Image-Based Modeling and Rendering

Richard Szeliski

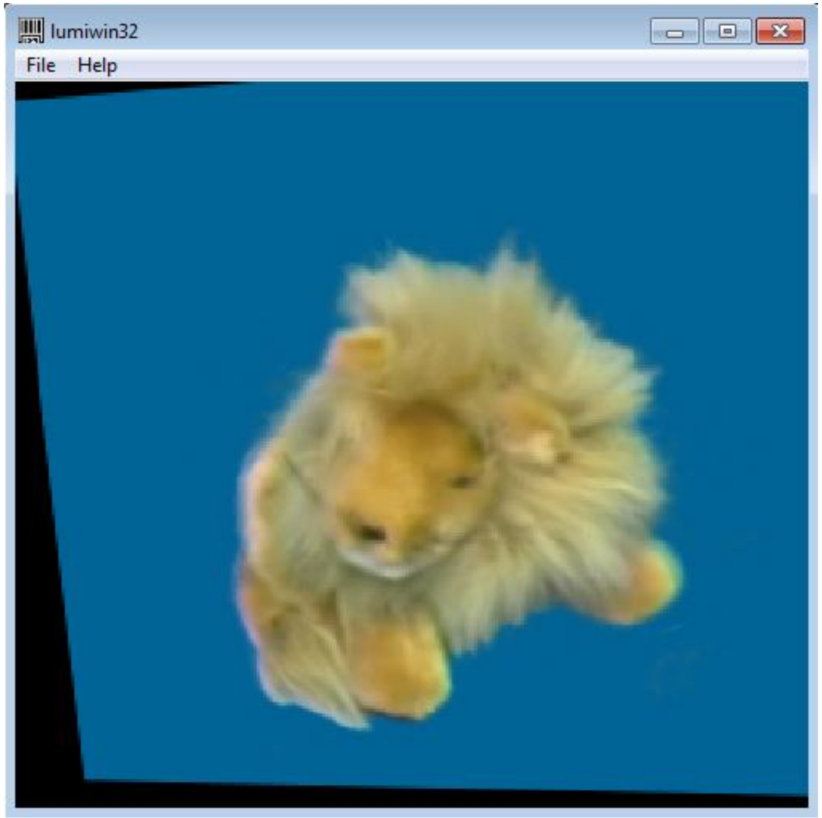
Microsoft Research

IPAM Graduate Summer School: Computer Vision

July 26, 2013

How far have we come?

Light Fields / Lumigraph - 1996



Environment matting - 2001

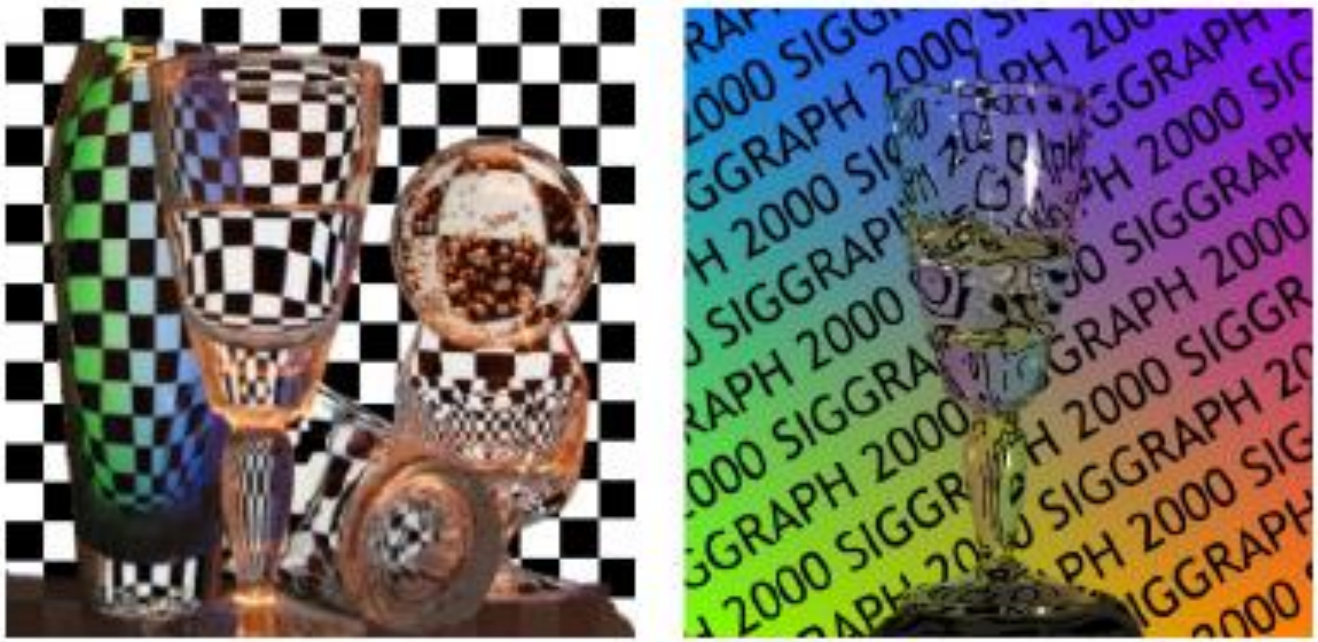


Figure 1 Sample composite images constructed with the techniques of this paper: slow but accurate on the left, and a more restricted example acquired at video rates on the right.

Photo Tourism - 2006

Ambient Point Clouds - 2010

Photo Tours - 2012

[Kushal *et al.*, 3DIMPVT 2012]

The Visual Turing Test

Figure 5. Visual Turing test. In each image pair, the ground truth image is on the left and our result is on the right.

[Shan *et al.*, 3DV 2013]

Where are we going?

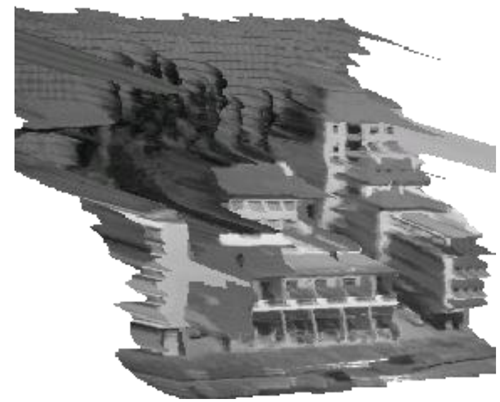
A personal retrospective

- Panoramas and 360° Video Walkthroughs
- Light Fields and Lumigraphs
- LDIs, Sprites, and Layered Video
- Image-Based Modeling
- Environment Mattes and Matting
- Photo Tourism and Photosynth
- Point-Based Rendering and NPR
- Reflections and Transparency [2012]

Pre-history (?): view interpolation

- Depth maps and images: view extrapolation [Matthies,Szeliski,Kanade'89]





input *depth image* *novel view*

- View interpolation: warp and dissolve [Beier & Neely '92; Chen & Williams '93]

Panoramic image stitching

Convert image sequence into a cylindrical image

+

+ … +

=

Panoramic image stitching

Key technical breakthroughs:

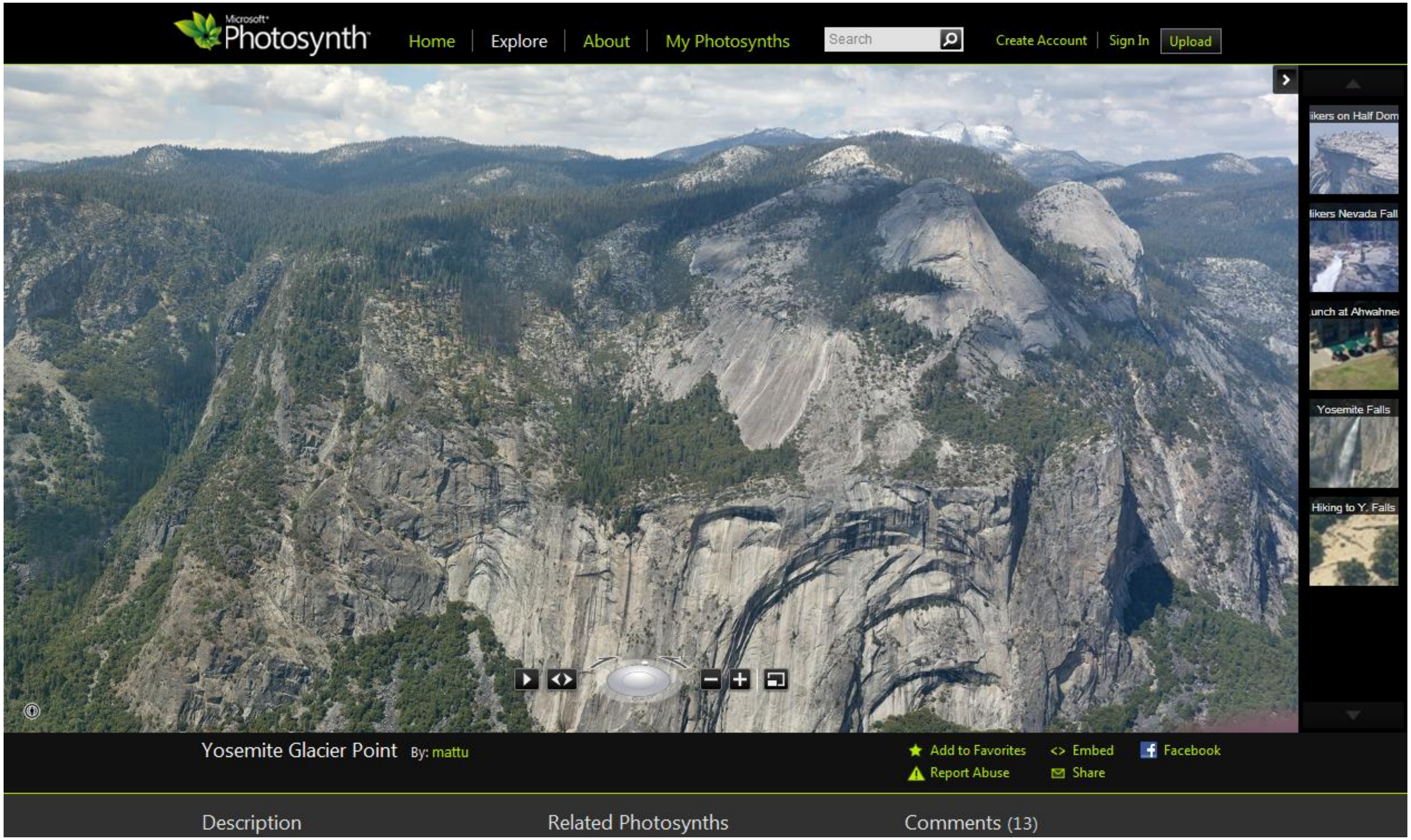
- ✓ interactive viewing [(Lippman'80), Chen '95]
- ✓ general camera motions [1997…]
- ✓ parallax compensation [1998…]
- ✓ fully automated feature detection and matching [Brown & Lowe '03]
- ✓ seam selection [1975 (1998), 2001…]
- ✓ Laplacian and Poisson blending [1984, 2003]

Early example of <u>*computational photography*</u>

“VR” today: Gigapan

“VR” today: 360 cities

“VR” today: Photosynth



Moving beyond single (or linked) panoramas…

Panoramic Video-Based Tours

[Uyttendaele *et al.* 2004]

2D $\rightarrow$ “3D Light Field” (?)

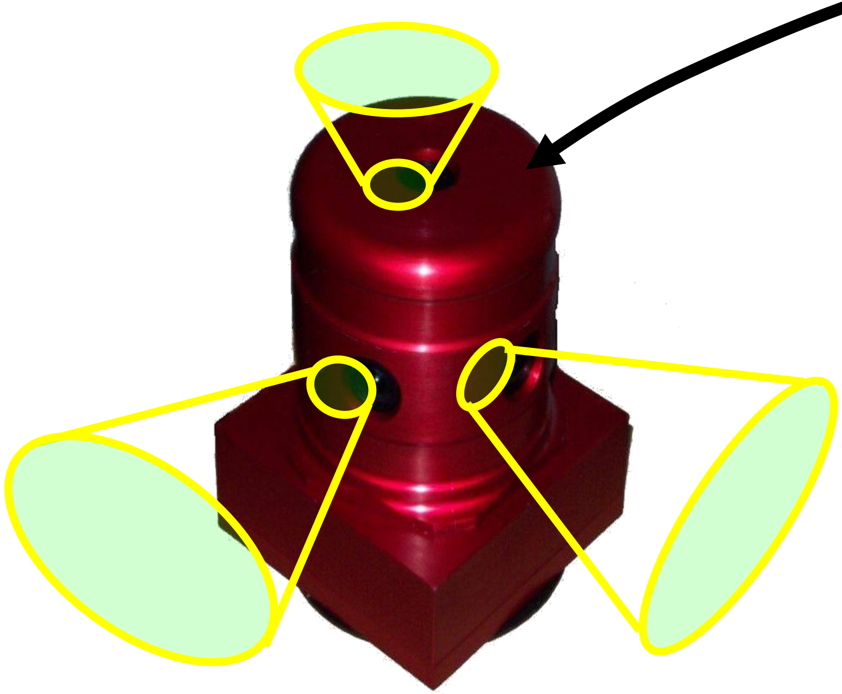
Video-Based Walkthroughs

Move camera along a rail ("dolly track") and play back a 360° video

Applications:

- Homes and architecture
- Outdoor locations (tourist destinations)

Surround video acquisition system



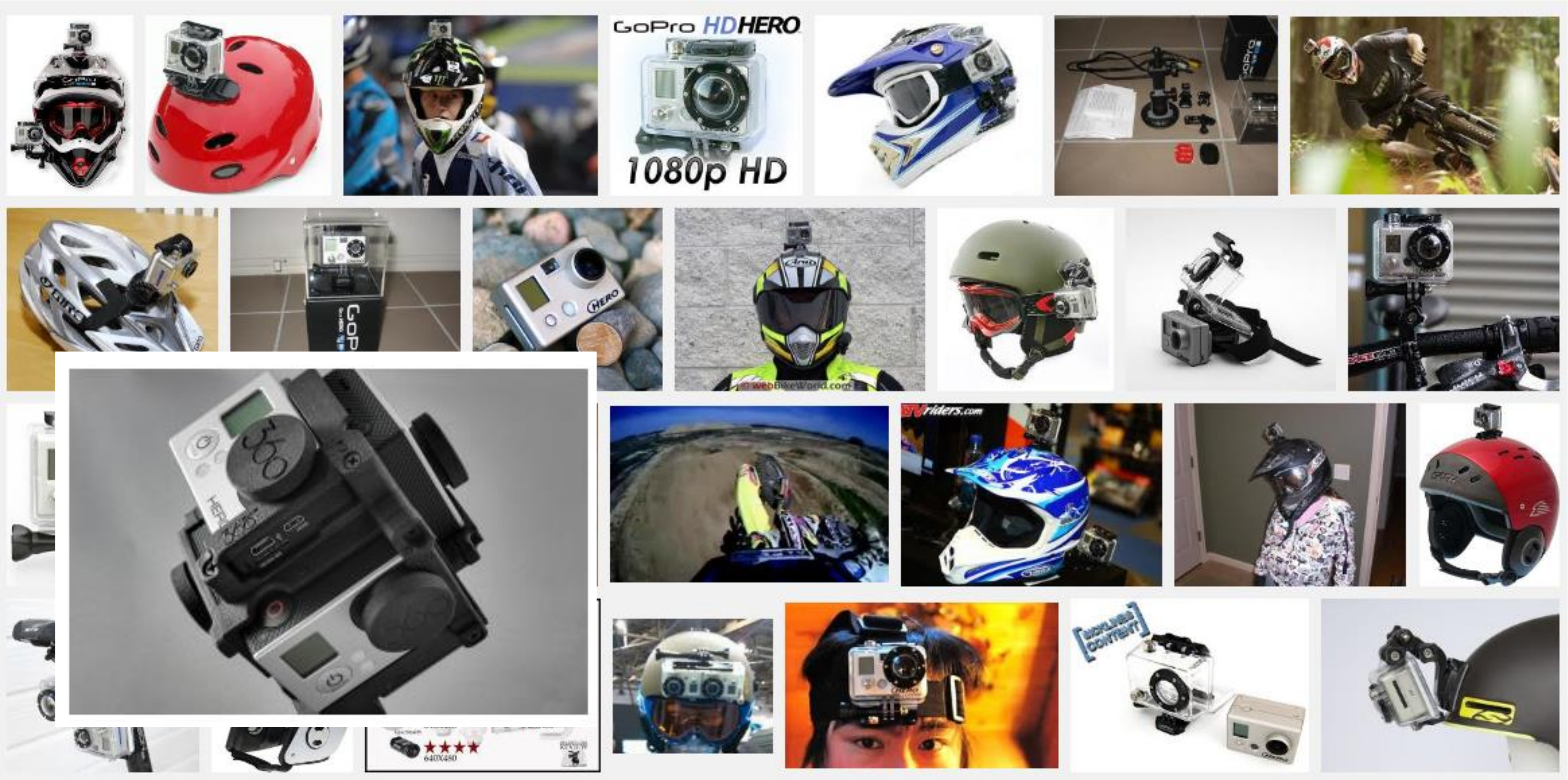
OmniCam (six-camera head)
[Point Grey Ladybug]

Acquisition platforms (then)

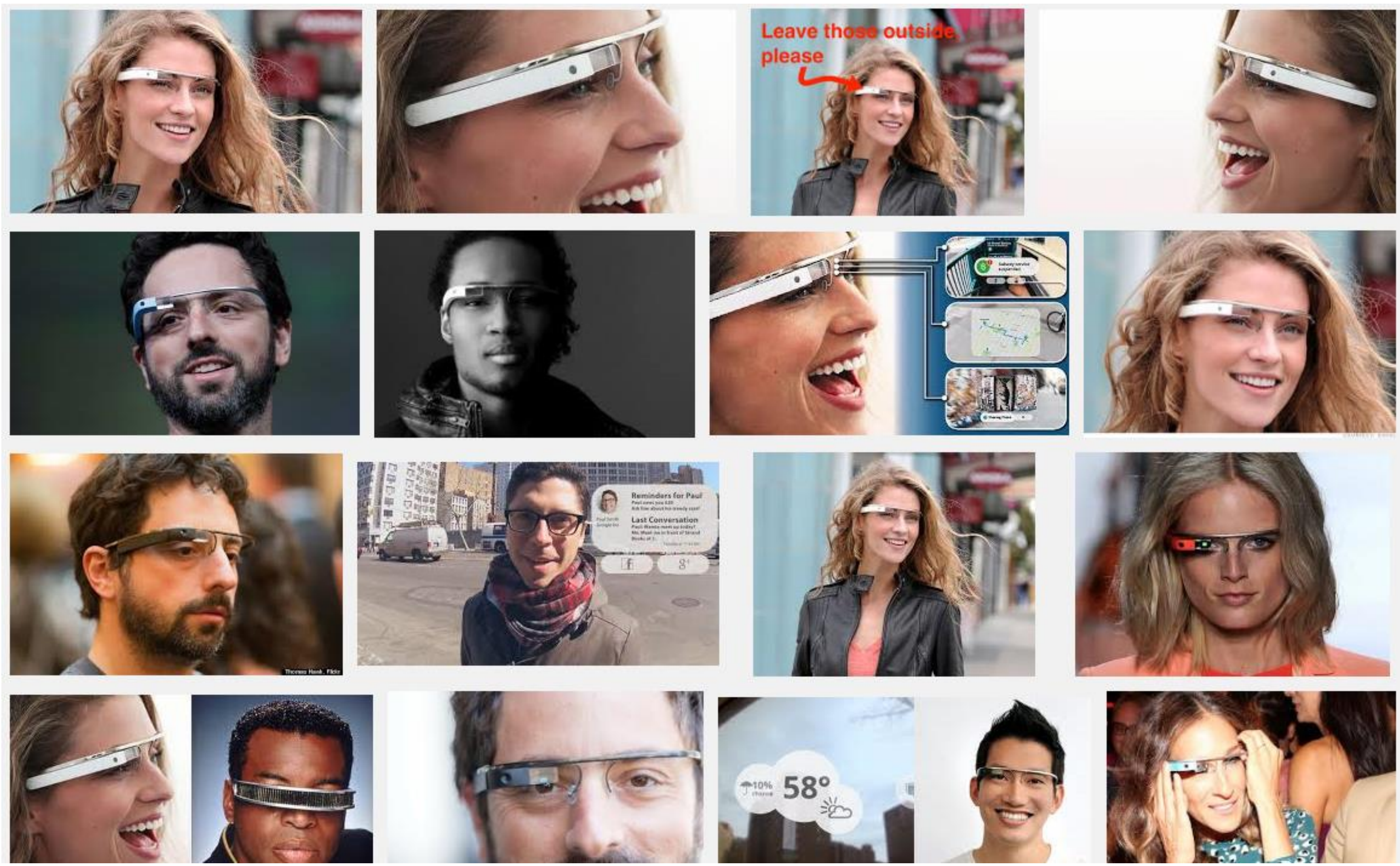
Robotic cart

Wearable

Acquisition platforms (today)

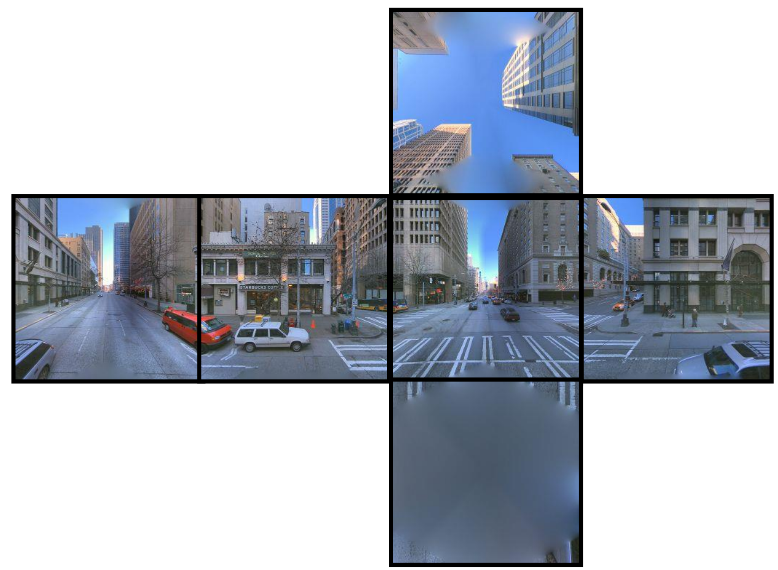


Acquisition platforms (future?)



Acquisition platforms (now)

- 10 camera panoramic rig
- One bubble every 2 meters
- Reproject into cube maps



Lightfields and Lumigraphs

"True 4D"

(with lots of slides from Michael Cohen)

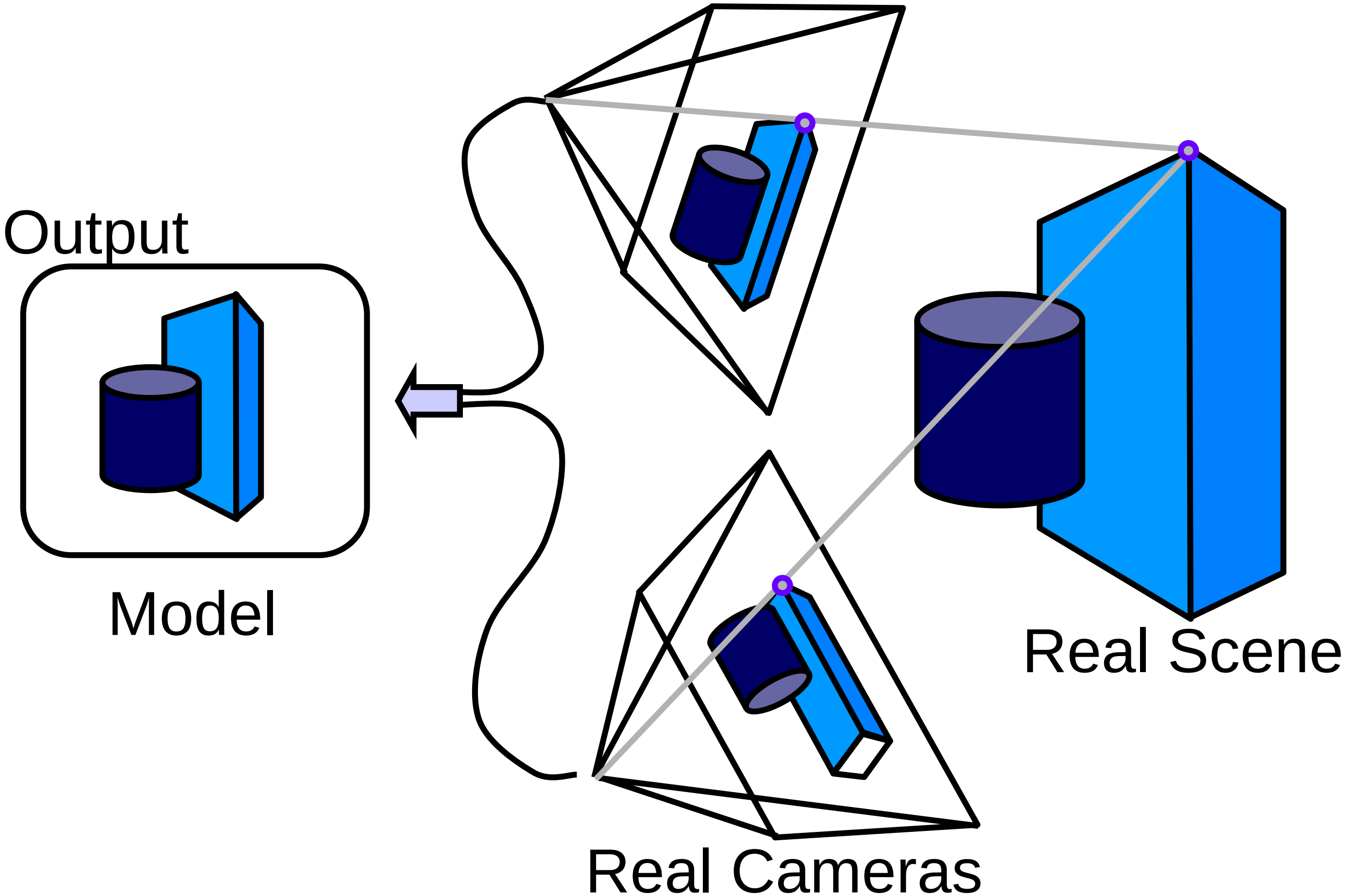
Modeling light

How do we generate new scenes and animations from existing ones?

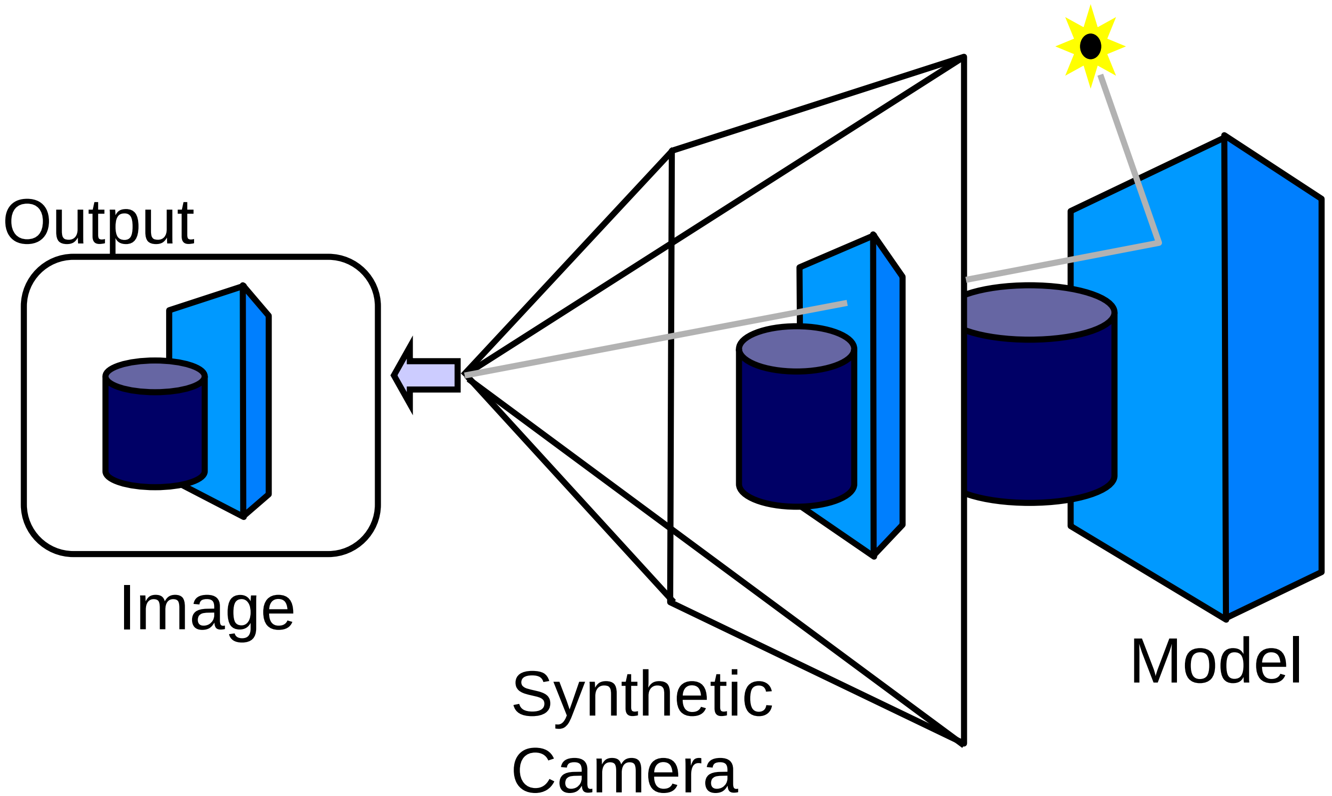
Classic “3D Vision + Graphics”:

- take (lots of) pictures
- recover camera pose
- build 3D model
- extract texture maps / BRDFs
- synthesize new views
- … tons of great work and demos as 3DIMPVT

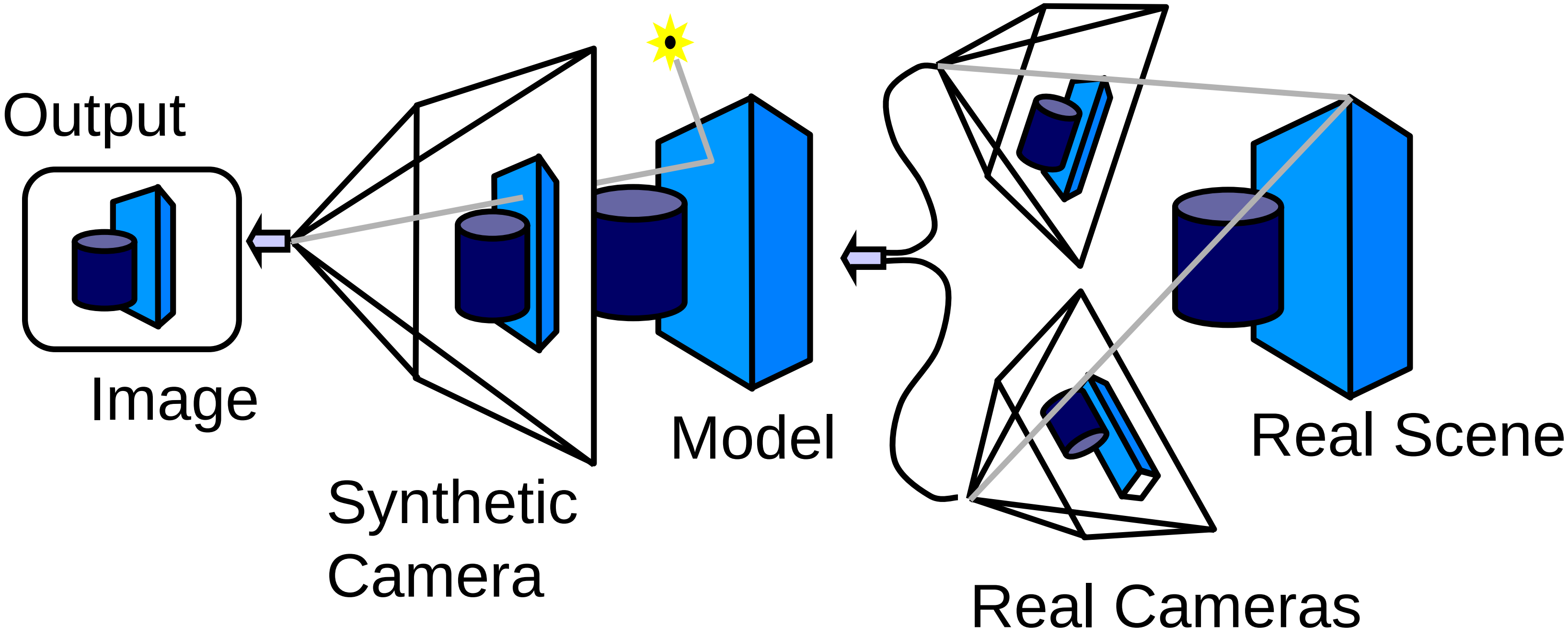
Computer Vision



Computer Graphics

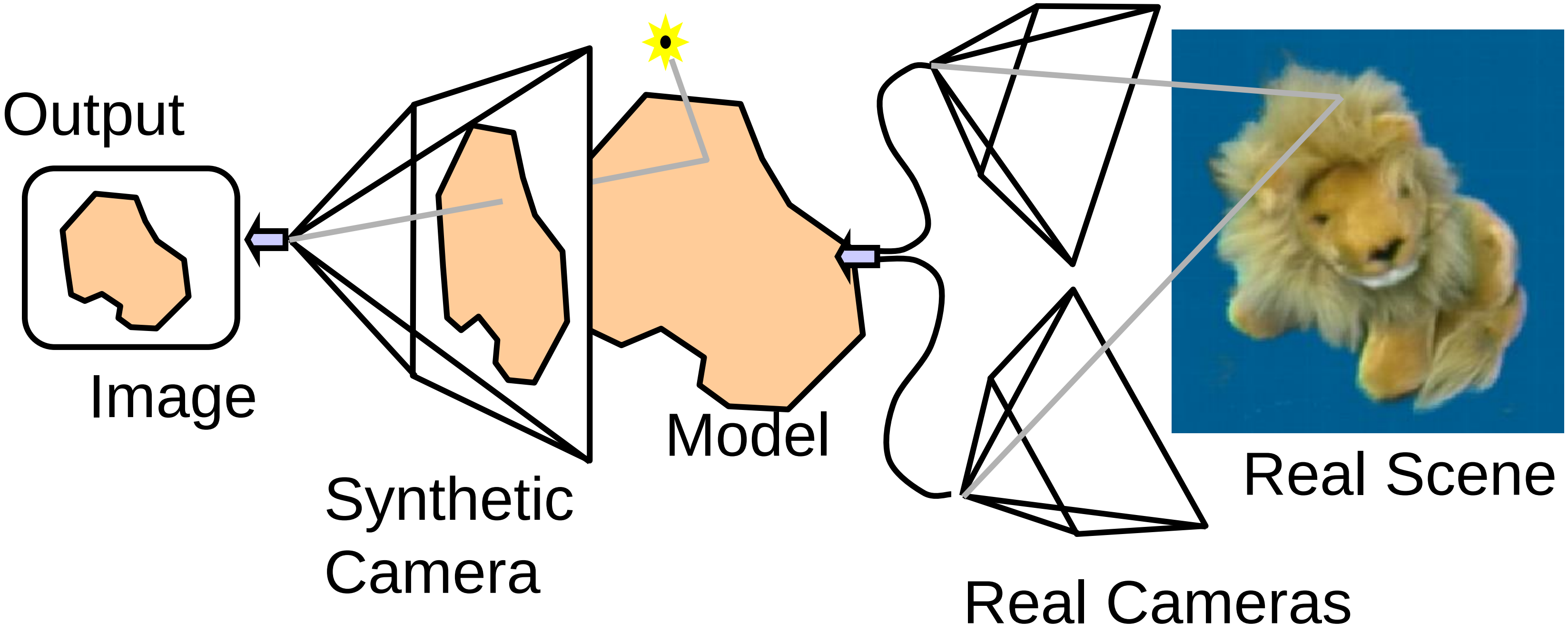


Combined



Output
Image
Synthetic
Camera
Model
Real Scene
Real Cameras

But, vision technology falls short



… and so does graphics.

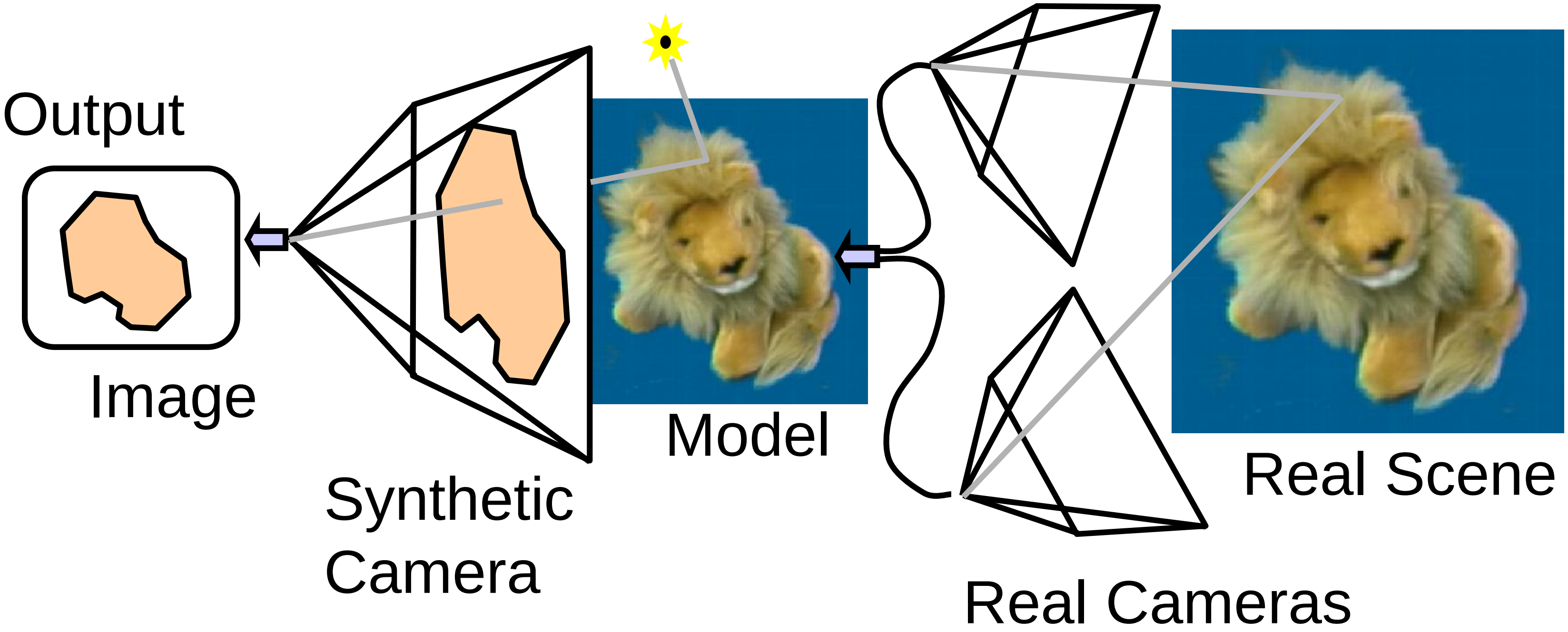
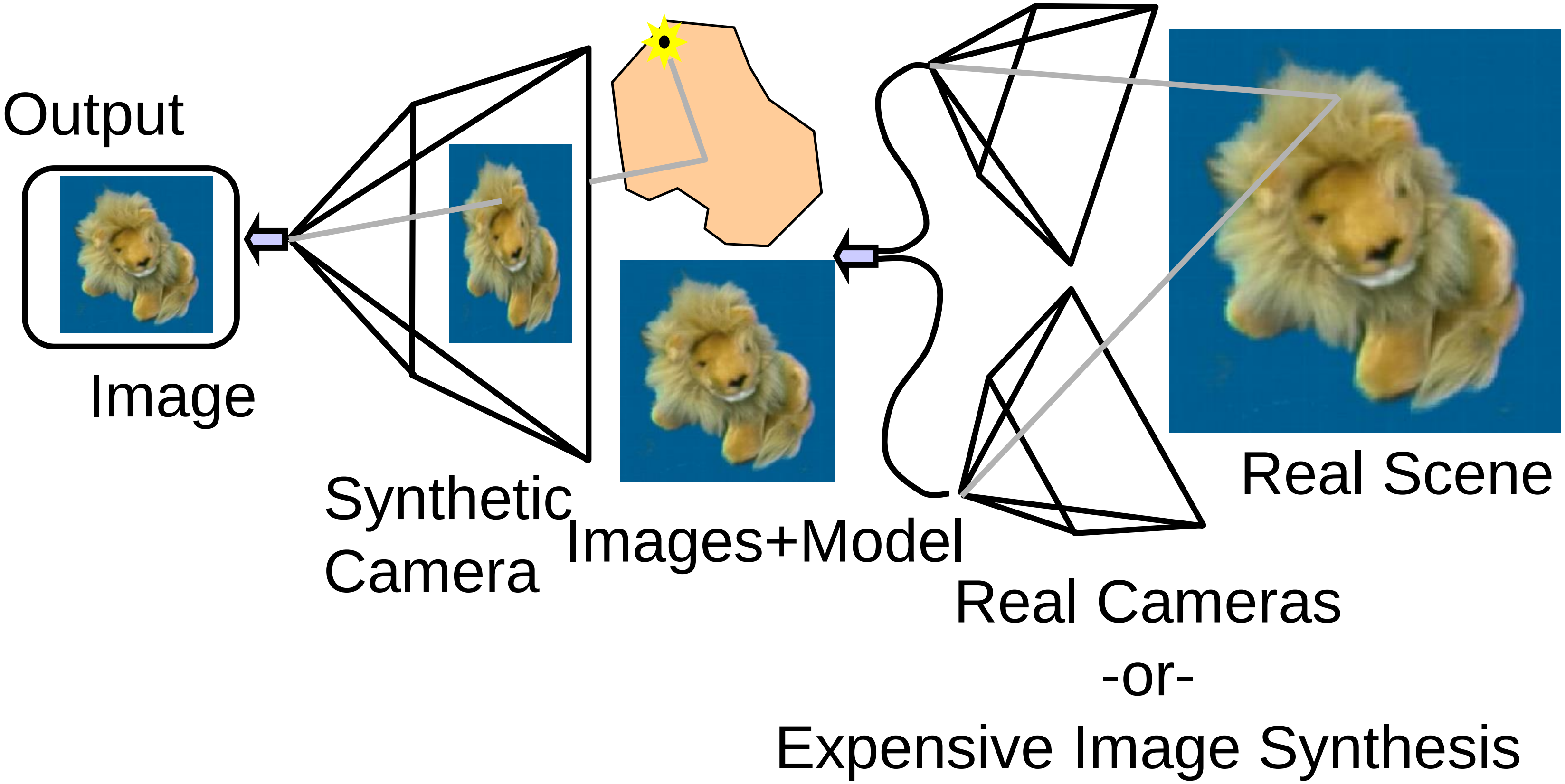
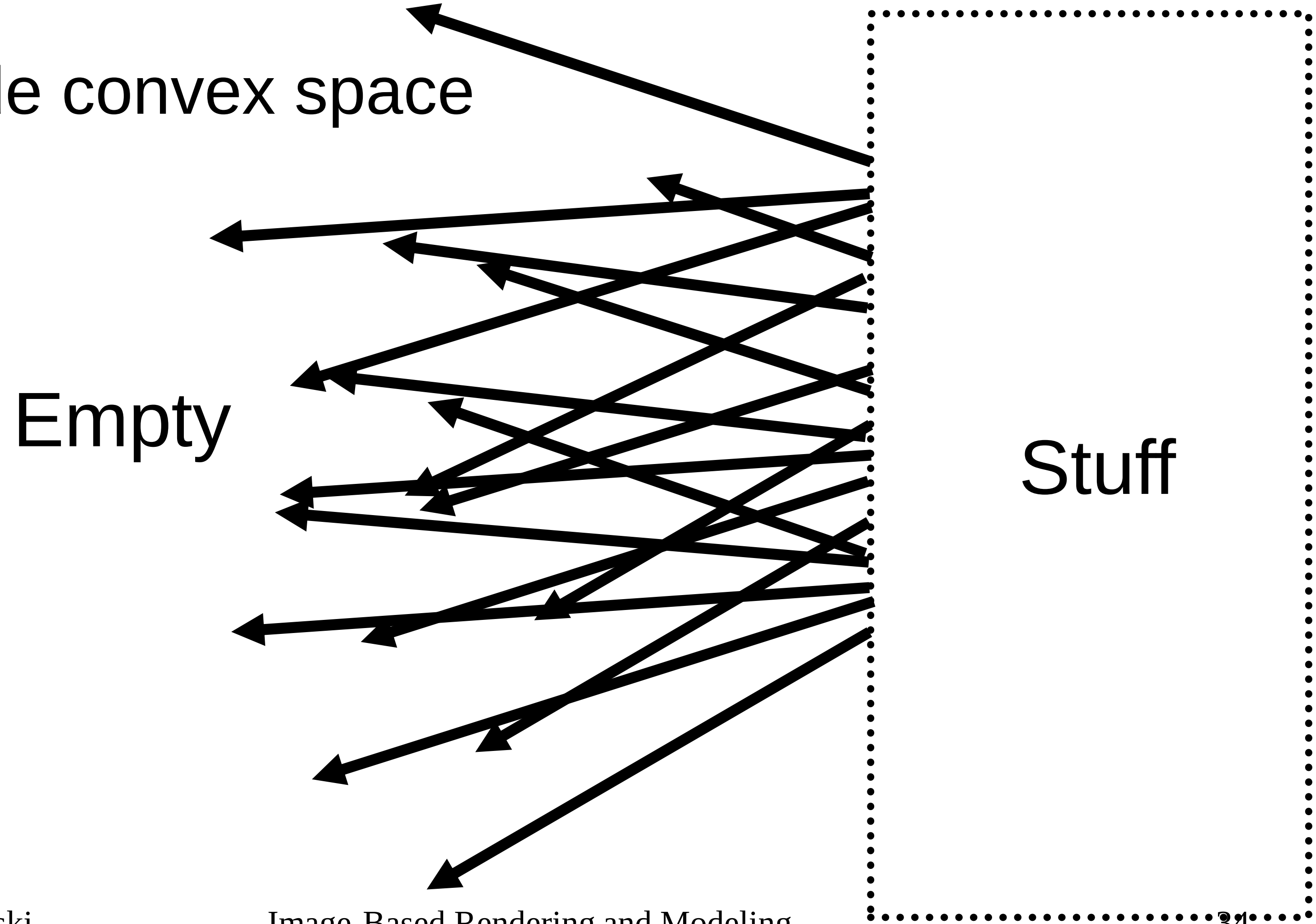


Image Based Rendering



Lumigraph / Lightfield

Outside convex space

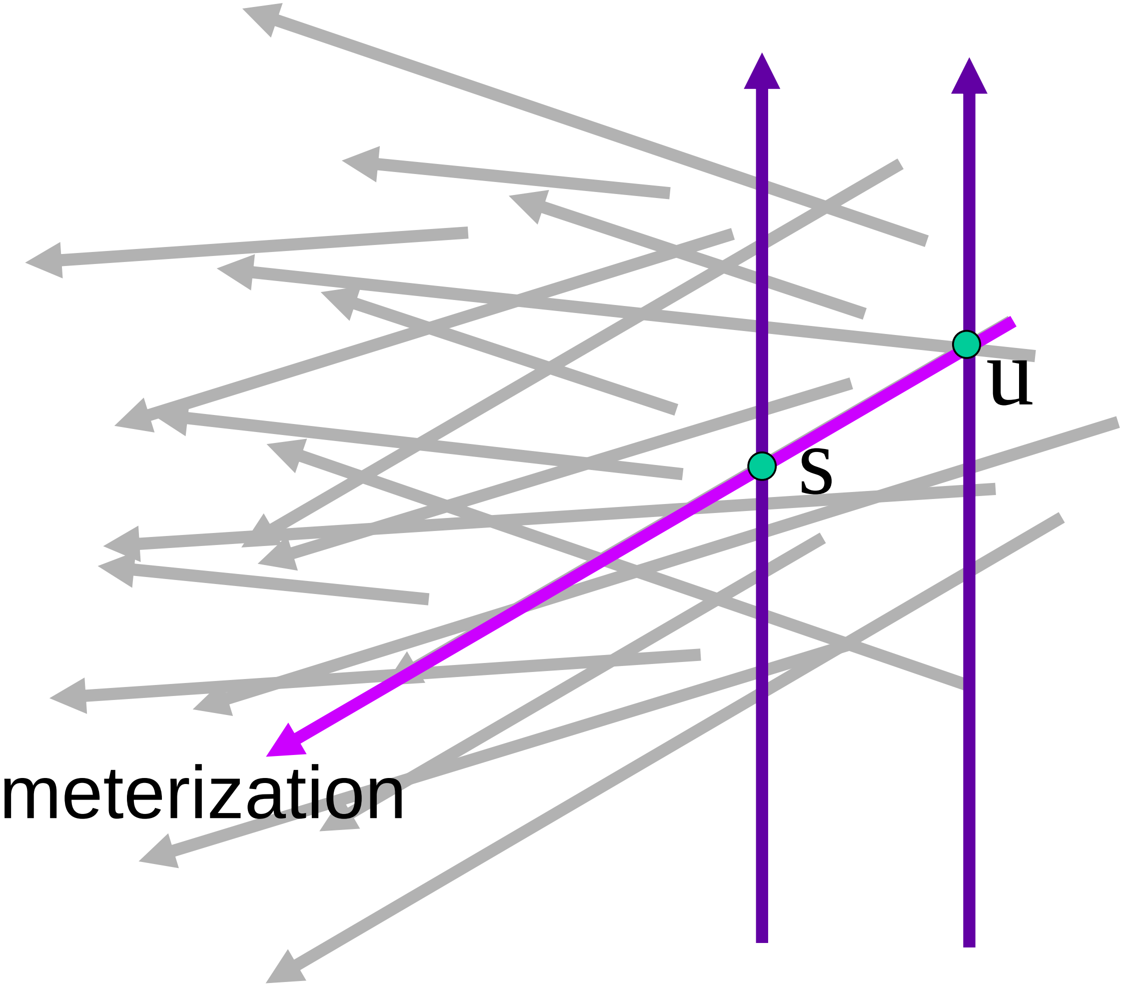


4D

Lumigraph - Organization

2D position

2D position

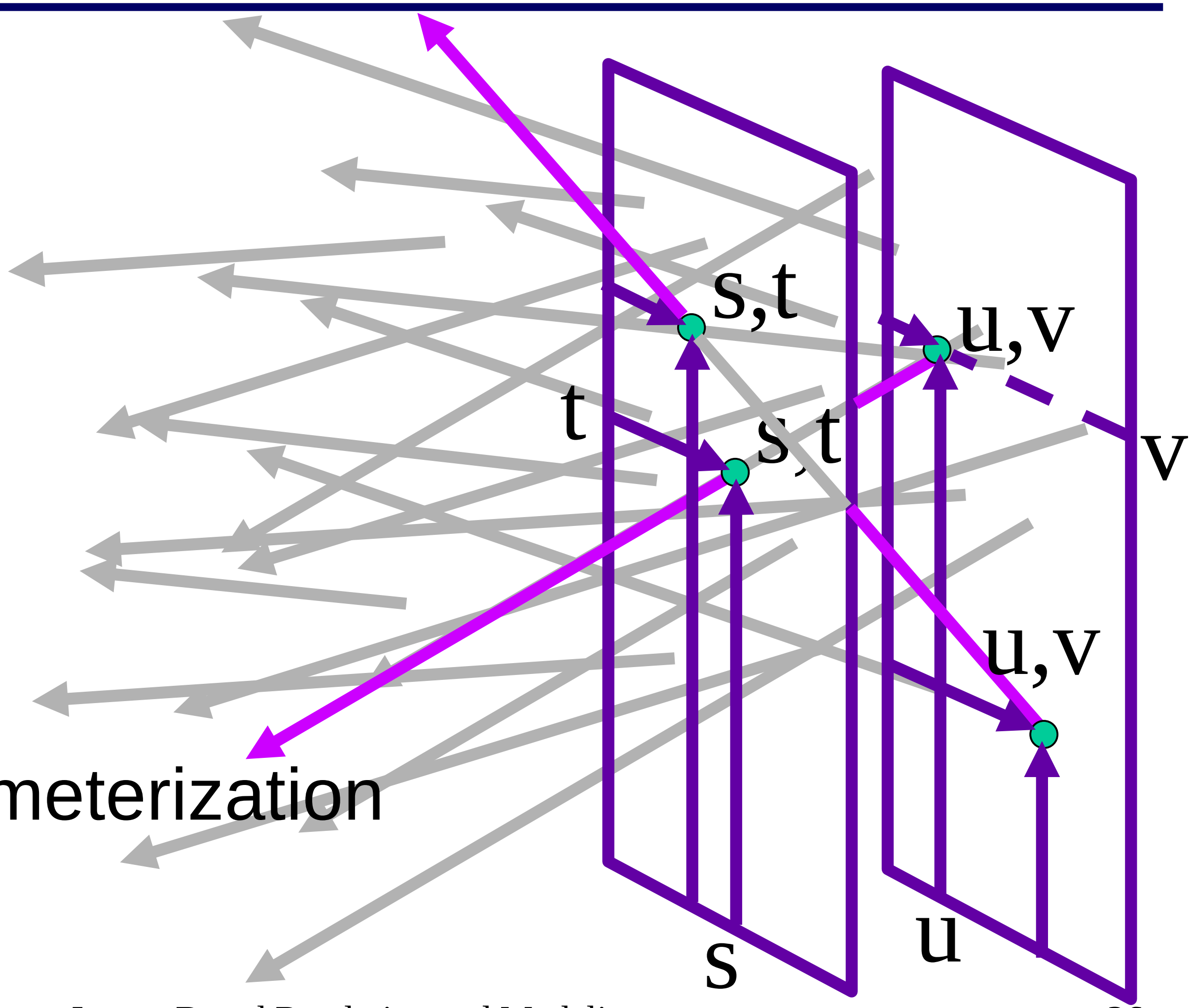


2 plane parameterization

Lumigraph - Organization

2D position

2D position

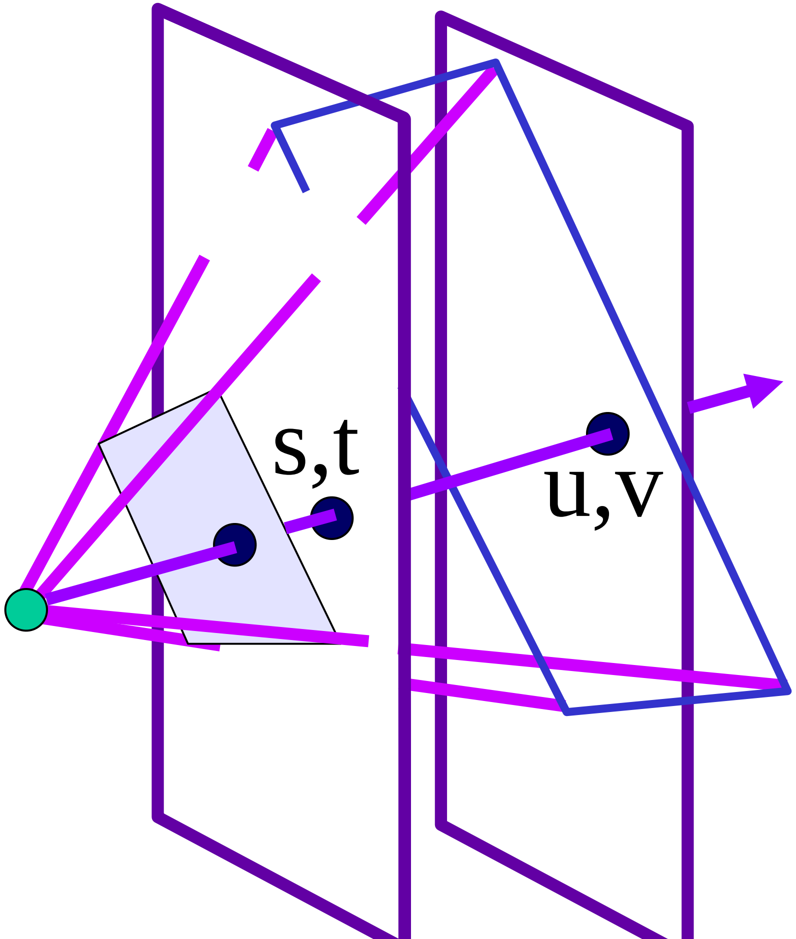


2 plane parameterization

Lumigraph - Rendering

For each output pixel

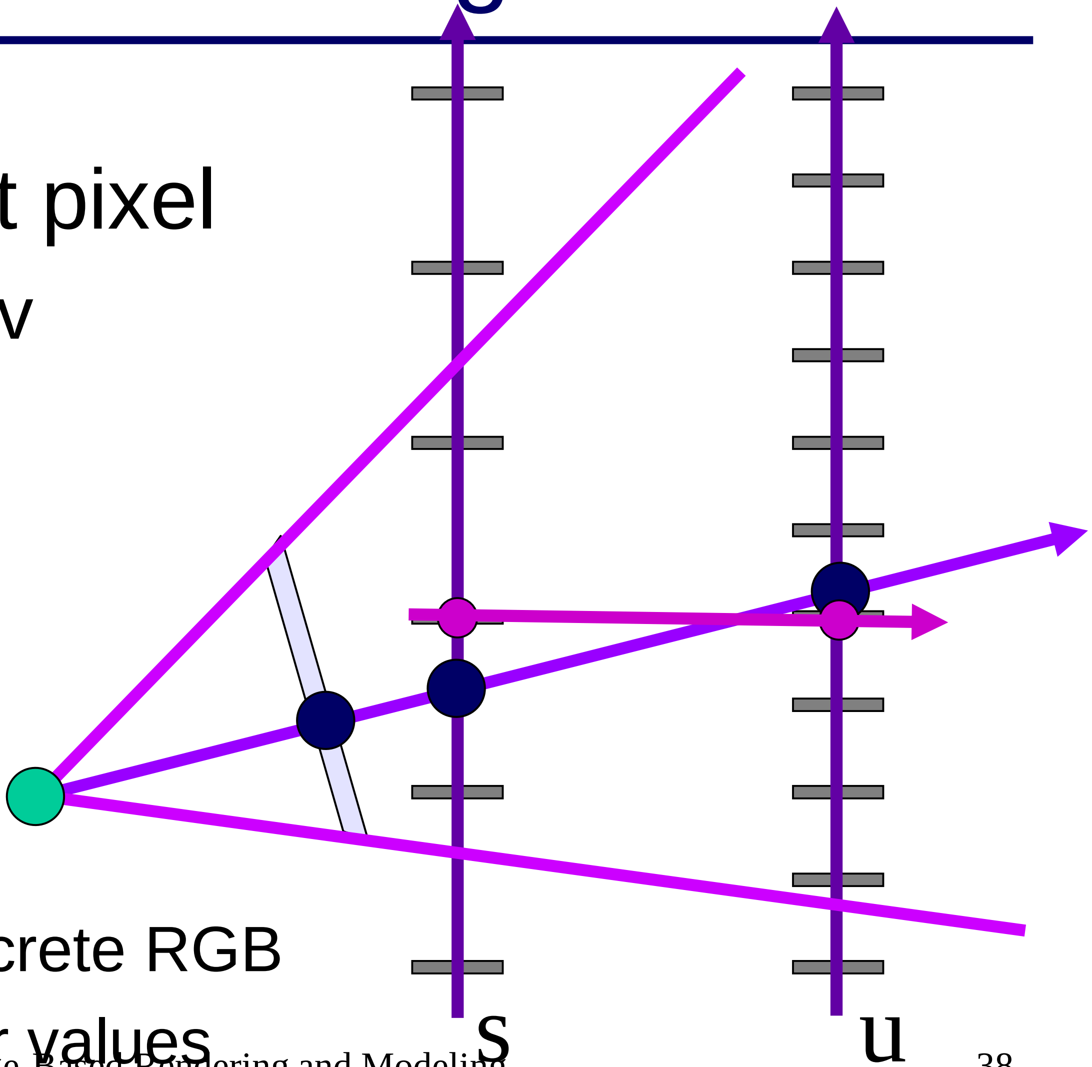
- determine s,t,u,v
- either
  - find closest discrete RGB
  - interpolate near values



Lumigraph - Rendering

For each output pixel

- determine s,t,u,v



- either
  - use closest discrete RGB
  - interpolate near values

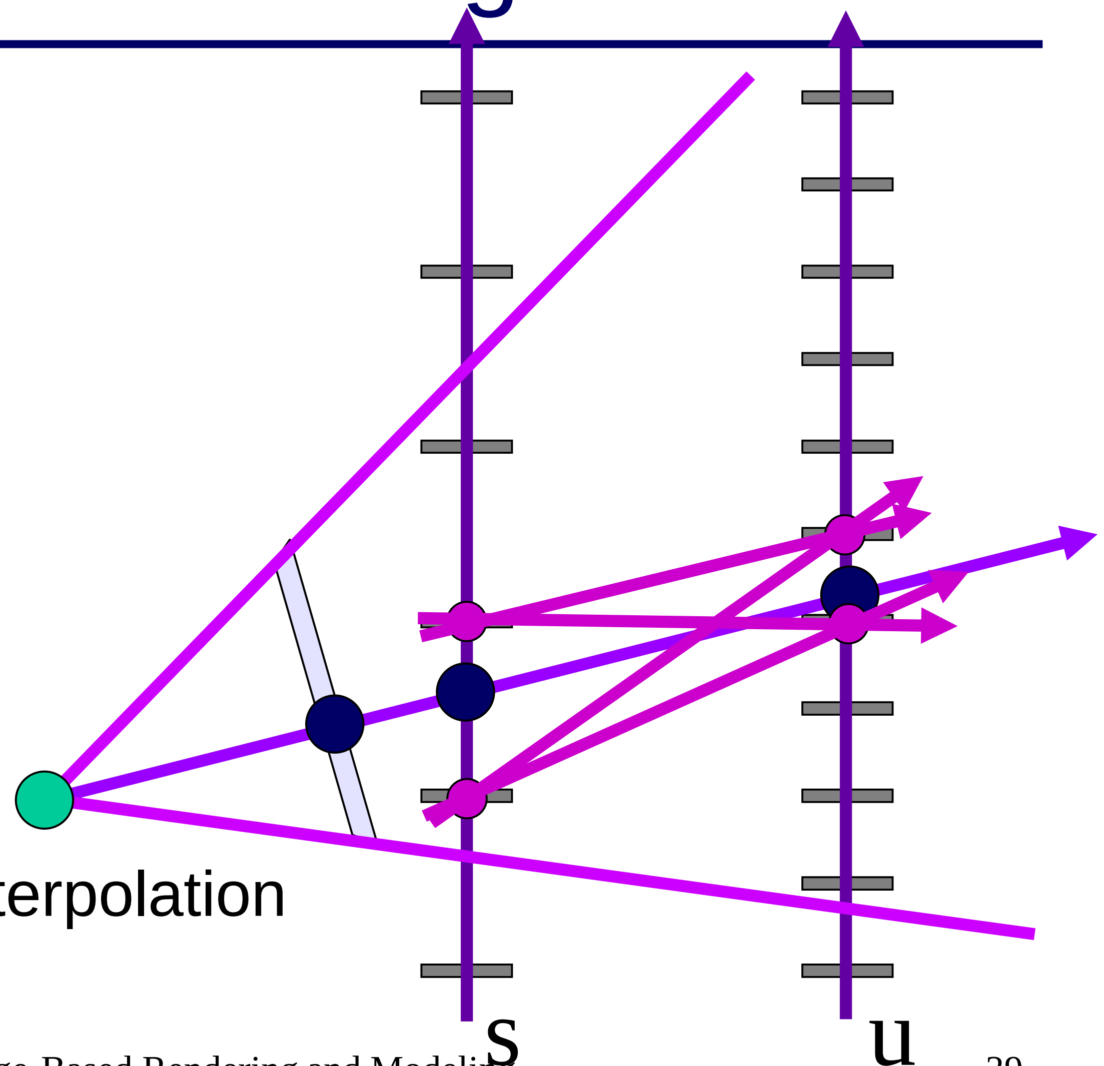
Lumigraph - Rendering

Nearest

- closest s
- closest u
- draw it

Blend 16 nearest

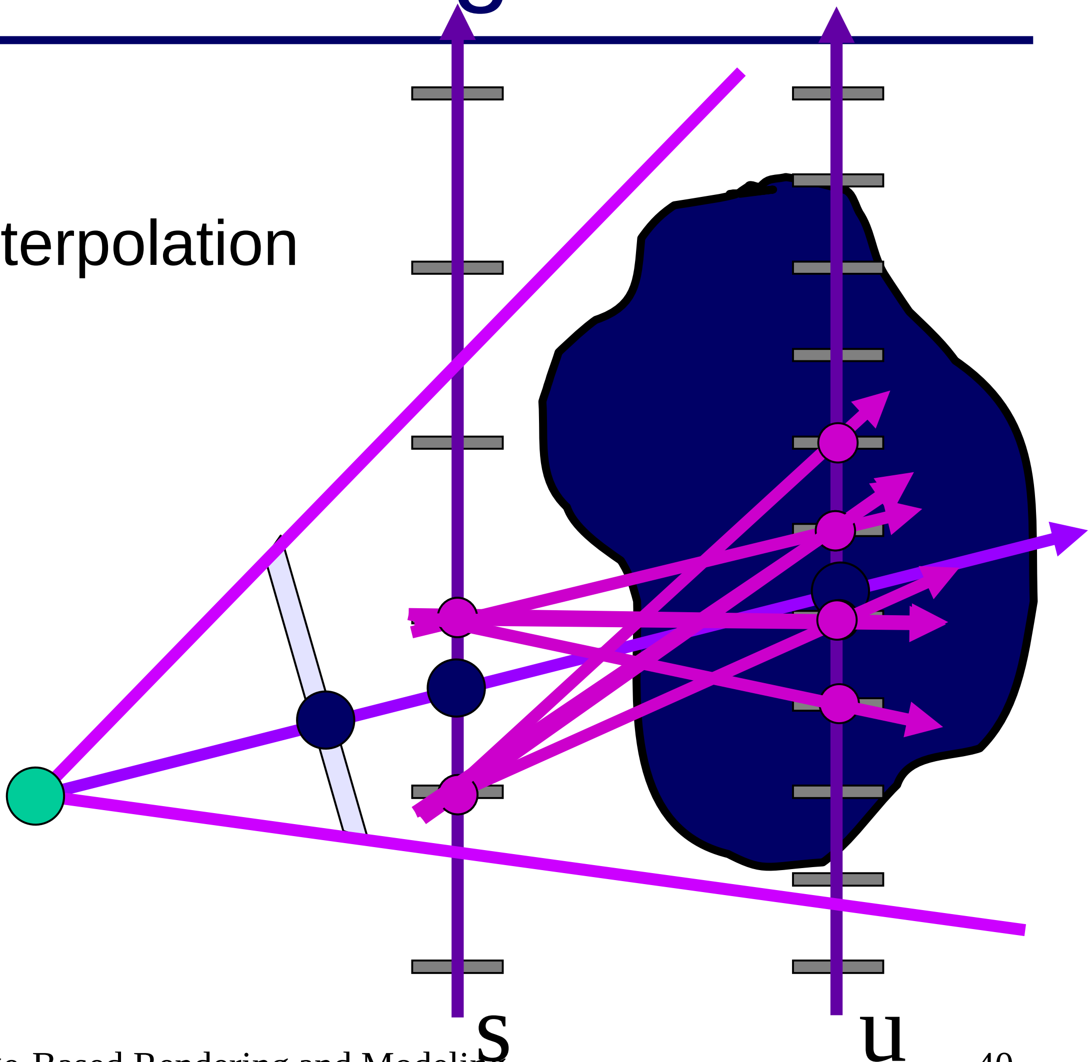
- quadrilinear interpolation



Lumigraph - Rendering

Depth Correction

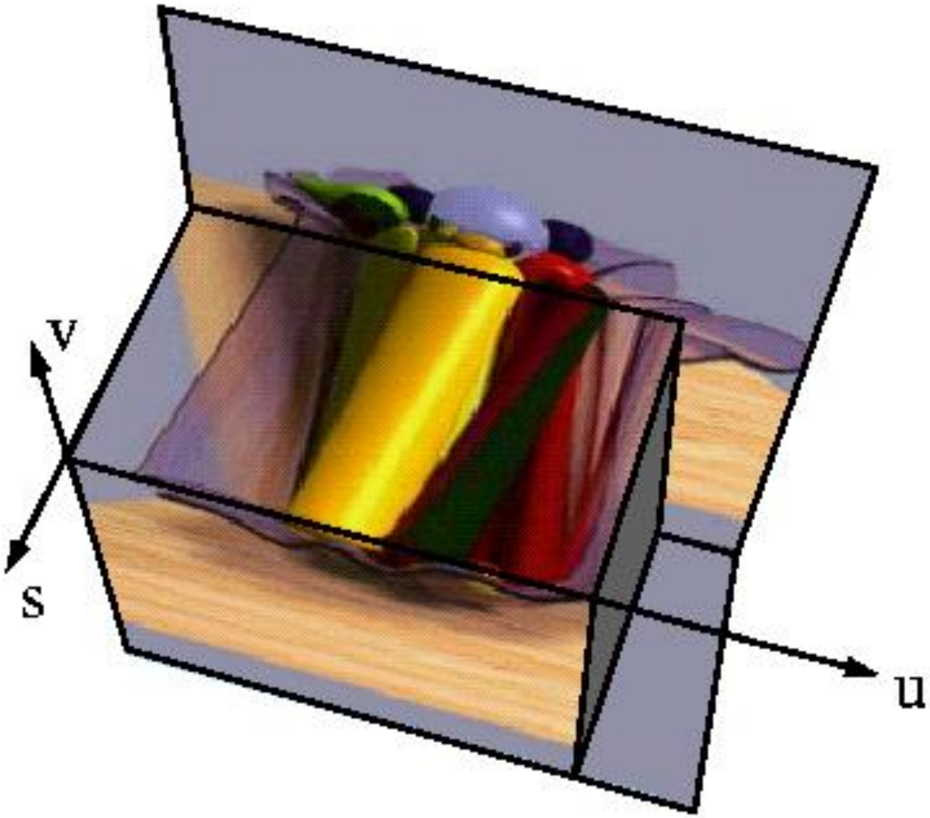
- quadralinear interpolation
- new “closest”
- like focus



Lumigraph – Image Effects

Image effects:

- parallax
- occlusion
- transparency
- highlights

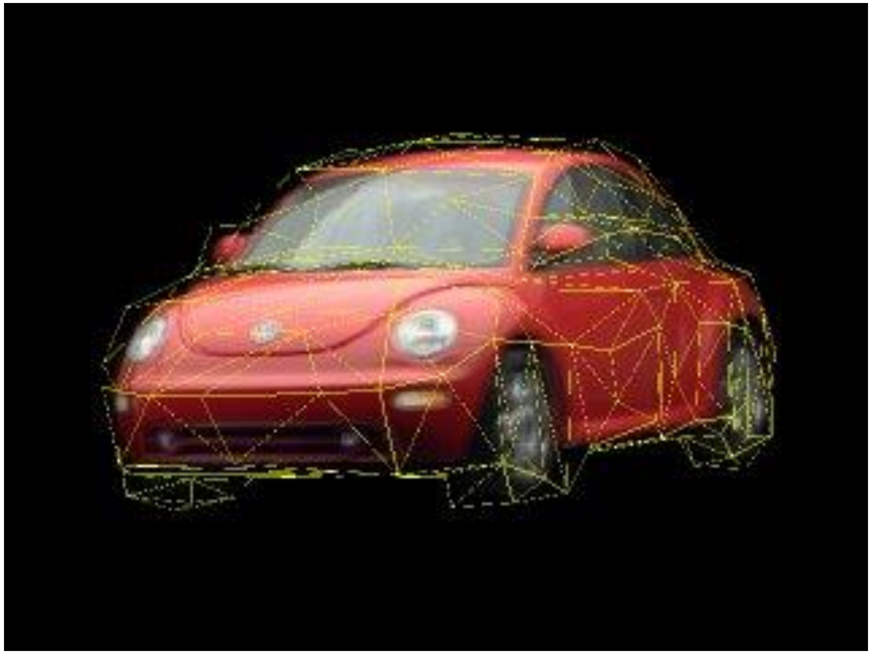


Unstructured Lumigraph

What if the images aren't sampled on a regular 2D grid?

- can still re-sample rays
- ray weighting becomes more complex

[Buehler *et al.*, SIGGRAPH'2000]



Surface Light Fields

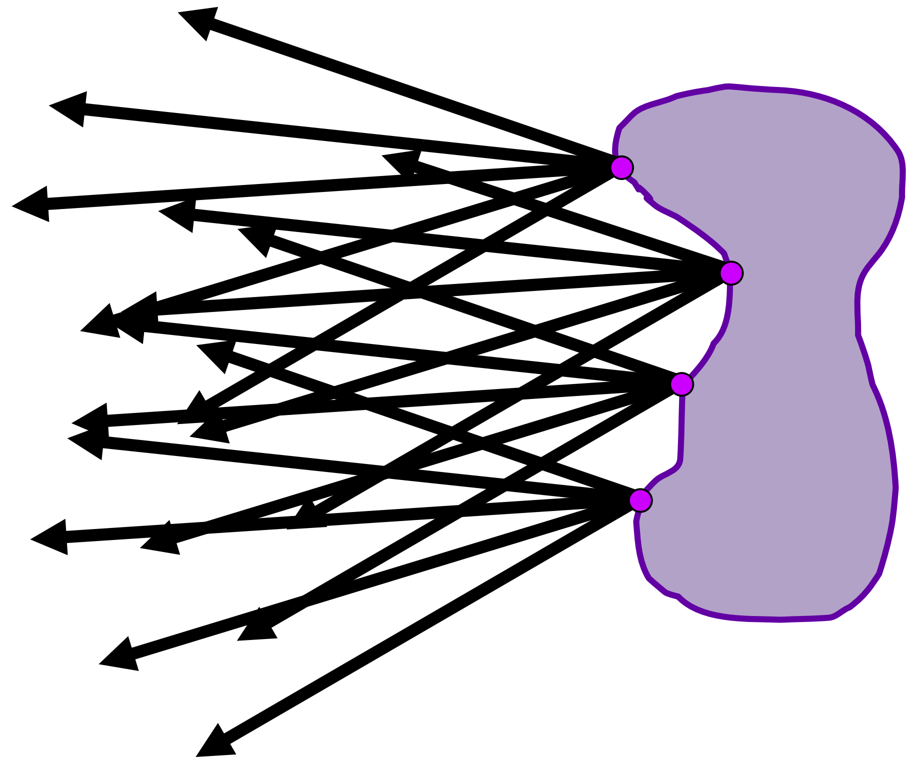
[Wood et al, SIGGRAPH 2000]

Turn 4D parameterization around:

- image @ every surface pt.

Leverage coherence:

- compress radiance fn (BRDF * illumination) after rotation by ***n***



Surface Light Fields

[Wood et al, SIGGRAPH 2000]

2.5D Representations

Images (and panoramas) are 2D

Lumigraph is 4D

3D

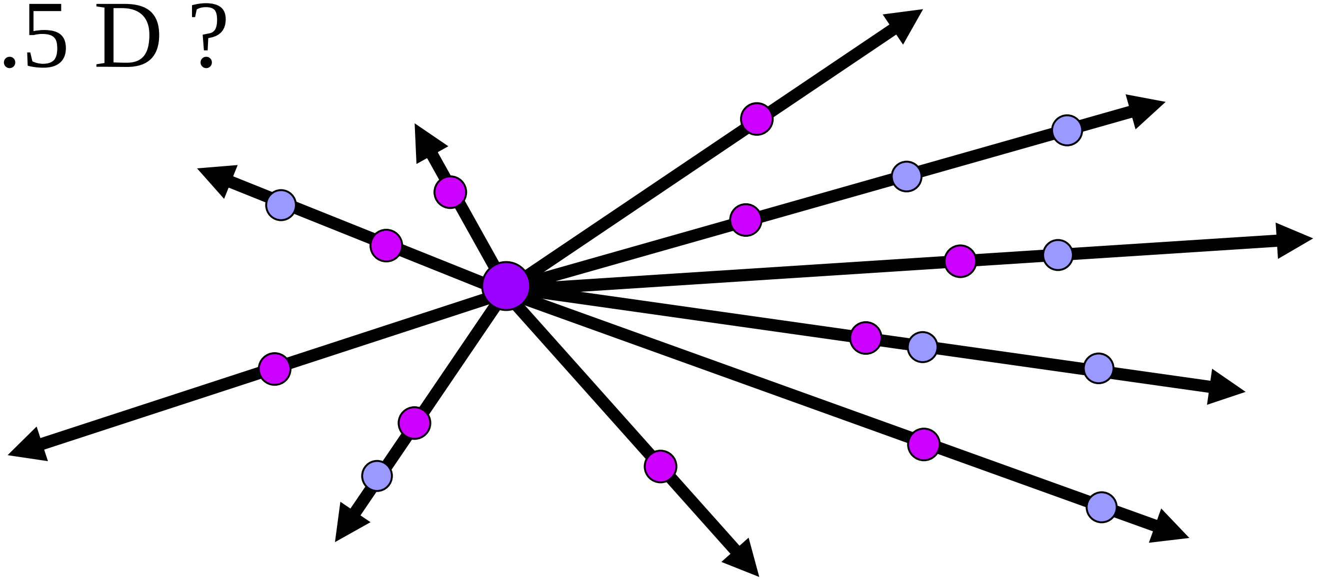
- 3D Lumigraph subsets (360 video tours)
- Concentric mosaics

2.5D

- Layered Depth Images
- Sprites with Depth (impostors)
- Layered (Virtual Viewpoint) Video
- View Dependent Surfaces (see Façade)

Layered Depth Image

2.5 D ?

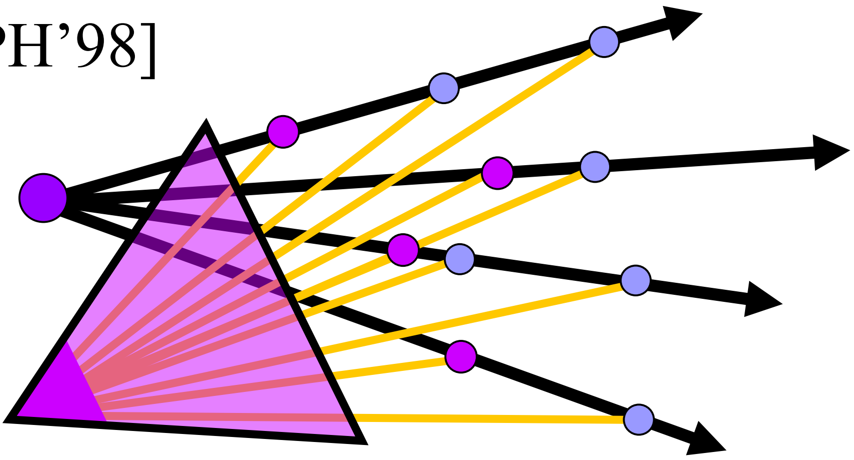


Layered Depth Image

Layered Depth Image

- Rendering from LDI

  [Shade *et al.*, SIGGRAPH'98]



- Incremental in LDI X and Y
- Guaranteed to be in back-to-front order

Sprites with Depth

Represent scene as collection of cutouts with depth (planes + parallax)

Render back to front with fwd/inverse warping [Shade *et al.*, SIGGRAPH'98]

Basis of Virtual Viewpoint Video [Zitnick *et al.* 2004]

Virtual Viewpoint (3D) Video

[Zitnick *et al*., SIGGRAPH 2004]

Scenarios for 3D video

Sports events, e.g., CMU's 30-camera "EyeVision" system at SuperBowl XXXV)

Concert performances, plays, circus acts

Games

Instructional video, e.g., golf, skating, martial arts

Interactive (Internet) video

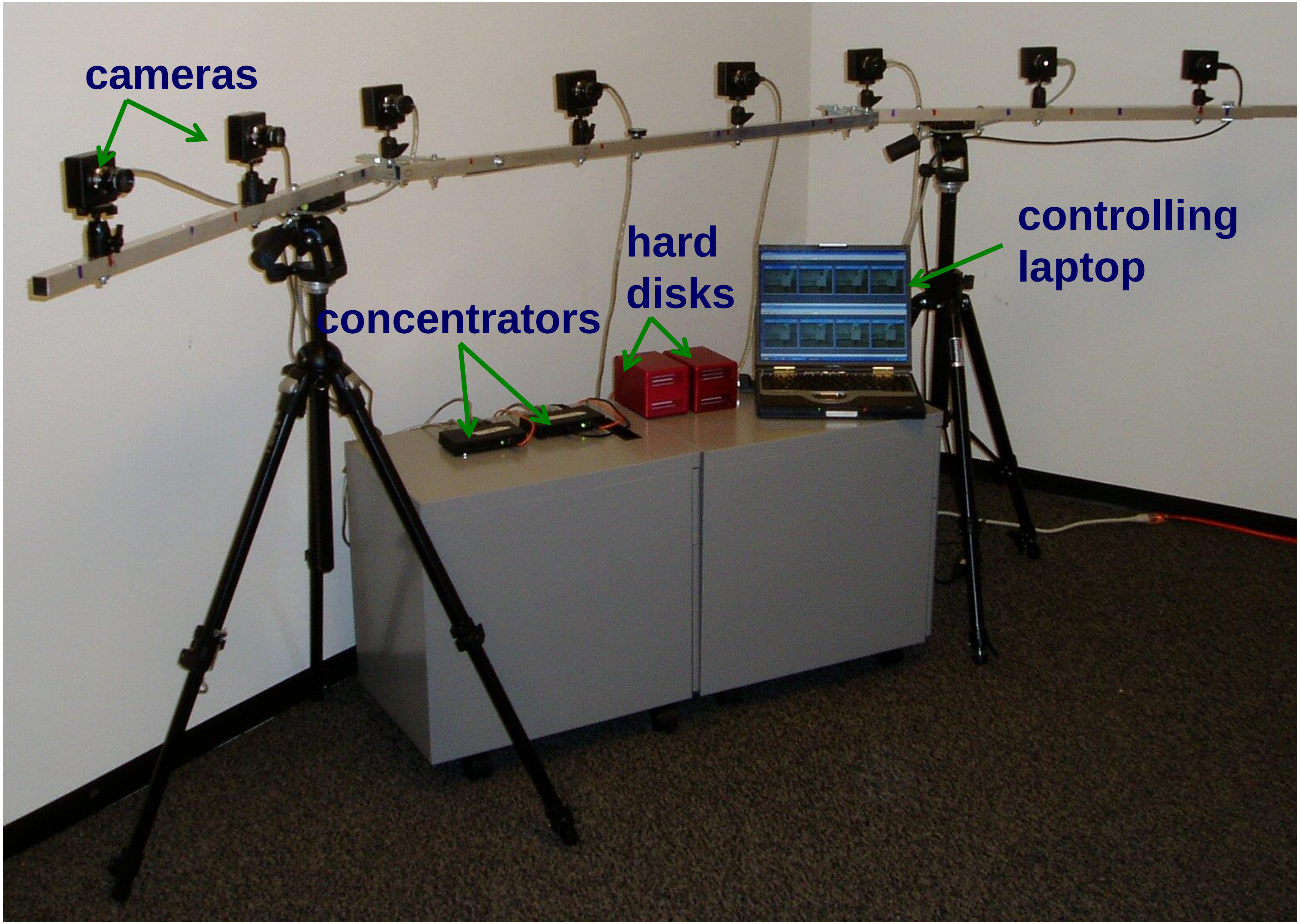
3D video: Challenges

Capture: record multiple synchronized camera streams

Processing and representation: ensure <u>*photorealism*</u>

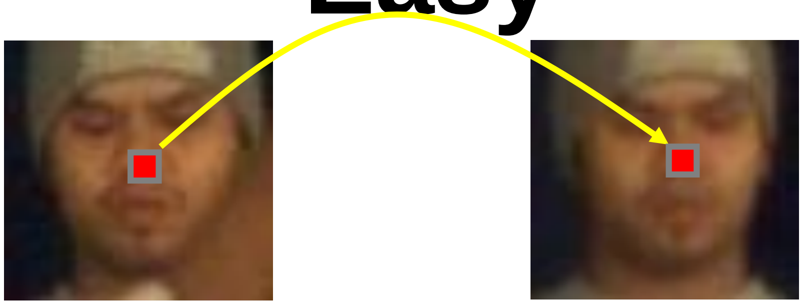
Compression: exploit redundancy

Rendering: provide interactive viewpoint control

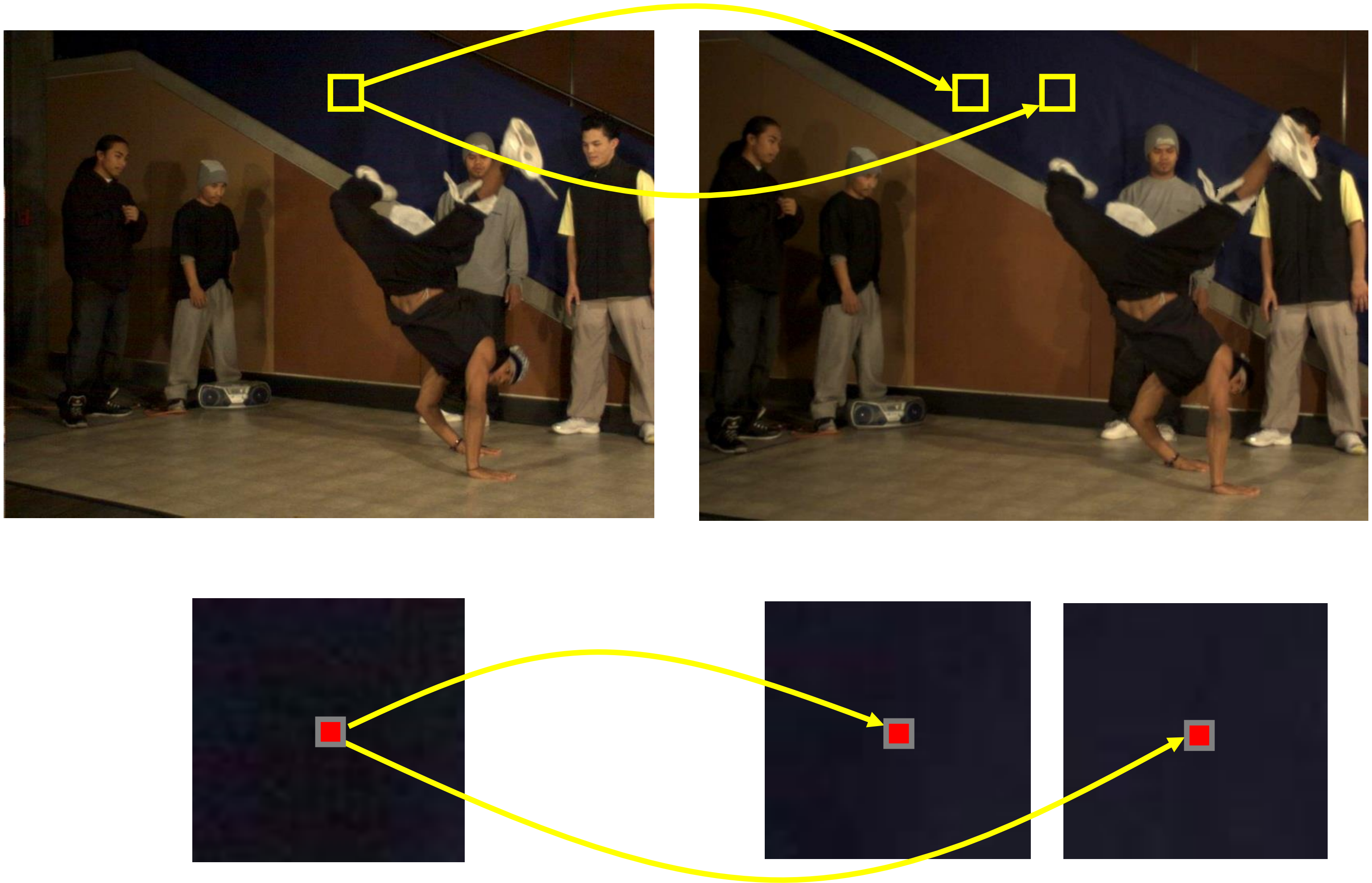

cameras
controlling
laptop
hard
disks
concentrators

Matching pixels

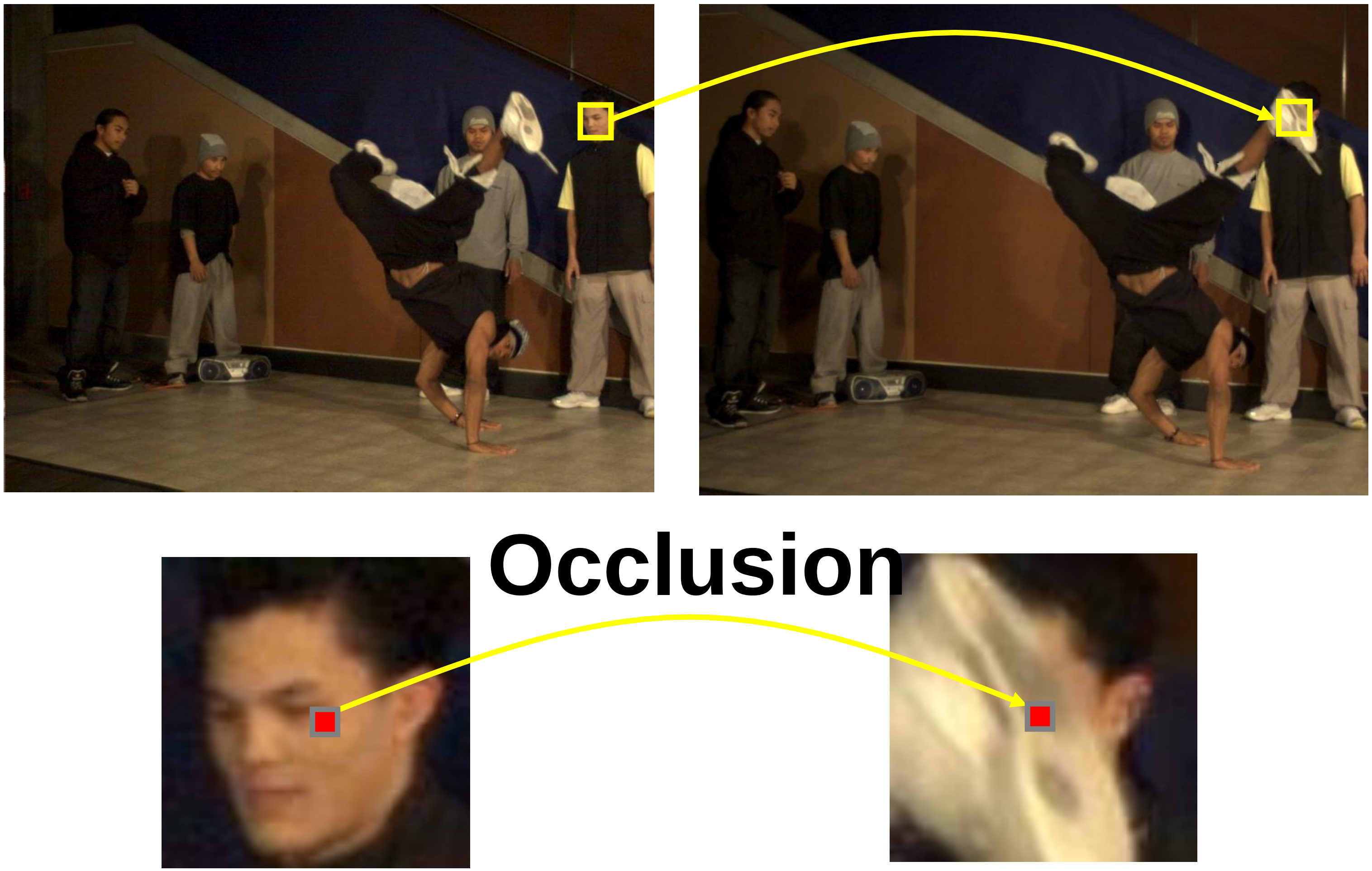
Easy



Harder

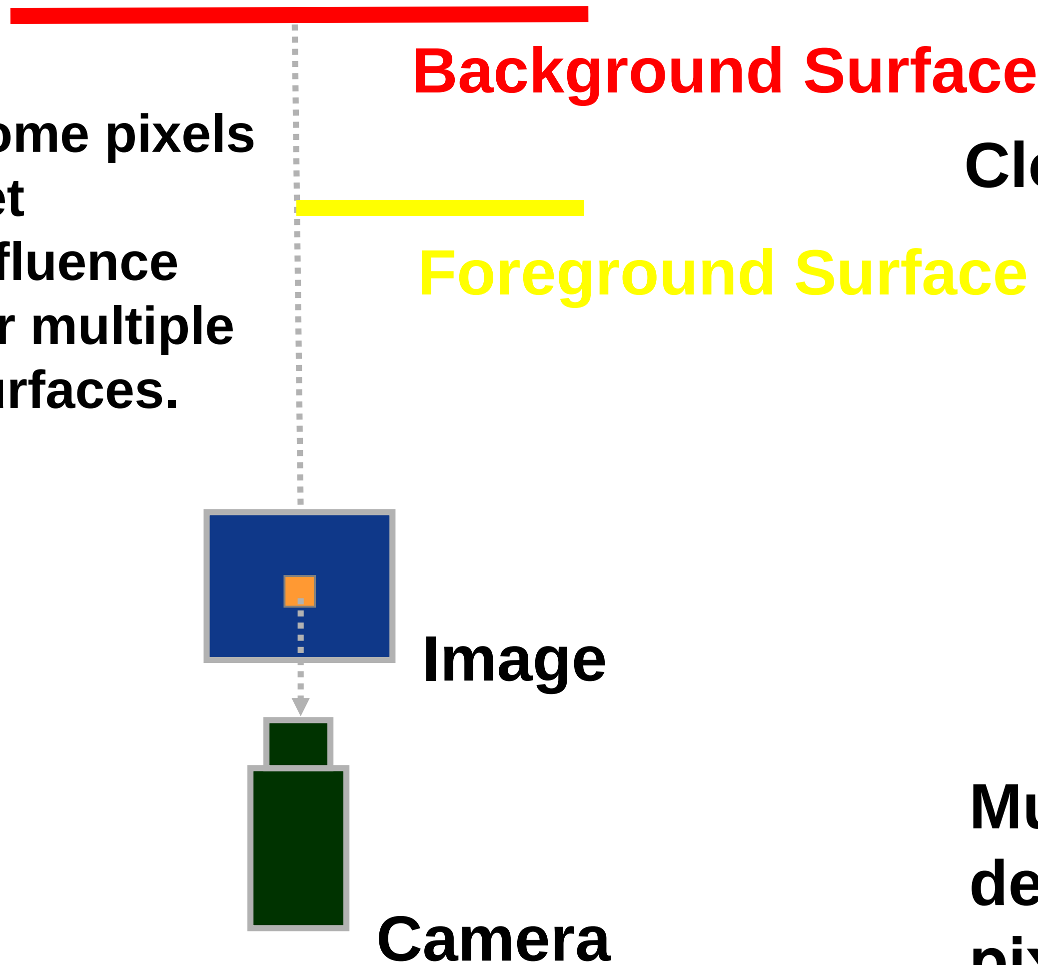


Really Really Hard



Matting

Some pixels get influence for multiple surfaces.

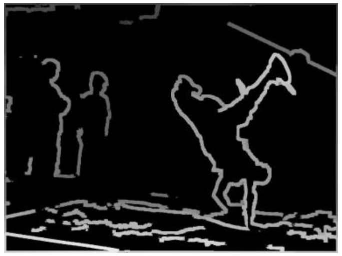


Close up of real image:

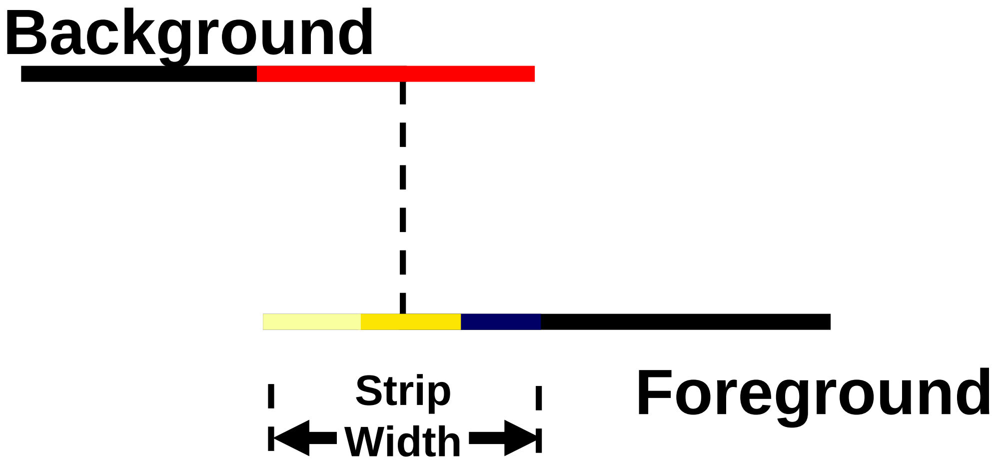
Multiple colors and depths at boundary pixels…

Find matting information:

1. **Find boundary strips using depth.**



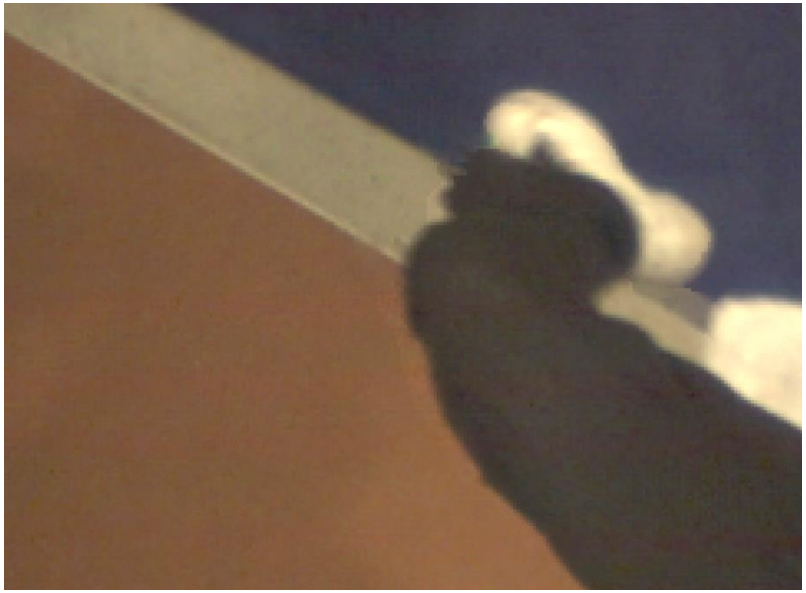
2. **Within boundary strips compute the colors and depths of the foreground and background object.**



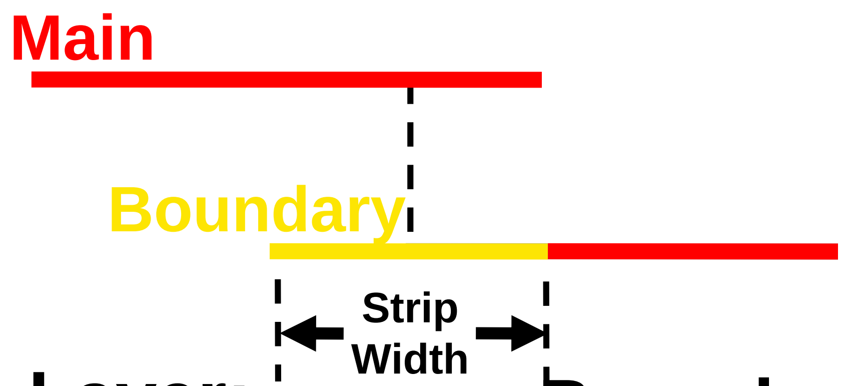
Why matting is important

No Matting

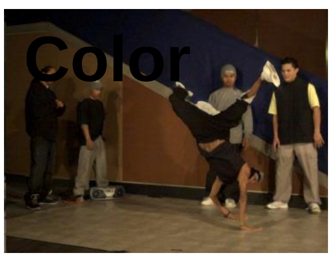
Matting



Representation



Main Layer:



Boundary Layer:

Massive Arabesque

Outline

- Panoramas and 360° Video Walkthroughs
- Light Fields and Lumigraphs
- LDIs, Sprites, and Layered Video
- ➢ Image-Based Modeling
- Environment Mattes and Matting
- Photo Tourism and Photosynth
- Point-Based Rendering and NPR
- Reflections and Transparency [2012]

Image-Based Modeling

Image-Based Modeling

Modeling and Rendering Architecture from Photographs

Debevec, Taylor, and Malik 1996

Original photograph with marked edges | Recovered model | Model edges projected onto photograph | Synthetic rendering

Façade

1. Select building blocks
2. Align them in each image
3. Solve for camera pose and block parameters (using constraints)

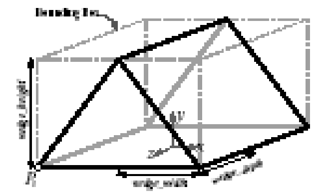


Figure 3: *A wedge block with its parameters and bounding box.*

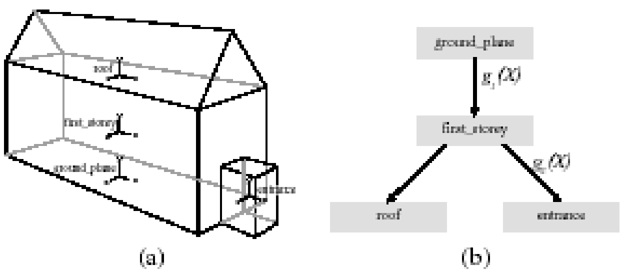
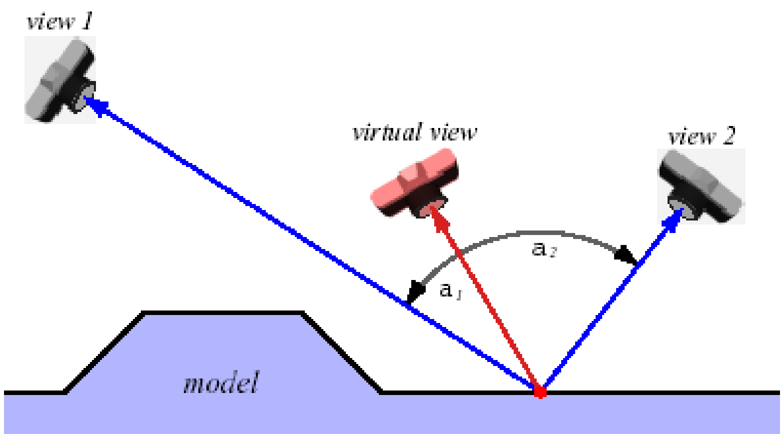
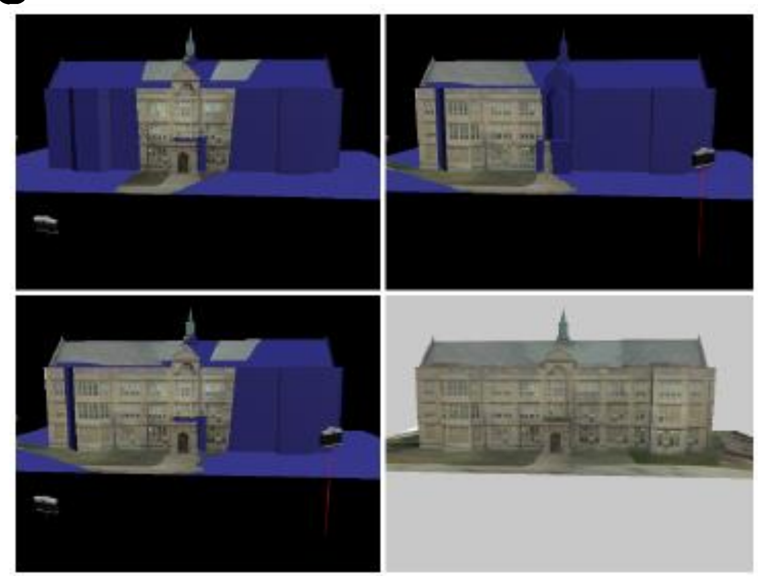


Figure 4: **(a)** *A geometric model of a simple building.* **(b)** *The model's hierarchical representation. The nodes in the tree represent parametric primitives (called* blocks*) while the links contain the spatial relationships between the blocks.*

View-dependent texture mapping

1. Determine visible cameras for each surface element
2. Blend textures (images) depending on distance between original camera and novel viewpoint



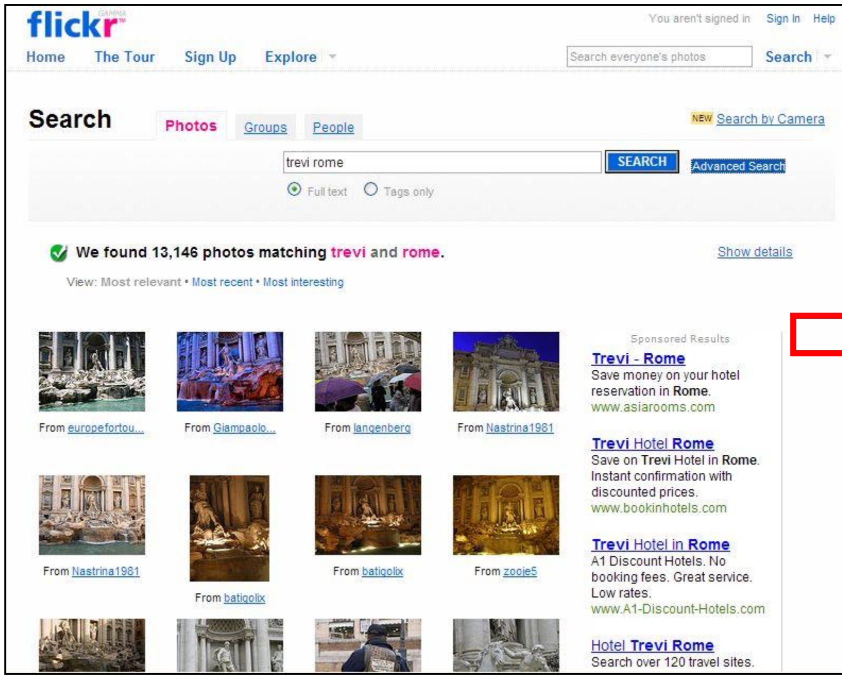


Graz reconstruction

Fully automated multi-view stereo, ca. 2009

Non-photorealistic experiences

Photo Tourism



Images on the Internet

Computed 3D structure

[Snavely, Seitz, Szeliski, SIGGRAPH 2006]

Photo Tourism

Internet Computer Vision

Use the Internet as an image and/or annotation source to solve challenging vision problems.

Scene Completion Using Millions of Photographs
James Hays and **Alexei Efros**.
Transactions on Graphics (SIGGRAPH 2007). August 2007, vol. 26, No. 3.

Project Page, **SIGGRAPH Paper**, **CACM Paper**, **CACM Technical Perspective by Marc Levoy**, **Bibtex**

Abstract: What can you do with a million images? In this paper we present a new image completion algorithm powered by a huge database of photographs gathered from the Web. The algorithm patches up holes in images by finding similar image regions in the database that are not only seamless but also semantically valid. Our chief insight is that while the space of images is

Internet Computer Vision

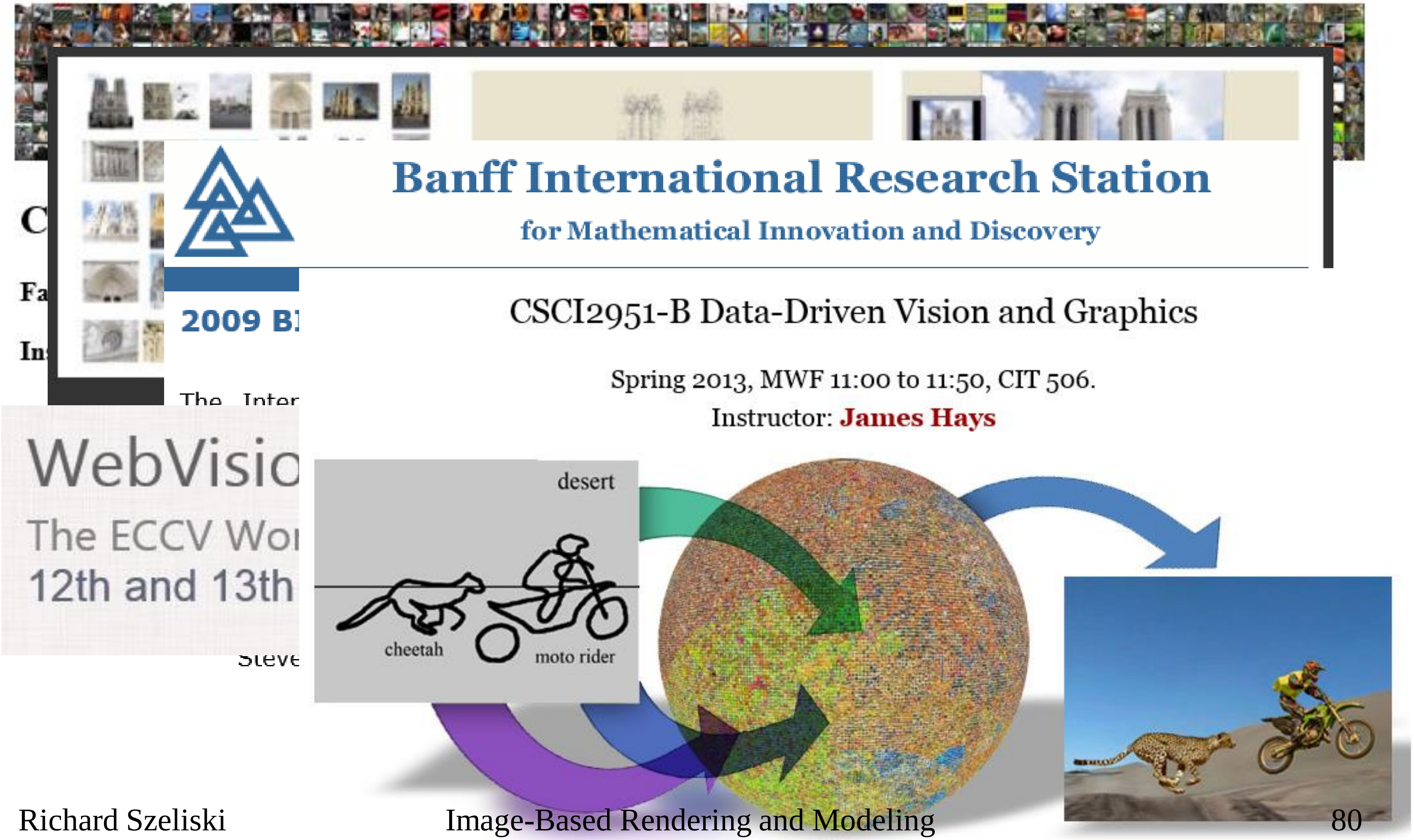
Photo Tourism and IM2GPS:

3D Reconstruction and Geolocation of Internet Photo Collections

CVPR 2009 half day course, June 20, 2009 (morning)

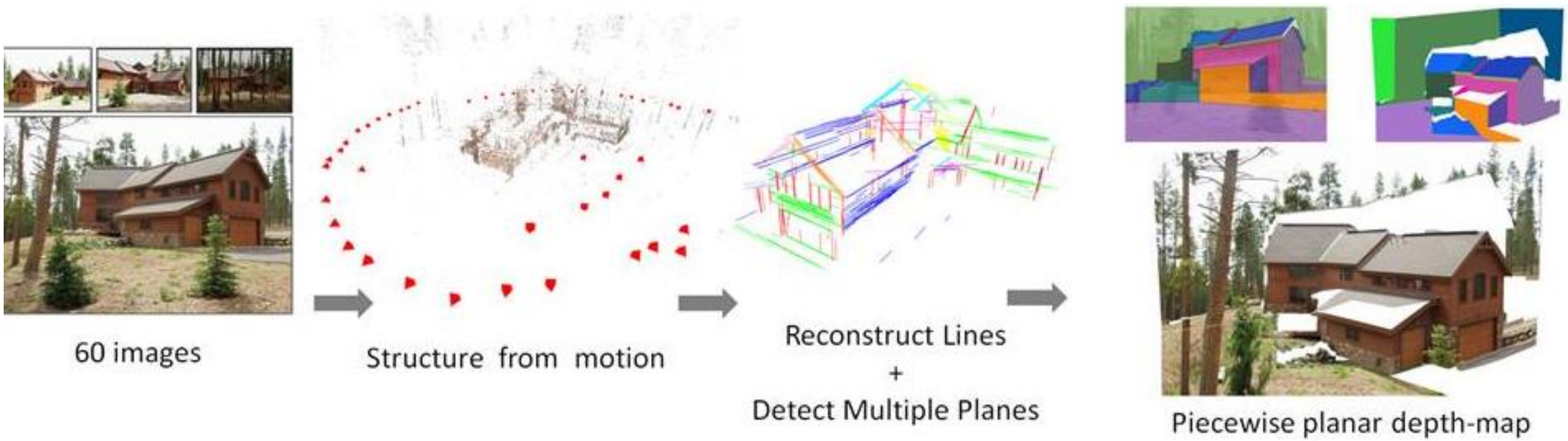
Course organizers: Noah Snavely, Cornell University, and James Hays, CMU

Internet Computer Vision



Recent research

Piecewise planar proxies



[Sinha, Steedly, Szeliski ICCV'09]

Building Interiors

[Furukawa *et al.*, ICCV'09]

Internet-Scale Stereo

[Furukawa *et al.*, CVPR'10]

What's missing?

Reflections

Gloss

…?

Reflections

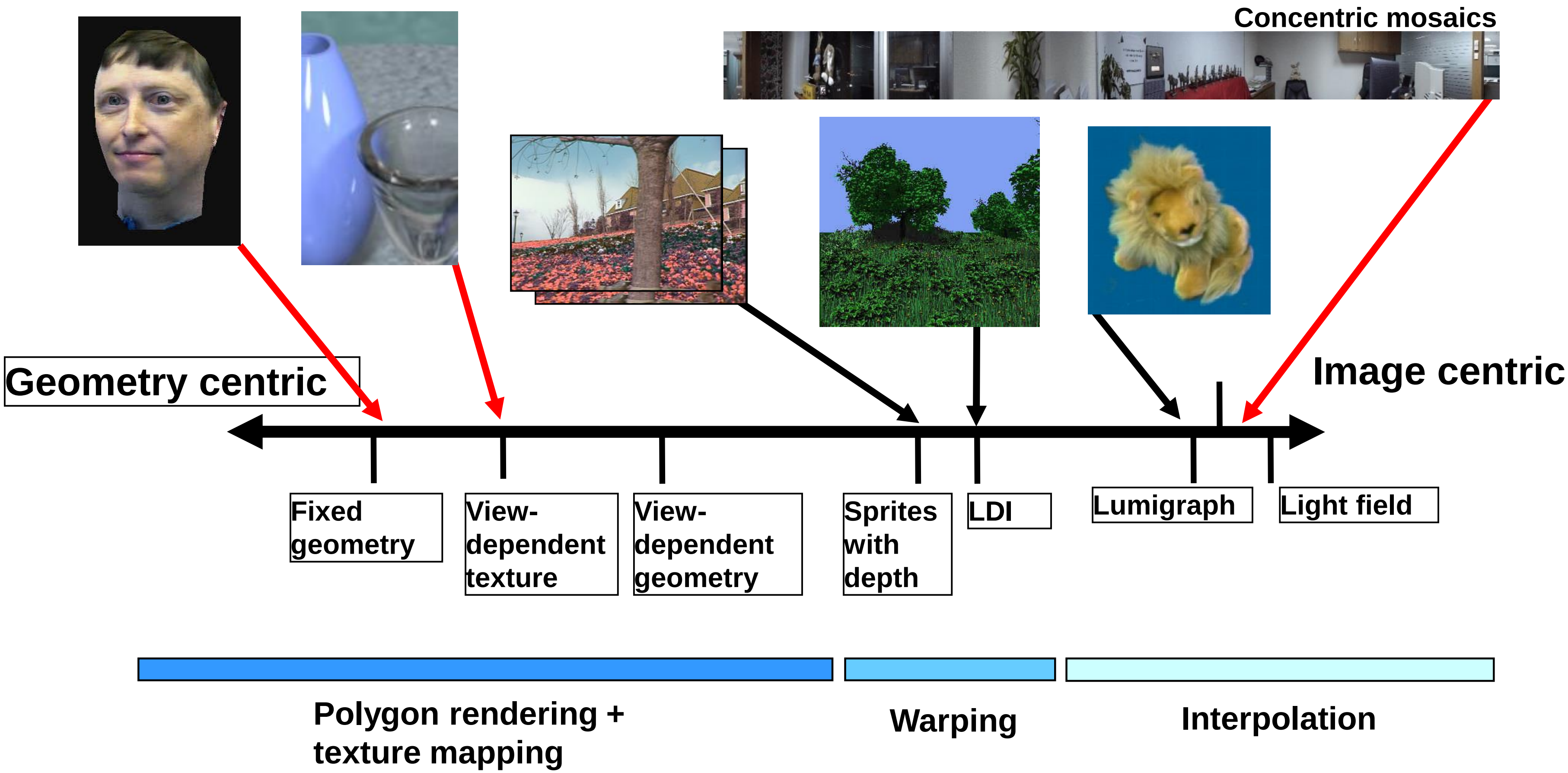
Image-Based Rendering for Scenes with Reflections

Sudipta N. Sinha
Johannes Kopf
Michael Goesele
Daniel Scharstein
Richard Szeliski

SIGGRAPH2012

Wrapping up…

Graphics/Imaging Continuum



What works? What doesn’t?

- ✓ Automatic 3D pose estimation
- ✓ Aerial (+ active ground level) modeling
- Accurate boundaries & matting
- Reflections and transparency
- User-generated content
  - Casually acquired photos and videos

New sensors

  - Active sensing (Kinect…);   Plenoptic cameras
- Integration with recognition

What about Computational Photography?

CSCI 129 Computational Photography

Fall 2012, MWF 11:00 to 11:50, CIT 477 (Lubrano)
Instructor: **James Hays**
HTA: Sam Birch and GTA: Emanuel Zgraggen

Course Description

Course Catalog Entry

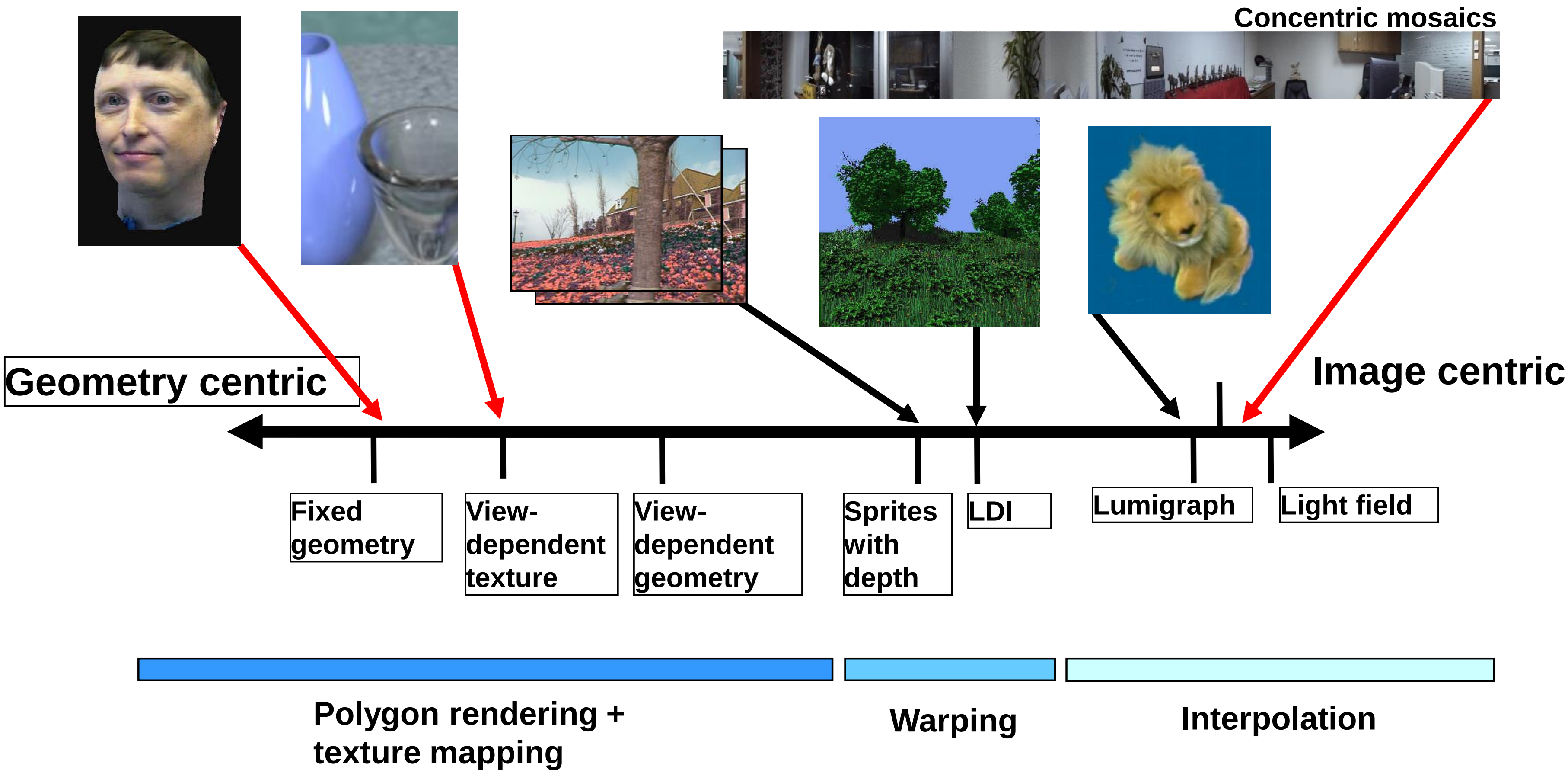
Computational Photography describes the convergence of computer graphics, computer vision, and the Internet with photography. Its goal is to overcome the limitations of traditional photography using computational techniques to enhance the way we capture, manipulate, and interact with visual media. In this course we will study many interesting, recent image based algorithms and implement them to the degree that is possible. We will cover topics such as

- Cameras and image formation
- Human visual perception
- Image processing (filtering, pyramids)
- Image blending and compositing
- Image retargeting
- Texture synthesis and transfer
- Image completion / inpainting
- Super-resolution, deblurring, and denoising.
- Image based lighting and rendering
- High dynamic range
- Depth and defocus
- Flash / no flash photography
- Coded aperture photography
- Single / multi view reconstruction
- Photo quality assessment
- Non photorealistic rendering
- Modeling and synthesis using Internet data

Image-Based Rendering

- Panoramas and 360° Video Walkthroughs
- Light Fields and Lumigraphs
- Layered Representations and Matting
- Image-Based Modeling
- Environment Mattes and Matting
- Photo Tourism and Photosynth
- Point-Based Rendering and NPR
- Reflections and Transparency
- Where next?
- How to use in *your* applications?

Graphics/Imaging Continuum



(Questions)

[The End]