. The entire hardware/software setup can conceivably be packaged into a self-contained device no larger than existing digital cameras.

Keywords: non-photorealistic rendering, image enhancement, depth edges

1 Introduction

Our goal is to create stylized images that facilitate viewer comprehension of the shape contours of the objects depicted. Non-photorealistic rendering (NPR) techniques aim to outline the shapes of objects, highlight the moving parts to illustrate action, and re-

duce visual clutter such as shadows and texture details [Gooch and Gooch 2001]. The result is useful for imaging low contrast and geometrically complex scenes such as mechanical parts (Figure 1), plants or the internals of a patient (in endoscopy).

When a rich 3D model of the scene is available, rendering subsets of view-dependent contours is a relatively well-understood task in NPR [Saito and Takahashi 1990]. Extending this approach to real scenes by first creating 3D scene models, however, remains difficult. In this paper, we show that it is possible to bypass geometry acquisition, and directly create stylized renderings from images. In the place of expensive, elaborate equipment for geometry acquisition, we propose using a camera with a simple extension: multiple strategically positioned flashes. Instead of having to estimate the full 3D coordinates of points in the scene, and then look for depth discontinuities, our technique reduces the general 3D problem of depth edge recovery to one of intensity step edge detection.

Exploiting the imaging geometry for rendering results in a simple and inexpensive solution for creating stylized images from real scenes. We believe that our camera will be a useful tool for professional artists and photographers, and we expect that it will also enable the average user to easily create stylized imagery.

1.1 Overview

Our approach is based on taking successive photos of a scene, each with a different light source close to and around the camera's center of projection. We use the location of the shadows abutting depth discontinuities as a robust cue to create a depth edge map in both static and dynamic scenes.

Contributions Our main contribution is a set of techniques for detecting and rendering shape contours of scenes with low-contrast or high geometric complexity. Our technical contributions include the following.

- A robust edge classification scheme to distinguish depth edges from texture edges
- A collection of rendering and reconstruction techniques for creating images highlighting shape boundaries from 2D data without creating 3D representations, using qualitative depths
- An image re-synthesis scheme that allows abstraction of textured regions while preserving geometric features
- A technique to detect depth edges in dynamic scenes

*e-mail: [raskar,tan]@merl.com,[rferis,turk]@cs.ucsb.edu

†email: jingyi@graphics.csail.mit.edu



Figure 2: Traditional image enhancement by improving (Left) brightness and (Right) contrast. Low contrast depth edges remain difficult to perceive.

We introduce the concept of a **self-contained stylized imaging device**, a ‘non-photorealistic camera’, which can directly generate images highlighting contours of geometric shapes in a scene. It contains a traditional camera and embedded flashes, and can be readily and inexpensively built. We attempt to address two important issues in NPR [Gooch and Gooch 2001] [Strothotte and Schlechtweg 2002], detecting shape contours that should be enhanced and identifying features that should be suppressed. We propose a new approach to take image-stylization beyond the processing of a photograph, to actively changing how the photographs are taken.

The output images or videos can be rendered in many ways, e.g., technical illustration, line art or cartoon-like style. We highlight depth discontinuities, suppress material and illumination transitions, and create renderings with large, smoothly colored regions outlined with salient contours [Durand 2002]. We describe several applications: imaging complex mechanical parts, improving images for endoscopes, anatomical drawings and highlighting changes in a scene. Our approach shares the **disadvantages** of NPR: relevant details may be lost as an image is simplified, so tunable abstraction is needed (Section 3.3), and the usefulness of the output is often difficult to quantify.

1.2 Related Work

NPR from images, rather than 3D geometric models has recently received a great deal of attention. The majority of the available techniques for image stylization involve **processing a single image** as the input applying morphological operations, image segmentation, edge detection and color assignment. Some of them aim for stylized depiction [DeCarlo and Santella 2002] [Hertzmann 1998] while others enhance legibility. Interactive techniques for stylized rendering such as rotoscoping have been used as well [Waking Life 2001; Avenue Amy 2002], but we aim to automate tasks where meticulous manual operation was previously required. Our work belongs to an emerging class of techniques to create an enhanced image from multiple images, where the images are captured from the same viewpoint but under different conditions, such as under different illumination, focus or exposure [Cohen et al. 2003; Akers et al. 2003; Raskar et al. 2004].

Aerial imagery techniques find **shadow** evidence by thresholding a single intensity image, assuming flat ground and uniform albedo to infer building heights [Huertas and Nevatia 1988; Irvin and McKeown 1989; Lin and Nevatia 1998]. Some techniques improve shadow capture with novel shadow extraction techniques to compute new shadow mattes [Chuang et al. 2003] or remove them to improve scene segmentation [Toyama et al. 1999]. Some other techniques remove shadows without explicitly detecting them, such as using intrinsic images [Weiss 2001].

Stereo techniques including passive and active illumination are generally designed to compute depth values or surface orientation

rather than to detect depth edges. Depth discontinuities present difficulties for traditional stereo: it fails due to *half-occlusions*, i.e., occlusion of scene points in only one of the two views, which confuse the matching process [Geiger et al. 1992]. Few techniques try to model the discontinuities and occlusions directly [Birchfield 1999; Kang et al. 2001; Scharstein and Szeliski 2002]. Active illumination methods, which generally give better results, have been used for depth extraction, shape from shading, shape-time stereo and photometric stereo but are unfortunately unstable around depth discontinuities [Sato et al. 2001]. An interesting technique has been presented to perform logical operations on detected intensity edges, captured under widely varying illumination, to preserve shape boundaries [Shirai and Tsuji 1972] but it is limited to uniform albedo scenes. Using photometric stereo, it is possible to analyze the intensity statistics to detect high curvature regions at **occluding contours** or *fold*s [Huggins et al. 2001]. But the techniques assume that the surface is locally smooth which fails for a flat foreground object like a leaf or piece of paper, or view-independent edges such as corner of a cube. They detect regions near occluding contours but not the contours themselves.

Techniques for **shape from shadow** (or darkness) build a continuous representation (*shadowgram*) from a moving light source from which continuous depth estimates are possible [Raviv et al. 1989; Langer et al. 1995; Daum and Dudek 1998]. However, it involves a difficult problem of estimating continuous heights and requires accurate detection of start and end of shadows. Good reviews of shadow-based shape analysis methods are available in [Yang 1996] [Kriegman and Belhumeur 2001] [Savarese et al. 2001].

A common limitation of existing active illuminations methods is that the light sources need to surround the object, in order to create significant shading and shadow variation from (estimated or known 3D) light positions. This necessitates a **fixed lighting rig**, which limits the application of these techniques to industrial settings, and they are impossible to build into a self-contained camera.

We believe our proposed method for extracting depth edges is complementary with many existing methods for computing depth and 3D surface shape, as depth edges often violate smoothness assumptions inherent in many techniques. If the locations of depth discontinuities can be reliably detected and supplied as input, we believe that the performance of many 3D surface reconstruction algorithms can be significantly enhanced.

To find depth edges, we avoid the dependence on solving a correspondence problem or analyzing pixel intensity statistics with moving lights, and we do not attempt to estimate any continuous value. In our search, we have not seen a photometric or other type of stereo method successfully applied to complex scenes where the normals change rapidly— such as a potted plant, or a scene with high depth complexity or low intensity changes, such as a car engine or bone.

1.3 Outline

Our method for creating a stylized image of a static scene consists of the following.

- ▷ Capture a series of images of the scene under shifted light positions
- ▷ Process these images to automatically detect depth edges
- ▷ Identify the subset of intensity edges that are illumination and texture edges
- ▷ Compute qualitative depth relationships
- ▷ Enhance or simplify detected features for rendering
- ▷ Insert processed scene appearance for stylization

We use the term *depth edges* to refer to the C0 discontinuities in a depth map. Depth edges correspond to internal or external occluding contours (or silhouettes) or boundaries of physical objects. The depth edges recovered are *signed*: in the local neighborhood,

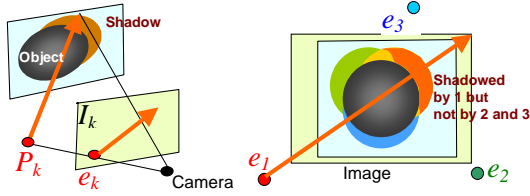


Figure 3: Imaging geometry. Shadows of the gray object are cast along the epipolar ray. We ensure that depth edges of all orientations create shadow in at least one image while the same shadowed points are lit in some other image.

the side with lower depth value, *foreground*, is considered positive while the opposite side is *background* and negative. *Texture edges* are reflectance changes or material discontinuities. Texture edges typically delineate textured regions.

In Section 2, we describe our approach to capturing important features using a multi-flash setup. In Section 3, we discuss methods to use the information to render the images in novel styles. In Section 4, we address the problem of extending the technique to dynamic scenes. We describe our results in Section 5 and conclude with discussion of limitations and future directions.

2 Capturing Edge Features

The image capturing process consists of taking successive pictures of a scene with a point light source *close* to the camera’s center of projection (COP). Due to a small *baseline* distance between the camera COP and the light source, a narrow sliver of shadow appears abutting each edge in the image with depth discontinuities; its width depends on the distance from the object edge to the background surface. By combining information about abutting cast shadow from two or more images with distinct light source positions, we can find the depth edges.

2.1 Depth Edges

The method for detecting depth edges is the foundation for our approach. The idea is very simple, in retrospect. It allows us to classify other edges by a process of elimination.

Our method is based on two observations regarding epipolar shadow geometry, as shown in Figure 3. The image of the point light source at P_k is at pixel e_k in the camera image, and is called the *light epipole*. The images of the pencil rays originating at P_k are the *epipolar rays* originating at e_k . (When P_k is behind the camera center, away from the image plane, the epipolar rays wrap around at infinity.) First, note that, a shadow of a depth edge pixel is constrained to lie along the epipolar ray passing through that pixel. Second, the shadow is observed if and only if the background pixel is on the side of the depth edge opposite the epipole *along the epipolar ray*. Hence, in general, if two light epipoles lie on opposite sides of an edge, a cast shadow will be observed at the depth edge in one image but not the other.

We detect shadows in an image by taking a ratio of the image with the maximum composite of all the images. The ratio image accentuates shadows, which abut the depth edges, and de-emphasizes texture edges. During epipolar traversal in the ratio image, the entry point of a shadowed region indicates a depth edge. The basic algorithm is as follows: Given n light sources positioned at P_1, P_2, \dots, P_n ,

- Capture ambient image I_0
- Capture n pictures I_k^+ , $k = 1..n$ with a light source at P_k
- Compute $I_k = I_k^+ - I_0$
- For all pixels x , $I_{max}(x) = \max_k(I_k(x))$, $k = 1..n$
- For each image k ,

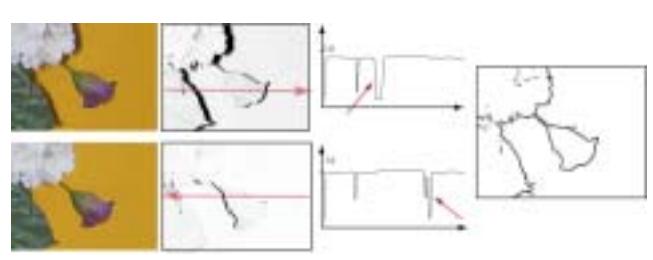


Figure 4: Detecting depth edges. (a) Photo (b) Ratio image (c) Plot along an epipolar ray, the arrows indicate negative transitions (d) Detected edges

▷ Create a ratio image, R_k , where $R_k(x) = I_k(x)/I_{max}(x)$

- For each image R_k

▷ Traverse each epipolar ray from epipole e_k

▷ Find pixels y with step edges with negative transition

▷ Mark the pixel y as a depth edge

With a number of light sources (minimum 2, but typically 4 to 8 are used) placed strategically around the camera, depth edges of all orientation with sufficient depth differences can be detected. In each image, as long as the epipolar ray at a depth edge pixel is not parallel to the image-space orientation of the depth edge, a step edge with negative transition (from lit part to shadowed part) will be detected. If the depth edge is oriented along the epipolar ray, the step edge cannot be detected.

Let us look at the algorithm in detail. Note that, the image I_k has ambient component removed, i.e., $I_k = I_k^+ - I_0$, where I_0 is an image taken with only ambient light and none of the n light sources on. The base image is the maximum composite image, I_{max} , which is an approximation of the image with light source at the camera COP, and in general has no shadows from any of the n light sources. The approximation is close if the n light sources are evenly distributed around the camera COP, have the same magnitude and the baseline is sufficiently smaller than the depth of the scene being imaged.

Consider the image of a 3D point X , given in camera coordinate system, imaged at pixel x . The intensity, $I_k(x)$, if X is lit by the light source at P_k , under lambertian assumption, is given by

$$I_k(x) = \mu_k \rho(x) (\hat{L}_k(x) \cdot N(x))$$

Otherwise, $I_k(x)$ is zero. The scalar μ_k is the magnitude of the light intensity and $\rho(x)$ is the reflectance at X . $\hat{L}_k(x)$ is the normalized light vector $L_k(x) = P_k - X$, and $N(x)$ is the surface normal, all in the camera coordinate system.

Thus, when X is seen by P_k , the ratio is as follows.

$$R_k(x) = \frac{I_k(x)}{I_{max}(x)} = \frac{\mu_k (\hat{L}_k(x) \cdot N(x))}{\max_i (\mu_i (\hat{L}_i(x) \cdot N(x)))}$$

It is clear that, for diffuse objects with nonzero albedo $\rho(x)$, $R_k(x)$ is independent of the albedo $\rho(x)$ and only a function of the local geometry. Further, if the light source-camera baseline $|P_k|$ is small compared to the distance to the point, i.e., $|X| \gg |P_k|$, then this ratio is approximately $\mu_k / \max_i (\mu_i)$, which is a constant for a set of omni-directional light sources in the imaging setup.

The ratio values in $(R_k = I_k/I_{max})$ are close to 1.0 in areas lit by light source k and close to zero in shadowed regions. (In general, the values are not zero due to interreflections). The intensity profile along the epipolar ray in the ratio image shows a sharp negative transition at the depth edge as we traverse from non-shadowed foreground to shadowed background, and a sharp positive transition as we traverse from shadowed to non-shadowed region on

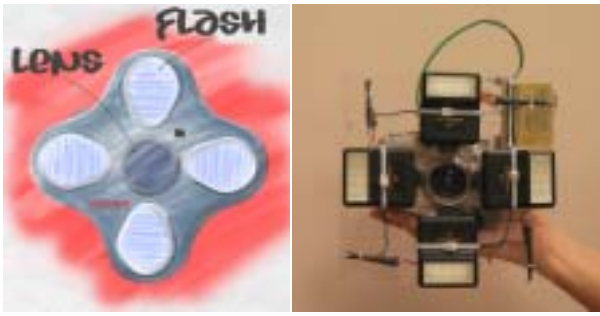


Figure 5: A stylized imaging camera to capture images under four different flash conditions and our prototype.

background (Figure 4). This reduces the depth edge detection problem to an intensity step edge detection problem. A 1D edge detector along the epipolar ray detects both positive and negative transitions, and we mark the negative transitions as depth edges. As mentioned earlier, since we are detecting a transition and not a continuous value, noise and interreflections only affect the accuracy of the position but not the detection of presence of the depth edge.

In summary, there are essentially three steps: (a) create a ratio image where the values in shadowed regions are close to zero; (b) carry out intensity edge detection on each ratio image along epipolar rays marking negative step edges as depth edges (c) combine the edge maps from all n images to obtain the final depth edge map.

Self-contained Prototype An ideal setup should satisfy the constraint that each depth pixel be imaged in both conditions, the negative side of the edge is shadowed at least in one image and not shadowed in at least one other image. We propose using the following configuration of light sources: four flashes at left, right, top and bottom positions (Figure 5).

This setup makes the epipolar ray traversal efficient. If the light source is in the plane parallel to the image plane that contains the center of projection, the light epipole is at infinity and the corresponding epipolar rays are parallel in the image plane. In addition, we place the epipoles such that the epipolar rays are aligned with the camera pixel grid. For the left-right pair, the ray traversal is along horizontal scan lines and for the top-bottom pair, the traversal is along vertical direction.

2.2 Material Edges

In addition to depth edges, we also need to consider illumination and material edges in the image. Illumination edges are boundaries between lit and shadowed regions due to ambient light source(s), rather than the flashes attached to our camera. Since the individual images I_k , are free of ambient illumination, they are free of ambient illumination edges. In general, since material edges are independent of illumination direction, they can be easily classified by a process of elimination. Material edges are intensity edges of I_{max} minus the depth edges.

This edge classification scheme works well and involves a minimal number of parameters for tuning. The only parameters we need are those for intensity edge detection of ratio images and I_{max} image, to detect depth and material edges, respectively.

2.3 Issues

The technique we presented to detect depth edges is surprisingly robust and reliable. We discuss the few conditions in which the basic algorithm fails: a false negative when a negative transition at a

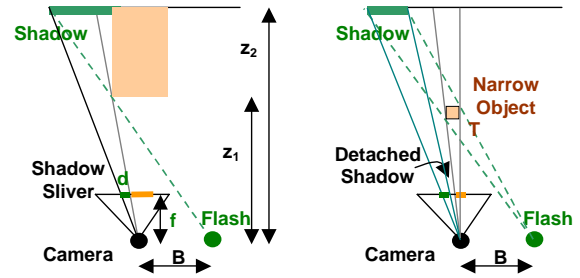


Figure 6: (a) Relationship between baseline and width of shadow (b) Condition where shadow detaches

depth edge cannot be detected in the ratio image R_k or a false positive when other conditions create spurious transitions in R_k . The depth edges can be **missed** due to detached shadows, lack of background, low albedo of background, holes and valleys, or if depth edges lie in shadowed region. The low albedo of background makes it difficult to detect increase in radiance due to a flash, but this problem can be reduced with a higher intensity flash. The problems due to holes/valleys or shadowed depth edges, where the visible background is shadowed for a majority of the flashes, are rare and further reduced when the flash baseline is small. Below, we only discuss the problem due to detached shadows and lack of background. Some pixels may be **misclassified** as depth edge pixels due to specularities or near silhouettes of curved surfaces. We discuss both these issues. We have studied these problems in detail and the solutions will be provided in a technical report. Here we describe the main ideas.

Curved surfaces The silhouettes on curved surfaces vary smoothly with change in viewpoint and the ratio $R_k(x)$ is very low near depth edges when the 3D contours corresponding to silhouettes with respect to neighboring flash positions are sufficiently different. This is because the dot product $(\hat{L}_k(x) \cdot N(x)) \approx 0$ and the dot product for light sources on the 'opposite' side will be larger $(\hat{L}_i(x) \cdot N(x)) > (\hat{L}_k(x) \cdot N(x))$. Thus $R_k(x)$ decreases rapidly even though the pixel is not in a shadowed region. However, as seen in examples shown here, this is not a major issue and simply results in a lower slope at the negative transition in R_k . Unlike the problems below, it does not lead to a reversal of intensity gradient along the epipolar ray.

Tradeoff in choosing the baseline A larger baseline distance between the camera and the flash is better to cast a wider detectable shadow in the image, but a smaller baseline is needed to avoid separation of shadow from the associated depth edge.

The width of the abutting shadow in the image is $d = fB(z_2 - z_1)/(z_1z_2)$, where f is the focal length, B is baseline in

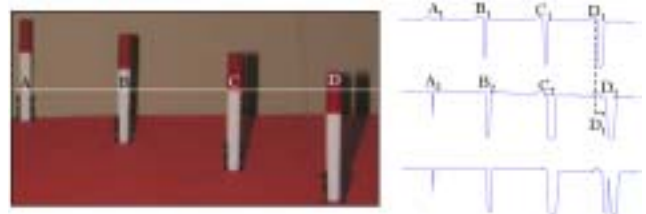


Figure 7: (Left) Minimum composite of image with flash F_S and F_L . (Right) Plot of intensity along a scanline due to F_S , F_L and $\min(I_S, I_L)$.

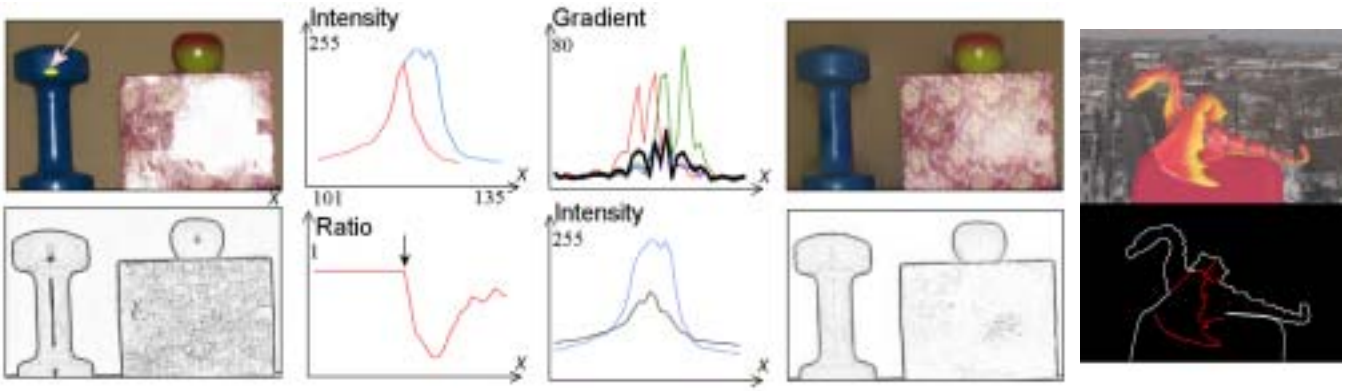


Figure 8: Specularities and lack of background. First column: I_{max} and corresponding result showing artifacts. Second column: For the yellow line marked on dumbbell ($x=101:135$); Top plot, I_{left} (red) with I_{max} (light blue). Bottom plot, ratio R_{left} . Note the spurious negative transition in R_{left} , at the arrow, which gets falsely identified as a depth edge. Third column: Top plot, gradient of I_{left} (red), I_{right} (green), I_{top} (blue) and Median of these gradients (black). Bottom plot, reconstructed intrinsic image (black) compared with I_{max} (light blue). Fourth column: Top, intrinsic image. Bottom, resulting depth edge map. Fifth column: Top, Scene without a background to cast shadow. Bottom, Edges of I_0/I_{max} , in white plus detected depth edges in red.

mm, and z_1, z_2 are depths, in mm, to the shadowing and shadowed edge. (See Figure 6)

Shadow detachment occurs when the width, T , of the object is smaller than $(z_2 - z_1)B/z_2$. So a smaller baseline, B , will allow narrower objects (smaller T) without shadow separation. Fortunately, with rapid miniaturization and sophistication of digital cameras, we can choose small baseline while increasing the pixel resolution (proportional to f), so that the product fB remains constant, allowing depth detection of narrow objects.

When camera resolutions are limited, we can exploit a **hierarchical baseline** method to overcome this tradeoff. We can detect small depth discontinuities (with larger baselines) without creating shadow separation at narrow objects (using narrow baselines). In practice, we found two different baselines were sufficient. We, however, now have to deal with spurious edges due to shadow separation in the image with larger baseline flash F_L . The image with smaller baseline flash, F_S , may miss small depth discontinuities. How can we combine the information in those two images? There are essentially four cases we need to consider at depth edges (Figure 7) (a) F_S creates an undetectable narrow shadow, F_L creates a detectable shadow (b) F_S creates a detectable small width shadow and F_L creates a larger width shadow. (c) F_S creates detectable shadow but F_L creates a detached shadow that overlaps with F_S shadow and (iv) same as (d) but the shadows of F_S and F_L do not overlap.

Our strategy is based on simply taking the minimum composite of the two images. In the first three cases, this conveniently increases the effective width of the abutting shadow without creating any artifacts, and hence can be treated using the basic algorithm without modifications. For the fourth case, a non-shadow region separates the two shadows in the min composite, so that the shadow in F_L appears spurious.

Our solution is as follows. We compute the depth edges using F_S and F_L (Figure 7). We then traverse the epipolar ray. If the depth edge appears in F_S (at D_1) but not in F_L we traverse the epipolar ray in F_L until the next detected depth edge. If this depth edge in F_L , there is no corresponding depth edge in F_S , we mark this edge as a spurious edge.

The solution using min-composite, however, will fail to detect minute depth discontinuities where even F_L does not create a detectable shadow. It will also fail for very thin objects where even F_S creates a detached shadow.

Specularities Specular highlights that appear at a pixel in one image but not others can create spurious transitions in the ratio im-

ages as seen in Figure 8. Although methods exist to detect specularities in a single image [Tan et al. 2003], detecting them reliably in textured regions is difficult.

Our method is based on the observation that specular spots shift according to the shifting of light sources that created them. We need to consider three cases of how specular spots in different light positions appear in each image: (i) shiny spots remain distinct (e.g., on highly specular surface with a medium curvature) (ii) some spots overlap and (iii) spots overlap completely (e.g., on a somewhat specular, fronto-parallel planar surface). Case (iii) does not cause spurious gradients in ratio images.

We note that although specularities overlap in the input images, the boundaries (intensity edges) around specularities in general do not overlap. The main idea is to exploit the gradient variation in the n images at a given pixel (x,y) . If (x,y) is in specular region, in cases (i) and (ii), the gradient due to specular boundary will be high in only one or a minority of the n images under different lighting. The **median of the n gradients** at that pixel will remove this outlier(s). Our method is motivated by the intrinsic image approach by [Weiss 2001], where the author removes shadows in outdoor scenes by noting that shadow boundaries are not static. We reconstruct the image by using median of gradients of input images as follows.

- Compute intensity gradient, $G_k(x,y) = \nabla I_k(x,y)$
- Find median of gradients, $G(x,y) = \text{median}_k(G_k(x,y))$
- Reconstruct image I' which minimizes $|\nabla I' - G|$

Image reconstruction from gradients fields, an approximate invertibility problem, is still a very active research area. In R^2 , a modified gradient vector field G may not be integrable. We use one of the direct methods recently proposed [Elder 1999] [Fattal et al. 2002]. The least square estimate of the original intensity function, I' , so that $G \approx \nabla I'$, can be obtained by solving the Poisson differential equation $\nabla^2 I' = \text{div } G$, involving a Laplace and a divergence operator. We use the standard full multigrid method [Press et al. 1992] to solve the Laplace equation. We pad the images to square images of size the nearest power of two before applying the integration, and then crop the result image back to the original size [Raskar et al. 2004]. We use a similar gradient domain technique to simplify several rendering tasks as described later.

The resultant intrinsic image intensity, $I'(x,y)$ is used as the denominator for computing the ratio image, instead of the max composite, $I_{max}(x,y)$. In specular regions, the ratio $I_k(x,y)/I'(x,y)$ now is larger than 1.0. This is clamped to 1.0 so that the negative transitions in the ratio image do not lie in specular parts.



Figure 9: (a) A edge rendering with over-under style. (b) Rendering edges with width influenced by orientation. (c) and (d) Normal Interpolation fortoon rendering exploiting over-under mattes.

Lack of Background Thus far we assumed that depth edges casting shadows on a background are within a finite distance. What if the background is significantly far away or not present? This turns out to be a simple situation to solve because in these cases only the outermost depth edge, the edge shared by foreground and distant background, is missed in our method. This can be easily detected with a foreground-background estimation technique. In I_{max} image the foreground pixels are lit by at least one of the flashes but in the ambient image, I_0 , neither the foreground nor the background is lit by any flash. Hence, the ratio of I_0/I_{max} is near 1 in background and close to zero in interior of the foreground. Figure 8 shows intensity edges of this ratio image combined with internal depth edges.

3 Image Synthesis

Contour-based comprehensible depiction is well explored for 3D input models [DeCarlo et al. 2003] but not for photographs. In the absence of a full 3D representation of the scene, we exploit the following 2D cues to develop novel rendering algorithms.

- (a) The sign of the depth edge,
- (b) Relative depth difference based on shadow width,
- (c) Color near the signed edges, and
- (d) Normal of a smooth surface at the occluding contour

We aim to automate tasks for stylized rendering where meticulous manual operation was originally required, such as image editing or rotoscoping [Waking Life 2001].

3.1 Rendering Edges

We create a vectorized polyline representation of the depth edges by linking the depth edge pixels into a contour. The polyline is smoothed and allows us to stylize the width and color of the contour maintaining spatial coherency. While traversing the marked depth edge pixels to create a contour, at T-junctions, unlike traditional methods that choose the next edge pixel based on orientation similarity, we use the information from the shadows to resolve the connected component. Two edge pixel are connected only if they are connected in the intensity edges of all the n ratio images.

Signed edges At the negative transition along the epipolar ray in the ratio image, R_k , the side of edge with higher intensity is the foreground and lower intensity (corresponding to shadowed region) is background. This qualitative depth relationship can be used to clearly indicate foreground-background separation at each edge. We emulate the over-under style used by artists in mattes. The foreground side is white while the background side is black. Both are rendered by displacing depth contour along the normal (Figure 9(a)).

Light direction We use a commonly known method to convey light direction by modifying the width of edges depending on the

edge orientation. Since the edge orientation in 3D is approximately the same as the orientation of its projection in image plane, the thickness is simply proportional to the dot product of the image space normal with a desired light direction (Figure 9(b)).

Color variation We can indicate color of original object by rendering the edges in color. From signed edges, we pick up a foreground color along the normal at a fixed pixel distance, without crossing another depth or intensity edge. The foreground colored edges can also be superimposed onto a segmented source image as seen in Figure 10(c).

3.2 Color Assignment

Since there is no 3D model of the scene, rendering non-edge pixels requires different ways of processing captured 2D images.

Normal interpolation For smooth objects, the depth edge corresponds to the occluding contour where the surface normal is perpendicular to the viewing direction. Hence the normals at depth edges lie in the plane of the image and we can predict normals at other pixels. We solve this sparse interpolation problem by solving a 2D Poisson differential equation. Our method is inspired by the Lumo [Johnston 2002] where the over-under mattes are manually created. In our case, signed depth edges allow normal interpolation while maintaining normal discontinuity at depth edges.

Image attenuation We accentuate the contrast at shape boundaries using an image attenuation maps (Figure 10(a)) as follows. Depth edges are in white on a black background. We convolve with a filter that is the gradient of an edge enhancement filter. Our filter is a Gaussian minus an impulse function. When we perform a 2D integration on the convolved image, we get a sharp transition at the depth edge.

Depicting Change Some static illustrations demonstrate action e.g., changing oil in a car, by making moving parts in the foreground brighter. Foreground detection via intensity-based schemes, however, is difficult when the colors are similar and texture is lacking, e.g., detecting hand gesture in front of other skin colored parts (Figure 11). We take two separate sets of multi-flash shots, without and with the hand in front of the face to capture the reference and changed scene. We note that any change in a scene is bounded by new depth edges introduced. Without explicitly detecting foreground, we highlight interiors of regions that contribute to new depth edges.

We create a gradient field where pixels marked as depth edges in changed scene but not in reference, are assigned a unit magnitude gradient. The orientation matches the image space normal to the depth edge. The gradient at other pixels is zero. The reconstructed image from 2D integration is a pseudo-depth map – least squared error solution via solving Poisson equation. We threshold this map at 1.0 to get the foreground mask which is brightened. Note, the shadow width along the epipolar ray is proportional to the ratio of depth values on two sides of the edge. Hence instead of a unit magnitude gradient, we could assign a value proportional



Figure 10: Color assignment. (a) Attenuation Map (b) Attenuated Image (c) Colored edges on de-emphasized texture

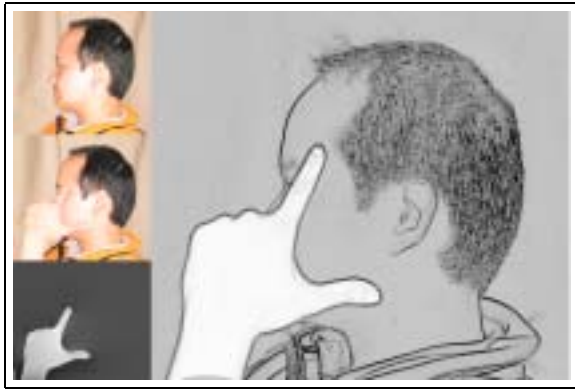


Figure 11: Change Detection. (Left column) Reference image, changed image, and pseudo depth map of new depth edges (Right) Modified depth edge confidence map.

to the logarithm of the shadow width along the epipolar ray to get a higher quality pseudo-depth map. Unfortunately, we found that the positive transition along the ray is not strong due to the use of a non-point light source and interreflections. In principle, estimated shadow widths could be used for say, tunable abstraction to eliminate edges with small depth difference.

3.3 Abstraction

One way to reduce visual clutter in an image and emphasize object shape is to simplify details not associated with the shape boundaries (depth edges) of the scene, such as textures and illumination variations [Gooch and Gooch 2001]. Our goal is to create large flat colored regions separated by strokes denoting important shape boundaries. Traditional NPR approaches based on image segmentation achieve this by assigning a fixed color to each segment [DeCarlo and Santella 2002]. However, image segmentation may miss a depth edge leading to merger of foreground and background near this edge into a single colored object. Although image segmentation can be guided by the computed depth edges, the segmentation scheme places hard constraint on closed contours and does not support small gaps in contours. We propose a method that is conceptually simple and easy to implement.

Our method reconstructs image from gradients without those at texture pixels. No decision need to be made about what intensity values to use to fill in holes, and no feathering and blurring need be done, as is required with conventional pixel-based systems. We use

a mask image, γ , to attenuate the gradients away from depth edges. The mask image is computed as follows.

$$\begin{aligned} \gamma(x,y) &= a \text{ if } (x,y) \text{ is a texture edge pixel} \\ &= a \cdot d(x,y) \text{ if } (x,y) \text{ is a featureless pixel} \\ &= 1.0 \text{ if } (x,y) \text{ is a depth edge pixel} \end{aligned}$$

The factor $d(x,y)$ is the ratio of the distance field of texture pixels by the distance field of depth edge pixels. The distance field value at a pixel is the Euclidean distance to the nearest (texture or depth) edge pixel. As shown in Figure 12, the parameter a controls the degree of abstraction, and textures are suppressed for $a = 0$. The procedure is as follows.

- Create a mask image $\gamma(x,y)$
- Compute intensity gradient $\nabla I(x,y)$
- Modify masked gradients $G(x,y) = \nabla I(x,y)\gamma(x,y)$
- Reconstruct image I' to minimize $|\nabla I' - G|$
- Normalize $I'(x,y)$ colors to closely match $I(x,y)$

The image reconstruction follows the solution of a Poisson equation via a multi-grid approach as in the specularly attenuation technique in Section 2.



Figure 12: Tunable abstraction for texture de-emphasis. Depth edge followed by abstraction with $a = 1$, $a = 0.5$ and $a = 0$.

4 Dynamic Scenes

Our method for capturing geometric features thus far requires taking multiple pictures of the same static scene. We examine the **lack of simultaneity** of capture for scenes with moving objects or a moving camera. Again, a large body of work exists for estimating motion in image sequences, and a sensible approach is to use the results from the static algorithm and apply motion compensation techniques to correct the artifacts introduced. Finding optical flow and motion boundaries, however, is a challenging problem especially in textureless regions [Papademetris and Belhumeur 1996; Birchfield 1999]. Fortunately, by exploiting properties of our unique imaging setup, in most cases, movement of depth edges in dynamic scenes can still be detected by observing the corresponding movement in shadowed regions. As in the static case, we bypass

the hard problem of finding the rich per-pixel motion representation and focus directly on finding the discontinuities i.e., depth edges in motion. The setup is similar to the static case with n flashes around the camera, but triggered in a rapid cyclic sequence, one flash per frame. We find depth edges in a given frame and connect edges found in adjacent frames into a complete depth edge map.

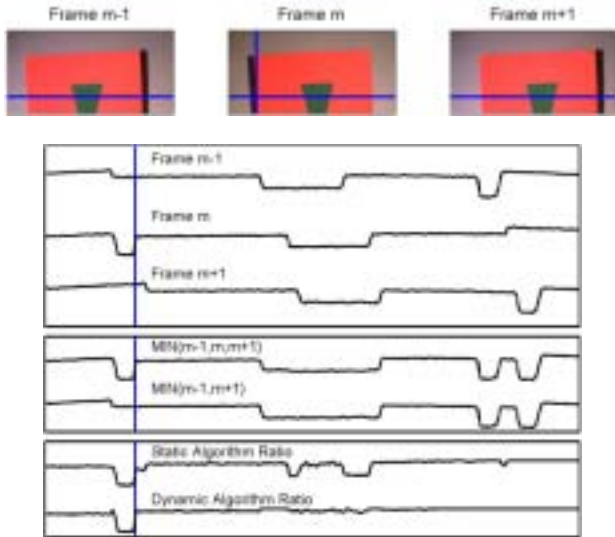


Figure 13: Depth edge detection for dynamic scenes. (Top) Three frames from multi-flash sequence of a toy example showing a red square with a green triangle texture moving from left to right. We are interested in detecting the depth edge in frame m . A single scan line shown in blue is used for the plots. (Middle) The three scan lines plots. The position of the correct depth edge position is indicated with a vertical blue line. (Bottom) Plot of minimum composite and ratio images computed using the static and dynamic algorithms. The motion induced unwanted edges in the static ratio image but not in the dynamic ratio image. The correct depth edge can then be detected from the ratio image using the same traversal procedure as before.

4.1 Depth Edges in Motion

To simplify the discussion, consider using just the left and right flashes to find vertical depth edges. Images from three frames, I_{m-1} , I_m and I_{m+1} , from a toy example are shown in Figure 13. In the sequence, a red square with a green triangle texture is shown moving from left to right, and the three frames are captured under left, right, and left flashes, as can be easily inferred from the cast shadows.

In presence of scene motion, it is difficult to reliably find shadow regions since the base image to compare with, e.g., the max composite, I_{max} , exhibits misaligned features. A high speed camera can reduce the amount of motion between frames but the lack of simultaneity cannot be assumed.

We make two simplifying assumptions (a) motion in image space is monotonic during the image capture from the start of frame $m-1$ to the end of frame $m+1$ and (b) the motion is also small enough that the depth and texture edges in the frames do not cross, i.e., the motion is restricted to the spacing between adjacent edges on the scan line.

Due to the left-right switch in illumination, a shadow near a depth edge disappears in alternate frame images, I_{m-1} and I_{m+1} , while a moving texture edge appears in all three frames. Monotonicity of motion without crossing over edges means

$\min(I_{m-1}, I_{m+1})$ or $\max(I_{m-1}, I_{m+1})$ will both have a flat region around the depth edge in frame m . Similarly, images $\min(I_{m-1}, I_m, I_{m+1})$ and $\max(I_{m-1}, I_m, I_{m+1})$ both are bound to have a flat region around texture edge in frame m . Since the cast shadow region at the depth edge in frame m is darker than the foreground and background objects in the scene, the shadow is preserved in $\min(I_{m-1}, I_m, I_{m+1})$ but not in $\max(I_{m-1}, I_m, I_{m+1})$. This leads to the following algorithm:

- Compute shadow preserving $I_t = \min(I_{m-1}, I_m, I_{m+1})$
- Compute shadow free $I_d = \max(I_{m-1}, I_m, I_{m+1})$
- Compute ratio image, R_m , where $R_m = I_t/I_d$
- Traverse along epipolar ray from e_m and mark negative transition

This ratio image is free of unwanted transitions and the same epipolar ray traversal method can be applied to localize the depth edges.

Figure 13 shows the algorithm in action. We tested the algorithm with synthetic sequences to investigate the set of conditions under which the algorithm is able to correctly localize the depth edges and also experimented with this algorithm in real dynamic scenes. An example frame from a dynamic sequence is shown in Figure 14. A full stylized example with human subjects can be seen in the accompanying video. While we are very encouraged by the simplicity of the algorithm as well as the results we were able to achieve with it, the simplifying assumptions made about the monotonicity and magnitude of motion are still fairly restrictive. For thin objects or objects with high frequency texture, large motions between successive frames creates spurious edges. We plan to continue our investigation in this area and designing algorithms that require fewer assumptions and work under a wider range of conditions.



Figure 14: (Left) A frame from a video sequence, shadows due to left flash. (Right) Detected depth edges merged from neighboring frames.

4.2 Edges and Colors

The depth edges in a given frame, m , are incomplete since they span only limited orientations. In a dynamic scene a union of depth edges from all n successive frames may not line up creating discontinuous contours. We match signed depth edges corresponding to the same flash i.e., m and $m+n$ and interpolate the displacement for intermediate frames. To assign colors, we take the maximum of three successive frames. Our video results can also be considered as tools for digital artists who traditionally use rotoscoping for finding shape boundaries in each frame.

5 Implementation

Our basic prototype makes use of a 4 MegaPixel Canon Power-shot G3 digital camera. The dynamic response in the images is linearized. The four booster (slaved Quantarray MS-1) 4ms duration flashes are triggered by optically coupled LEDs turned on sequentially by a PIC microcontroller, which in turn is interrupted by the

hot-shoe of the camera. Our video camera is a PointGrey Dragon-Fly camera at 1024x768 pixel resolution, 15 fps which drives the attached 5W LumiLeds LED flashes in sequence. We used a *Lumina Wolf* endoscope with 480x480 resolution camera.

It takes 2 seconds to capture each image. Our basic algorithm to detect depth edges executes in 5 seconds in C++ on a Pentium4 3GHz PC. The rendering step for 2D Poisson takes about 3 minutes.

6 Results

We show a variety of examples of real scenes, from millimeter scale objects to room sized environments.



Figure 15: Room sized scene: Right flash image and depth edge map.

Objects and room sized scenes We examine imaging a mechanical (car engine, Figure 1(b)), organic (plant, Figure 1(d)) and anatomical (bone, Figure 9) object. For organic objects, such as flower plant, the geometric shape is complex with specular high-lights, probably challenging for many shape-from-x algorithms. Note the individual stems and leafs that are clear in the new synthesis. The white **bone** with complex geometry, is enhanced with different shape contour styles. In all these scenes, intensity edge detection and color segmentation produce poor results because the objects are almost uniformly colored. The method can be easily used with room-sized scenes (Figure 15).

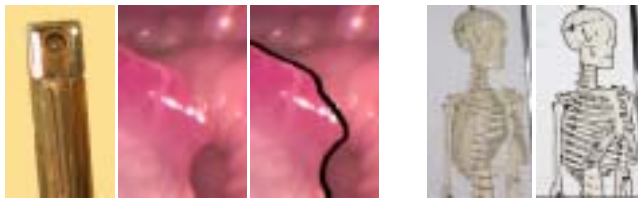


Figure 16: (Left) Enhanced endoscope, with only left lights turned on; input image and depth edge superimposed image. (Right) Skeleton and depth edge superimposed image.

Milli-scale Scene Medical visualization can also benefit from multi-flash imaging. We manipulated the two light sources available near the tip of an endoscopic camera. The baseline is 1mm for 5mm wide endoscope (Figure 16.left). From our discussions with medical doctors and researchers who with such images, extension to video appears to be a promising aid in examination [Tan et al. 2004]. A similar technique can also be used in boroscopes that are used to check for gaps and cracks inside inaccessible mechanical parts - engines or pipes.

Comparison with other strategies We compared our edge rendering technique for comprehension with intensity edge detection using Canny operator, and segmentation. We also compared with active illumination stereo 3D scanning methods, using a state of the art 3Q scanner. Edges captured via **intensity edge detection** are sometimes superimposed on scenes to improve comprehension. While this works in high contrast imagery, sharp changes in image



Figure 17: (Left) Intensity edge detection (Canny) for engine of Figure 1(a). (Right Top) Depth map from 3Q scanner, notice the jagged depth edges on the neck. (Right Bottom) Depth edge confidence map using our technique.

values do not necessarily imply object boundaries, and vice versa [Forsyth and Ponce 2002]. The Canny edge detection or segmentation based NPR approaches unfortunately also fail in low-contrast areas e.g., in the plant, bone or engine (Figure 17.left) example. The 3D scanner output is extremely high quality in the interior of objects as well as near the depth edges. But due to partial occlusions, the depth edges are noisy (Figure 17).

7 Discussion

Feature capture For comprehensible imagery, **other shape cues** such as high curvature regions (ridges, valleys and creases) and self-shadowing boundaries from external point light sources are also useful, and are not captured in our system. Our method is highly dependent on being able to detect the scene radiance contributed by the flash, so bright outdoors or distant scenes are a problem. Given the dependence on shadows of opaque objects, our method cannot handle transparent, translucent, luminous, and mirror like objects.

Many **hardware improvements** are possible. Note that the depth edge extraction scheme could be used for spectrums other than visible light that create 'shadows', e.g., in infrared, sonar, X-rays and radars imaging. Specifically, we envision the video-rate camera to be fitted with infrared light sources invisible to humans so the resulting flashes are not distracting. In fact, one can use a frequency division multiplexing scheme to create a **single shot** multi-flash photography. The flashes simultaneously emit four different colors (wavelength) and the Bayer mosaic like pattern of filters on the camera imager decodes the four separate wavelengths.

Applications of depth edges Detecting depth discontinuity is fundamental to image understanding and can be used in many applications [Birchfield 1999]. Although current methods rely primarily on outermost silhouettes of objects, we believe a complete depth edge map can benefit problems in visual hull, segmentation, layer resolving and aspect graphs. Aerial imaging techniques [Lin and Nevatia 1998] can improve building detection by looking at *multiple* time-lapsed images of cast shadows from known sun directions before and after local noon. In addition, effects such as depth of field effect during post-processing, synthetic aperture using camera array and screen matting for virtual sets (with arbitrary background) require high quality *signed* depth edges.

Edge-based or area-based stereo correspondence can be improved by matching signed depth edges, constraining dynamic pro-

gramming to segments within depth edges and modifying correlation filters to deal with partial occlusions [Scharstein and Szeliski 2002]. Edge classification can provide confidence map to assist color and texture segmentation in low-contrast images. Shape contours can also improve object or gesture recognition [Feris et al. 2004].

8 Conclusion

We have presented a simple yet effective method to convey shape boundaries by rendering new images and videos of real world scenes. We exploit the epipolar relationship between light sources and cast shadows to extract geometric features from multiple images of a scene. By making use of image space discontinuity rather than relying on 3D scene reconstruction, our method can robustly capture the underlying primitives for rendering in different styles.

We have presented basic prototypes, related feature capturing and rendering algorithms, and demonstrated applications in technical illustration and video processing. Finally, since a depth edge is such a basic primitive, we have suggested ways in which this information can be used in applications beyond NPR.

Minor modification to camera hardware enables this method to be implemented in a self-contained device no larger than existing digital cameras. We have proposed one possible approach to leveraging the increasing sophistication of digital cameras to easily produce useful and interesting stylized images.

Acknowledgements We thank the anonymous reviewers for useful comments and guidance. We thank Adrian Ilie, Hongcheng Wang, Rebecca Xiong, Paul Beardsley, Darren Leigh, Paul Dietz, Bill Yerazunis and Joe Marks for stimulating discussions, James Kobler (MEEI), Takashi Kan and Keiichi Shiotani for providing motivating applications, Narendra Ahuja and Beckman Institute Computer Vision and Robotics Lab for suggestions and support, and many members of MERL for help in reviewing the paper.

References

AKERS, D., LOSASSO, F., KLINGNER, J., AGRAWALA, M., RICK, J., AND HANRAHAN, P. 2003. Conveying Shape and Features with Image-Based Relighting. In *IEEE Visualization*.

AVENUE AMY, 2002. Curious Pictures.

BIRCHFIELD, S. 1999. *Depth and Motion Discontinuities*. PhD thesis, Stanford University.

CHUANG, Y.-Y., GOLDMAN, D. B., CURLESS, B., SALESIN, D. H., AND SZELISKI, R. 2003. Shadow matting and compositing. *ACM Trans. Graph.* 22, 3, 494–500.

COHEN, M. F., COLBURN, A., AND DRUCKER, S. 2003. Image stacks. Tech. Rep. MSR-TR-2003-40, Microsoft Research.

DAUM, M., AND DUDEK, G. 1998. On 3-D Surface Reconstruction using Shape from Shadows. In *CVPR*, 461–468.

DECARLO, D., AND SANTELLA, A. 2002. Stylization and Abstraction of Photographs. In *Proc. Siggraph 02, ACM Press*.

DECARLO, D., FINKELSTEIN, A., RUSINKIEWICZ, S., AND SANTELLA, A. 2003. Suggestive contours for conveying shape. *ACM Trans. Graph.* 22, 3, 848–855.

DURAND, F. 2002. An Invitation to Discuss Computer Depiction. In *Proceedings of NPAR 2002*.

ELDER, J. 1999. Are Edges Incomplete?. *International Journal of Computer Vision* 34, 2/3, 97–122.

FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient Domain High Dynamic Range Compression. In *Proceedings of SIGGRAPH 2002, ACM SIGGRAPH*, 249–256.

FERIS, R., TURK, M., RASKAR, R., TAN, K., AND OHASHI, G. 2004. Exploiting Depth Discontinuities for Vision-based Fingerspelling Recognition. In *IEEE Workshop on Real-time Vision for Human-Computer Interaction (in conjunction with CVPR'04)*.

FORSYTH, AND PONCE. 2002. *Computer Vision, A Modern Approach*.

GEIGER, D., LADENDORF, B., AND YUILLE, A. L. 1992. Occlusions and Binocular Stereo. In *European Conference on Computer Vision*, 425–433.

GOOCH, B., AND GOOCH, A. 2001. *Non-Photorealistic Rendering*. A K Peters, Ltd., Natick.

HERTZMANN, A. 1998. Painterly Rendering with Curved Brush Strokes of Multiple Sizes. In *ACM SIGGRAPH*, 453–460.

HUERTAS, A., AND NEVATIA, R. 1988. Detecting buildings in aerial images. *Computer Vision, Graphics and Image Processing* 41, 2, 131–152.

HUGGINS, P., CHEN, H., BELHUMEUR, P., AND ZUCKER, S. 2001. Finding Folds: On the Appearance and Identification of Occlusion. In *IEEE CVPR*, vol. 2, 718–725.

IRVIN, R., AND MCKEOWN, D. 1989. Methods for exploiting the relationship between buildings and their shadows in aerial imagery. *IEEE Transactions on Systems, Man and Cybernetics* 19, 6, 1564–1575.

JOHNSTON, S. F. 2002. Lumo: Illumination for cel animation. In *Proceedings of NPAR*, ACM Press, 45–52.

KANG, S. B., SZELISKI, R., AND CHAI, J. 2001. Handling occlusions in dense multi-view stereo. In *IEEE CVPR*, vol. 1, 102–110.

KRIEGMAN, D., AND BELHUMEUR, P. 2001. What Shadows Reveal About Object Structure. *Journal of the Optical Society of America*, 1804–1813.

LANGER, M., DUDEK, G., AND ZUCKER, S. 1995. Space Occupancy using Multiple Shadow Images. *International Conference on Intelligent Robots and Systems*, 390–396.

LIN, C., AND NEVATIA, R. 1998. Building detection and description from a single intensity image. *Computer Vision and Image Understanding: CVIU* 72, 2, 101–121.

PAPADEMETRIS, X., AND BELHUMEUR, P. N. 1996. Estimation of motion boundary location and optical flow using dynamic programming. In *Proc. Int. Conf. on Image Processing*.

PRESS, W. H., TEUKOLSKY, S., VETTERLING, W. T., AND FLANNERY, B. P. 1992. *Numerical Recipes in C: The Art of Scientific Computing*. Pearson Education.

RASKAR, R., ILIE, A., AND YU, J. 2004. Image Fusion for Context Enhancement and Video Surrealism. In *Proceedings of NPAR*.

RAVIV, D., PAO, Y., AND LOPARO, K. A. 1989. Reconstruction of Three-dimensional Surfaces from Two-dimensional Binary Images. In *IEEE Transactions on Robotics and Automation*, vol. 5(5), 701–710.

SAITO, T., AND TAKAHASHI, T. 1990. Comprehensible Rendering of 3-D Shapes. In *ACM SIGGRAPH*, 197–206.

SATO, I., SATO, Y., AND IKEUCHI, K. 2001. Stability issues in recovering illumination distribution from brightness in shadows. *IEEE Conf. on CVPR*, 400–407.

SAVARESE, S., RUSHMEIER, H., BERNARDINI, F., AND PERONA, P. 2001. Shadow Carving. In *ICCV*.

SCHARSTEIN, D., AND SZELISKI, R. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *International Journal of Computer Vision*, vol. 47(1), 7–42.

SHIRAI, Y., AND TSUJI, S. 1972. Extraction of the Line Drawing of 3-Dimensional Objects by Sequential Illumination from Several Directions. *Pattern Recognition* 4, 4, 345–351.

STROTHOTTE, T., AND SCHLECHTWEG, S. 2002. *NonPhotorealistic Computer Graphics: Modeling, Rendering and Animation*. Morgan Kaufmann, San Francisco.

TAN, P., LIN, S., QUAN, L., AND SHUM, H.-Y. 2003. Highlight Removal by Illumination-Constrained Inpainting. In *Ninth IEEE International Conference on Computer Vision*.

TAN, K., KOBLER, J., DIETZ, P., FERIS, R., AND RASKAR, R. 2004. Shape-Enhanced Surgical Visualizations and Medical Illustrations with Multi-Flash Imaging. In *MERL TR/38*.

TOYAMA, K., KRUMM, J., BRUMITT, B., AND MEYERS, B. 1999. Wallflower: Principles and Practice of Background Maintenance. In *ICCV*, 255–261.

WAKING LIFE, 2001. Waking Life, the movie.

WEISS, Y. 2001. Deriving intrinsic images from image sequences. In *Proceedings of ICCV*, vol. 2, 68–75.

YANG, D. K.-M. 1996. *Shape from Darkness Under Error*. PhD thesis, Columbia University.

Table-top Computed Lighting for Practical Digital Photography

Ankit Mohan^{1†}, Jack Tumblin¹, Bobby Bodenheimer², Cindy Grimm³, Reynold Bailey³

¹Northwestern University, ²Vanderbilt University, ³Washington University in St. Louis

Abstract

We apply simplified image-based lighting methods to reduce the equipment, cost, time, and specialized skills required for high-quality photographic lighting of desktop-sized static objects such as museum artifacts. We place the object and a computer-steered moving-head spotlight inside a simple foam-core enclosure, and use a camera to quickly record low-resolution photos as the light scans the box interior. Optimization guided by interactive user sketching selects a small set of frames whose weighted sum best matches the target image. The system then repeats the lighting used in each of these frames, and constructs a high resolution result from re-photographed basis images. Unlike previous image-based relighting efforts, our method requires only one light source, yet can achieve high resolution light positioning to avoid multiple sharp shadows. A reduced version uses only a hand-held light, and may be suitable for battery-powered, field photography equipment that fits in a backpack.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture I.3.3 [Computer Graphics]: Picture/Image Generation

1. Introduction

Modern digital cameras have made picture-taking much easier and more interactive. However, lighting a scene for good photography is still difficult, and practical methods to achieve good lighting have scarcely changed at all. We show that sketch-guided optimization and simplified forms of image-based lighting can substantially reduce the cost, equipment, skill, and patience required for small-scale studio-quality lighting.

Good studio lighting is difficult because it is a 4D inverse problem that photographers must solve by making successive approximations guided by years of experience. For non-experts, good studio lighting can be surprisingly frustrating. Most people can specify the lighting they want in screen space (e.g., “get rid of this obscuring highlight, make some shadows to reveal rough texture here, but fill in the shadows there”), but determining what kind of lights to use, where to place them, and how to orient them is never easy.

We are especially interested in camera-assisted lighting for human-scale, desktop-sized static objects. We want lighting that accurately reveals the shape, texture, materials, and most visually meaningful features of the photographed item. In particular, we seek a method to help museum curators as they gather digital photographic archives of their vast collections of items.

Pioneering work in image-based lighting [DHT*00, HCD01, DWT*02, MPDW03] offers promising approaches that can help with the photographic lighting problem. Unfortunately, most require too many precise measurements and adjustments for day-to-day use outside the laboratory. Precision is required to address more ambitious goals such as recovering shape, BRDF, and appearance under arbitrary viewing and lighting conditions. For the much smaller, yet more widespread problem of photographic lighting, we need a method that requires less time, expense, and complexity, yet allows users who are not lighting experts to quickly find the lighting they want.

This paper offers three contributions. We extend existing

[†] ankit@cs.northwestern.edu

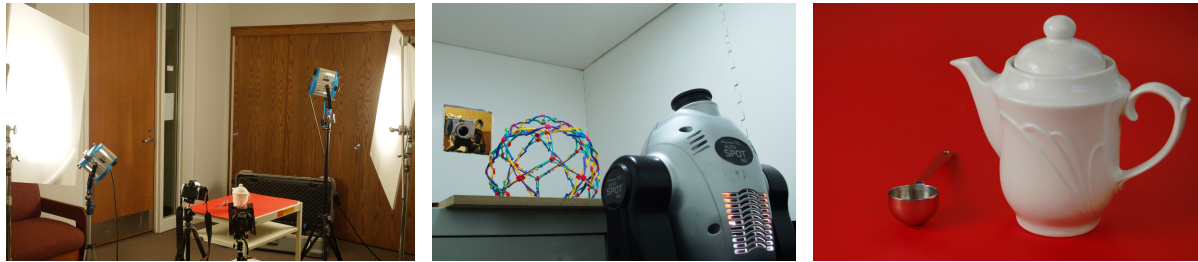


Figure 1: Light placement for obtaining high quality photographs can be extremely tedious and time consuming (left). Our system use a simple setup with a steerable spotlight and an uncalibrated enclosure (center) to obtain results comparable to professional lighting even when used by novice users (right).

image-based lighting ideas to reduce the required equipment to a single light source and single camera; we replace trial-and-error light repositioning with optimization and on-screen painting; and we reduce the need for high dynamic range photography, thus reducing the capture time. The result is a novel and inexpensive system that a novice can use to intuitively describe and obtain the desired lighting for a photograph.

2. Related Work

Lighting has long been recognized as a hard problem in computer graphics and many papers have explored optimization for light placement and other parameters [SDS*93, KPC93, PF95, CSF99, SL01]. Some of these systems used painting interfaces to specify desired lighting in a 3D scene [SDS*93, PF95, PRJ97], and we use a similar approach to make lighting for photography more intuitive. The system by Shacked et al. [SL01] was even able to provide fully automatic lighting by applying image quality metrics. Marschner et al. [MG97] used inverse rendering techniques to estimate and alter the directional distribution of incident light in a photograph. However, all these systems require 3D information unavailable in our photographic application.

Several commercial photographic products have also used lighting enclosures similar to ours, but they achieve very soft lighting with limited user controls. Moreover, they do not help users solve light placement problems. These systems include diffusive tents [Pho], photo-boxes [MK] and translucent back-lit platforms with an array of individually dimmed light sources [Ast].

Image-based methods have also been used to permit arbitrary relighting of well-measured objects. Most methods, including ours, perform relighting using a weighted sum of differently lit basis images, done first by [NSD94]. However, prior efforts used more elaborate and expensive equipment because their goals were different from ours. These include measurement of a 4D slice of the reflectance field of

the human face [DHT*00], museum artifacts measured by a rotating-arm light stage [HCD01], an ingenious but extensive system by Debevec et al. [DWT*02] for real-time video playback and measurement of light fields, a dome of electronic flashes for real time image relighting [MGW01], a free form light stage to enable portable gathering of light-field data with some calibration [MDA02], and full 4D incident light measurements by Masselus et al. [MPDW03]. In all of these cases, data-gathering required either customized equipment or collection times much longer than would be practical for photographic lighting.

Three recent systems also offered novel sketch guided relighting from basis images. Akers et al. [ALK*03] used a robotic light-positioning gantry to gather precisely lit images, and like us, provided a painting interface to guide relighting. But unlike us they used spatially varying weights that could produce physically impossible lighting. Digital Photomontage [ADA*04] used sketch guided graph-cut segmentation coupled with gradient domain fusion to seamlessly merge several photographs. They demonstrated merging differently lit photographs to create novel illumination conditions. Though their interaction scheme worked well for a small number of images (~ 10), it may be impractical for the hundreds of images required for complete control over lighting directions. Also, their system does nothing to help the user with light placement, and may produce physically unrealizable results. Anrys and Dutre [AD04] used a Debevec-style light stage with around 40 fixed, low powered light sources and a painting interface to guide lighting. Their optimization only found light intensities, and light placement was still left up to the user. Also, their point light sources could cause multiple shadows and highlights which might be undesirable for archival purposes. The data capture time was high since they captured high-dynamic-range (HDR) photos for every light location.

Unlike previous attempts, our system does not require users to decide on correct or complete light source placement. This is possible because our capture process is significantly dif-

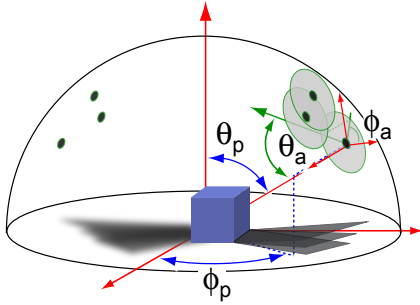


Figure 2: All possible lighting angles parameterized by light position (θ_p, ϕ_p) and direction (θ_a, ϕ_a) . Point light sources (on the left side of the hemisphere) result in multiple hard shadows, while overlapping area (on the right) light sources can be used to simulate a larger light source.

ferent, and better suited for the task of photography. We require less than five minutes to complete the initial image capture and a few more minutes to get the final result. The equipment required is minimal and portable, and our handheld version can be carried in a backpack. Also, HDR capture is reduced to a minimum in our system.

3. Simplifications: HDR and 2D lighting

Our goal is to do what a good photographer does, but with computational help. We want to light a scene for a particular photograph, *not* build a calibrated 4D data set to reconstruct every possible form of illumination. Photographers make consistent choices about which types of lights to use, how to adjust them, and where to place them. We will show how our streamlined image-based method follows these same choices.

Like most previous image-based lighting methods, we apply the observations formalized by Nimeroff [NSD94] that lights and materials interact linearly. If a fixed camera makes an image I_i from a fixed scene lit only by a light L_i , then the same scene lit by many lights scaled by weights w_i will make an image $I_{out} = \sum_i w_i I_i$. Adjusting weights lets us “relight” the image, as if the weights modulate the lights rather than the images. As we collect more images I_i , we can simulate more lighting possibilities.

How many images do we really need to gather? We only need enough images to span the kind of lighting a skilled photographer might explore to get good results in a photo studio. Several common practices in studio lighting can help us.

First, professional photographers choose lamps with broad, nearly uniform beams of light, often with a reflector and lens to help direct more light forward. Second, they adjust light placement angles carefully, but not their distances from the

object. Distance to the light affects foreshortening of shadow shapes, but these effects are subtle and rarely noticed. Third, they adjust lights to control shadow softness versus sharpness. Light sources (or more accurately, the shadows they form) become ‘softer’ by increasing the angular extent as measured from the lit object. Fourth, they seek out lighting arrangements that produce a simple set of shadows and highlights that best reveal the object’s shape, position, and surface qualities. They avoid complex overlapping shadows, lack of shadows due to overly-soft light, and contrast extremes due to large specular highlights or very dark shadows. Simpler shadows usually mean fewer lights, and thus fewer basis images.

Accordingly, we use commercially available light sources instead of custom or special-purpose devices. We place light sources at a moderate distance (typically around 1 meter) from the object. We use small-to-moderate area ‘soft’ light sources instead of the much sharper point-like sources often used in earlier approaches. Overlapped soft shadows blend far less noticeably than sharp shadows from the same light positions (as shown in Figure 2), thus requiring fewer images to avoid multiple shadow artifacts. Also, overlapping area light sources can be combined to produce a larger area light source.

Note that we do not need to know the light positions or their absolute intensities for our images; we select weights w_i and images I_i by their ability to match the lighting target images a user sketches for us. Instead of calibration, we only need consistency in the aiming direction of a single, commercially available steerable light, and consistency in the light response curve of a commercially available digital camera.

We also avoid the use of HDR photographs where possible, as these typically require multiple calibrated exposures and computation to merge them [DM97]. Instead, we rely on the camera’s automatic exposure adjustments to capture what we call *light-aiming images* suitable for interactive lighting design. We photograph high resolution basis images afterwards, for construction of the output image, and only resort to HDR capture methods for a basis image with large over-exposed regions. Under-exposed regions can be ignored, as their contributions are already invisible, and are further reduced as their weights are less than one ($w_i \leq 1$).

Formally, arbitrary external illumination is four-dimensional for a desktop scene: $L(\theta_p, \phi_p, \theta_a, \phi_a) = L(\Theta)$. Suppose that the photographed object receives all its light from a hemisphere of tiny, invisible, inward-pointing video projectors, each at a distance r from the object. Each projector’s position in desktop polar coordinates is (θ_p, ϕ_p) . Each projector’s centermost pixel $P(\theta_a = 0, \phi_a = 0)$ forms a ray that illuminates the center point of our desktop, and in the projector’s polar coordinates the other pixels are $P(\theta_a, \phi_a)$, as shown in Figure 2. All projectors’ light output is the 4-D incident light field, and describes all possible lighting. To simulate all possible lighting, we would need a new image I_i to capture

light from each pixel of each video projector! Instead, we use only broad beams of light ($P(\theta_a, \phi_a) \cong \cos(\theta_a)\cos(\phi_a)$), regular sampling of light placement angles (θ_p, ϕ_p) , and specify ‘softer’ to ‘sharper’ shadows by varying the angular extent (θ_p, ϕ_p) as measured from the lit object. This angular extent should not be confused with the lamp’s beam width (θ_a, ϕ_a) ; in our ‘hemisphere of video projectors’ analogy, beam width sets the image from a projector, but angular extent sets the number of adjacent projectors that emit this same image.

In summary, rather than recreate arbitrary 4D incident light fields, we use weighted sums of basis images that represent the type of lighting used by professional photographers. This method is much more practical and efficient, with little, if any, loss of useful generality.

4. Method

We construct a high quality user-guided picture in three steps. First the system automatically captures low-resolution light-aiming photos for densely sampled lighting angles around the photographed object. These quick photos are used only to guide the lighting design, not to form the final output. Second, the user iteratively paints the desired lighting by simple lighten-darken operations to generate a target image. The system finds weights w_i for each light-aiming photo such that their weighted sum matches the target image in the least-squares sense. Finally the system takes a few selected high resolution basis images by relighting the scene from light source positions that have weights w_i greater than a threshold. A weighted sum of these high resolution images gives the final result. If the result is not satisfactory, the user can sketch on the current result for use as the next iteration’s target image.

4.1. Enclosed Light Source & Aiming Images

Freed from photometric and angular calibration requirements as discussed in Section 3, we are able to build a much simpler and cost-effective controlled light source. We place the object and a gimbal-mounted moving-head spotlight inside an enclosure of almost any convenient size, shape and material. The powerful computer-aimed light pivots to any desired pan and tilt angle with good repeatability ($\leq \pm 0.5^\circ$) to light any desired spot inside our enclosure. The enclosure acts as a reflector, and effectively provides a controllable 2D area light source around the object. The size and shape of the enclosure is almost irrelevant as long as the light is close enough to the object to keep parallax low, and the light is powerful enough for the camera to get a reasonable exposure.

We built a $1 \times 1 \times 1.5m^3$ sized box of white 1/2” foam-core board as our enclosure, and chose an inexpensive moving-head spotlight. The 150-watt *American DJ Auto Spot 150* disco-light, shown in Figure 1 can tilt 270° , pan 540° , and includes 9 color filters, gobos and several other fun features.

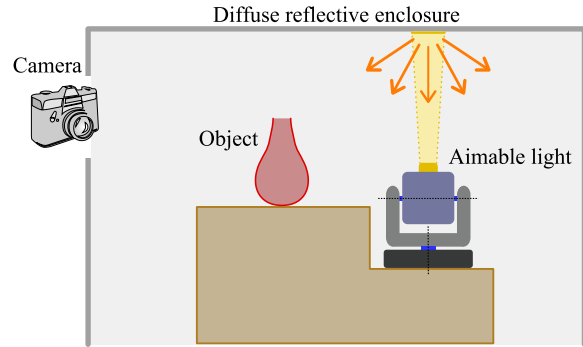


Figure 3: The disco-light setup. The object and disco light are both enclosed in a white foam box, with the camera looking in through a window in the enclosure wall farthest from the light.

Computer control by the DMX512 protocol is easy to program with the SoundLight USB DMX controller. Our foam-core enclosure resembles a hemi-cube around a pair of tables. We place the gimbal light on a small table that lowers its rotation center to the plane of an adjacent taller table holding the photographed object, as shown in Figure 3. Using adjacent but separate tables reduces vibration, permits gimbal angles to approximate hemisphere angles, and separates the object from the swiveling lamp. We place the camera behind a small opening cut in the enclosure wall on the end farthest from the light source.

The system gathers aiming images rapidly and automatically. Through the DMX512 controller we direct the gimbal light to scan the upper hemisphere of light aiming directions in equal-angle increments as we record low-resolution aiming images, either by collecting viewfinder video ($320 \times 240@10\text{Hz}$) or by individual computer-triggered photographs using auto-exposure. We are able to record hundreds of individual aiming images per minute, and can complete all the data gathering in less than five minutes using a Pentium 2GHz computer, and a Canon Powershot G3 camera.

To the best of our knowledge, no other image-based lighting work exploits these movable and controllable lights. Enclosed pivoting lights retain many advantages of the more sophisticated lighting systems, avoid multiple sharp shadows, can offer variable ‘softness’ by spot size adjustment, and are much simpler and cheaper to construct. Of course, they do not easily provide accurate lighting direction calibration or point-light illumination, but these features are not needed for our goals.

After recording, we linearize each captured frame (RGB) by applying the camera’s inverse response curve, recovered by the method of Debevec et al. [DM97], and converted to luminance values. Linear response ensures weighted sums of

whole images are accurate representations of physically realizable lighting. We then down-sample the linearized aiming image dataset to 64×64 for use as the *aiming basis set* for the following optimization step.

4.2. Sketch-Guided Lighting Optimization

After gathering aiming images, users can interactively specify and refine lighting by sketching the desired intensity on a *target intensity image*. This grey-scale image (examples in Figure 5) approximates the final output image the user would like to see. For editing the target image, the user starts off either with a simple grey wash (such as uniform grey, or light grey fading to dark grey across the image, etc.), or the previous iteration’s result. The user then carries out a series of lighten and darken operations in the different regions of the image to approximate the desired results. The process is extremely simple and intuitive, and takes a few of minutes at most.

Given a target image, the optimization finds weights w_i for each aiming image that produces the best match to the target image. We take a constrained least-squares approach, solving for weights w_i for each of the small, luminance-only aiming basis images. Let N be the number of images in the aiming image set, each of size $m \times n$. We formulate the optimization problem as follows:

$$\min_w |Aw - t|^2$$

subject to $0 \leq w_i \leq 1 \quad \forall i \in (1 \dots N)$

where w is the N -dimensional vector of weights, A is an $(m \times n) \times N$ matrix of basis images (that is, each basis image is treated as a vector), t is the $(m \times n)$ vector representing the target image painted by the user, and $|\cdot|$ represents the L^2 norm of the vector. We solve this bound constrained quadratic optimization problem using an active set method [NW99]. The optimization is quite fast and takes around 1-2 minutes on a 2GHz Pentium 4 desktop machine.

The result is a least-squares optimal match to the supplied target image. As the objective function is quadratic, weights for images with weak contributions are rapidly driven to zero. In our experience, the number of significant nonzero weights is consistently small (5 – 15). This greatly reduces the number of images needed for the final lighting solution.

After finding the w_i weights, we apply them to the linearized color aiming images, then re-apply the camera response function to display a preview of the output image. The user then has the option of replacing the target with a grayscale version of this result and can repeat the sketching and optimization cycle until satisfied with the color preview of the output image.



Figure 4: Light source with attached foam-core diffuse reflector used for hand-held data gathering.

4.3. Output Assembly

The user now has the desired visually pleasing, but low-resolution, image that is a weighted sum of a small subset of the linearized aiming images. For high-quality results, we wish to replace each of these aiming images with an image taken at the maximum resolution available from the camera. We re-take *just* those photos that correspond to the aiming images with significant weights w_i , again using auto-exposure on the camera, and record a set of high-resolution photos called *basis images*. Recall that we can exactly replicate the lighting using the gimbaled spotlight; the only things that change are the camera settings.

We capture HDR photographs for images that contain large over-exposed regions as a result of the camera’s autoexposure. As discussed in Section 3, under-exposed regions do not require HDR photos. We then linearize each basis image to remove effects of the camera response curve. As before, we construct a linear output image as a weighted sum of basis images, using the weights determined by the optimization to match the target image. Finally, we re-apply the camera’s response function to the linear output image to get the desired high resolution result.

5. Portable, Hand-held Method

Even a foam-core box and a moving-head spotlight are impractical to carry around everywhere. However, the ‘Free-form light-stage’ [MDA02] showed that it is possible to gather calibrated image sets suitable for 2D relighting with nothing more than four small light-probe-like spheres, a digital camera on a tripod, a hand-held point-light source, possibly battery-powered, and approximately 30 minutes of time to take several hundred digital photographs. Pang et al. [PWH04] also used a similar approach by mounting a

camera on the light source and used camera calibration techniques to estimate lighting directions with reasonable accuracy. While these methods try to meet the ambitious goal of incident light field capture, they would tax anyone's patience to record more than just a few items. We present a faster and simpler variant that serves our purposes better.

In the method of Section 4, we required repeatable light source positioning. However, if we record all of our 'aiming images' at the final output resolution, and if we either ignore over-exposed specular highlights or record high dynamic range images when needed, then *repeatability is not needed*. This allows us to use a hand-held light source instead. As shown in Figure 4, we use a small 250W hand-held light intended for television news cameras, attached to a diffuse reflector (foam core again), and limit the beam width with barn-doors to form a well-defined area light source.

To gather all photos, we hold the light outstretched and "dance" (see video). We sample the hemisphere of lighting directions by a polar-coordinate scan in ϕ -major order as the camera takes sequential photographs. A Nikon D70 camera, takes a steady stream of photos at about 3 frames per second using autoexposure for each frame. The user stands facing the object, and holds the light at arms' length while moving the lamp in an arc that passes directly over the object. The user moves the lamp from one side of the table to the other, scanning by π radians in θ axis with constant ϕ , and the natural alignment of their shoulders helps aim the light's centerline directly at the object. After each pass over the object with the light, the user steps sideways to change the ϕ angle for the next scan, and makes enough of these passes to cover $0 \leq \phi < \pi$ radians. In practice the user can be more careless with the light, as long as the hemisphere of light directions is well-sampled and the images are not over-exposed. After the image capture *dance* is complete, we downsample all images to construct aiming photos, and proceed with the sketch guided lighting design as before.

We find this process is quite simple and pleasing, and in under three minutes we can gather around 150 high-quality aiming/basis photos. An experienced user might not need to scan the whole hemisphere, but can quickly illuminate just from the useful and interesting lighting directions.

6. Results

Images in Figure 5 show results from our sketch guided lighting system. Both the moving-head light and the hand-held methods are equally successful at creating arbitrary cleanly-lit images of desktop-sized objects. The data sets gathered by either method is sufficiently dense to allow easy lighting design. Additionally, our system yields reasonable results even when presented with unrealistic targets or highly reflective objects.

Figure 5(a), demonstrates a user interaction sequence with the system. Starting from a uniform grayscale image as the

target, the user guides the optimization, iteratively improving the target until she gets the desired output. Figure 5(b) shows how simple approximate sketching on the target image can give an interesting sidelighting effect. Figure 5(c) shows how the highlight can bring out the underlying texture in a surface.

Figure 5(d) shows lighting for a highly specular object. Good lighting for such smooth, highly reflective objects is always difficult, as the light source itself is visible in the reflection. Our system produces results similar to the target image without large, objectionable saturated regions. In future systems we may hide the enclosure seams by constructing wide smooth rounded corners resembling a photographer's 'cyc'.

Figure 5(f) shows results from the handheld method of Section 5. The data gathering time was under 3 minutes, and the results are comparable to the moving-head light method. While the handheld method is not practical for photographing a large collection of objects, it can be an invaluable tool for well-lit photography in the field.

7. Discussion and Future Work

The ability to have large area light sources is crucial for photographing highly specular objects. Light source size also affects the sharpness of shadows and highlights. Our system has a unique advantage in that larger area light sources can be simulated by combining pictures illuminated with overlapping light sources. We could extend our optimization to penalize each distinct light source cluster, thus preventing disjoint highlights. The softness of the light can also be controlled by varying the beam width between a point-source and a large area source as it quickly sweeps over the hemisphere of lighting directions. More advanced moving-head spotlights usually provide controllable spot sizes suitable for this purpose.

Even though our system is aimed primarily at non-professional photographers, a few simple additions can make it a flexible tool for a creative expert to experiment with different lighting designs more easily. For example, the user might specify a simple weighting mask to set the importance of different image regions and influence the optimization process. While weighting masks would make the system more flexible, they would complicate the target sketching process. We do not know yet if the results would warrant the increase in complexity. Also, tools to directly tweak the light position and size on a virtual hemisphere around the object might also aid expert users.

There are several possible ways of dealing with the ambient light in the reflective enclosure. Underexposing all images using exposure compensation on the camera, using a larger enclosure or one made of materials with special reflective properties would greatly minimize the ambient component. Finally, it might also be possible to explicitly subtract the ambient term from the basis images.

This paper takes the problem of good lighting for desktop photography and finds a simple and practical solution using image-based relighting techniques. More sophisticated image-based measurements might also be achievable while maintaining the simplicity and elegance of the system. For example, we could estimate the incoming light direction by calibrating the ad-hoc enclosure setup with a light-probe, or by using dimensionality reduction [WMTG05] for the hand-held case. Combined with surface normals, such calibration might suffice for image-based estimates of BRDF.

8. Acknowledgements

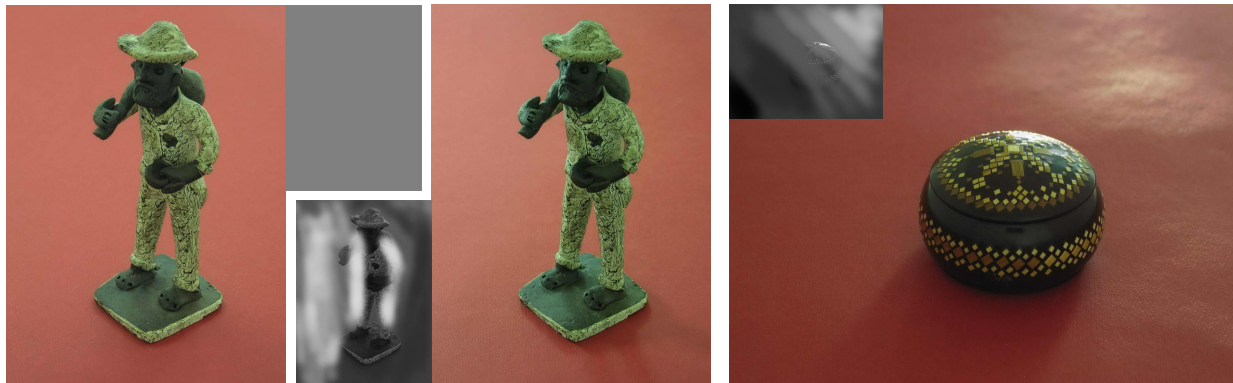
We would like to thank the computer graphics group at Northwestern University for their support and feedback and the anonymous reviewers for their helpful comments and suggestions. Thanks to Jingjing Meng and Xing Hu for help in early stages of the project, Kelli Johnson and Nathan Matsuda for help with the construction of the enclosure, Vincent Masselus and Amy Gooch for help with Figure 2, and Holger Winnemöller for help with the video. This research was funded in part by NSF grants 0238062 and 0237621.

References

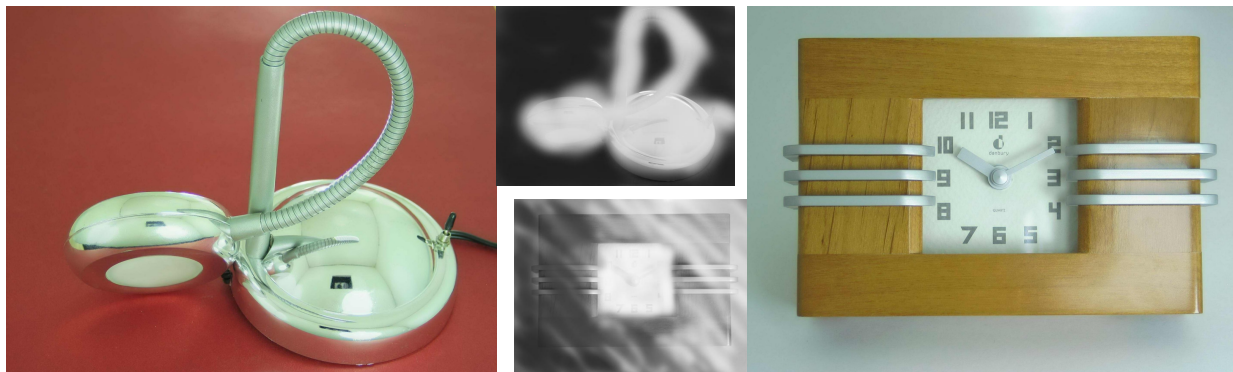
- [AD04] ANRYS F., DUTRÉ P.: Image based lighting design. In *The 4th IASTED International Conference on Visualization, Imaging, and Image Processing* (2004). 2
- [ADA*04] AGARWALA A., DONTCHEVA M., AGRAWALA M., DRUCKER S., COLBURN A., CURLESS B., SALESIN D., COHEN M.: Interactive digital photomontage. In *SIGGRAPH* (2004), vol. 23, pp. 294–302. 2
- [ALK*03] AKERS D., LOSASSO F., KLINGNER J., AGRAWALA M., RICK J., HANRAHAN P.: Conveying shape and features with image-based relighting. In *IEEE Visualization* (2003). 2
- [Ast] ASTRON SYSTEMS: Orbiculight lighting system. <http://www.astronsys.com>. 2
- [CSF99] COSTA A. C., SOUSA A. A., FERREIRA F. N.: Lighting design: a goal based approach using optimisation. In *Proceedings of Eurographics Workshop on Rendering* (1999), vol. 10, pp. 317–328. 2
- [DHT*00] DEBEVEC P., HAWKINS T., TCHOU C., DUIKER H.-P., SAROKIN W., SAGAR M.: Acquiring the reflectance field of a human face. In *SIGGRAPH* (2000), pp. 145–156. 1, 2
- [DM97] DEBEVEC P. E., MALIK J.: Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH* (1997), vol. 31, pp. 369–378. 3, 4
- [DWT*02] DEBEVEC P., WENGER A., TCHOU C., GARDNER A., WAESE J., HAWKINS T.: A lighting reproduction approach to live-action compositing. In *SIGGRAPH* (2002), pp. 547–556. 1, 2
- [HCD01] HAWKINS T., COHEN J., DEBEVEC P.: A photometric approach to digitizing cultural artifacts. In *Proceedings of conference on Virtual reality, archeology, and cultural heritage* (2001), pp. 333–342. 1, 2
- [KPC93] KAWAI J. K., PAINTER J. S., COHEN M. F.: Radioptimization: goal based rendering. In *SIGGRAPH* (1993), pp. 147–154. 2
- [MDA02] MASSELUS V., DUTRÉ P., ANRYS F.: The free-form light stage. In *Proceedings of the 13th Eurographics workshop on Rendering* (2002), pp. 247–256. 2, 5
- [MG97] MARSCHNER S. R., GREENBERG D. P.: Inverse lighting for photography. In *Proceedings of IS&T/SID Fifth Color Imaging Conference* (1997), pp. 262–265. 2
- [MGW01] MALZBENDER T., GELB D., WOLTERS H.: Polynomial texture maps. In *SIGGRAPH* (2001), ACM Press, pp. 519–528. 2
- [MK] MK DIGITAL: Photo light box. <http://www.mkdigitaldirect.com>. 2
- [MPDW03] MASSELUS V., PEERS P., DUTRÉ P., WILLEMS Y. D.: Relighting with 4d incident light fields. In *SIGGRAPH* (2003), vol. 22, pp. 613–620. 1, 2
- [NSD94] NIMEROFF J. S., SIMONCELLI E., DORSEY J.: Efficient re-rendering of naturally illuminated environments. In *Proceedings of the Fifth Eurographics Workshop on Rendering* (1994), pp. 359–373. 2, 3
- [NW99] NOCEDAL J., WRIGHT S. J.: *Numerical Optimization*. Springer-Verlag New York, 1999. 5
- [PF95] POULIN P., FOURNIER A.: Painting surface characteristics. In *Proceedings of Eurographics Workshop on Rendering* (1995), pp. 160–169. 2
- [Pho] PHOTEK: Digital lighthouse. <http://www.photekusa.com>. 2
- [PRJ97] POULIN P., RATIB K., JACQUES M.: Sketching shadows and highlights to position lights. In *Proceedings of Conference on Computer Graphics International* (1997), pp. 56–63. 2
- [PWH04] PANG W.-M., WONG T.-T., HENG P.-A.: Estimating light vectors in real time. *IEEE Computer Graphics and Applications* 24, 3 (2004), 36–43. 5
- [SDS*93] SCHOENEMAN C., DORSEY J., SMITS B., ARVO J., GREENBURG D.: Painting with light. In *SIGGRAPH* (1993), pp. 143–146. 2
- [SL01] SHACKED R., LISCHINSKI D.: Automatic lighting design using a perceptual quality metric. *Computer Graphics Forum* 20, 3 (2001), 215–226. 2
- [WMTG05] WINNEMÖELLER H., MOHAN A., TUMBLIN J., GOOCH B.: Light waving: Estimating light positions from photographs alone. *Computer Graphics Forum* 24, 3 (2005), to appear. 7



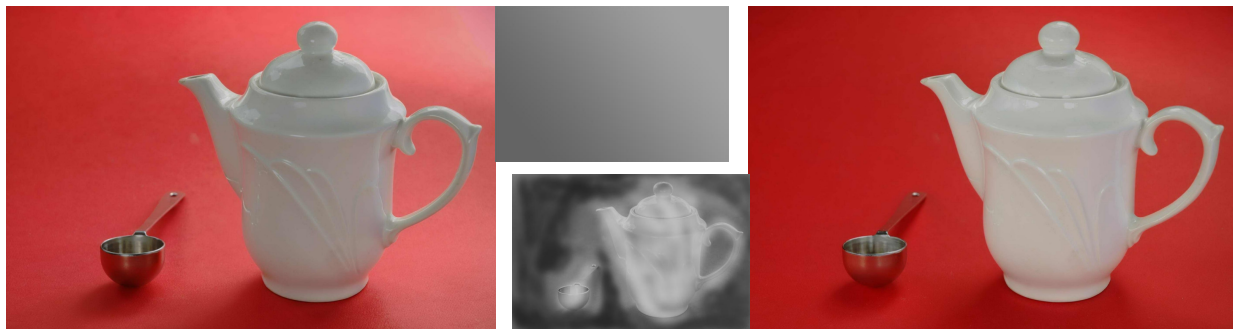
(a) Sequence showing successive sketching/optimization iterations to get the desired lighting. The first result uses a constant grayscale target, while the others use previous results as starting points for the target image.



(b) Strategic placement of highlights in the target result in an interesting side-lit image. (c) Positioning of highlights reveals underlying texture in the surface.



(d) Lighting a highly specular object by forcing the background to be dark. (e) Target results in image suggesting illumination from the right.



(f) Data captured by the handheld method. Image on the left uses a smooth grayscale gradient as the target image.

Figure 5: Sample target images and lit photographs.

Assorted Pixels: Multi-Sampled Imaging with Structural Models

Shree K. Nayar and Srinivasa G. Narasimhan

Department of Computer Science, Columbia University
New York, NY 10027, U.S.A.
{nayar, srinivas}@cs.columbia.edu

Abstract. Multi-sampled imaging is a general framework for using pixels on an image detector to simultaneously sample multiple dimensions of imaging (space, time, spectrum, brightness, polarization, etc.). The mosaic of red, green and blue spectral filters found in most solid-state color cameras is one example of multi-sampled imaging. We briefly describe how multi-sampling can be used to explore other dimensions of imaging. Once such an image is captured, smooth reconstructions along the individual dimensions can be obtained using standard interpolation algorithms. Typically, this results in a substantial reduction of resolution (and hence image quality). One can extract significantly greater resolution in each dimension by noting that the light fields associated with real scenes have enormous redundancies within them, causing different dimensions to be highly correlated. Hence, multi-sampled images can be better interpolated using local structural models that are learned offline from a diverse set of training images. The specific type of structural models we use are based on polynomial functions of measured image intensities. They are very effective as well as computationally efficient. We demonstrate the benefits of structural interpolation using three specific applications. These are (a) traditional color imaging with a mosaic of color filters, (b) high dynamic range monochrome imaging using a mosaic of exposure filters, and (c) high dynamic range color imaging using a mosaic of overlapping color and exposure filters.

1 Multi-Sampled Imaging

Currently, vision algorithms rely on images with 8 bits of brightness or color at each pixel. Images of such quality are simply inadequate for many real-world applications. Significant advances in imaging can be made by exploring the fundamental trade-offs that exist between various dimensions of imaging (see Figure 1). The relative importances of these dimensions clearly depend on the application at hand. In any practical scenario, however, we are given a finite number of pixels (residing on one or more detectors) to sample the imaging dimensions. Therefore, it is beneficial to view imaging as the judicious assignment of resources (pixels) to the dimensions of imaging that are relevant to the application.

Different pixel assignments can be viewed as different types of samplings of the imaging dimensions. In all cases, however, more than one dimension is simultaneously sampled. In the simplest case of a gray-scale image, image brightness and image space are sampled, simultaneously. More interesting examples result from using image detectors made of an assortment of pixels, as shown in Figure

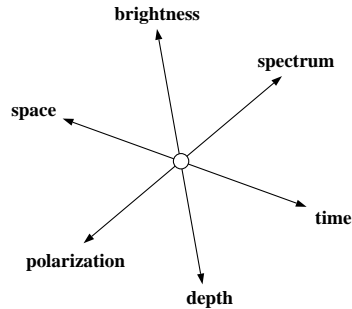


Fig. 1. A few dimensions of imaging. Pixels on an image detector may be assigned to multiple dimensions in a variety of ways depending on the needs of the application.

2. Figure 2(a) shows the popular Bayer mosaic [Bay76] of red, green and blue spectral filters placed adjacent to pixels on a detector. Since multiple color measurements cannot be captured simultaneously at a pixel, the pixels are assigned to specific colors to trade-off spatial resolution for spectral resolution. Over the last three decades various color mosaics have been suggested, each one resulting in a different trade-off (see [Dil77], [Dil78], [MOS83], [Par85], [KM85]).

Historically, multi-sampled imaging has only been used in the form of color mosaics. Only recently has the approach been used to explore other imaging dimensions. Figure 2(b) shows the mosaic of neutral density filters with different transmittances used in [NM00] to enhance an image detector’s dynamic range. In this case, spatial resolution is traded-off for brightness resolution (dynamic range). In [SN01], spatially varying transmittance and spectral filters were used with regular wide FOV mosaicing to yield high dynamic range and multi-spectral mosaics. Figure 2(c) shows how space, dynamic range and color can be sampled simultaneously by using a mosaic of filters with different spectral responses and transmittances. This type of multi-sampling is novel and, as we shall show, results in high dynamic range color images. Another example of assorted pixels was proposed in [BE00], where a mosaic of polarization filters with different orientations is used to estimate the polarization parameters of light reflected by scene points. This idea can be used in conjunction with a spectral mosaic, as shown in Figure 2(d), to achieve simultaneous capture of polarization and color.

Multi-sampled imaging can be exploited in many other ways. Figures 2(e) shows how temporal sampling can be used with exposure sampling. This example is related to the idea of sequential exposure change proposed in [MP95], [DM97] and [MN99] to enhance dynamic range. However, it is different in that the exposure is varied as a periodic function of time, enabling the generation of high dynamic range, high framerate video. The closest implementation appears to be the one described in [GHZ92] where the electronic gain of the camera is varied periodically to achieve the same effect. A more sophisticated implementation may sample space, time, exposure and spectrum, simultaneously, as shown in Figure 2(f).

The above examples illustrate that multi-sampling provides a general framework for designing imaging systems that extract information that is most per-

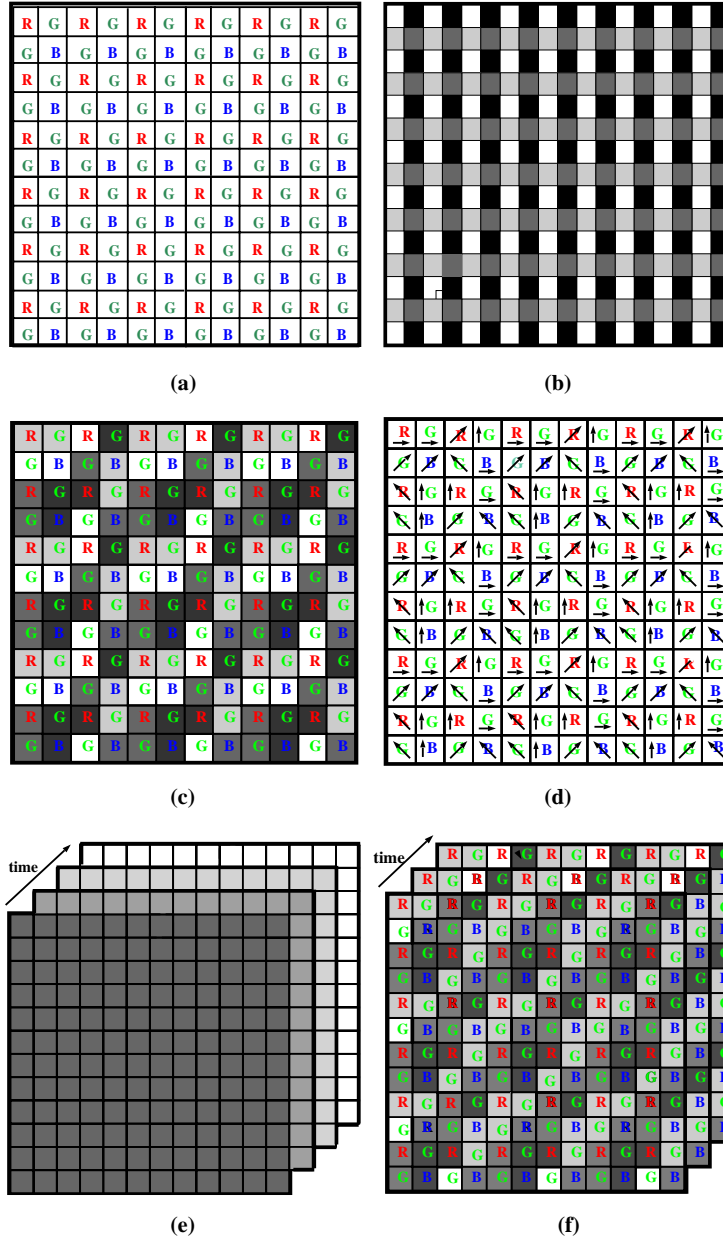


Fig. 2. A few examples of multi-sampled imaging using assorted pixels. (a) A color mosaic. Such mosaics are widely used in solid-state color cameras. (b) An exposure mosaic. (c) A mosaic that includes different colors and exposures. (d) A mosaic using color and polarization filters. (e), (f) Multi-sampling can also involve varying exposure and/or color over space and time.

herent to the application. Though our focus is on the visible light spectrum, multi-sampling is, in principle, applicable to any form of electromagnetic radiation. Therefore, the pixel assortments and reconstruction methods we describe in this paper are also relevant to other imaging modalities such as X-ray, magnetic resonance (MR) and infra-red (IR). Furthermore, the examples we discuss are two-dimensional but the methods we propose are directly applicable to higher-dimensional imaging problems such as ones found in tomography and microscopy.

2 Learned Structural Models for Reconstruction

How do we reconstruct the desired image from a captured multi-sampled one? Nyquist’s theory [Bra65] tells us that for a continuous signal to be perfectly reconstructed from its discrete samples, the sampling frequency must be at least twice the largest frequency in the signal. In the case of an image of a scene, the optical image is sampled at a frequency determined by the size of the detector and the number of pixels on it. In general, there is no guarantee that this sampling frequency satisfies Nyquist’s criterion. Therefore, when a traditional interpolation technique is used to enhance spatial resolution, it is bound to introduce errors in the form of blurring and/or aliasing. In the case of multi-sampled images (see Figure 2), the assignment of pixels to multiple dimensions causes further undersampling of scene radiance along at least some dimensions. As a result, conventional interpolation methods are even less effective.

Our objective is to overcome the limits imposed by Nyquist’s theory by using prior models that capture redundancies inherent in images. The physical structures of real-world objects, their reflectances and illuminations impose strong constraints on the light fields of scenes. This causes different imaging dimensions to be highly correlated with each other. Therefore, a local mapping function can be learned from a set of multi-sampled images and their corresponding correct (high quality) images. As we shall see, it is often beneficial to use multiple mapping functions. Then, given a novel multi-sampled image, these mapping functions can be used to reconstruct an image that has enhanced resolution in each of the dimensions of interest. We refer to these learned mapping functions as *local structural models*.

The general idea of learning interpolation functions is not new. In [FP99], a probabilistic Markov network is trained to learn the relationship between sharp and blurred images, and then used to increase spatial resolution of an image. In [BK00], a linear system of equations is solved to estimate a high resolution image from a sequence of low resolution images wherein the object of interest is in motion. Note that both these algorithms are developed to improve *spatial* resolution, while our interest is in resolution enhancement along multiple imaging dimensions.

Learning based algorithms have also been applied to the problem of interpolating images captured using color mosaics. The most relevant among these is the work of Wober and Soini [WS95] that estimates an interpolation kernel from training data (high quality color images of test patterns and their corresponding color mosaic images). The same problem was addressed in [Bra94] using a Bayesian method.

We are interested in a general method that can interpolate not just color mosaic images but any type of multi-sampled data. For this, we propose the use of a structural model where each reconstructed value is a polynomial function of the image brightnesses measured within a local neighborhood. The size of the neighborhood and the degree of the polynomial vary with the type of multi-sampled data being processed. It turns out that the model of Wober and Soini [WS95] is a special instance of our model as it is a first-order polynomial applied to the specific case of color mosaic images. As we shall see, our polynomial model produces excellent results for a variety of multi-sampled images. Since it uses polynomials, our method is very efficient and can be easily implemented in hardware. In short, it is simple enough to be incorporated into any imaging device (digital still or video camera, for instance).

3 Training Using High Quality Images

Since we wish to learn our model parameters, we need a set of high quality training images. These could be real images of scenes, synthetic images generated by rendering, or some combination of the two. Real images can be acquired using professional grade cameras whose performance we wish to emulate using lower quality multi-sampling systems. Since we want our model to be general, the set of training images must adequately represent a wide range of scene features. For instance, images of urban settings, landscapes and indoor spaces may be included. Rotated and magnified versions of the images can be used to capture the effects of scale and orientation. In addition, the images may span the gamut of illumination conditions encountered in practice, varying from indoor lighting to overcast and sunny conditions outdoor. Synthetic images are useful as one can easily include in them specific features that are relevant to the application.



Fig. 3. Some of the 50 high quality images (2000 x 2000 pixels, 12 bits per color channel) used to train the local structural models described in Sections 4, 5 and 6.

Some of the 50 high quality images we have used in our experiments are shown in Figure 3. Each of these is a 2000 x 2000 color (red, green, blue) image with 12 bits of information in each color channel. These images were captured using a film camera and scanned using a 12-bit slide scanner. Though the total

number of training images is small they include a sufficiently large number of local (say, 7×7 pixels) appearances for training our structural models.

Given such high quality images, it is easy to generate a corresponding set of low-quality multi-sampled images. For instance, given a 2000×2000 RGB image with 12-bits per pixel, per color channel, simple downsampling in space, color, and brightness results in a 1000×1000 , 8 bits per pixel multi-sampled image with the sampling pattern shown in Figure 2(c). We refer to this process of generating multi-sampled images from high quality images as *downgrading*.

With the high quality images and their corresponding (downgraded) multi-sampled images in place, we can learn the parameters of our structural model. A structural model is a function f that relates measured data $\mathbf{M}(x, y)$ in a multi-sampled image to a desired value $\mathbf{H}(i, j)$ in the high quality training image:

$$\mathbf{H}(i, j) = f(\mathbf{M}(1, 1), \dots, \mathbf{M}(x, y), \dots, \mathbf{M}(X, Y)) \quad (1)$$

where, X and Y define some neighborhood of measured data around, or close to, the high quality value $\mathbf{H}(i, j)$. Since our structural model is a polynomial, it is linear in its coefficients. Therefore, the coefficients can be efficiently computed from training data using linear regression.

Note that a single structural model may be inadequate. If we set aside the measured data and focus on the type of multi-sampling used (see Figure 2), we see that pixels can have different types of neighborhood sampling patterns. If we want our models to be compact (small number of coefficients) and effective we cannot expect them to capture variations in scenes as well as changes in the sampling pattern. Hence, we use a single structural model for each type of local sampling pattern. Since our imaging dimensions are sampled in a uniform manner, in all cases we have a small number of local sampling patterns. Therefore, only a small number of structural models are needed. During reconstruction, given a pixel of interest, the appropriate structural model is invoked based on the pixel's known neighborhood sampling pattern.

4 Spatially Varying Color (SVC)

Most color cameras have a single image detector with a mosaic of red, green and blue spectral filters on it. The resulting image is hence a widely used type of multi-sampled image. We refer to it as a spatially varying color (SVC) image. When one uses an NTSC color camera, the output of the camera is nothing but an interpolated SVC image. Color cameras are notorious for producing inadequate spatial resolution and this is exactly the problem we seek to overcome using structural models. Since this is our first example, we will use it to describe some of the general aspects of our approach.

4.1 Bayer Color Mosaic

Several types of color mosaics have been implemented in the past [Bay76], [Dil77], [Dil78], [MOS83], [Par85], [KM85]. However, the most popular of these is the Bayer pattern [Bay76] shown in Figure 4. Since the human eye is more sensitive to the green channel, the Bayer pattern uses more green filters than it does red

and blue ones. Specifically, the spatial resolutions of green, red and blue are 50%, 25% and 25%, respectively. Note that the entire mosaic is obtained by repeating the 2×2 pattern shown on the right in Figure 4. Therefore, given a neighborhood size, all neighborhoods in a Bayer mosaic must have one of four possible sampling patterns. If the neighborhood is of size 3×3 , the resulting patterns are **p1**, **p2**, **p3** and **p4** shown in Figure 4.

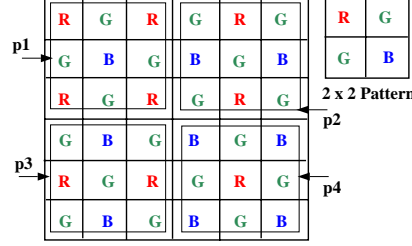


Fig. 4. Spatially varying color (SVC) pattern on a Bayer mosaic. Given a neighborhood size, all possible sampling patterns in the mosaic must be one of four types. In the case of a 3×3 neighborhood, these patterns are **p1**, **p2**, **p3** and **p4**.

4.2 SVC Structural Model

From the measured SVC image, we wish to compute three color values (red, green and blue) at each pixel, even though each pixel in the SVC image provides a single color measurement. Let the measured SVC image be denoted by \mathbf{M} and the desired high quality color image by \mathbf{H} . A structural model relates each color value in \mathbf{H} to the measured data within a small neighborhood in \mathbf{M} . This neighborhood includes measurements of different colors and hence the model implicitly accounts for correlations between different color channels.

As shown in Figure 5, let $\mathbf{M}_{\mathbf{p}}$ be the measured data in a neighborhood with sampling pattern \mathbf{p} , and $\mathbf{H}_{\mathbf{p}}(i + 0.5, j + 0.5, \lambda)$ be the high quality color value at the center of the neighborhood. (The center is off-grid because the neighborhood is an even number of pixels in width and height.) Then, a polynomial structural model can be written as:

$$\mathbf{H}_{\mathbf{p}}(i + 0.5, j + 0.5, \lambda) = \sum_{(x,y) \in \mathbf{S}_{\mathbf{p}}(i,j)} \sum_{(k \neq x, l \neq y) \in \mathbf{S}_{\mathbf{p}}(i,j)} \sum_{n=0}^{N_{\mathbf{p}}} \sum_{q=0}^{N_{\mathbf{p}}-n} \mathbf{C}_{\mathbf{p}}(a, b, c, d, \lambda, n) \mathbf{M}_{\mathbf{p}}^n(x, y) \mathbf{M}_{\mathbf{p}}^q(k, l). \quad (2)$$

$\mathbf{S}_{\mathbf{p}}(i, j)$ is the neighborhood of pixel (i, j) , $N_{\mathbf{p}}$ is the order of the polynomial and $\mathbf{C}_{\mathbf{p}}$ are the polynomial coefficients for the pattern \mathbf{p} . The coefficient indices (a, b, c, d) are equal to $(x - i, y - j, k - i, l - j)$.

The product $\mathbf{M}_{\mathbf{p}}(x, y) \mathbf{M}_{\mathbf{p}}(k, l)$ explicitly represents the correlation between different pixels in the neighborhood. For efficiency, we have not used these cross-terms in our implementations. We found that very good results are obtained

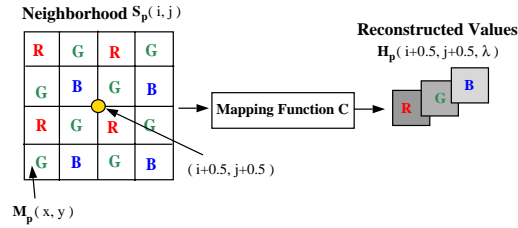


Fig. 5. The measured data $\mathbf{M}_{\mathbf{p}}$ in the neighborhood $\mathbf{S}_{\mathbf{p}}(i, j)$ around pixel (i, j) are related to the high quality color values $\mathbf{H}_{\mathbf{p}}(i + 0.5, j + 0.5, \lambda)$ via a polynomial with coefficients $\mathbf{C}_{\mathbf{p}}$.

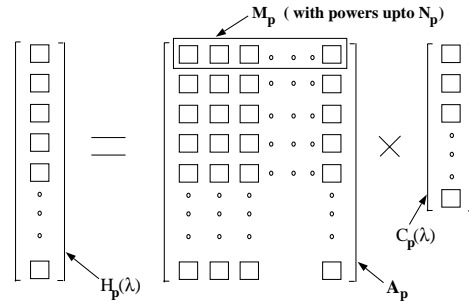


Fig. 6. The mapping function in (3) can be expressed as a linear system using matrices and vectors. For a given pattern \mathbf{p} and color λ , $\mathbf{A}_{\mathbf{p}}$ is the measurement matrix, $\mathbf{C}_{\mathbf{p}}(\lambda)$ is the coefficient vector and $\mathbf{H}_{\mathbf{p}}(\lambda)$ is the reconstruction vector.

even when each desired value is expressed as just a sum of polynomial functions of the individual pixel measurements:

$$\mathbf{H}_{\mathbf{p}}(i + 0.5, j + 0.5, \lambda) = \sum_{(x, y) \in \mathbf{S}_{\mathbf{p}}(i, j)} \sum_{n=0}^{N_{\mathbf{p}}} \mathbf{C}_{\mathbf{p}}(a, b, \lambda, n) \mathbf{M}_{\mathbf{p}}^n(x, y). \quad (3)$$

The mapping function (3), for each color λ and each local pattern type \mathbf{p} , can be conveniently rewritten using matrices and vectors, as shown in Figure 6:

$$\mathbf{H}_{\mathbf{p}}(\lambda) = \mathbf{A}_{\mathbf{p}} \mathbf{C}_{\mathbf{p}}(\lambda). \quad (4)$$

For a given pattern type \mathbf{p} and color λ , $\mathbf{A}_{\mathbf{p}}$ is the measurement matrix. The rows of $\mathbf{A}_{\mathbf{p}}$ correspond to the different neighborhoods in the image that have the pattern \mathbf{p} . Each row includes all the relevant powers (up to $N_{\mathbf{p}}$) of the measured data $\mathbf{M}_{\mathbf{p}}$ within the neighborhood. The vector $\mathbf{C}_{\mathbf{p}}(\lambda)$ includes the coefficients of the polynomial mapping function and the vector $\mathbf{H}_{\mathbf{p}}(\lambda)$ includes the desired high quality values at the off-grid neighborhood centers. The estimation of the model parameters $\mathbf{C}_{\mathbf{p}}$ can then be posed as a least squares problem:

$$\mathbf{C}_{\mathbf{p}}(\lambda) = (\mathbf{A}_{\mathbf{p}}^{\mathbf{T}} \mathbf{A}_{\mathbf{p}})^{-1} \mathbf{A}_{\mathbf{p}}^{\mathbf{T}} \mathbf{H}_{\mathbf{p}}(\lambda), \quad (5)$$

When the signal-to-noise characteristics of the image detector are known, (5) can be rewritten using weighted least squares to achieve greater accuracy [Aut01].

4.3 Total Number of Coefficients

The number of coefficients in the model (3) can be calculated as follows. Let the neighborhood size be $u \times v$, and the polynomial order corresponding to each pattern \mathbf{p} be $N_{\mathbf{p}}$. Let the number of distinct local patterns in the SVC image be P and the number of color channels be Λ . Then, the total number of coefficients needed for structural interpolation is:

$$|\mathbf{C}| = \left(P + u * v * \sum_{\mathbf{p}=1}^P N_{\mathbf{p}} \right) * \Lambda . \quad (6)$$

For the Bayer mosaic, $P = 4$ and $\Lambda = 3$ (R, G, B). If we use $N_{\mathbf{p}} = 2$ and $u = v = 6$, the total number of coefficients is 876. Since these coefficients are learned from real data, they yield greater precision during interpolation than standard interpolation kernels. In addition, they are very efficient to apply. Since there are $P = 4$ distinct patterns, only 219 (a quarter) of the coefficients are used for computing the three color values at a pixel. Note that the polynomial model is linear in the coefficients. Hence, structural interpolation can be implemented in real-time using a set of linear filters that act on the captured image and its powers (up to $N_{\mathbf{p}}$).

4.4 Experiments

A total of 30 high quality training images (see Figure 3) were used to compute the structural model for SVC image interpolation. Each image is downgraded to obtain a corresponding Bayer-type SVC image. For each of the four sampling patterns in the Bayer mosaic, and for each of the three colors, the appropriate image neighborhoods were used to compute the measurement matrix $\mathbf{A}_{\mathbf{p}}$ and the reconstruction vector $\mathbf{H}_{\mathbf{p}}(\lambda)$. While computing these, one additional step was taken; each measurement is normalized by the energy within its neighborhood to make the structural model insensitive to changes in illumination intensity and camera gain. The resulting $\mathbf{A}_{\mathbf{p}}$ and $\mathbf{H}_{\mathbf{p}}(\lambda)$ are used to find the coefficient vector $\mathbf{C}_{\mathbf{p}}(\lambda)$ using linear regression (see (5)). In our implementation, we used the parameter values $P = 4$ (Bayer), $N_{\mathbf{p}} = 2$, $u = v = 6$ and $\Lambda = 3$ (R, G, B), to get a total of 876 coefficients.

The above structural model was used to interpolate 20 test SVC images that are different from the ones used for training. In Figure 7(a), a high quality (8-bits per color channel) image is shown. Figure 7(b) shows the corresponding (downgraded) SVC image. This is really a single channel 8-bit image and its pixels are shown in color only to illustrate the Bayer pattern. Figure 7(c) shows a color image computed from the SVC image using bi-cubic interpolation. As is usually done, the three channels are interpolated separately using their respective data in the SVC image. The magnified image region clearly shows that bi-cubic interpolation results in a loss of high frequencies; the edges of the tree branches and the squirrels are severely blurred. Figure 7(d) shows the result of applying structural interpolation. Note that the result is of high quality with minimal loss of details.

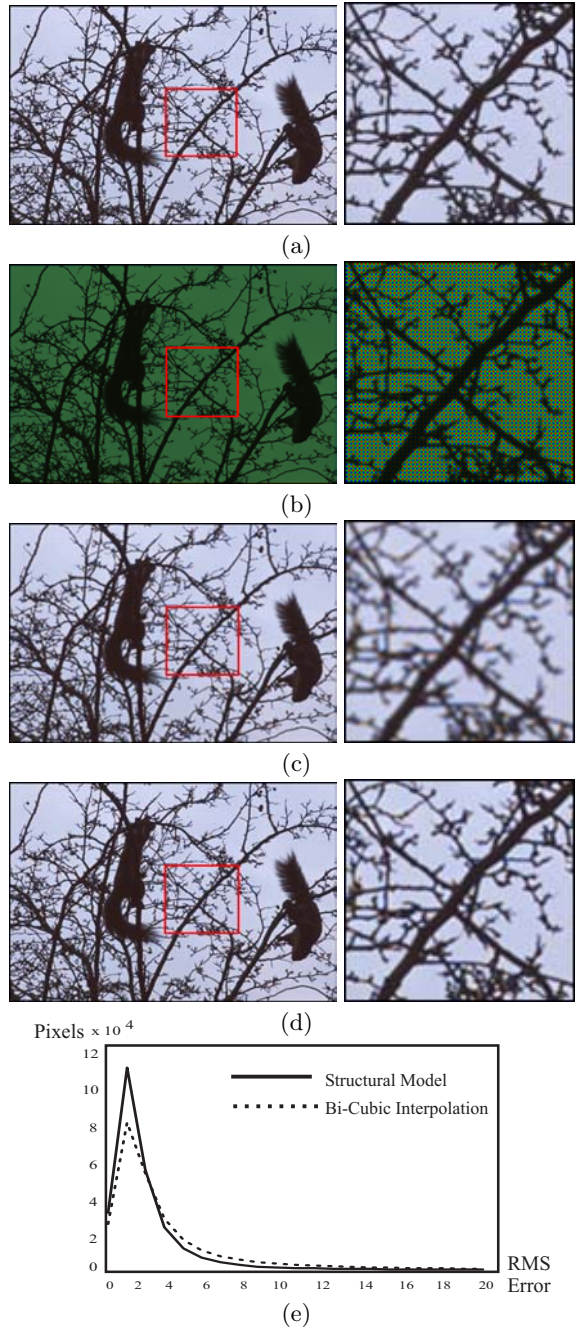


Fig. 7. (a) Original (high quality) color image with 8-bits per color channel. (b) SVC image obtained by downgrading the original image. The pixels in this image are shown in color only to illustrate the Bayer mosaic. Color image computed from the SVC image using (c) bi-cubic interpolation and (d) structural interpolation. (e) Histograms of luminance error (averaged over 20 test images). The RMS error is 6.12 gray levels for bi-cubic interpolation and 3.27 gray levels for structural interpolation.

We have quantitatively verified our results. Figure 7(e) shows histograms of the luminance error for bi-cubic and structural interpolation. These histograms are computed using **all 20 test images** (not just the one in Figure 7). **The difference between the two histograms may appear to be small but is significant because a large fraction of the pixels in the 20 images belong to “flat” image regions that are easy to interpolate for both methods.** The RMS errors (computed over all 20 images) are 6.12 and 3.27 gray levels for bi-cubic and structural interpolation, respectively.

5 Spatially Varying Exposures (SVE)

In [NM00], it was shown that the dynamic range of a gray-scale image detector can be significantly enhanced by assigning different exposures (neutral density filters) to pixels, as shown in Figure 8. This is yet another example of a multi-sampled image and is referred to as a spatially varying exposure (SVE) image. In [NM00], standard bi-cubic interpolation was used to reconstruct a high dynamic range gray-scale image from the captured SVE image; first, saturated and dark pixels are eliminated, then all remaining measurements are normalized by their exposure values, and finally bi-cubic interpolation is used to find the brightness values at the saturated and dark pixels. As expected, the resulting image has enhanced dynamic range but lower spatial resolution. In this section, we apply structural interpolation to SVE images and show how it outperforms bi-cubic interpolation.

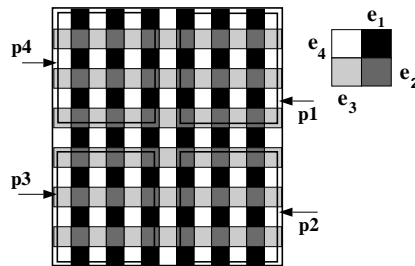


Fig. 8. The dynamic range of an image detector can be improved by assigning different exposures to pixels. In this special case of 4 exposures, any 6×6 neighborhood in the image must belong to one of four possible sampling patterns shown as **p1** . . . **p4**.

5.1 SVE Structural Model

As in the SVC case, let the measured SVE data be \mathbf{M} and the corresponding high dynamic range data be \mathbf{H} . If the SVE detector uses only four discrete exposures (see Figure 8), it is easy to see that a neighborhood of any given size can have only one of four different sampling patterns ($P = 4$). Therefore, for each sampling pattern \mathbf{p} , a polynomial structural model is used that relates the captured data $\mathbf{M}_{\mathbf{p}}$ within the neighborhood to the high dynamic range value $\mathbf{H}_{\mathbf{p}}$ at the center of the neighborhood:

$$\mathbf{H}_{\mathbf{p}}(i + 0.5, j + 0.5) = \sum_{(x,y) \in \mathbf{S}_{\mathbf{p}}(i,j)} \sum_{n=0}^{N_{\mathbf{p}}} \mathbf{C}_{\mathbf{p}}(a, b, n) \mathbf{M}_{\mathbf{p}}^n(x, y), \quad (7)$$

where, as before, $(a, b) = (x - i, y - j)$, $\mathbf{S}_{\mathbf{p}}(i, j)$ is the neighborhood of pixel (i, j) , $N_{\mathbf{p}}$ is the order of the polynomial mapping, and $\mathbf{C}_{\mathbf{p}}$ are the polynomial coefficients for the pattern \mathbf{p} . Note that there is only one channel in this case (gray-scale) and hence the parameter λ is omitted. The above model is rewritten in terms of a measurement matrix $\mathbf{A}_{\mathbf{p}}$ and a reconstruction vector $\mathbf{H}_{\mathbf{p}}$, and the coefficients $\mathbf{C}_{\mathbf{p}}$ are found using (5). The number of coefficients in the SVE structural model is determined as:

$$|\mathbf{C}| = P + u * v * \sum_{\mathbf{p}=1}^P N_{\mathbf{p}}. \quad (8)$$

In our implementation, we have used $P = 4$, $N_{\mathbf{p}} = 2$ and $u = v = 6$, which given a total of 292 coefficients. Since $P = 4$, only 73 coefficients are needed for reconstructing each pixel in the image.

5.2 Experiments

The SVE structural model was trained using 12-bit gray-scale versions of 6 of the images shown in Figure 3 and their corresponding 8-bit SVE images. Each SVE image was obtained by applying the exposure pattern shown in Figure 8 (with $e_4 = 4e_3 = 16e_2 = 64e_1$) to the original image, followed by a downgrade from 12 bits to 8 bits. The structural model was tested using 6 test images, one of which is shown in Figure 9. Figure 9(a) shows the original 12-bit image, Figure 9(b) shows the downgraded 8-bit SVE image, Figure 9(c) shows a 12-bit image obtained by bi-cubic interpolation of the SVE image, and Figure 9(d) shows the 12-bit image obtained by structural interpolation. The magnified images shown on the right are histogram equalized to bring out the details (in the clouds and walls) that are lost during bi-cubic interpolation but extracted by structural interpolation. Figure 9(e) compares the error histograms (computed using all 6 test images) for the two cases. The RMS errors were found to be 33.4 and 25.5 gray levels (in a 12-bit range) for bi-cubic and structural interpolations, respectively. Note that even though a very small number (6) of images were used for training, our method outperforms bi-cubic interpolation.

6 Spatially Varying Exposure and Color (SVEC)

Since we are able to extract high spatial and spectral resolution from SVC images and high spatial and brightness resolution from SVE images, it is natural to explore how these two types of multi-sampling can be combined into one. The result is the simultaneous sampling of space, color and exposure (see Figure 10). We refer to an image obtained in this manner as a spatially varying exposure and color (SVEC) image. If the SVEC image has 8-bits at each pixel, we would like to compute at each pixel three color values, each with 12 bits of precision. Since the same number of pixels on a detector are now being used to sample three different dimensions, it should be obvious that this is a truly challenging interpolation problem. We will see that structural interpolation does remarkably well at extracting the desired information.

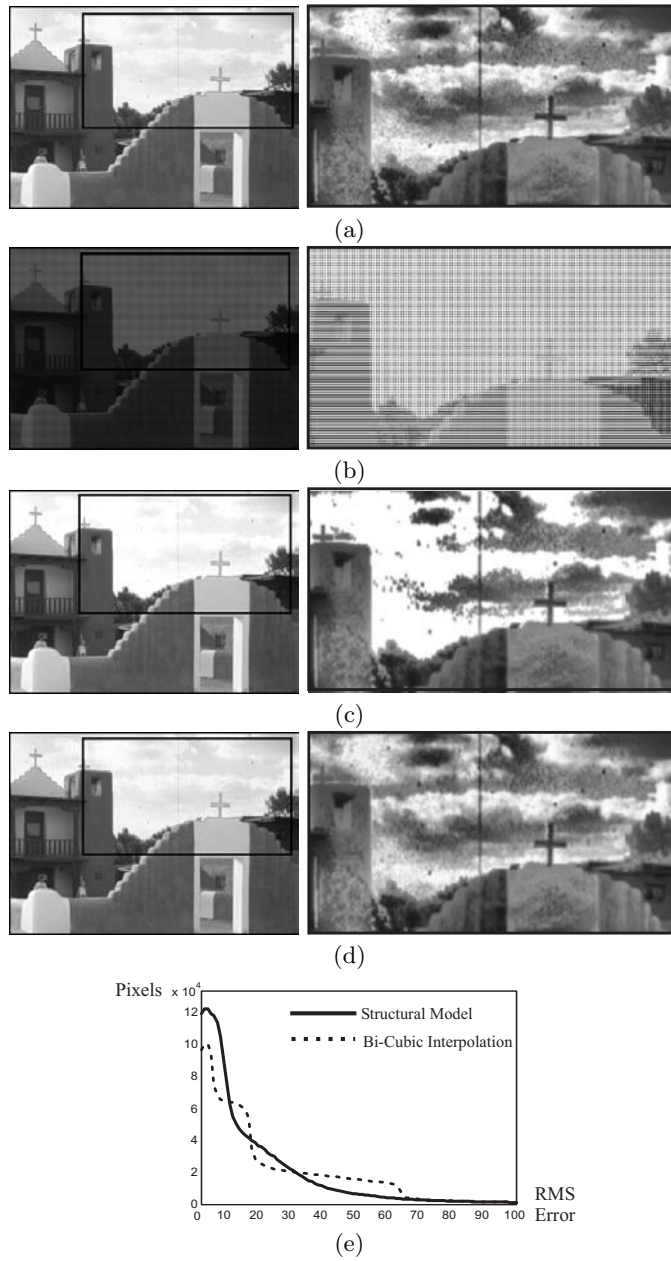


Fig. 9. (a) Original 12-bit gray scale image. (b) 8-bit SVE image. (c) 12-bit (high dynamic range) image computed using bi-cubic interpolation. (d) 12-bit (high dynamic range) image computed using structural models. (e) Error histograms for the two case (averaged over 6 test images). The RMS error for the 6 images are 33.4 and 25.5 gray levels (on a 12-bit scale) for bi-cubic and structural interpolation, respectively. The magnified image regions on the right are histogram equalized.

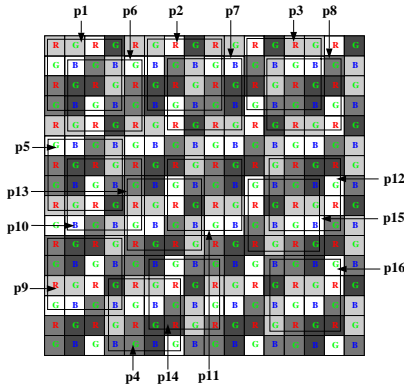


Fig. 10. A combined mosaic of 3 spectral and 4 neutral density filters used to simultaneously sample space, color and exposures using a single image detector. The captured 8-bit SVEC image can be used to compute a 12-bit (per channel) color image by structural interpolation. For this mosaic, for any neighborhood size, there are only 16 distinct sampling patterns. For a 4×4 neighborhood size, the patterns are **p1 . . . p16**.

Color and exposure filters can be used to construct an SVEC sensor in many ways. All possible mosaics that can be constructed from A colors and E exposures are derived in [Aut01]. Here, we will focus on the mosaic shown in Figure 10, where 3 colors and 4 exposures are used. For any given neighborhood size, it is shown in [Aut01] that only 16 different sampling patterns exist (see Figure 10).

6.1 SVEC Structural Model

The polynomial structural model used in the SVEC case is the same as the one used for SVC, and is given by (3). The only caveat is that in the SVEC case we need to ensure that the neighborhood size used is large enough to adequately sample all the colors and exposures. That is, the neighborhood size is chosen such that it includes all colors and all exposures of each color.

The total number of polynomial coefficients needed is computed the same way as in the SVC case and is given by (6). In our experiments, we have used the mosaic shown in Figure 10. Therefore, $P = 16$, $A = 3$ (R , G , and B), $N_{\mathbf{p}} = 2$ for each of the 16 patterns, and $u = v = 6$, to give a total of 3504 coefficients. Once again, at each pixel, for each color, only $3504/48 = 73$ coefficients are used. Therefore, even for this complex type of multi-sampling, our structural models can be applied to images in real-time using a set of linear filters.

6.2 Experiments

The SVEC structural model was trained using 6 of the images in Figure 3. In this case, the 12-bit color images in the training set were downgraded to 8-bit SVEC images. The original and SVEC images were used to compute the 3504 coefficients of the model. The model was then used to map 10 different test SVEC images to 12-bit color images. One of these results is shown in Figure 11. The original 12-bit image shown in Figure 11(a) was downgraded to obtain the 8-bit SVEC image shown in Figure 11(b). This image has a single channel

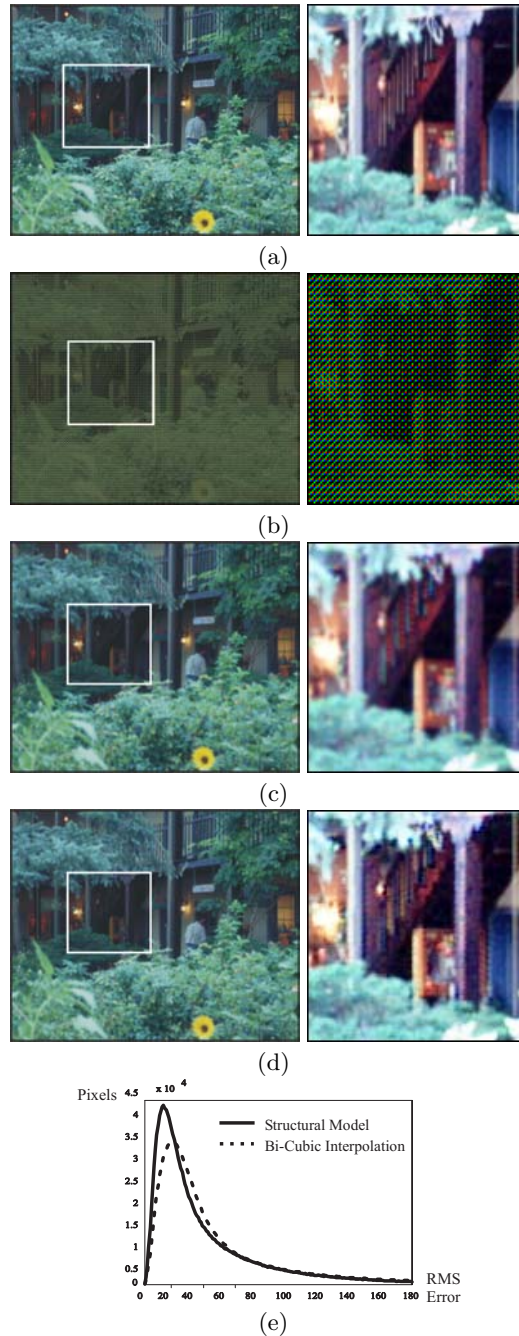


Fig. 11. (a) Original 12-bit color image. (b) 8-bit SVEC Image. 12-bit color images reconstructed using (c) bi-cubic interpolation and (d) structural interpolation. (e) Luminance error histogram computed using 10 test images. RMS luminance errors were found to be 118 and 80 (on a 12-bit scale) for bi-cubic and structural interpolation, respectively. The magnified images on the right are histogram equalized.

and is shown here in color only to illustrate the effect of simultaneous color and exposure sampling. Figures 11(c) and (d) show the results of applying bi-cubic and structural interpolation, respectively. It is evident from the magnified images on the right that structural interpolation yields greater spectral and spatial resolution. The two interpolation techniques are compared in Figure 11(e) which shows error histograms computed using all 10 test images. The RMS luminance errors were found to be 118 gray-levels and 80 gray-levels (on a 12 bit scale) for bi-cubic and structural interpolations, respectively.

Acknowledgements: This research was supported in parts by a ONR/ARPA MURI Grant (N00014-95-1-0601) and an NSF ITR Award (IIS-00-85864).

References

- [Aut01] Authors. Resolution Enhancement Along Multiple Imaging Dimensions Using Structural Models. *Technical Report, Authors' Institution*, 2001.
- [Bay76] B. E. Bayer. Color imaging array. *U.S. Patent 3,971,065*, July 1976.
- [BE00] M. Ben-Ezra. Segmentation with Invisible Keying Signal. In *CVPR*, 2000.
- [BK00] S. Baker and T. Kanade. Limits on Super-Resolution and How to Break Them. In *Proc. of CVPR*, pages II:372–379, 2000.
- [Bra65] R. N. Bracewell. *The Fourier Transform and Its Applications*. McGraw Hill, 1965.
- [Bra94] D. Brainard. Bayesian method for reconstructing color images from trichromatic samples. In *Proc. of IS&T 47th Annual Meeting*, pages 375–380, Rochester, New York, 1994.
- [Dil77] P. L. P. Dillon. Color imaging array. *U.S. Patent 4,047,203*, September 1977.
- [Dil78] P. L. P. Dillon. Fabrication and performance of color filter arrays for solid-state imagers. In *IEEE Trans. on Electron Devices*, volume ED-25, pages 97–101, 1978.
- [DM97] P. Debevec and J. Malik. Recovering High Dynamic Range Radiance Maps from Photographs. *Proc. of ACM SIGGRAPH 1997*, pages 369–378, 1997.
- [FP99] W. Freeman and E. Pasztor. Learning Low-Level Vision. In *Proc. of ICCV*, pages 1182–1189, 1999.
- [GHZ92] R. Ginosar, O. Hilsenrath, and Y. Zeevi. Wide dynamic range camera. *U.S. Patent 5,144,442*, September 1992.
- [KM85] K. Knop and R. Morf. A new class of mosaic color encoding patterns for single-chip cameras. In *IEEE Trans. on Electron Devices*, volume ED-32, 1985.
- [MN99] T. Mitsunaga and S. K. Nayar. Radiometric Self Calibration. In *Proc. of CVPR*, pages I:374–380, 1999.
- [MOS83] D. Manabe, T. Ohta, and Y. Shimidzu. Color filter array for IC image sensor. In *Proc. of IEEE Custom Integrated Circuits Conference*, pages 451–455, 1983.
- [MP95] S. Mann and R. Picard. Being ‘Undigital’ with Digital Cameras: Extending Dynamic Range by Combining Differently Exposed Pictures. *Proc. of IST’s 48th Annual Conf.*, pages 442–448, May 1995.
- [NM00] S. K. Nayar and T. Mitsunaga. High Dynamic Range Imaging: Spatially Varying Pixel Exposures. In *Proc. of CVPR*, pages I:472–479, 2000.
- [Par85] K. A. Parulski. Color filters and processing alternatives for one-chip cameras. In *IEEE Trans. on Electron Devices*, volume ED-32, 1985.
- [SN01] Y. Y. Schechner and S. K. Nayar. Generalized mosaicing. In *ICCV*, 2001.
- [WS95] M. A. Wober and R. Soini. Method and apparatus for recovering image data through the use of a color test pattern. *U.S. Patent 5,475,769*, Dec. 1995.

Code A: Matlab Code for Poisson Image Reconstruction from Image Gradients

```
% Read Input Gray Image
imgstr = 'test.png'; disp(sprintf('Reading Image %s',imgstr));
img = imread(imgstr);      [H,W,C] = size(img);      img = double(img);
% Find gradinets
gx = zeros(H,W); gy = zeros(H,W);      j = 1:H-1; k = 1:W-1;
gx(j,k) = (img(j,k+1) - img(j,k));      gy(j,k) = (img(j+1,k) - img(j,k));

% Reconstruct image from gradients for verification
img_rec = poisson_solver_function(gx,gy,img);

figure;imagesc(img);colormap gray;colorbar;title('Image')
figure;imagesc(img_rec);colormap gray;colorbar;title('Reconstructed');
figure;imagesc(abs(img_rec-img));colormap gray;colorbar;title('Abs error');
```

```
function [img_direct] = poisson_solver_function(gx,gy,boundary_image);
% function [img_direct] = poisson_solver_function(gx,gy,boundary_image)
% Inputs; Gx and Gy -> Gradients
% Boundary Image -> Boundary image intensities
% Gx Gy and boundary image should be of same size
[H,W] = size(boundary_image);
gxx = zeros(H,W); gyy = zeros(H,W);
f = zeros(H,W);      j = 1:H-1;      k = 1:W-1;

% Laplacian
gyy(j+1,k) = gy(j+1,k) - gy(j,k);      gxx(j,k+1) = gx(j,k+1) - gx(j,k);
f = gxx + gyy;      clear j k gxx gyy gyyd gxxd

% boundary image contains image intensities at boundaries
boundary_image(2:end-1,2:end-1) = 0;
disp('Solving Poisson Equation Using DST');      tic
j = 2:H-1;      k = 2:W-1;      f_bp = zeros(H,W);
f_bp(j,k) = -4*boundary_image(j,k) + boundary_image(j,k+1) +
            boundary_image(j,k-1) + boundary_image(j-1,k) + boundary_image(j+1,k);
clear j k

f1 = f - reshape(f_bp,H,W);% subtract boundary points contribution
clear f_bp f

% DST Sine Transform also starts here
f2 = f1(2:end-1,2:end-1);      clear f1
%compute sine transform
tt = dst(f2);      f2sin = dst(tt)';      clear f2

%compute Eigen Values
[x,y] = meshgrid(1:W-2,1:H-2);      denom = (2*cos(pi*x/(W-1))-2) + (2*cos(pi*y/(H-1)) - 2) ;

%divide
f3 = f2sin./denom;      clear f2sin x y

%compute Inverse Sine Transform
tt = idst(f3);      clear f3;      img_tt = idst(tt)';      clear tt

time_used = toc;      disp(sprintf('Time for Poisson Reconstruction = %f secs',time_used));

% put solution in inner points; outer points obtained from boundary image
img_direct = boundary_image;
img_direct(2:end-1,2:end-1) = 0;
img_direct(2:end-1,2:end-1) = img_tt;
```

```

function b=dst(a,n)
%DST Discrete sine transform (Used in Poisson reconstruction)
% Y = DST(X) returns the discrete sine transform of X.
% The vector Y is the same size as X and contains the
% discrete sine transform coefficients.
% Y = DST(X,N) pads or truncates the vector X to length N
% before transforming.
% If X is a matrix, the DST operation is applied to each
% column. This transform can be inverted using IDST.

error(nargchk(1,2,nargin));

if min(size(a))==1
    if size(a,2)>1
        do_trans = 1;
    else
        do_trans = 0;
    end
    a = a(:);
else
    do_trans = 0;
end
if nargin==1, n = size(a,1); end
m = size(a,2);

% Pad or truncate a if necessary
if size(a,1)<n,
    aa = zeros(n,m); aa(1:size(a,1),:) = a;
else
    aa = a(1:n,:);
end

y=zeros(2*(n+1),m); y(2:n+1,:)=aa; y(n+3:2*(n+1),:)=flipud(aa);
yy=fft(y); b=yy(2:n+1,:)/(-2*sqrt(-1));

if isreal(a), b = real(b); end
if do_trans, b = b.'; end

```

```

function b=idst(a,n)
%IDST Inverse discrete sine transform (Used in Poisson reconstruction)
%
% X = IDST(Y) inverts the DST transform, returning the
% original vector if Y was obtained using Y = DST(X).
% X = IDST(Y,N) pads or truncates the vector Y to length N
% before transforming.
% If Y is a matrix, the IDST operation is applied to
% each column.

if nargin==1
    if min(size(a))==1
        n=length(a);
    else
        n=size(a,1);
    end
end

nn=n+1; b=2/nn*dst(a,n);

```

Code B: Matlab Code for Graph Cuts on Images

```
% Read gray scale image
I = imread('test.png');    [H,W,C] = size(I);

% Find graph cut
Ncut = graphcuts(I,33,255);

figure;imagesc(I);colormap gray;title('Image');
figure;imagesc(Ncut);colormap gray;title('Segmentation');
```

```
function [Ncut] = graphcuts(I,pad,MAXVAL)
% function [Ncut] = graphcuts(I)
% Input: I image
%   pad: spatial connectivity; eg. 3
%   MAXVAL: maximum image value
% Output: Ncut: Binary map 0 or 1 corresponding to image segmentation

I = double(I);    [H,W] = size(I);

% Find weights between nodes I1 and I2, w = exp(a*abs(I1-I2));
% Set a to have a weight of 0.01 for diff = MAXVAL
a = log(0.01)/MAXVAL;    x = [0:MAXVAL/100:MAXVAL]';    y = exp(a*x);
figure;plot(x,y);xlabel('intensity diff');ylabel('weights');    title('weights')

ws = 2*pad + 1;
if(ws <= 3)    ws = 3;    end

%Build the weight matrix
disp('Building Weight Matrix');    close all;    tic

WM = zeros(H*W,H*W);    countWM = 0;
for kk = 1:W
    for jj = 1:H
        mask = logical(zeros(H,W));
        cs = kk-pad;    ce = kk+pad;    rs = jj-pad;    re = jj+pad;
        if(cs<1)    cs = 1;    end;
        if(ce>W)    ce = W;    end;
        if(rs<1)    rs = 1;    end;
        if(re>H)    re = H;    end;
        mask(rs:re,cs:ce) = 1;
        idx = find(mask==1);
        p = abs(I(idx) - I(jj,kk));    p = exp(a*p);
        countWM = countWM + 1;    WM(countWM,idx) = p(:)';
    end
end
ttime = toc; disp(sprintf('Time for generating weight matrix = %f',ttime)); clear countWM

% Weight between a node and itself is 0
for jj = 1:H*W    WM(jj,jj) = 0;    end;    WM = sparse(WM);

% Shi and Malik Algorithm: second smallest eigen vector
disp('Finding Eigen Vector');
d = sum(WM,2);    D = diag(d);    tic
B = (D-WM);    B = (B+B')/2;    OPTS.disp = 0;
[v,d,flag] = eigs(B,D,2,'SA',OPTS);    ttime = toc;
disp(sprintf('Time for finding eigen vector = %f',ttime));    clear OPTS
y = v(:,2);
Ncut = reshape(y,H,W);
Ncut = Ncut > 0;
```

Code C: Matlab Code for Bilateral Filtering on Images

```
function [img1] = bilateral_filtering(img,winsize,sigma)

% Bilateral Filtering(img,winsize,sigma)
% Input      -> Image img
%           -> winsize: spatial filter width
%           -> sigma for intensity diff gaussian filter
%           -> sigma for spatial filter = winsize/6
% Output     -> Filtered Image
% Author: Amit Agrawal, 2004

disp('Bilateral Filtering');

[H,W] = size(img);

%Gaussian spatial filter
g_filter = fspecial('gaussian',winsize,winsize/6);
padnum = (winsize-1)/2;

A = padarray(img, [padnum padnum], 'replicate', 'both');
img1 = zeros(size(img));

for jj = padnum+1:(padnum+1+H-1)
    for kk = padnum+1:(padnum+1+W-1)

        % Get a local neighborhood
        imgwin = A(jj-padnum:jj+padnum, kk-padnum:kk+padnum);

        % Find weights according to intensity diffs
        Wwin = exp(-abs(imgwin - imgwin(padnum+1,padnum+1))/sigma^2);

        % Find composite filter
        newW = Wwin.*g_filter;

        t = sum(sum(newW));
        if(t>0)
            newW = newW/t;
        end

        img1(jj-padnum, kk-padnum) = sum(sum(imgwin.*newW));
    end
end
end
```

Bibliography

Fusion of Images Taken by Varying Camera and Scene Parameters

General

AKERS, D., LOSASSO, F., KLINGNER, J., AGRAWALA, M., RICK, J., AND HANRAHAN, P. 2003. Conveying shape and features with image-based relighting. In *IEEE Visualization*, 349–354.

BURT, P., AND KOLCZYNSKI, R. 1993. Enhanced image capture through fusion. In *International Conference on Computer Vision (ICCV 93)*, 173–182.

LEVIN, A., ZOMET, A., PELEG, S., AND WEISS, Y. 2004. Seamless image stitching in the gradient domain. In *European Conference on Computer Vision (ECCV 04)*.

MASSEY, M., AND BENDER, W. 1996. Salient stills: Process and practice. *IBM Systems Journal* 35, 3&4, 557–574.

MUTTER, S., AND KRAUSE, M. 1992. *Surrational Images: Photomontages*. University of Illinois Press.

ROBINSON, H. P. 1869. *Pictorial Effect in Photography: Being Hints on Composition and Chiaroscuro for Photographers*. Piper & Carter.

MUYBRIDGE, E. 1955. *The human figure in motion*. Dover Publications, Inc.

AGARWALA, A., DONTCHEVA, AGRAWALA, M., DRUCKER, COLBURN, CURLESS, SALESIN AND COHEN, M. Interactive Digital Photomontage. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2004)*, 2004.

Time

BRAUN, M. 1992. *Picturing Time: The Work of Etienne-Jules Marey*. The University of Chicago Press.

FREEMAN, W. T., AND ZHANG, H. 2003. Shape-time photography. In *Conference on Computer Vision and Pattern Recognition (CVPR 03)*, 151–157.

Exposure

FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. *ACM Transactions on Graphics* 21, 3, 249–256.

REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. *ACM Transactions on Graphics* 21, 3, 267–276.

DEBEVEC, AND MALIK. 1997. Recovering high dynamic range radiance maps from photographs. In *Proc. SIGGRAPH*.

DURAND, AND DORSEY. 2002. Fast bilateral filtering for the display of high-dynamic-range images. *ACMTrans. on Graphics* 21, 3.

MANN, AND PICARD. 1995. Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proc. IS&T 46th ann. conference*.

TUMBLIN, AND TURK. 1999. LCIS: A boundary hierarchy for detail-preserving contrast reduction. In Proc. SIGGRAPH.

DICARLO, J., AND WANDELL, B. 2000. Rendering high dynamic range images. Proc. SPIE: Image Sensors 3965, 392–401.

Focus

HAEBERLI, P. 1994. Grafica Obscura web site. <http://www.sgi.com/grafica/>.

MORGAN MCGUIRE, MATUSIK, PFISTER, HUGHES, AND DURAND, Defocus Video Matting, ACM Transactions on Graphics, Vol 24, No 3, July 2005, (Proceedings of ACM SIGGRAPH 2005).

Passive Illumination

RASKAR, R., ILIE, A., AND YU, J. 2004. Image fusion for context enhancement and video surrealism. In NPAR 2004: Third International Symposium on Non- Photorealistic Rendering.

WEISS, Y. 2001. Deriving intrinsic images from image sequences. In International Conference On Computer Vision (ICCV 01), 68–75.

Polorization

Y. Y. SCHECHNER, S. G. NARASIMHAN and S. K. NAYAR, Instant Dehazing of Images Using Polarization, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, December 2001.

S. K. NAYAR, X. FANG, and T. E. BOULT, Removal of Specularities using Color and Polarization, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,

Wavelength

D. A. SOCOLINSKY, “Dynamic range constraints in image fusion and realization.” Proc. IASTED Int. Conf. Signal and Image Process, 349-354 (2000).

Y. Y. SCHECHNER and S. K. NAYAR , Uncontrolled Modulation Imaging, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Washington DC, June 2004.

Location

B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, M. Levoy, High-Performance Imaging Using Large Camera Arrays.. ACM Transactions on Graphics, Vol 24, No 3, July 2005, pp 765-776 (Proceedings of ACM SIGGRAPH 2005).

Matting

CHUANG, Y.-Y., CURLESS, B., SALESIN, D., AND SZELISKI, R. 2001. A Bayesian approach to digital matting. In Proceedings of Computer Vision and Pattern Recognition (CVPR 2001), vol. II, 264 – 271.

PORTER, T., AND DUFF, T. 1984. Compositing digital images. In Computer Graphics (Proceedings of ACM SIGGRAPH 84), vol. 18, 253–259.

SMITH, A. R., AND BLINN, J. F. 1996. Blue screen matting. In Proceedings of ACM SIGGRAPH 96, 259–268.

Jian SUN, Jiaya JIA, Chi-Keung TANG and Heung-Yeung SHUM, Poisson Matting, ACM Transactions on Graphics, also in SIGGRAPH 2004, vol. 23, no. 3, July 2004, pages 315-321.

Techniques

General

DANIELSSON, P.-E. 1980. Euclidean distance mapping. *Computer Graphics and Image Processing* 14, 227–248.

LUCAS, B. D., AND KANADE, T. 1981. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*, 674–679.

MORTENSEN, E. N., AND BARRETT, W. A. 1995. Intelligent scissors for image composition. In *Proceedings of SIGGRAPH 95, Computer Graphics Proceedings, Annual Conference Series*, 191–198.

Graph Cuts

BOYKOV, Y., VEKSLER, O., AND ZABIH, R. 2001. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 11, 1222–1239.

KWATRA, V., SCHÖDL, A., ESSA, I., TURK, G., AND BOBICK, A. 2003. Graphcut textures: Image and video synthesis using graph cuts. *ACM Transactions on Graphics* 22, 3, 277–286.

SHI, J., AND MALIK, J. Normalized Cuts and Image Segmentation. *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, June 1997, Puerto Rico

Gradient Domain

PEREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. *ACM Transactions on Graphics* 22, 3, 313–318.

Smoothing, Bilateral and Trilateral Filter

C. TOMASI, AND R. MANDUCHI, Bilateral Filtering of gray and colored images, *Proc. IEEE Intl. Conference on Computer Vision*, pp. 836-846, 1998.

CHOUDHURY, P., TUMBLIN, J., "The Trilateral Filter for High Contrast Images and Meshes", *Proc. of the Eurographics Symposium on Rendering*, Per. H. Christensen and Daniel Cohen eds., pp. 186-196, 2003

Feature Extraction

Shape/Material/Illumination, Surface normals

BASRI, R. JACOBS, D. Photometric stereo with general, unknown lighting, *Computer Vision and Pattern Recognition*, 2001

B. K. P. HORN, "Shape from shading: A method for obtaining the shape of a smooth opaque object from one view," *MIT Project MAC Int. Rep. TR-79 and MIT AI Lab, Tech. Rep. 232*, Nov. 1970.

TODD ZICKLER, PETER N. BELHUMEUR, AND DAVID J. KRIEGMAN, "Helmholtz Stereopsis: Exploiting Reciprocity for Surface Reconstruction." Proc. 7th European Conference on Computer Vision, May 2002. Vol. III, pp 869-884.

ANDREAS WENGER, A GARDNER, CHRIS TCHOU, J UNGER, T HAWKINS, P DEBEVEC, Postproduction Relighting and Reflectance Transformation With Time-Multiplexed Illumination, SIGGRAPH 2005

Depth edges

Ramesh RASKAR , Karhan TAN, Rogerio FERIS, Jingyi YU, Matthew TURK, Non-photorealistic Camera: Depth Edge Detection and Stylized Rendering Using a Multi-Flash Camera, SIGGRAPH 2004

Depth

S. K. NAYAR and Y. NAKAGAWA., Shape from Focus, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No. 8, pp. 824-831,

Transfer and denoising

Flash to no-flash

Elmar EISEMANN and Frédo DURAND, Flash Photography Enhancement Via Intrinsic Relighting, SIGGRAPH 2004

Georg PETSCHNIGG, Maneesh AGRAWALA, Hugues HOPPE, Richard SZELISKI, Michael COHEN, Kentaro TOYAMA. Digital Photography with Flash and No-Flash Image Pairs. ACM Transactions on Graphics (Proceedings of SIGGRAPH 2004), 2004.

DICARLO, J. M., XIAO, F., AND WANDELL, B. A. 2001. Illuminating illumination. In 9th Color Imaging Conference, 27–34.

Noise

P. MEER, J. JOLION, AND A. ROSENFELD, "A Fast Parallel Algorithm For Blind Estimation of Noise Variance," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 2, pp. 216-223, 1990.

Eric P. BENNETT and Leonard McMILLAN "Video Enhancement Using Per-Pixel Virtual Exposures" SIGGRAPH 2005

Geometric Operations

Panorama

DAVIS, J. 1998. Mosaics of scenes with moving objects. In Computer Vision and Pattern Recognition (CVPR 98), 354–360.

UYTTENDAELE, M., EDEN, A., AND SZELISKI, R. 2001. Eliminating ghosting and exposure artifacts in image mosaics. In Conference on Computer Vision and Pattern Recognition (CVPR 01), 509–516.

SZELISKI, R., AND SHUM, H.-Y. 1997. Creating full view panoramic mosaics and environment maps. In Proceedings of SIGGRAPH 97, Computer Graphics Proceedings, Annual Conference Series, 251–258.

Synthetic Aperture

Marc LEVOY, Billy CHEN, Vaibhav VAISH, Mark HOROWITZ, Ian MCDOWALL, Mark BOLAS, Synthetic Aperture Confocal Imaging. ACM SIGGRAPH 2004.

REN NG, Fourier Slice Photography, SIGGRAPH 2005

A. STERN and B. JAVIDI, "3-D computational synthetic aperture integral imaging (COMPSAII)," Opt. Express 11, 2446-2451 (2003),

C. OLIVER and S. QUEGAN, Understanding Synthetic Aperture Radar Images. London: Artech House, 1998.

Deblurring and Superresolution

M. BEN-EZRA AND S. K. NAYAR , Motion Deblurring using Hybrid Imaging, In Proc. IEEE Computer Vision and Pattern Recognition (CVPR), Wisconsin, June 2003.

ZHOUCHE LIN, HEUNG-YEUNG SHUM Fundamental Limits of Reconstruction-Based Superresolution Algorithms under Local Translation PAMI, January 2004 - (Vol. 26, No. 1) pp. 83-9

O. LANDOLT, A. MITROS, AND C. KOCH, "Visual Sensor with Resolution Enhancement by Mechanical Vibrations," Proc. 2001 Conf. Advanced Research in VLSI, pp. 249-264, 2001.

Smart, Unconventional Cameras

MEMS Technology

S. K. NAYAR, V. BRANZOI, AND T. BOULT. Programmable Imaging using a Digital Micromirror Array, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Washington DC, June 2004.

High Speed Imaging

B. WANDELL, P. CATRYSSSE, J. DICARLO, D. YANG AND A. EL GAMAL Multiple Capture Single Image Architecture with a CMOS Sensor , In Proceedings of the International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives, pp. 11-17, Chiba, Japan, October 21-22 1999. (Society of Multispectral Imaging of Japan.)

S. KLEINFELDER, S.H. LIM, X.Q. LIU AND A. EL GAMAL A 10,000 Frames/s CMOS Digital Pixel Sensor, In IEEE Journal of Solid State Circuits, Vol.36, No.12, Pages 2049-2059, December 2001

X.Q. LIU AND ABBAS EL GAMAL, Synthesis of High Dynamic Range Motion Blur Free Image From Multiple Captures, In IEEE Transactions on circuits and systems (TCASI), VOL. 50, NO. 4, pp 530-539, APRIL 2003

Ramesh RASKAR, Amit AGRAWAL, Jack TUMBLIN, Coded Exposure Photography: Motion Deblurring using Fluttered Shutter, ACM SIGGRAPH 2006.

Programmable SIMD

JOHANSSON, R., LINDGREN, L., MELANDER, J., AND MOLLER, B. 2003. A multi-resolution 100 gops 4 gpixels/s programmable cmos image sensor for machine vision. In Proceedings of the 2003 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors, IEEE.

Advanced, Programmable, Demodulating Cameras and Temporal Correlation

CANESTA Inc, 2004

PIXIM Inc, 2004

FOVEON Inc, 2004

JENOPTIK Inc, 2004

IVP Inc, Ranger Camera, 2004

F. XIAO, J. DICARLO, P. CATRYSSE AND B. WANDELL, Image Analysis using Modulated Light Sources, In Proceedings of the SPIE Electronic Imaging '2001 conference, Vol. 4306, San Jose, CA, January 2001.

ANDO, S., K. NAKAMURA, AND T. SAKAGUCHI. Ultrafast Correlation Image Sensor: Concept, Design, and Applications,. in Proc. IEEE CCD/AIS Workshop. 1997. Bruges, Belgium: IEEE.

ANDO, S. AND A. KIMACHI. Time-Domain Correlation Image Sensor: First CMOS Realization of Demodulator Pixels Array. in Proc. '99 IEEE CCD/AIS Workshop. 1999.

Computational Photography

SIGGRAPH 2006 Course 15 Notes

Organizers

Ramesh Raskar

Senior Research Scientist

MERL - Mitsubishi Electric Research Labs

201 Broadway

Cambridge, MA 02139

Email: raskar@merl.com

<http://www.merl.com/people/raskar/>

Jack Tumblin

Assistant Professor

Computer Science Dept., Northwestern University

1890 Maple Avenue, Evanston IL 60201,

Email: jet@cs.northwestern.edu

<http://www.cs.northwestern.edu/~jet>

Course Web Page

<http://www.merl.com/people/raskar/photo/>