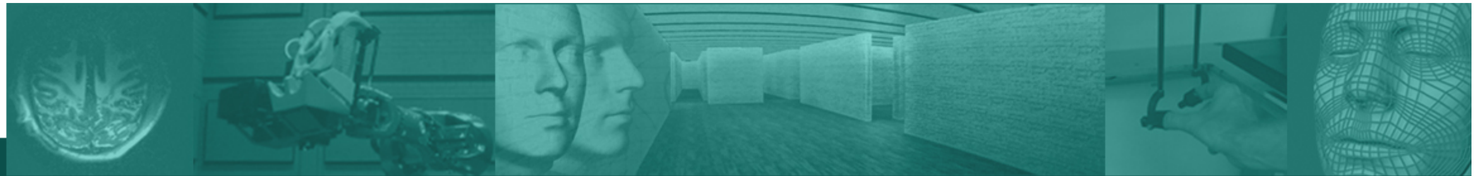
Perceptual Graphics:
Integrating Perception, Computer Graphics and Computer Vision

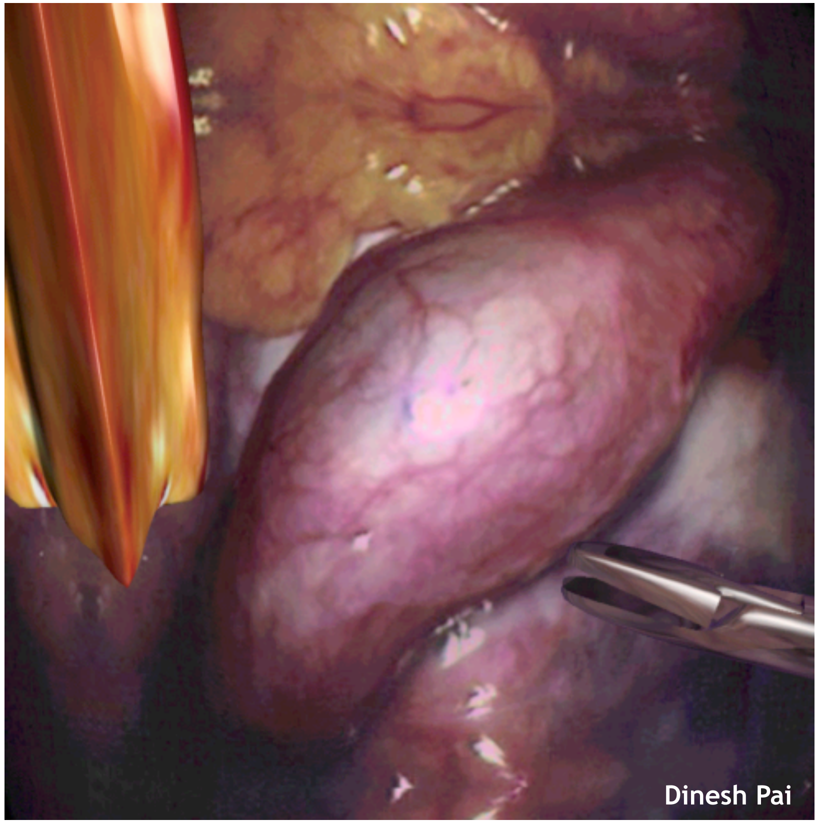
Heinrich H. Bülthoff
Roland W. Fleming
Max Planck Institute for Biological Cybernetics



Cognitive & Computational Psychophysics
Max Planck Institute for Biological Cybernetics

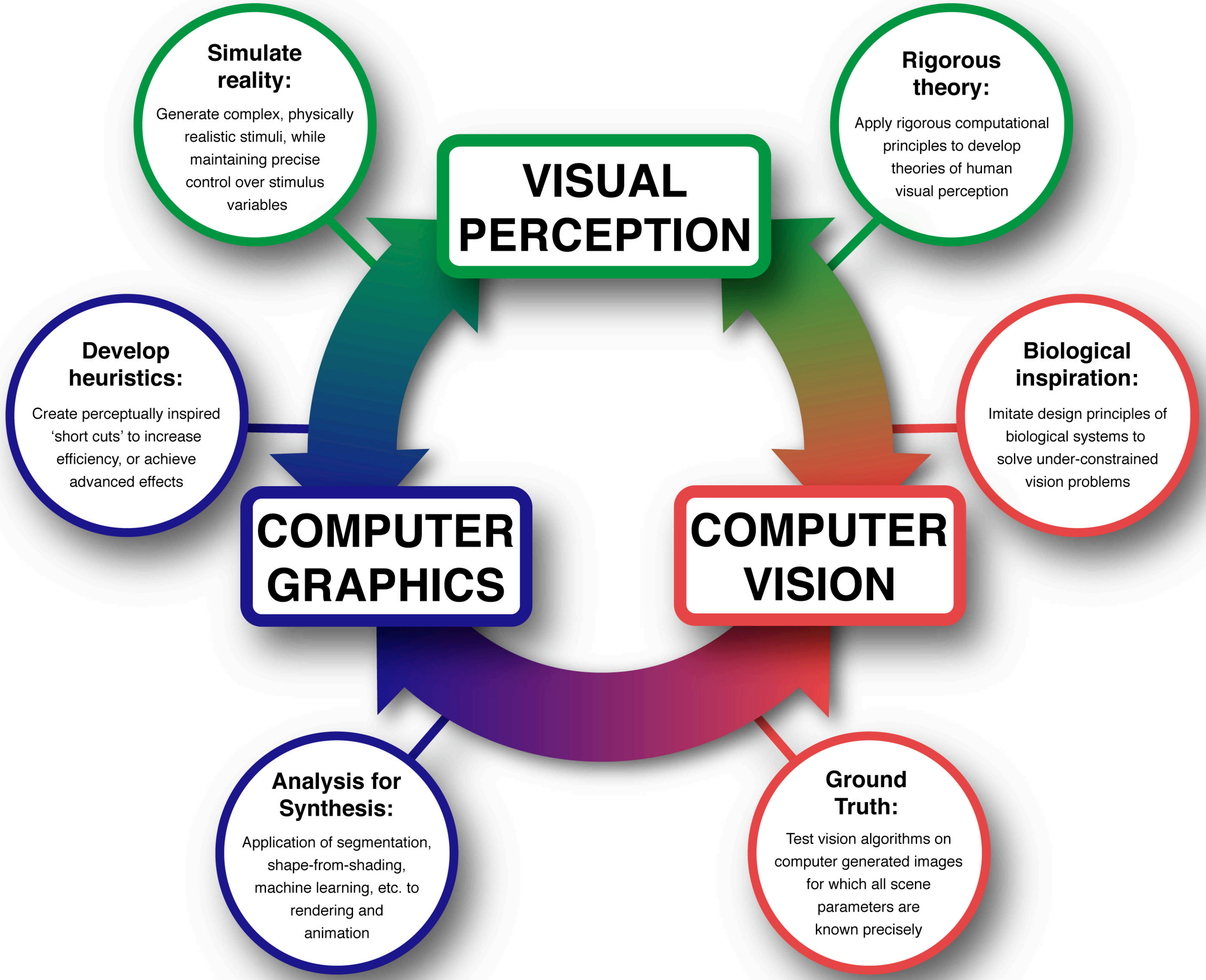
What is CG for ?


The Matrix


Dinesh Pai

movies, TV, games, art, architecture, CAD, data visualization, surgical, military and industrial training, remote operation, …

CG is for human observers


VISUAL PERCEPTION
COMPUTER GRAPHICS
COMPUTER VISION
Simulate reality:
Generate complex, physically realistic stimuli, while maintaining precise control over stimulus variables
Rigorous theory:
Apply rigorous computational principles to develop theories of human visual perception
Develop heuristics:
Create perceptually inspired 'short cuts' to increase efficiency, or achieve advanced effects
Biological inspiration:
Imitate design principles of biological systems to solve under-constrained vision problems
Analysis for Synthesis:
Application of segmentation, shape-from-shading, machine learning, etc. to rendering and animation
Ground Truth:
Test vision algorithms on computer generated images for which all scene parameters are known precisely

Outline

- **Using Computer Graphics to study human vision**
  - 3D shape perception
  - Perception of material properties (e.g. translucency, glossiness)

- **Visual Psychophysics in the service of CG**
  - Perceptual issues in inverse tonemapping of LDR -> HDR content

Outline

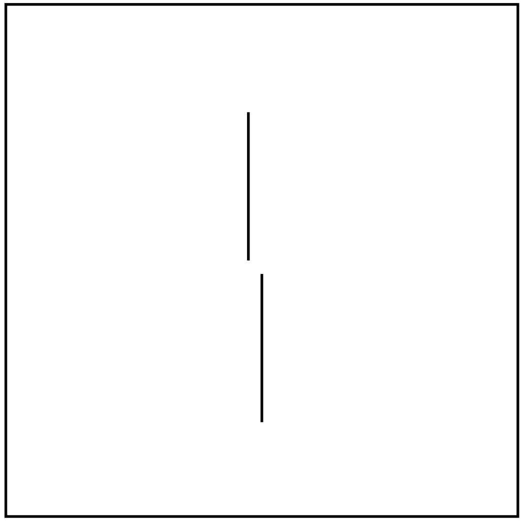
- **Exploiting limits of human perception to facilitate graphics**
  - BW-FIT Gaze-Contingent Display
  - Visual attention
- **Exploiting ambiguity in visual perception**
  - Visual Perception as inference
  - Image-based Material Editing
- **Closing the loop**
  - Interactivity and VR research at the MPI

Outline

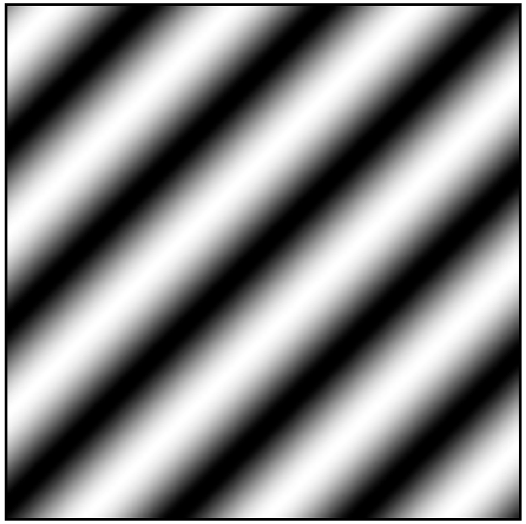
- **Using Computer Graphics to study human vision**
  - 3D shape perception
  - Perception of material properties (e.g. translucency, glossiness)
- **Visual Psychophysics in the service of CG**
  - Perceptual issues in inverse tonemapping of LDR -> HDR content

Traditional Psychophysics

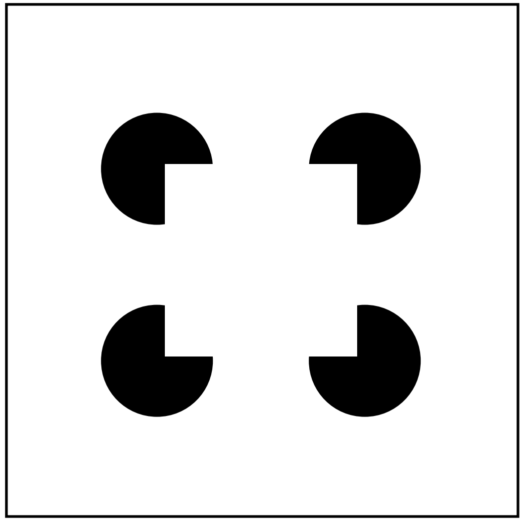
Vernier acuity



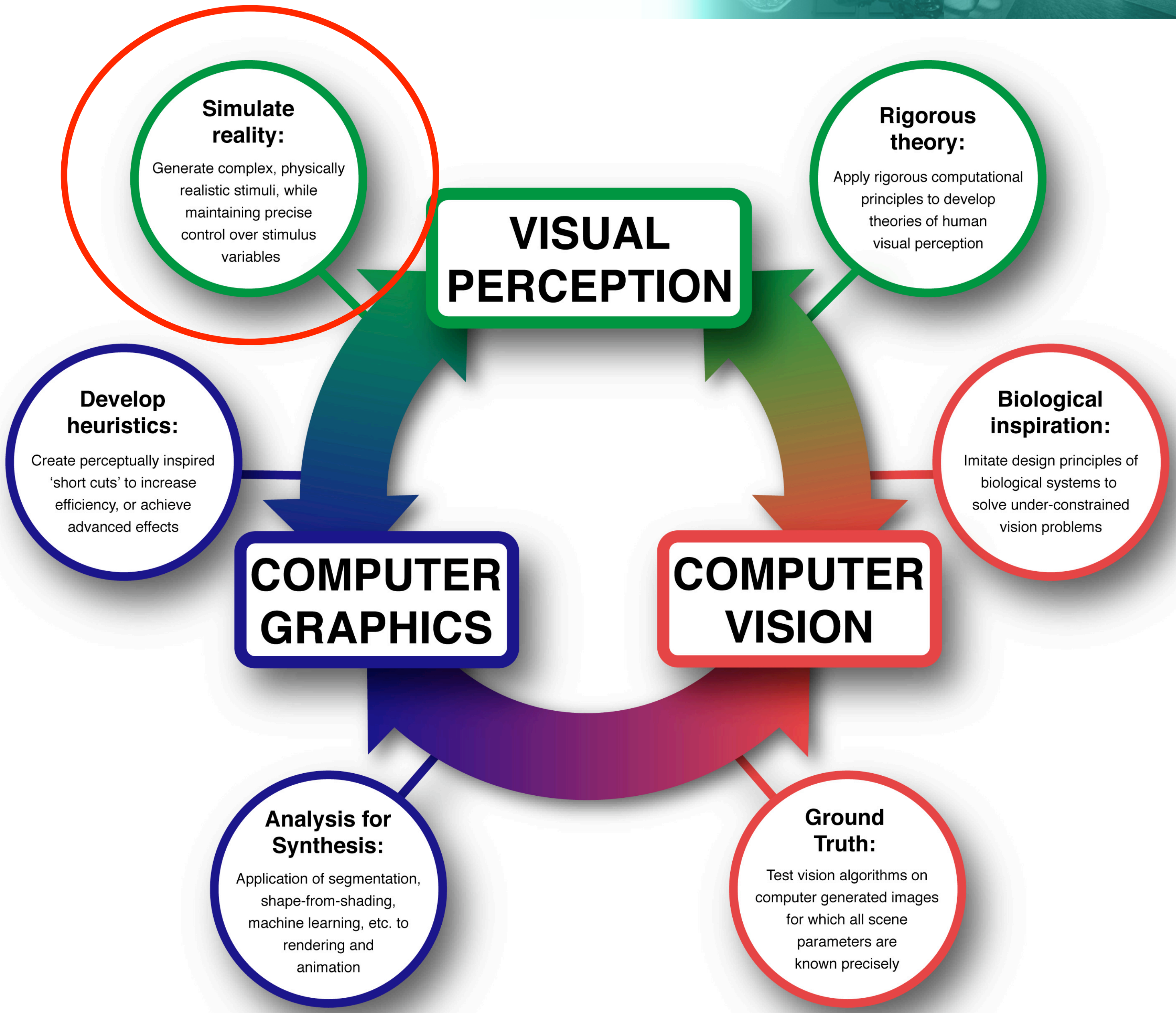
Sine wave



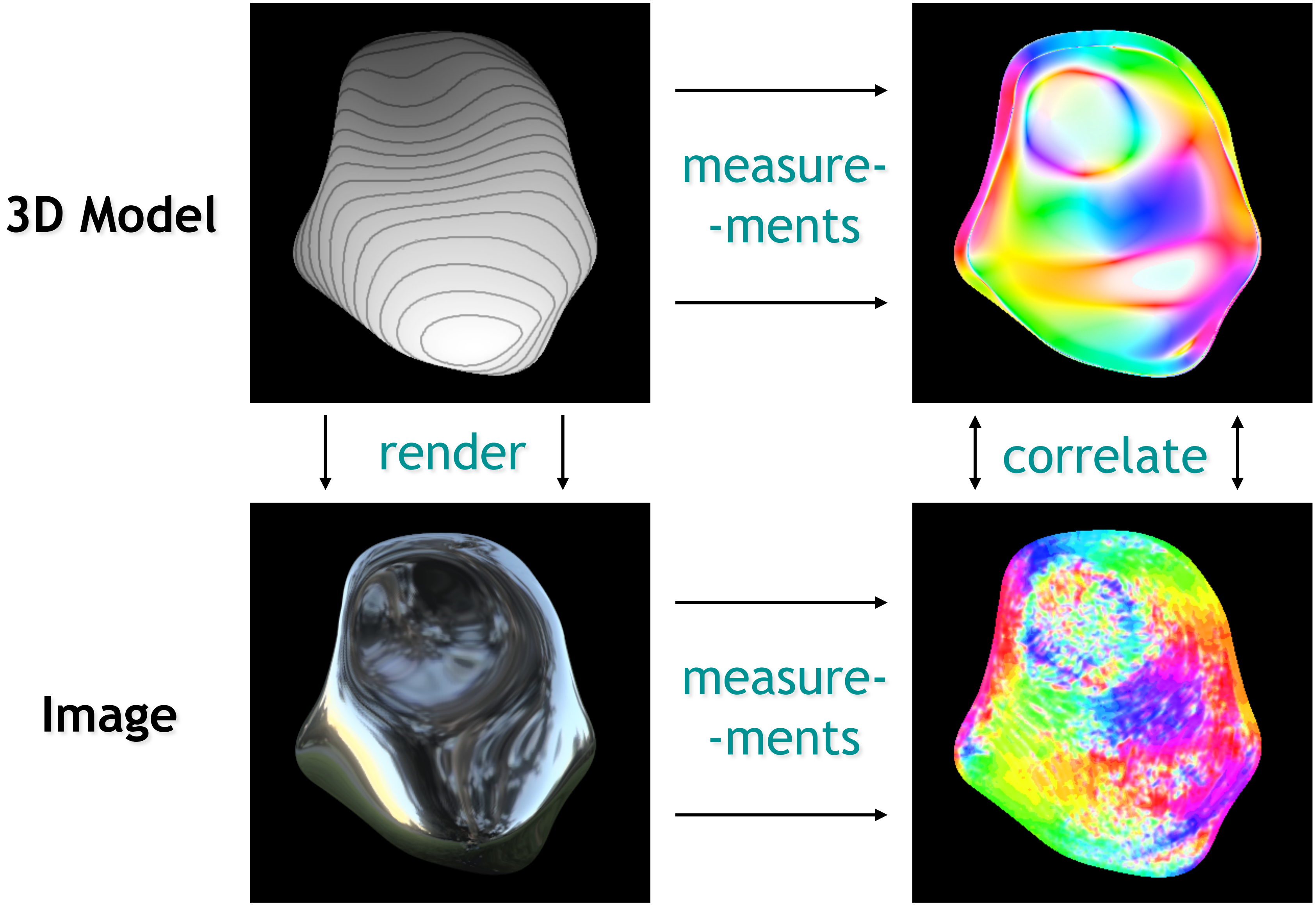
Kanizsa Square



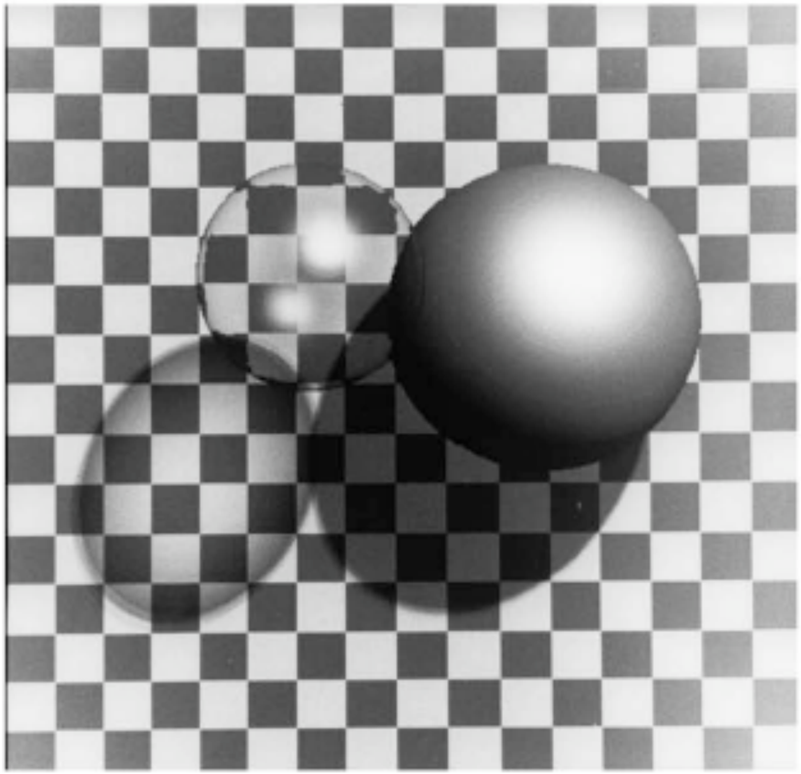
- **Traditional Psychophysics used simple stimuli**
  - Easy to generate and parametrically vary
  - Allow extremely precise control
  - Answered many important basic questions
  - **BUT: Can lead to studying vision in unnatural conditions**


Simulate reality:
Generate complex, physically realistic stimuli, while maintaining precise control over stimulus variables
VISUAL PERCEPTION
Rigorous theory:
Apply rigorous computational principles to develop theories of human visual perception
Develop heuristics:
Create perceptually inspired 'short cuts' to increase efficiency, or achieve advanced effects
Biological inspiration:
Imitate design principles of biological systems to solve under-constrained vision problems
COMPUTER GRAPHICS
COMPUTER VISION
Analysis for Synthesis:
Application of segmentation, shape-from-shading, machine learning, etc. to rendering and animation
Ground Truth:
Test vision algorithms on computer generated images for which all scene parameters are known precisely

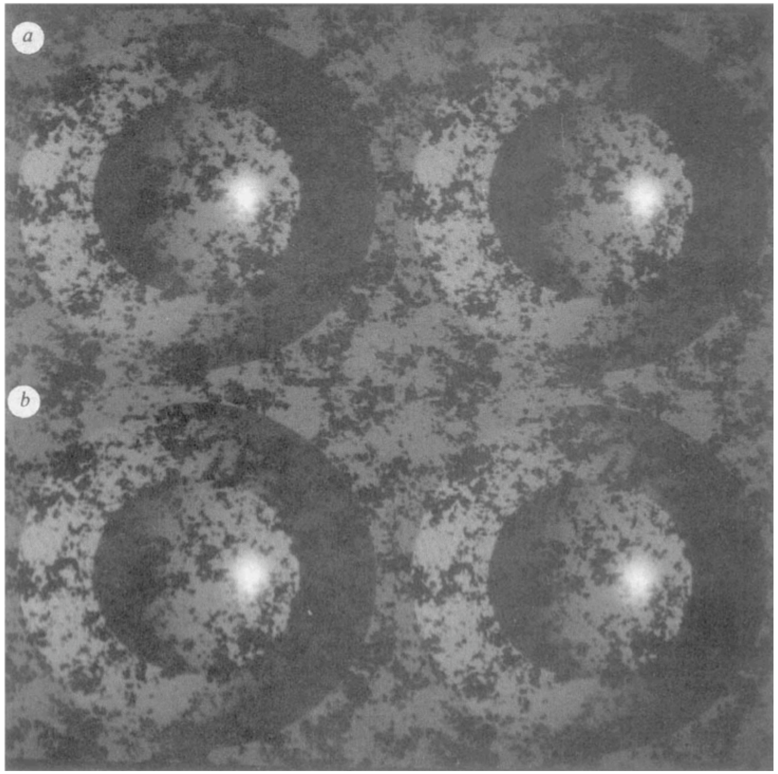
"Ground Truth"



Early Work (1984-1990)



Todd & Mingolla (1984). The simulation of curved surfaces from patterns of optical texture. *Journal of Experimental Psychology*



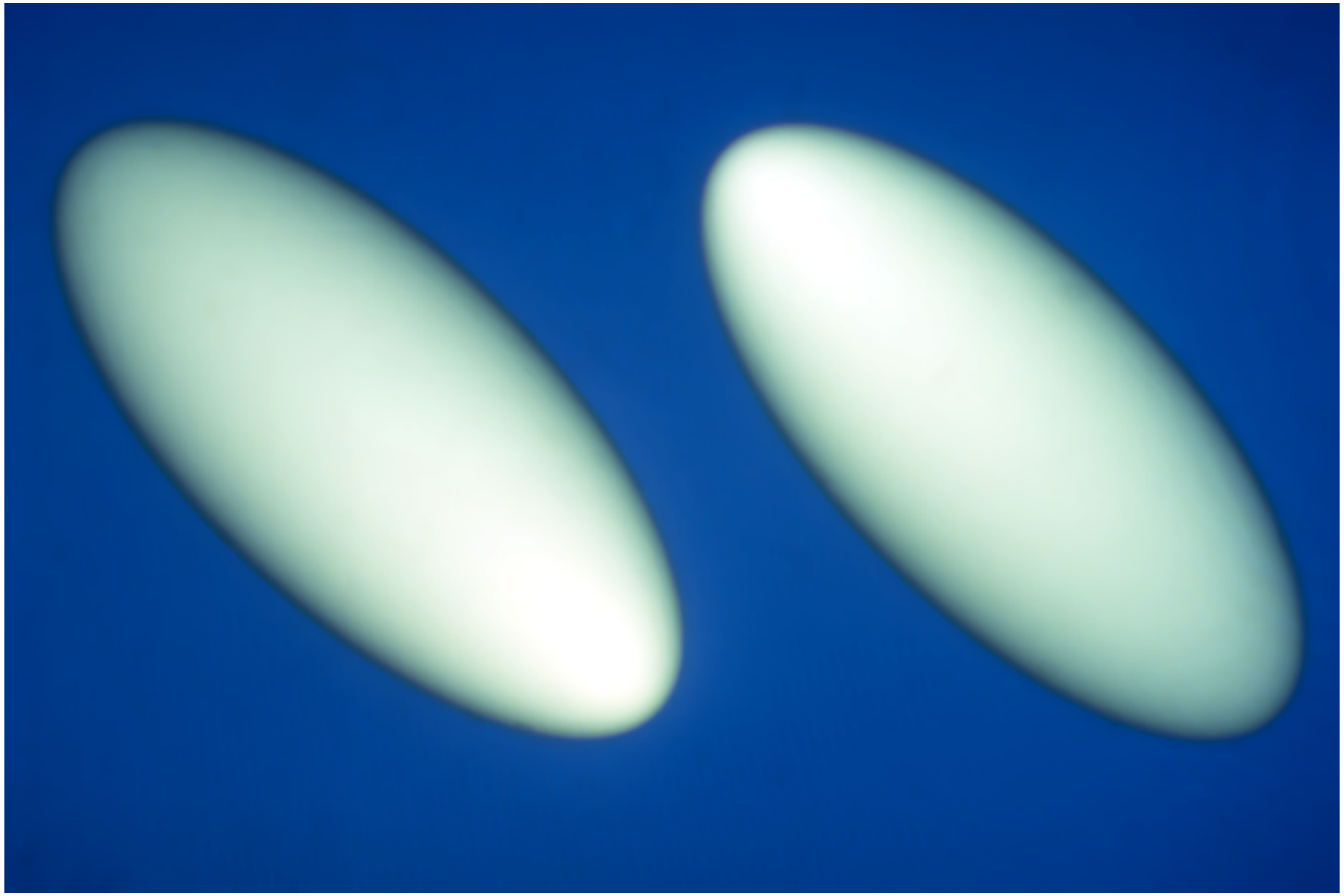
Blake & Bülthoff (1990). Does the brain know the physics of specular reflection? *Nature*

Shape from X

- Shading
- Texture
- Highlights
- Binocular Stereopsis
- Motion parallax
- Pictorial cues (perspective, etc.)
- …

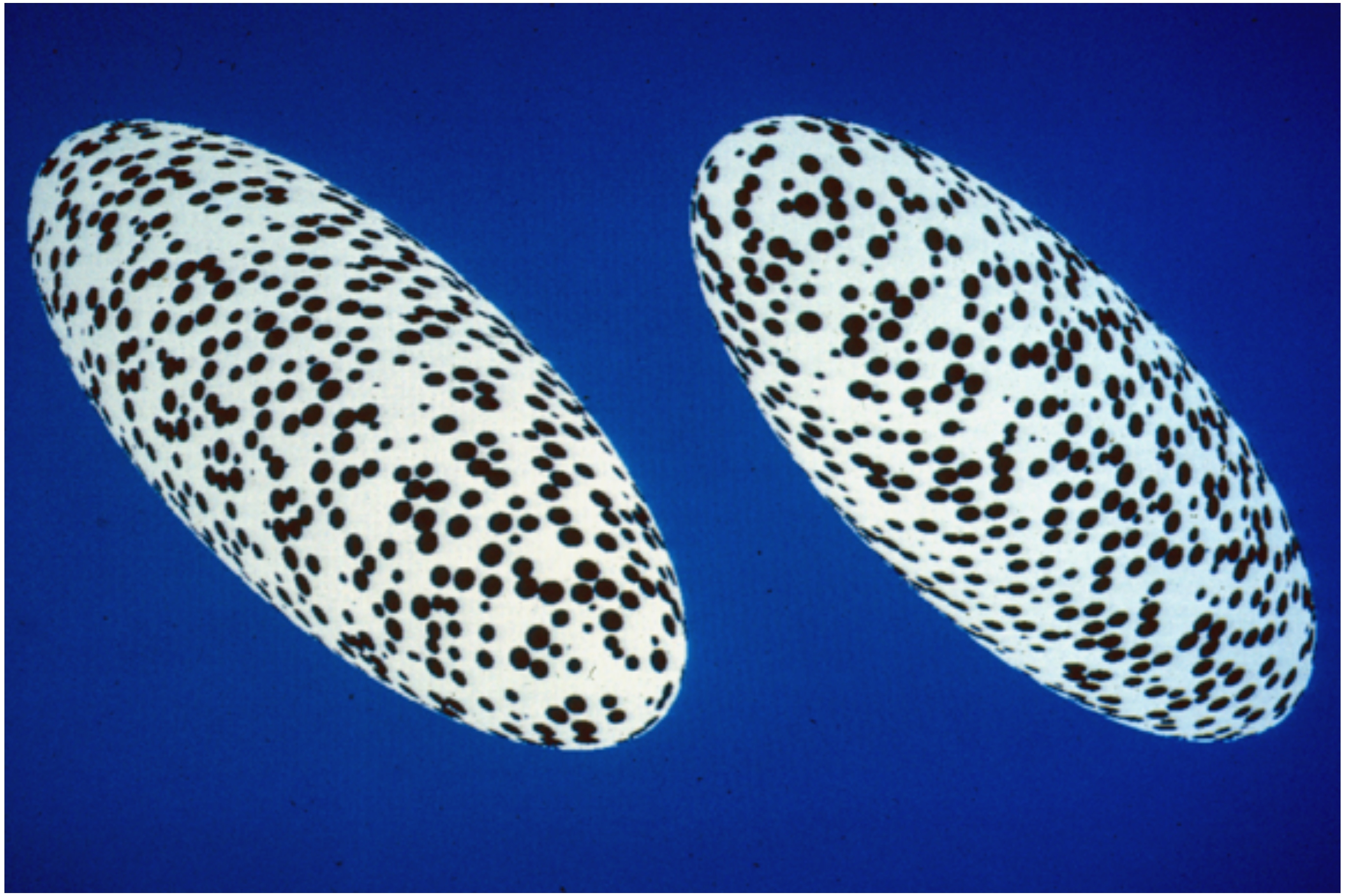
- How do we combine these different cues?

Shape-from-Shading



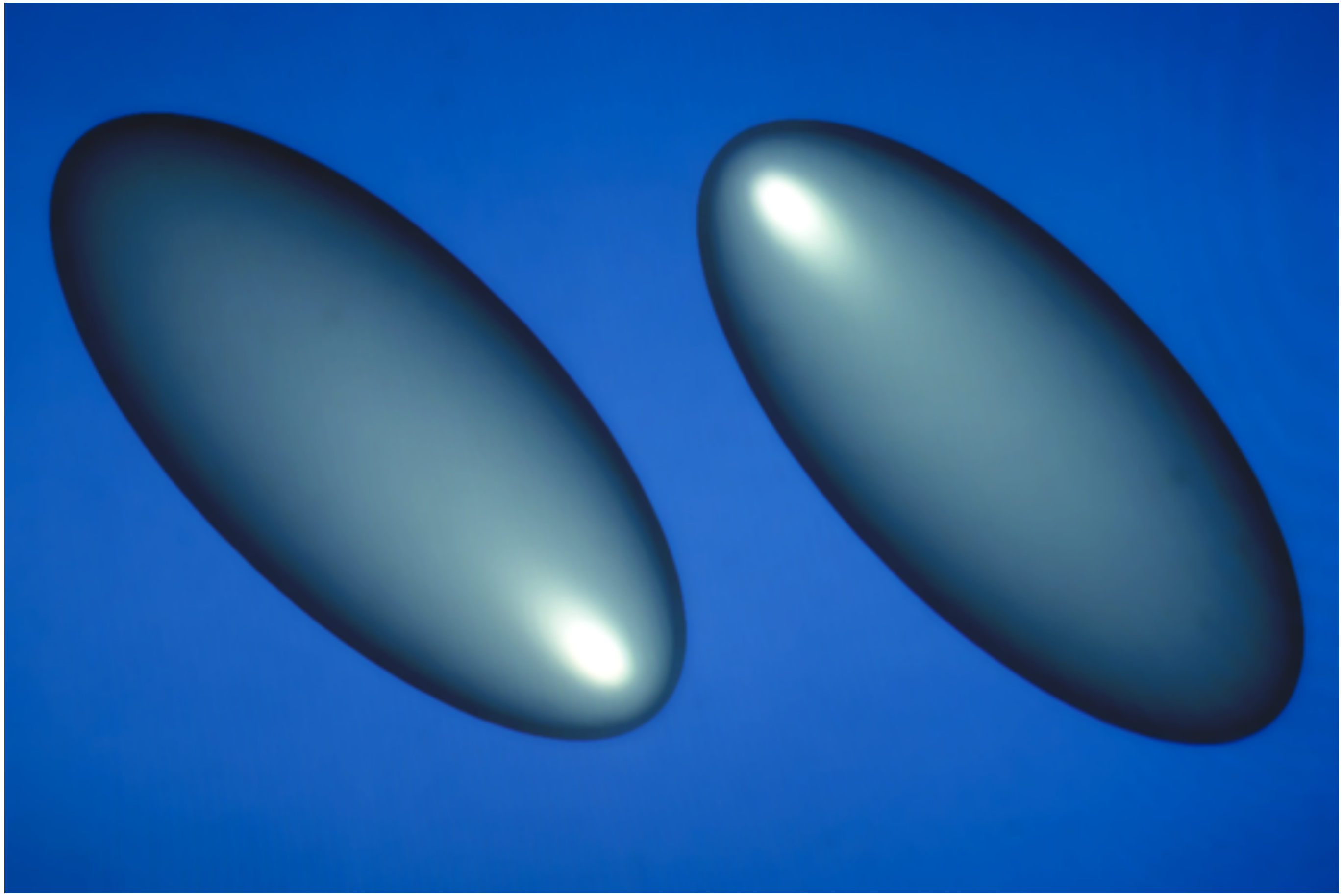
- **Shape-from-Shading is a weak cue to the perception of orientation**

Shape-from-Texture



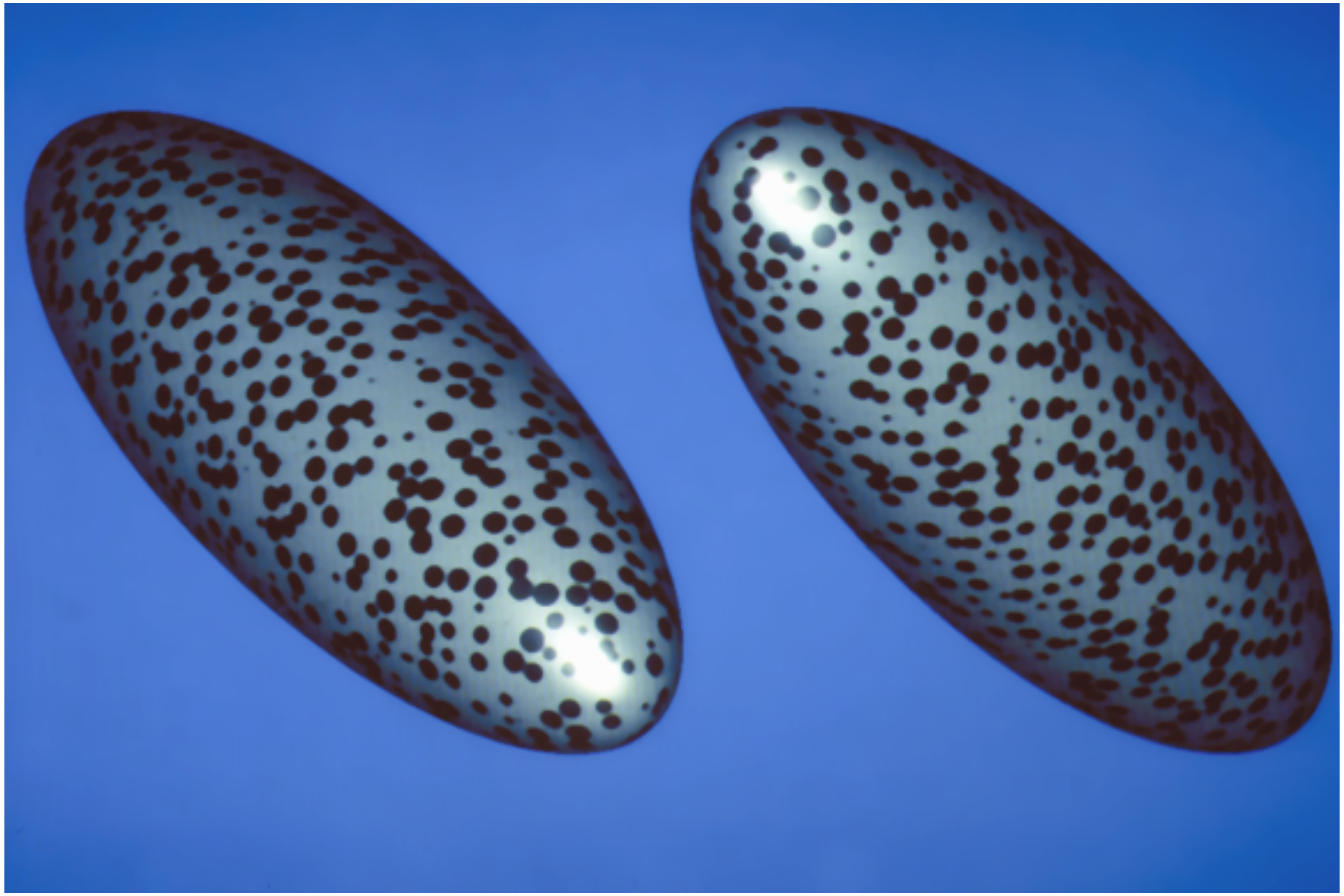
- **Shape-from-Texture is a weak cue to the perception of form and orientation**

Phong Shading



- **Better shading models improve the perception of form and orientation**

Shading + Texture



- **Integration of several cues provides good perception of form and orientation**
  ***Bülthoff and Mallot*, JOSA, 1988**
  ***Bülthoff and Yuille*, Theoretical Biology, 1991**

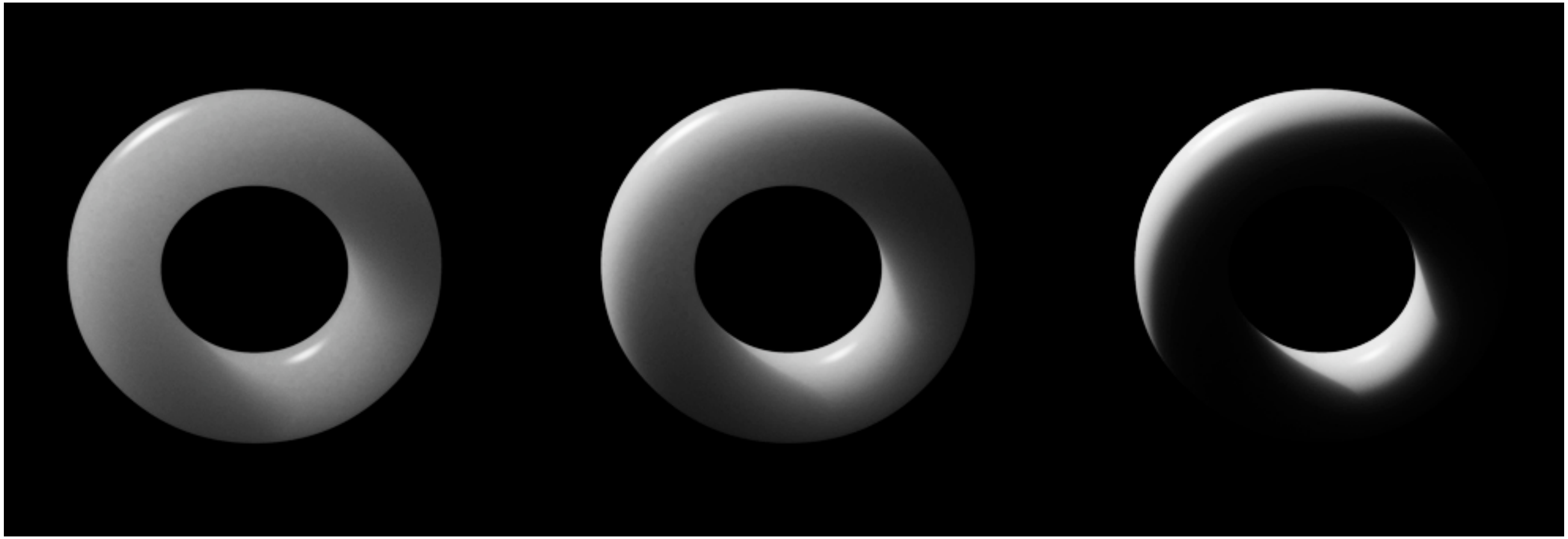
Perception of shape from shading

RADIANCE (Greg Ward) using image-based lighting

Perception of translucent materials

DALI (Henrik Wann Jensen) using BSSRDF shader

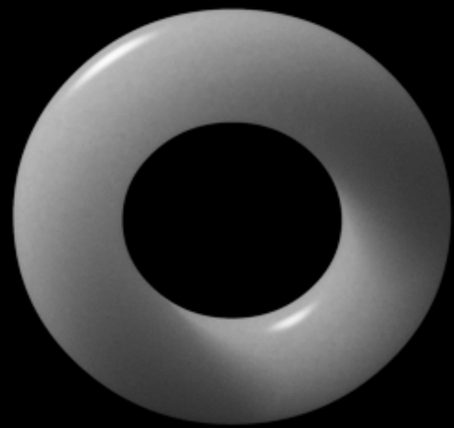
Realism with control



Identical geometry, identical lighting, identical viewpoint.
Only differ in degree of translucency

Asymmetric Matching Experiment

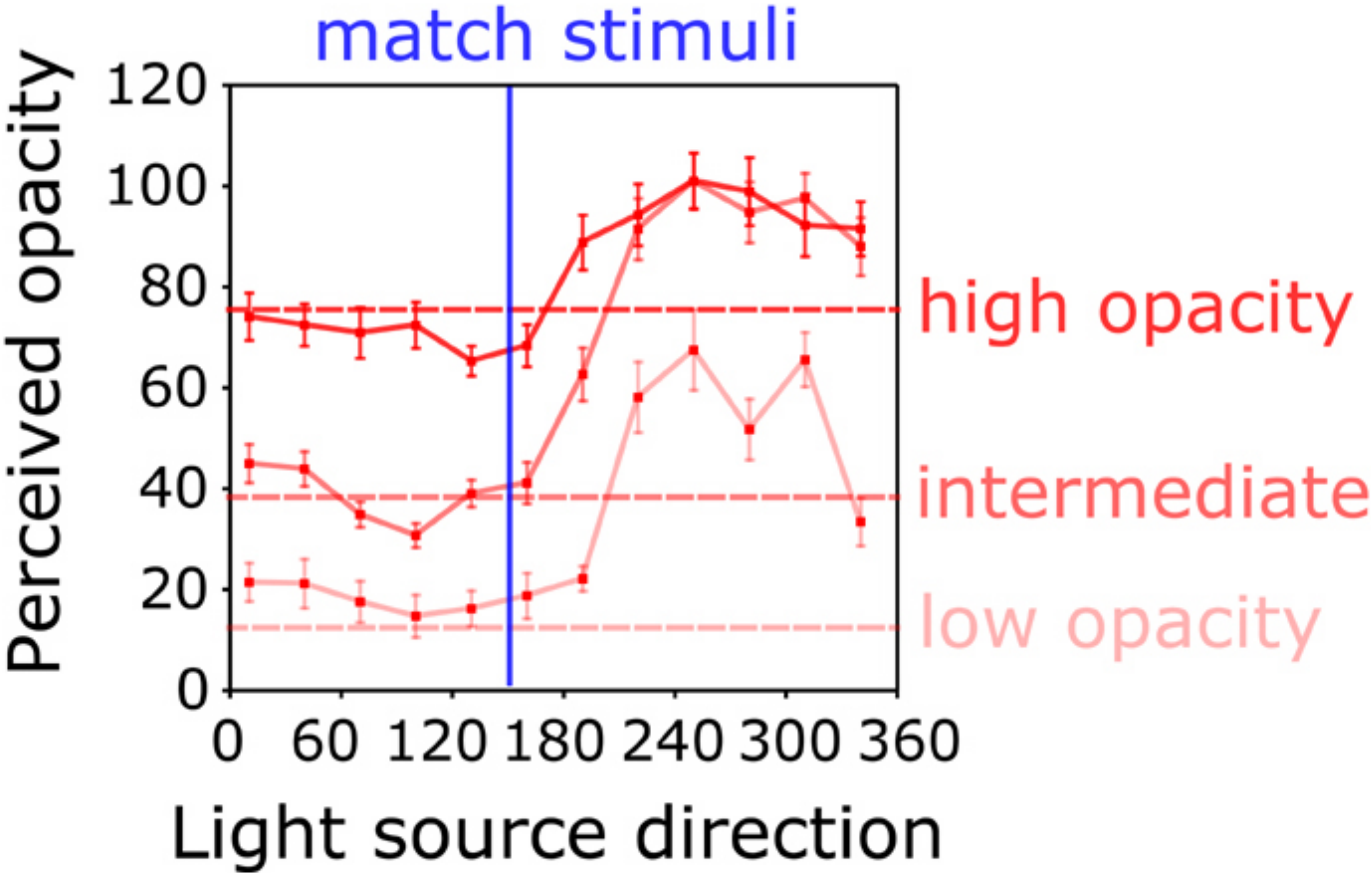
Test **Match**



Task: adjust the subsurface scattering coefficients of the “Match” stimulus, until the object appears to be made of the same material as the “Test” stimulus

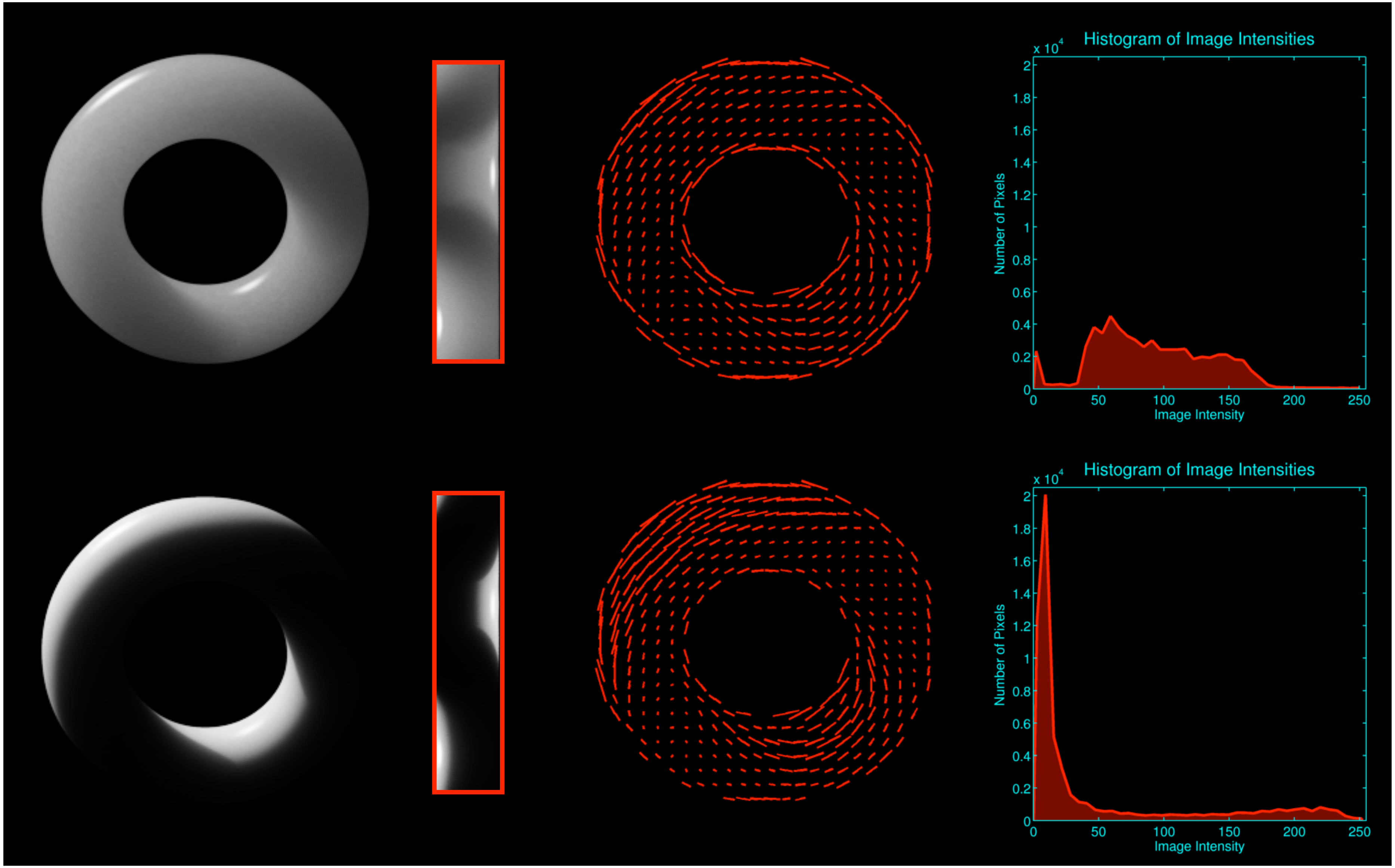
Fleming & Bülthoff (2005). ACM Trans. on Applied Perception.

Results

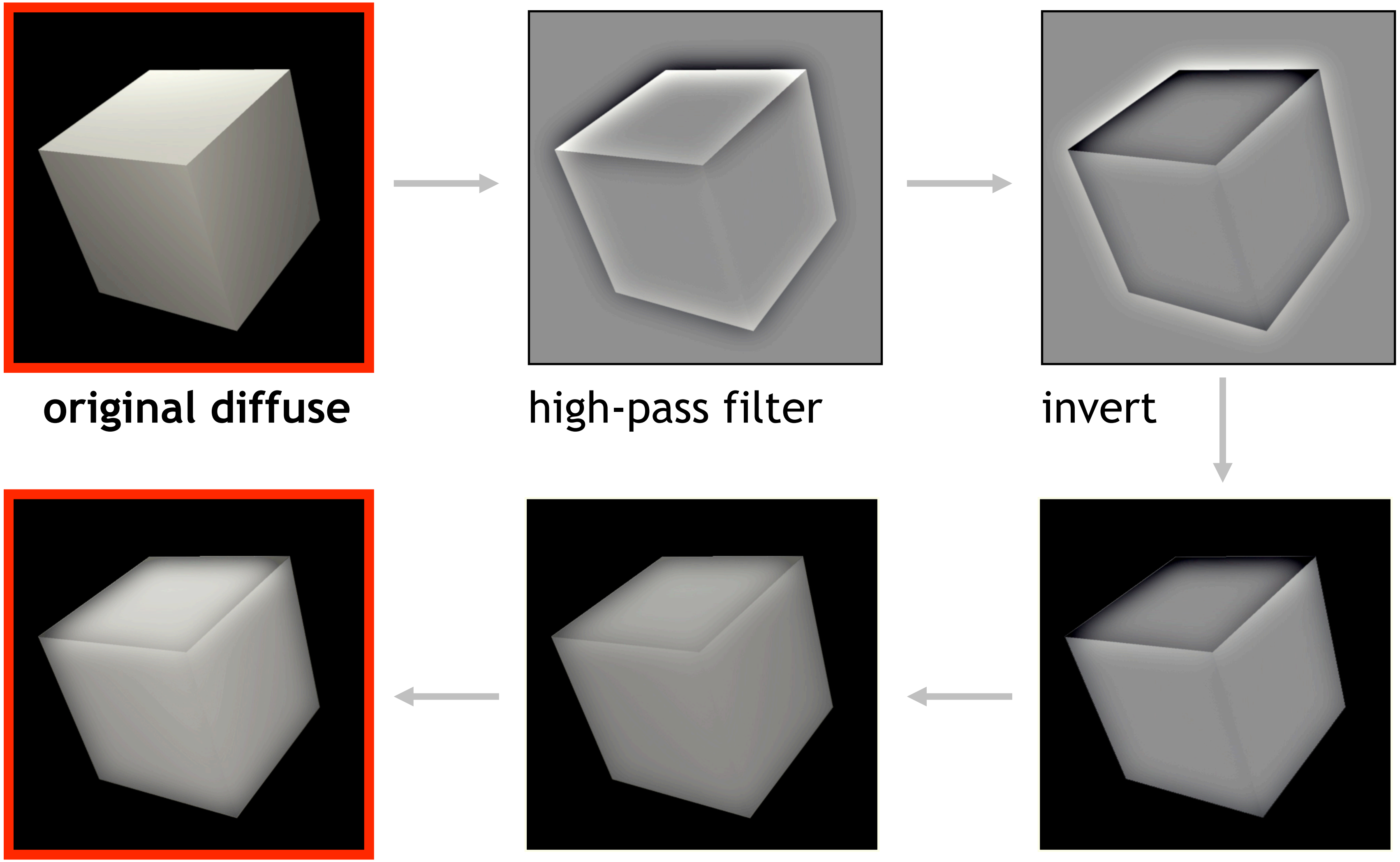


match stimuli
Perceived opacity
120
100
80
60
40
20
0
0
60
120
180
240
300
360
Light source direction
high opacity
intermediate
low opacity

Image Statistics



Perceptually Inspired Heuristics



Outline

- Using Computer Graphics to study human vision
  - 3D shape perception
  - Perception of material properties (e.g. translucency, glossiness)

- **Visual Psychophysics in the service of CG**
  - Perceptual issues in inverse tonemapping of LDR -> HDR content

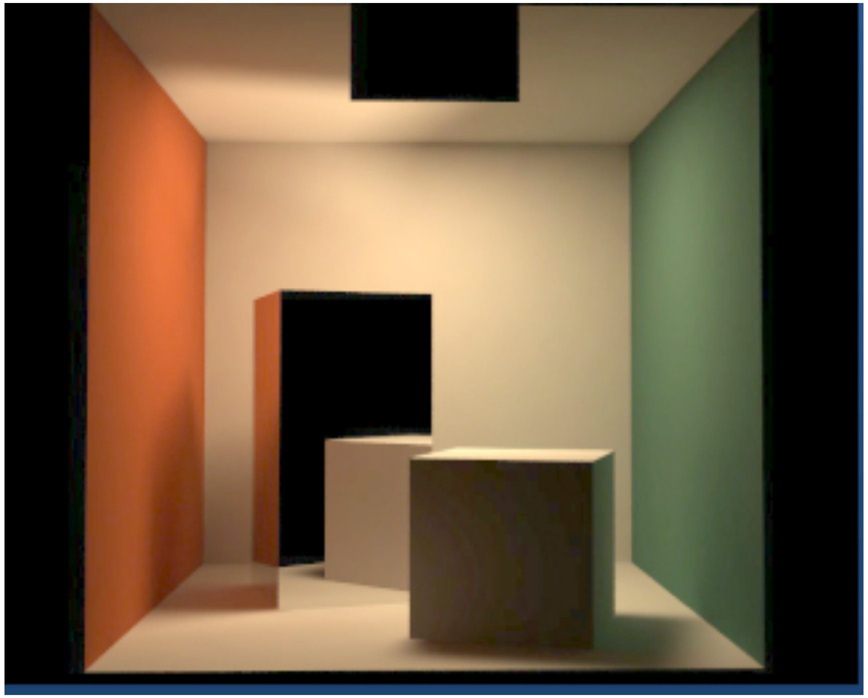
Psychophysical Methods

- **Visual Psychophysics provides a rigorous set of methods for:**
  - Evaluating the subjective visual appearance of images
  - Measuring the fidelity of a rendering, including detectability of shortcuts or approximations
  - Specifying perceptually meaningful parameters for systems

Cornell Box

real

simulated



Meyer, Rushmeier, Cohen, Greenberg and Torrance (1986). "An Experimental Evaluation of Computer Graphics Imagery." *ACM Transactions on Graphics*

Do HDR monitors support LDR content ?

Ahmet Oguz Akyuz, Roland W. Fleming, Bernhard Riecke,
Erik Reinhard & Heinrich H Bülthoff

SIGGRAPH2007

- HDR display technology is poised to hit the mass market
- Main limitation: “What do I do with my old (LDR) movies and photos ?”

wide aperture

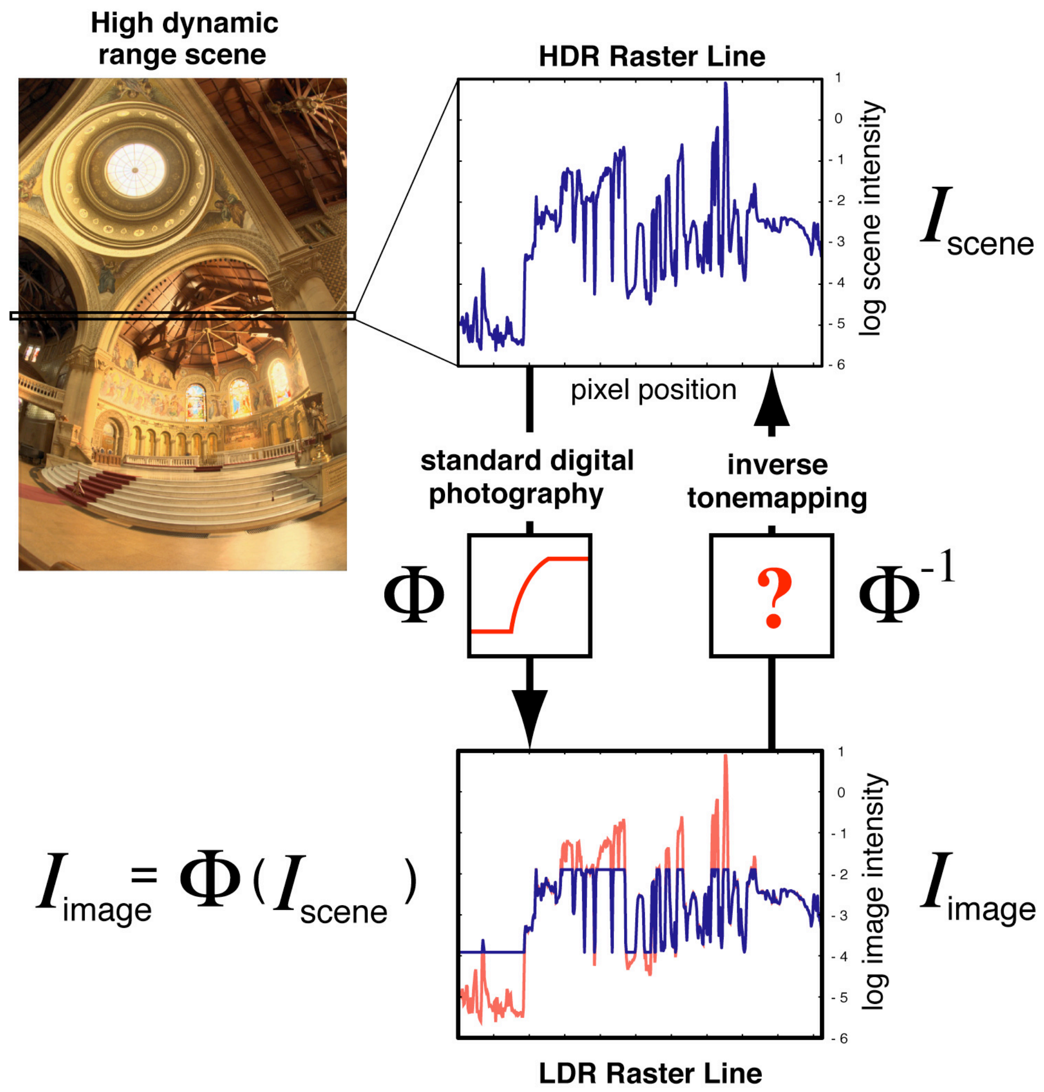
narrow aperture

tonemapped HDR

Do HDR monitors support LDR content ?

Ahmet Oguz Akyuz, Roland W. Fleming, Bernhard Riecke,
Erik Reinhard & Heinrich H Bülthoff

- "Inverse Tonemapping"
- A set of very tricky, mathematically ill-posed problems
- **Idea:** Do some psychophysics to set the specifications for inverse tonemapping



Do HDR monitors support LDR content ?

Ahmet Oguz Akyuz, Roland W. Fleming, Bernhard Riecke,
Erik Reinhard & Heinrich H Bülthoff

- Subjective ratings for :
  - Realism
  - Visual Appeal
  - Depth
- HDR is generally rated more appealing and more realistic than standard imagery
- **Surprise 1**: tonemapped images are often no better than the best single exposure from a bracketed sequence
- **Surprise 2**: Simple linear inverse tonemapping can produce images comparable to true HDR content

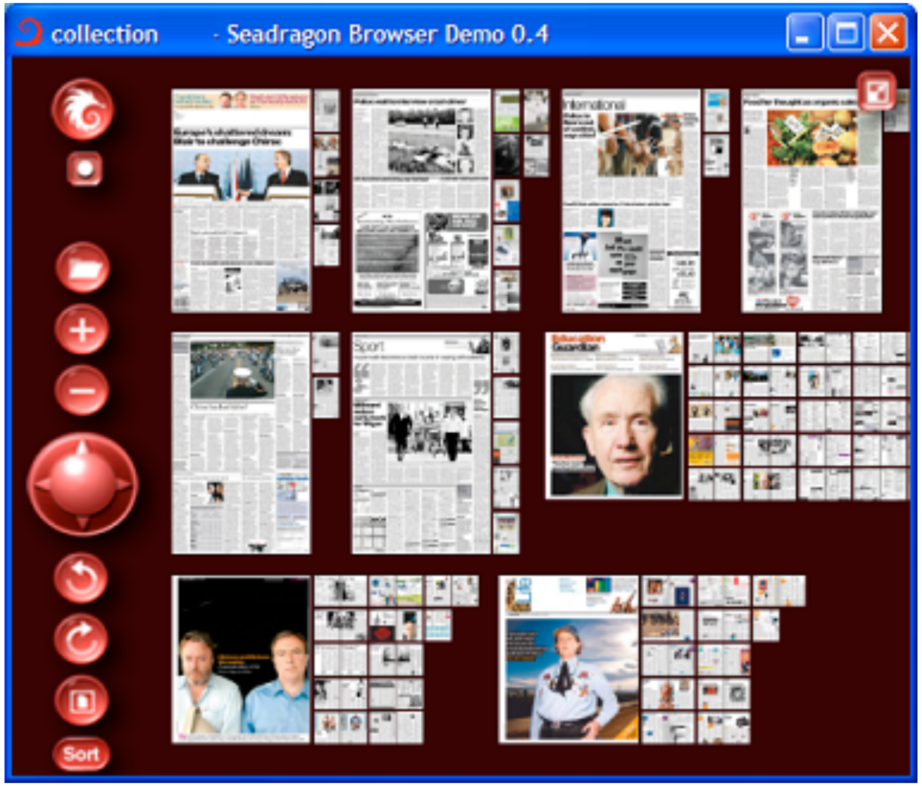
Brightness is more important than contrast in the HDR experience

Outline

- **Exploiting limits of human perception to facilitate graphics**
  - BW-FIT Gaze-Contingent Display
  - Change blindness demo
  - Explanation of top-down attention and bottom-up salience

- **Exploiting ambiguity in visual perception**
  - Visual Perception as inference, with some example ambiguities
  - Image-based Material Editing

Ultra-high resolution:
the future of digital imagery

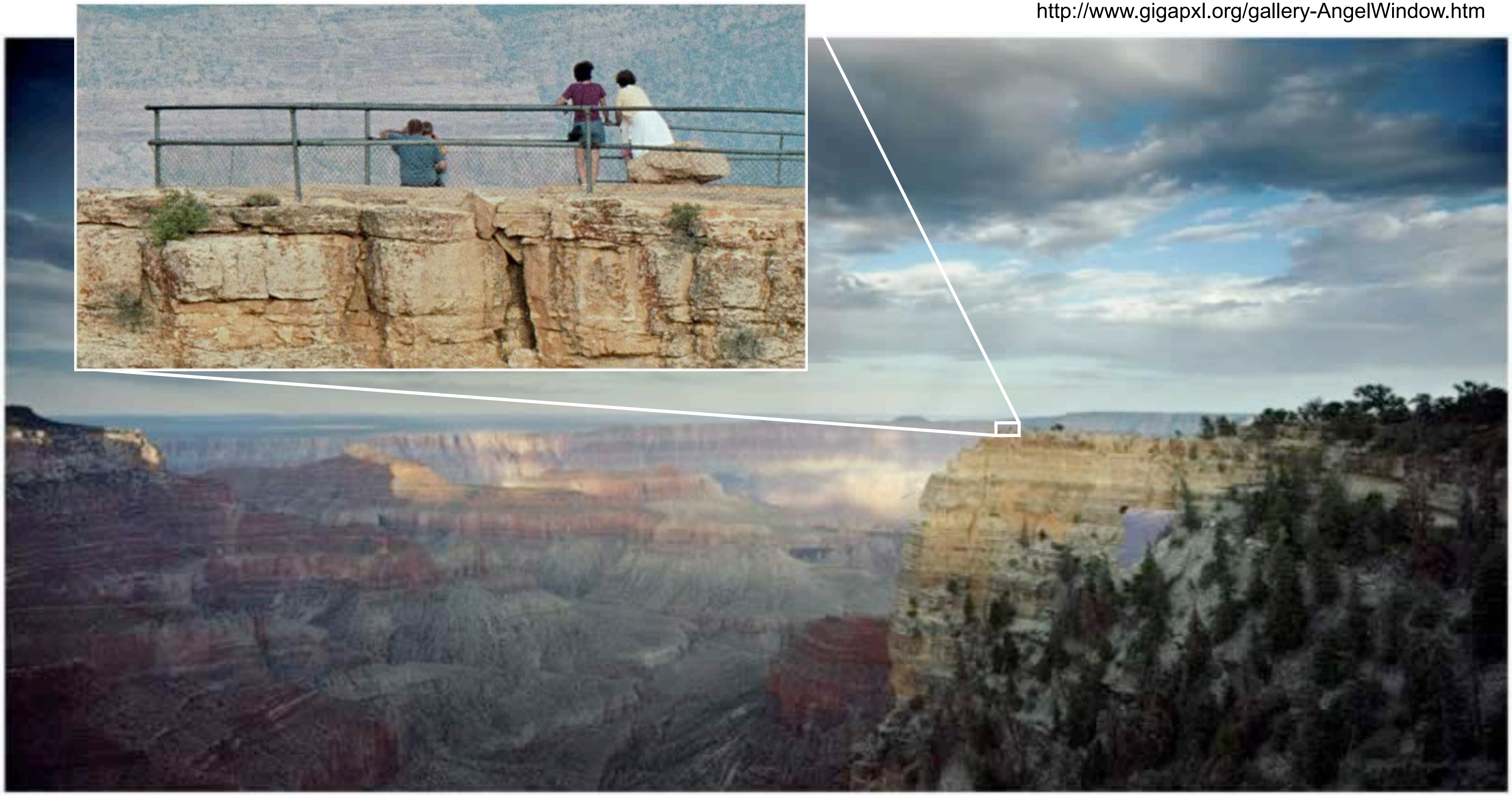


Microsoft Live Labs Seadragon

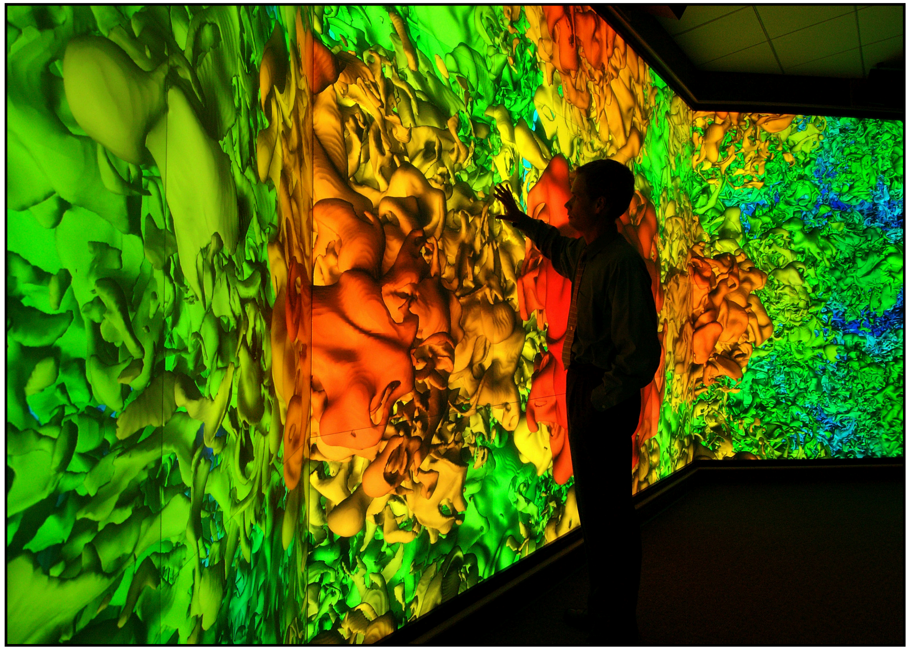
Microsoft Research HD View Beta

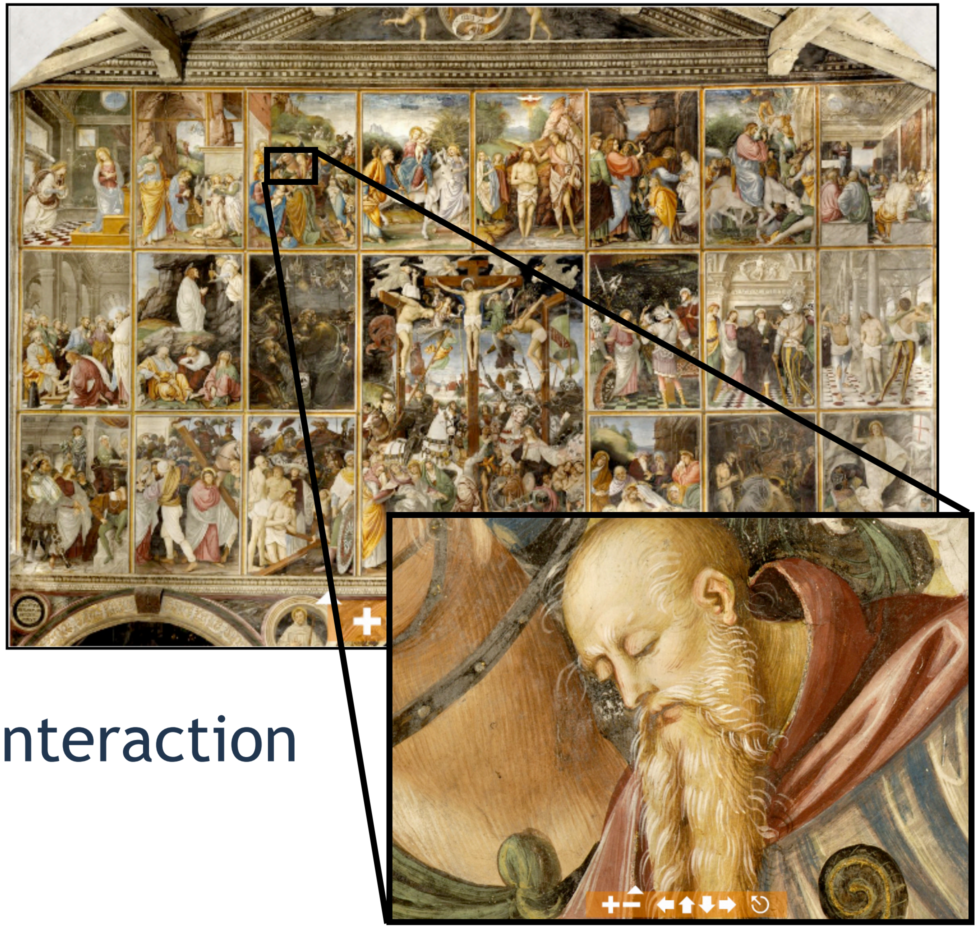
Ultra-high resolution:
the future of digital imagery

http://www.gigapxl.org/gallery-AngelWindow.htm



Gigapixel displays





8.5 Gigapixel image

http://haltadefinizione.deagostini.it/

- Enables new modes of user interaction
- Applications in:
  - Industrial data visualization
  - Scientific and medical imaging
  - Entertainments industry

Ultra-high resolution:
the future of digital imagery

Sony 4K Projector

The Red One:
4K film camera

Pixel-mania

Samsung SCH-B600

Seitz Roundshot D3

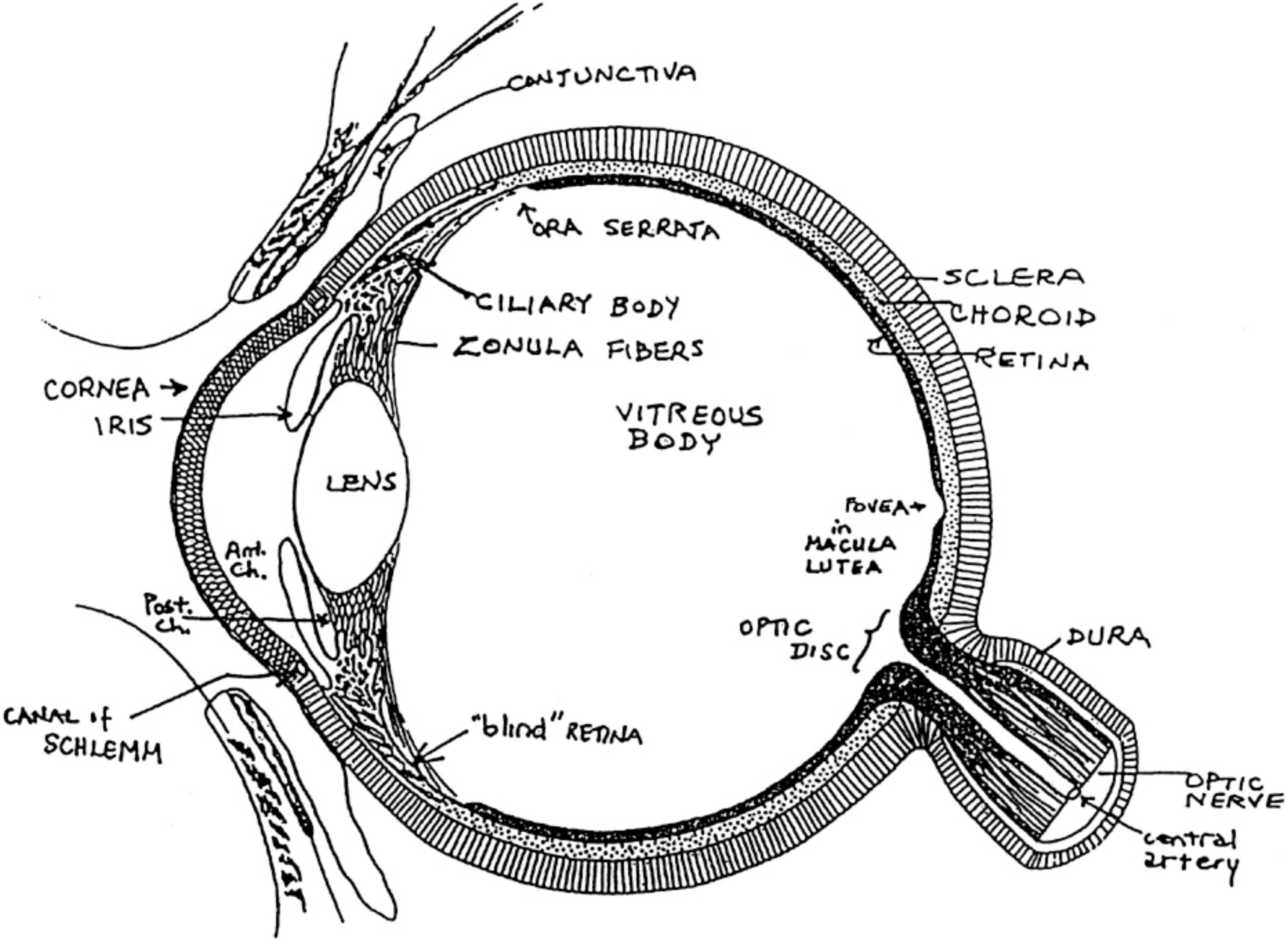
Pixel-mania

But how many pixels can the human eye actually see?

http://www.roundshot.ch/

Seitz Roundshot D3

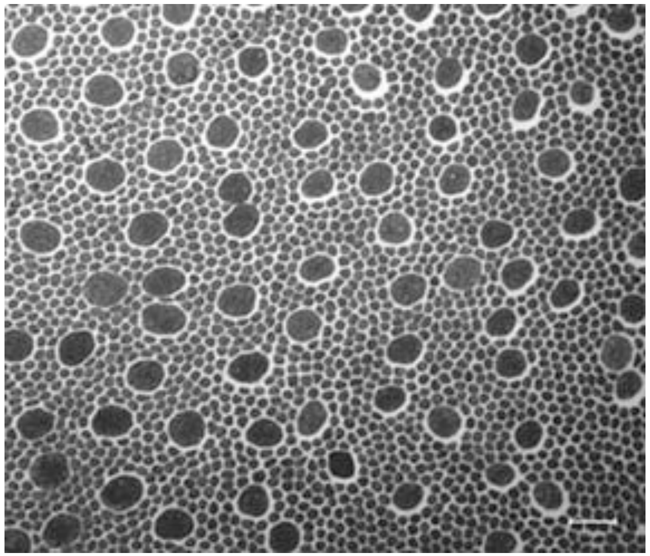
Inside the human eye



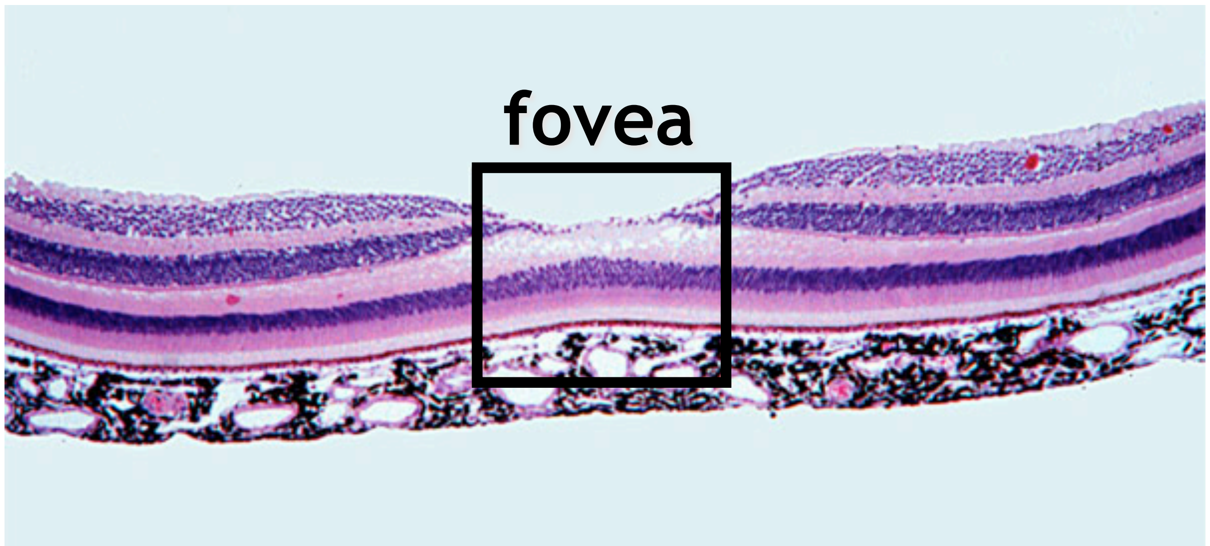
The retina

- The retina is like the **photographic film** of the human eye.
- The **fovea** is a special high-acuity “hot spot”

Photoreceptors:
The pixels of the eye



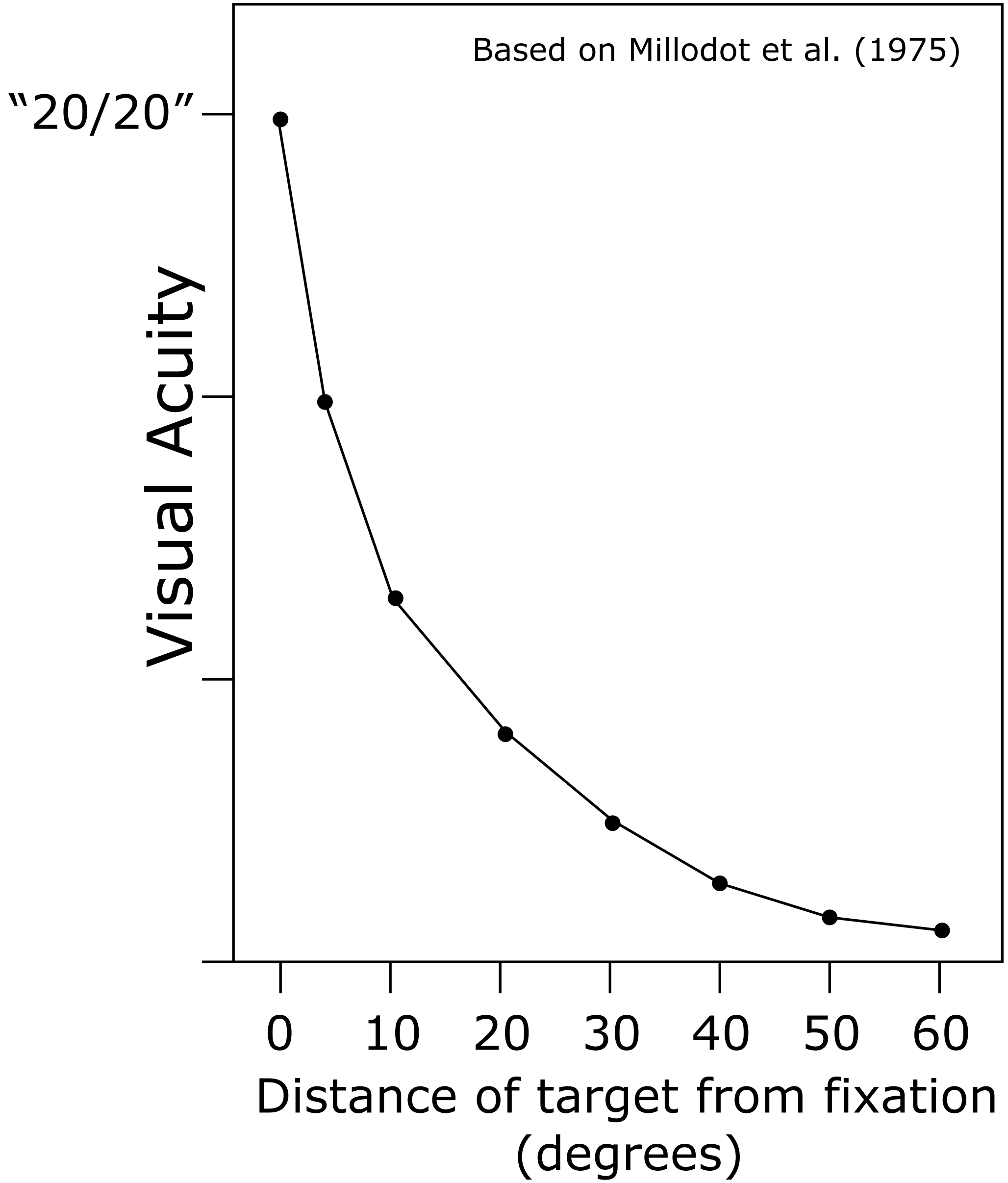
Cross-section through
the retina



Visual acuity

- The resolution of human vision **falls off dramatically** as you move away from the foveal "hot-spot".

- Your impression that the whole world is sharp and high resolution is really an **illusion**!

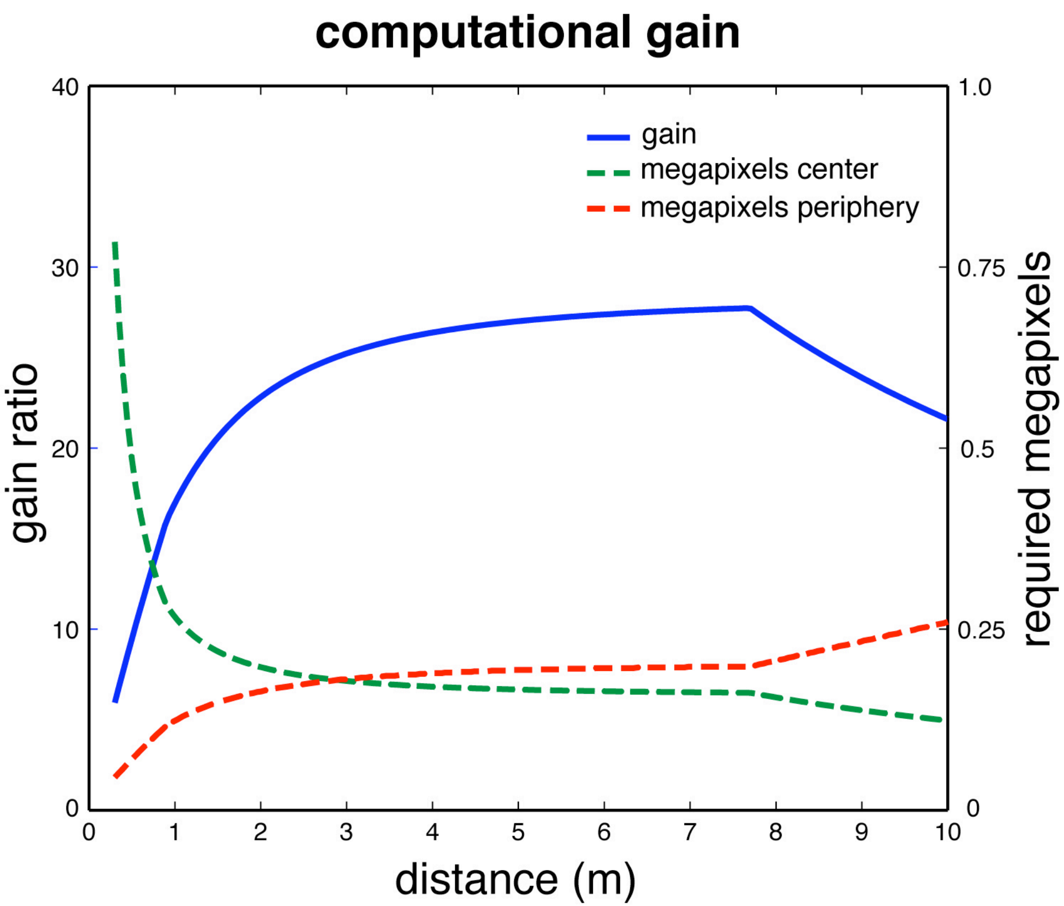


So, how many pixels?

- The number of pixels you can actually see at any single instant is **surprisingly low:** ~ 1Mpixel !

- If you are very close to the screen, you can see the pixels in front of you, but not in the periphery

- If you are far enough away to see the edges of the screen, you can’t resolve the individual pixels any more



Key idea

Modify the resolution of the display depending on where the viewer is standing and where they are looking: **"Gaze-contingent display"**

- **Technical Challenge:** Track the viewer's eyes, head and body to work out where they are and what they are looking at.
- **Why?** Enormous computational savings
- **Why?** Massively enhances user interaction

Multi-resolution display

- High resolution where the viewer is looking
- Low resolution everywhere else
- Seamless illusion of very high visual fidelity, but at low computational load.

Multi-resolution display

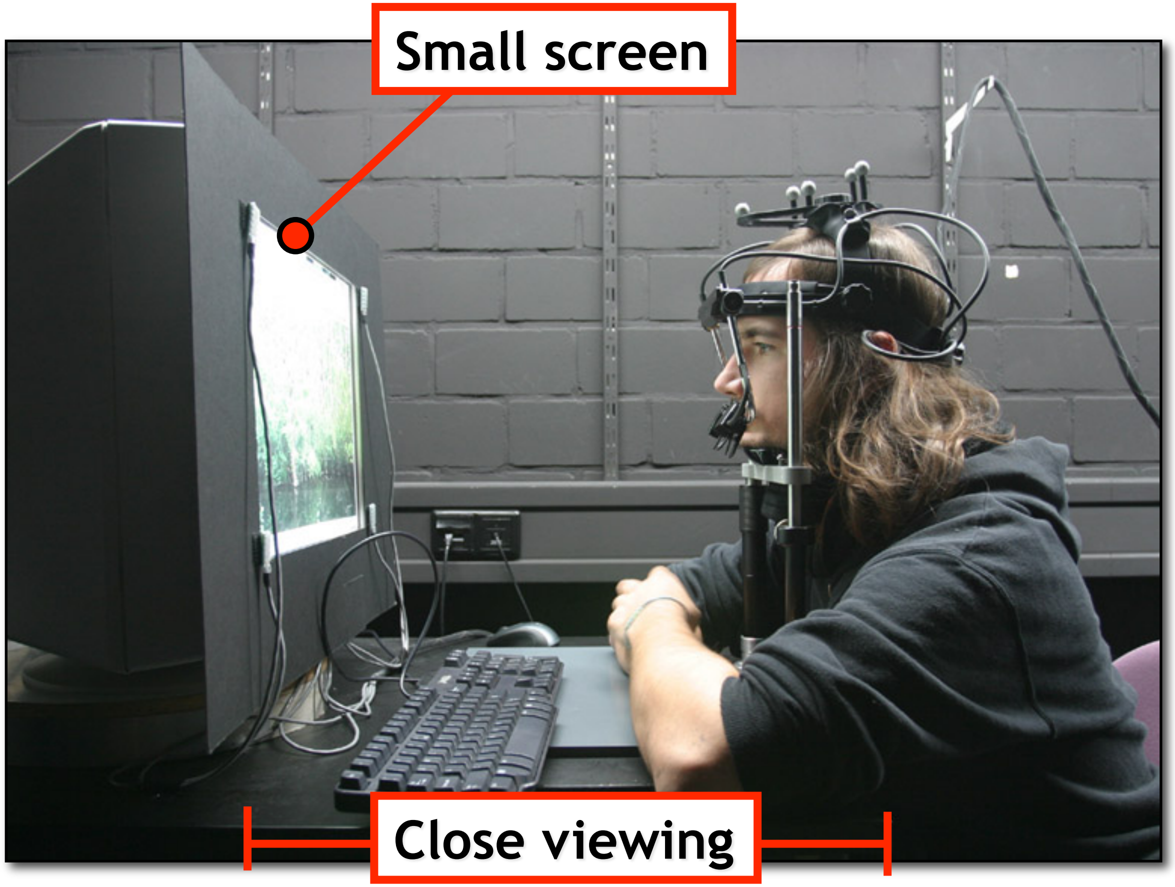
- High resolution where the viewer is looking
- Low resolution everywhere else
- Seamless illusion of very high visual fidelity, but at low computational load.

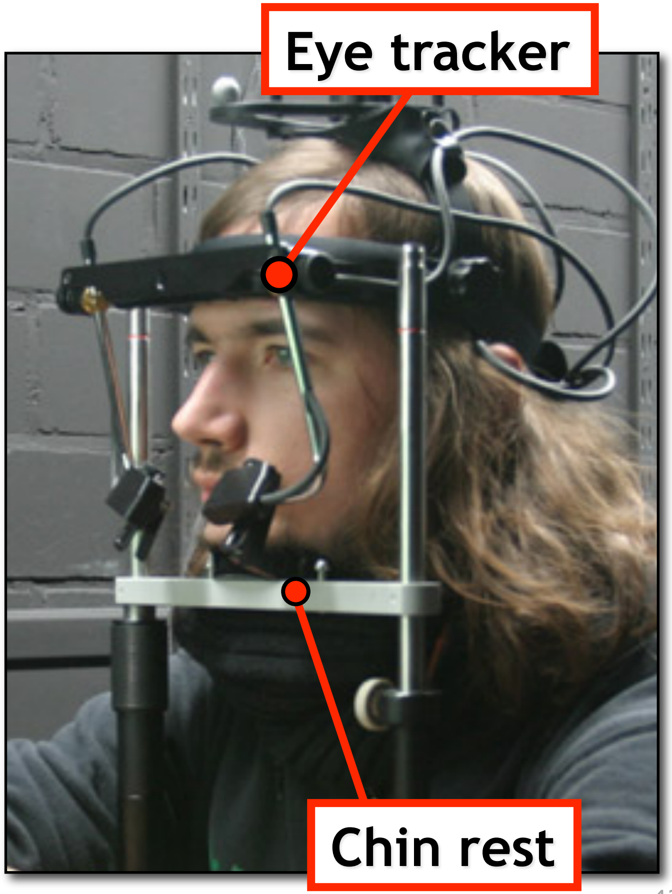
Multi-resolution display

- High resolution where the viewer is looking
- Low resolution everywhere else
- Seamless illusion of very high visual fidelity, but at low computational load.

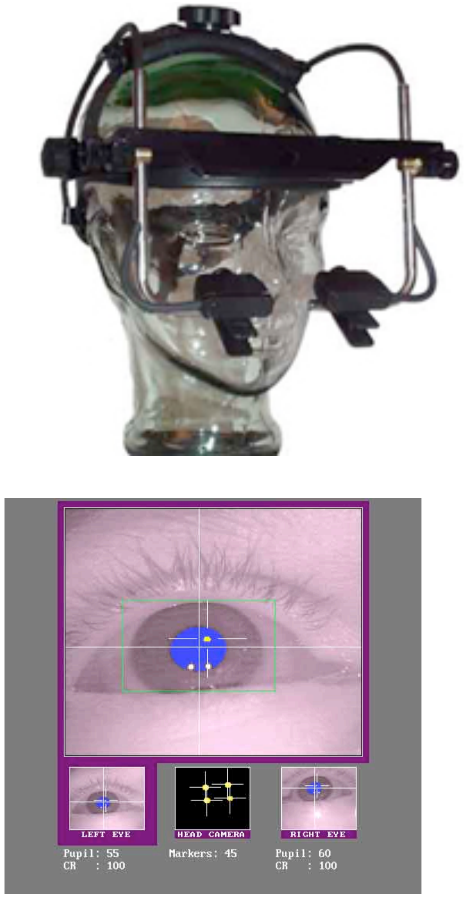
Traditional eye-tracking

- User must be seated with fixed head position
  - Static, unnatural viewing conditions
  - Limited field of view





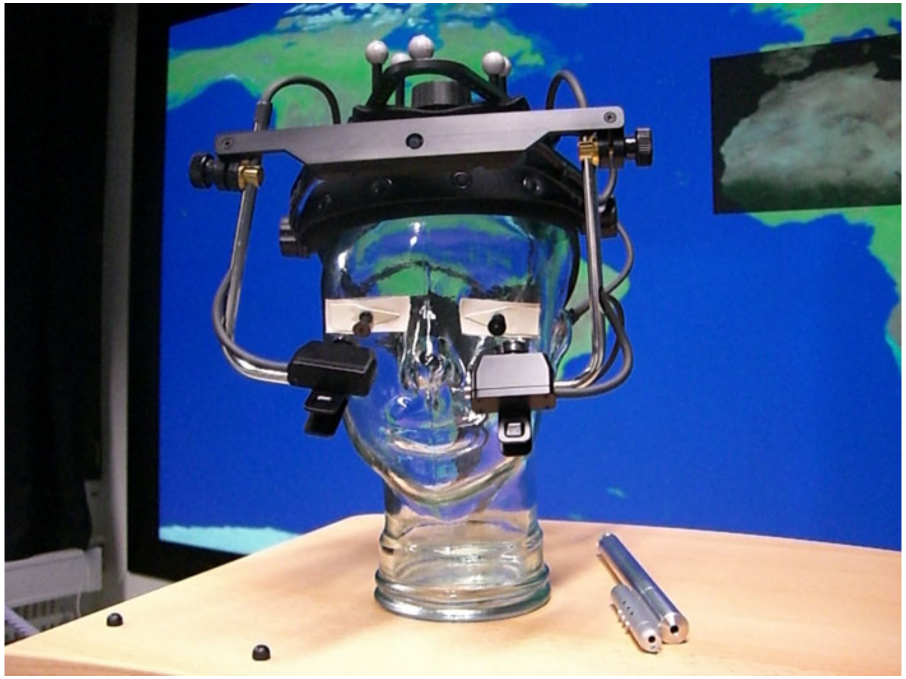
Technology



High speed eye tracker

Infra-red position trackers

Technology



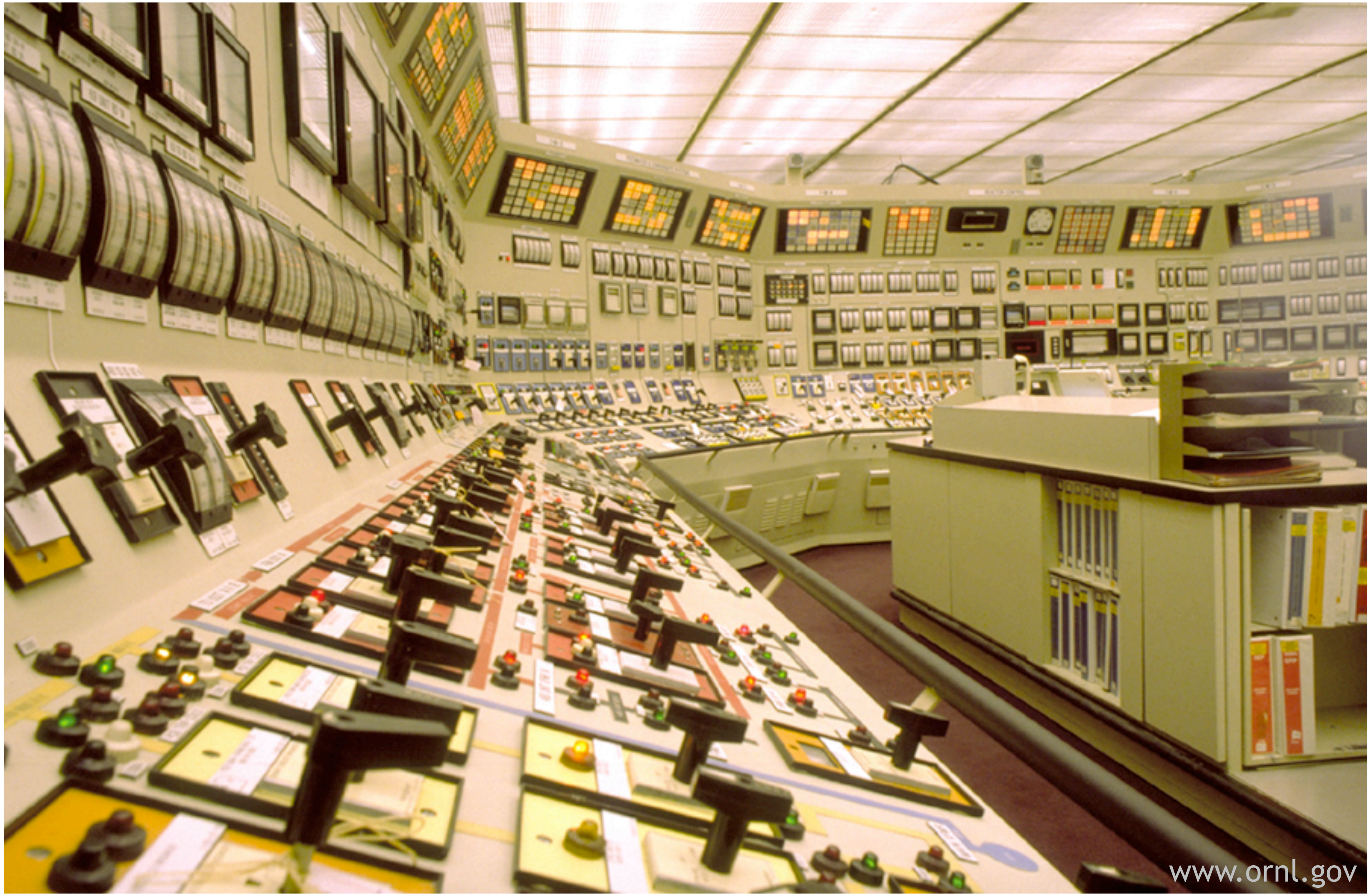
Combined eye- and
head-tracking system

System in use with a
freely moving user

- The user can **move around, explore** and **interact with** the data on the screen.

information at your fingertips

Visual Attention



Change Blindness

Ron Rensink, UBC Vancouver

Change Blindness – Trans-saccadic Memory

Ron Rensink, UBC Vancouver

Change Blindness

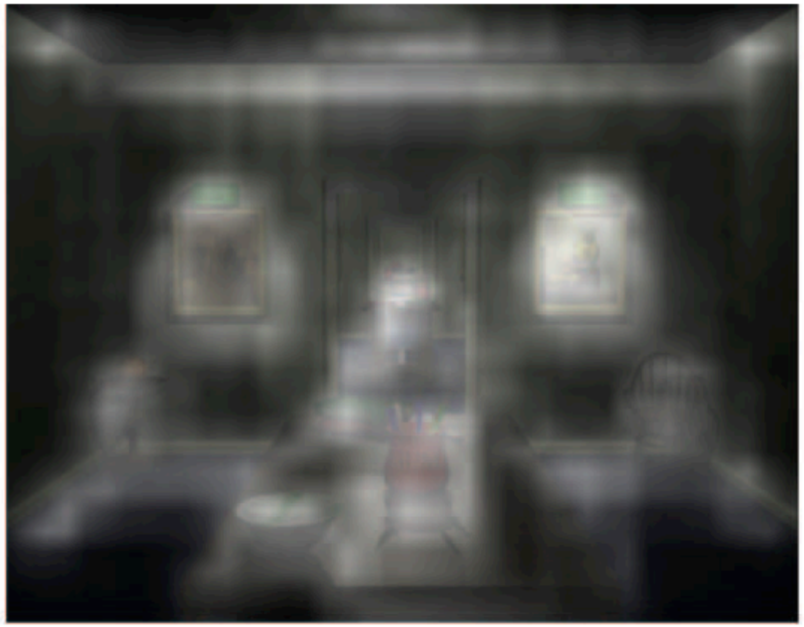
Ron Rensink, UBC Vancouver

Saliency and Rendering

Idea:

1. prerender a low-cost version of the scene
2. Evaluate salient (visually important) regions automatically
3. Render with high fidelity only where this is required

The 90th Frame



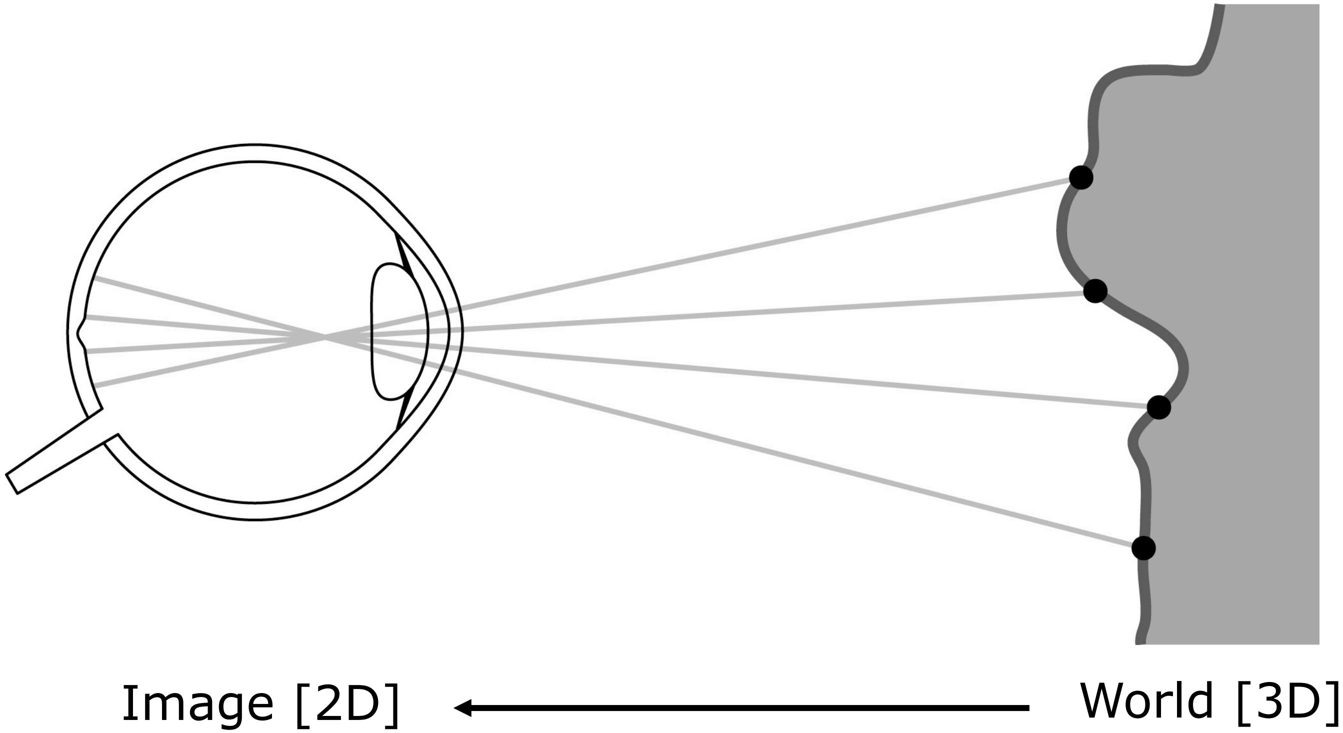
The Saliency map of the 90th Frame

Yang & Chalmers (2005).

Outline

- Exploiting limits of human perception to facilitate graphics
  - BW-FIT Gaze-Contingent Display
  - Change blindness demo
  - Explanation of top-down attention and bottom-up salience
- **Exploiting ambiguity in visual perception**
  - Visual Perception as inference
  - Image-based Material Editing

Perception as Inference

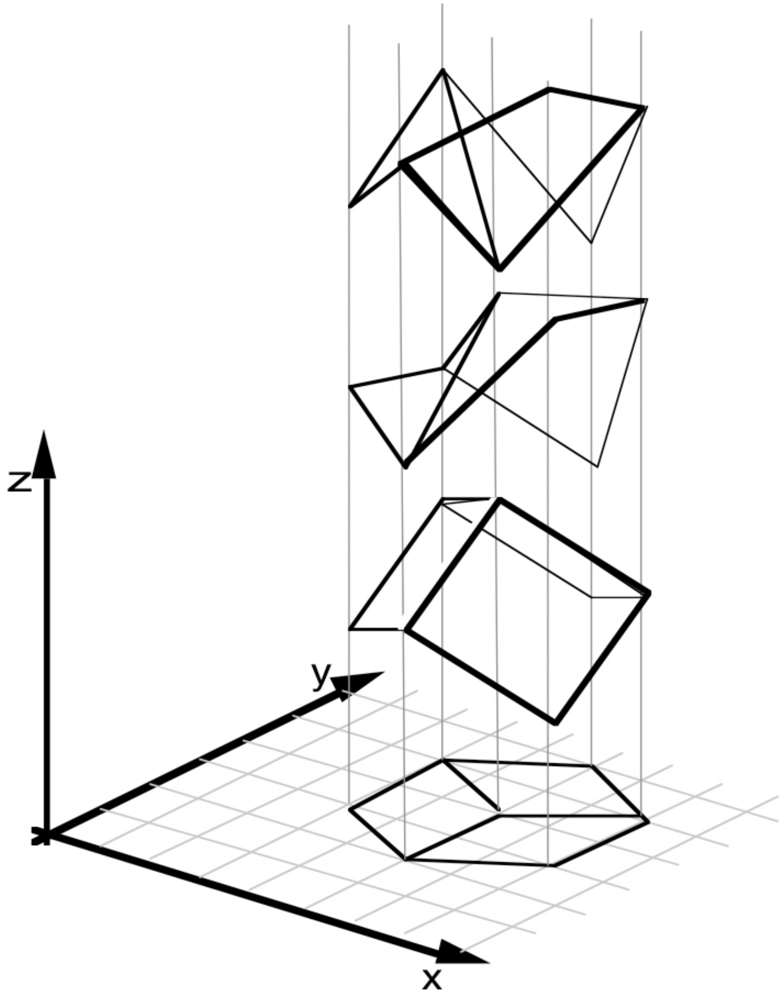


- The optics of the eye project the 3D world onto a 2D image plane on the retina.
- What we as behaving organisms care about is the 3D structure of the world. Unfortunately the projection from 3D to 2D is not invertible.

Perception as Inference

- **Goal of human visual perception:**
  - Reconstruct the complex, dynamic, 3D world around us.

- **Problem:**
  - Mathematically ‘ill-posed’
  - Retinal image is ambiguous

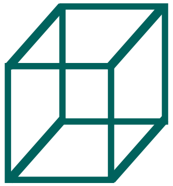


Perception is under-constrained

- Vision is an inverse problem
  - not well-posed in the sense of Hadamard (1902)
    - a solution exists
    - the solution is unique
    - the solution depends continuously on the data

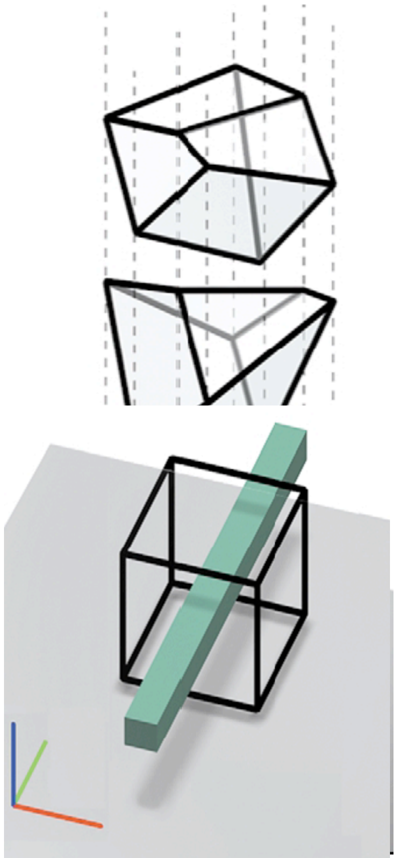
- A single image with many interpretations

The Necker Cube



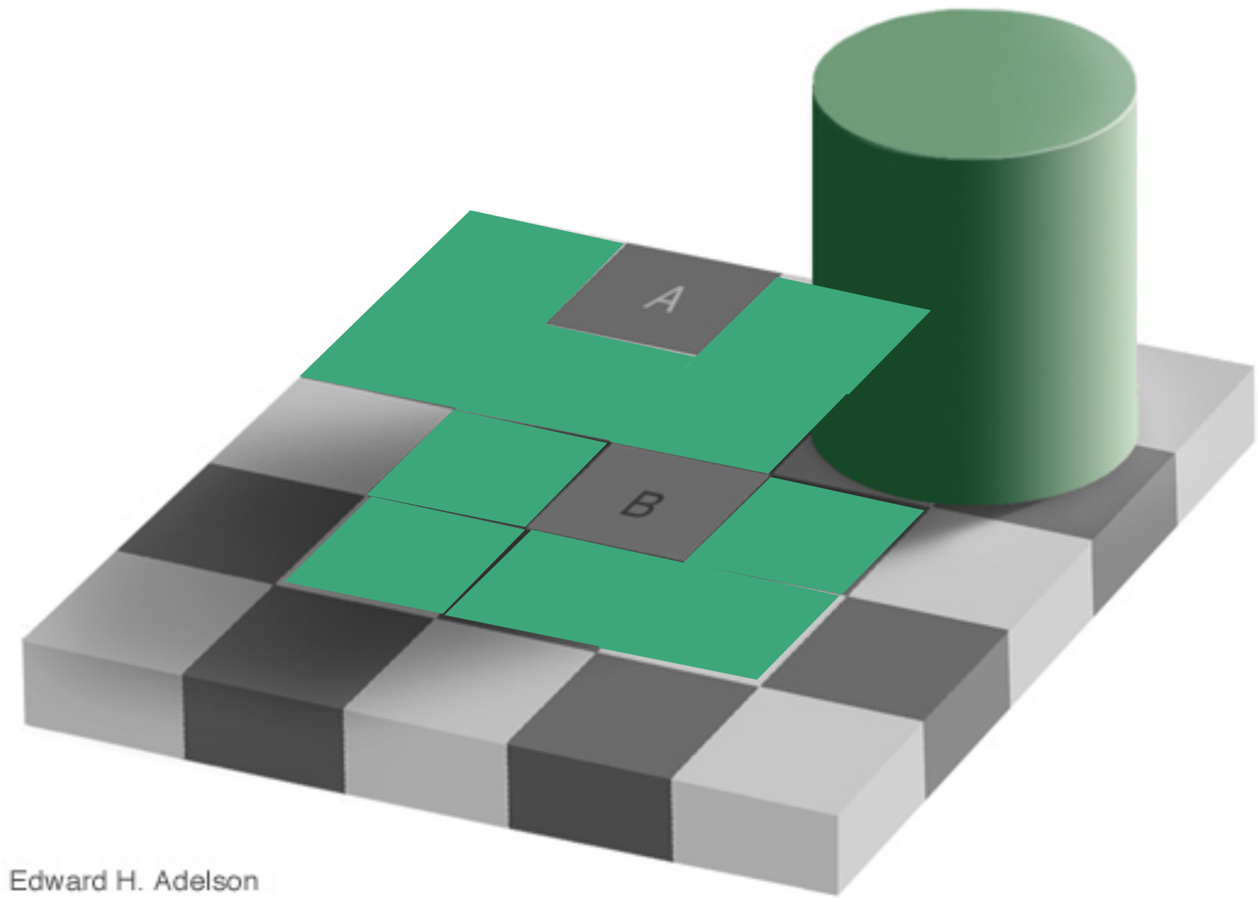
- The brain makes assumptions about the world in order to solve the inverse problem of vision
  - small angular variations
  - planar surfaces
  - compact form

- A single cue can disambiguate the 3D structure

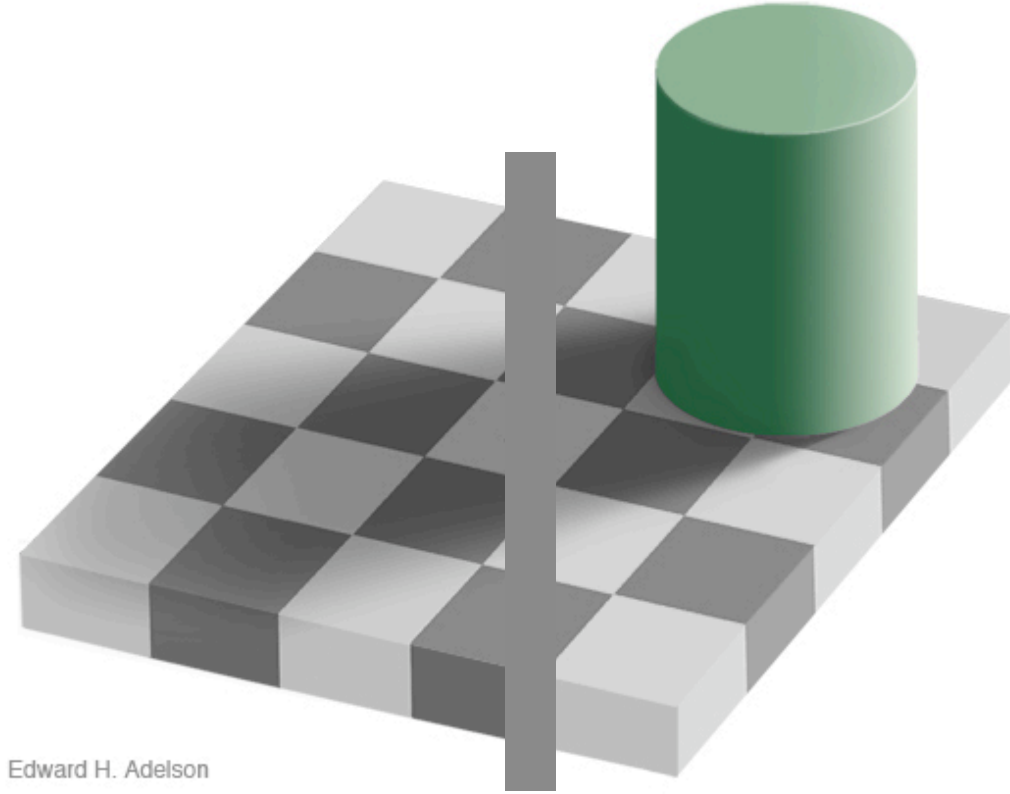


Ernst & Bülthoff, TICS ***8(4)***, 162-169 (2004)

Perception is not always “true”



Visual Illusions

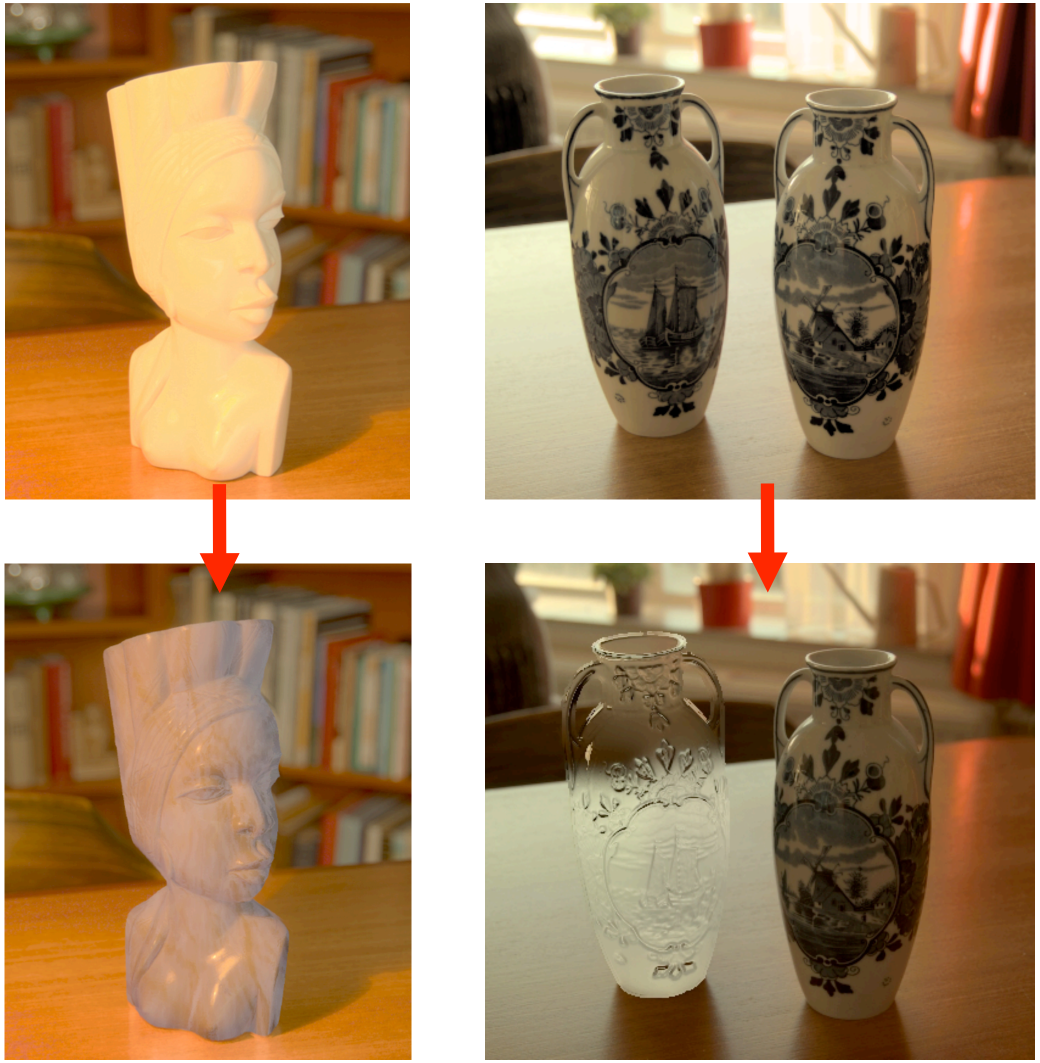


- *As with many so-called illusions, this effect really demonstrates the success rather than the failure of the visual system.*
- *Of course, it makes sense if the visual system is trying to infer properties of the world --- that is the shade of paint --- instead of trying to measure properties of the retinal image --- that is the retinal intensity ---, which are of no survival value.*

Image-based material editing

Erum Kahn, Erik Reinhard, **Roland Fleming** & **Heinrich Bülthoff**

- Given single photograph as input, modify material appearance of object.
- Physically correct solution not possible: aim for ‘perceptually correct’ solution.
- Exploit assumptions of human visual to develop heuristics.



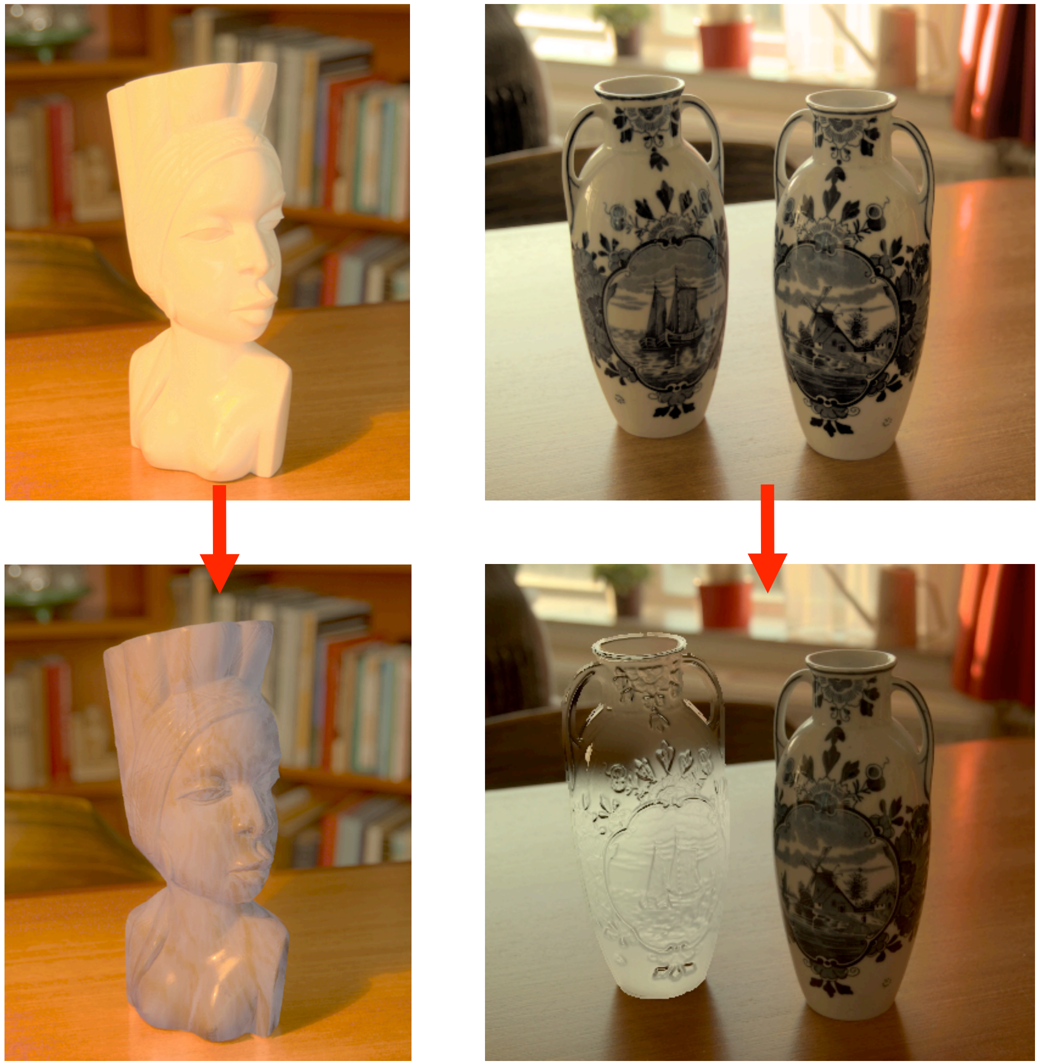
re-textured **transparency**

Presented at SIGGRAPH2006

Image-based material editing

Erum Kahn, Erik Reinhard, **Roland Fleming** & **Heinrich Bülthoff**

- Re-texturing
- Medium gloss to matte or glossy
- Opaque to transparent or translucent
- Arbitrary BRDFs



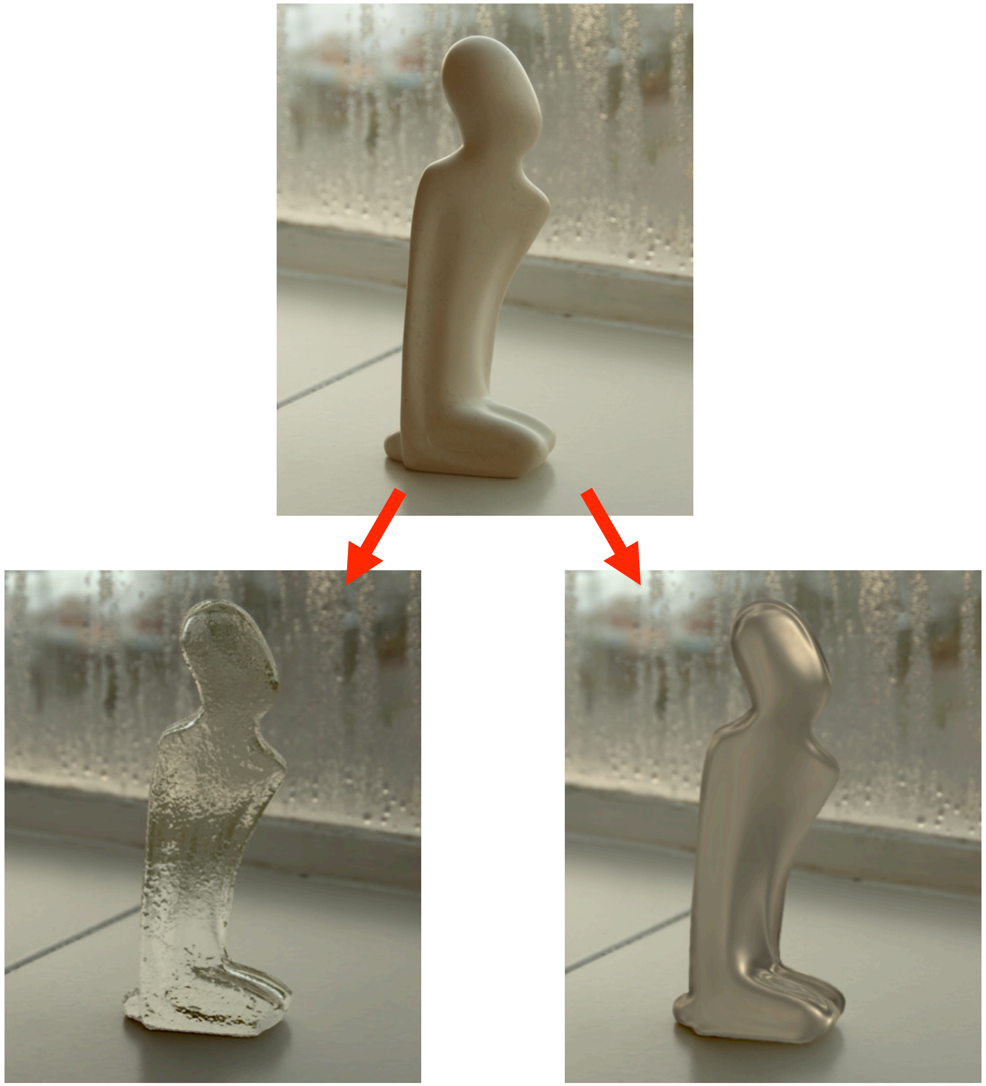
re-textured **transparency**

Presented at SIGGRAPH2006

Image-based material editing

Erum Kahn, Erik Reinhard, **Roland Fleming** & **Heinrich Bülthoff**

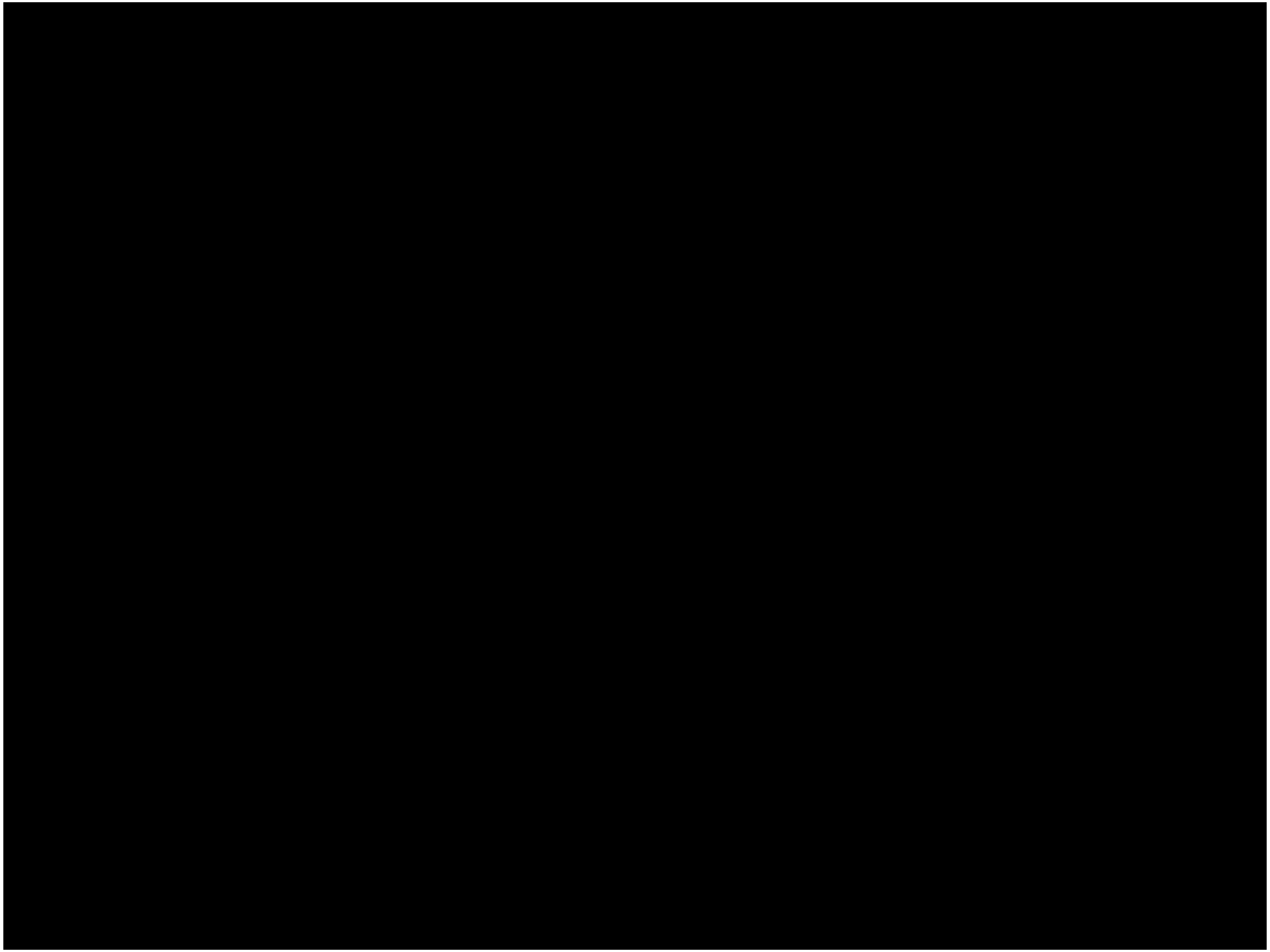
- Re-texturing
- Medium gloss to matte or glossy
- Opaque to transparent or translucent
- Arbitrary BRDFs



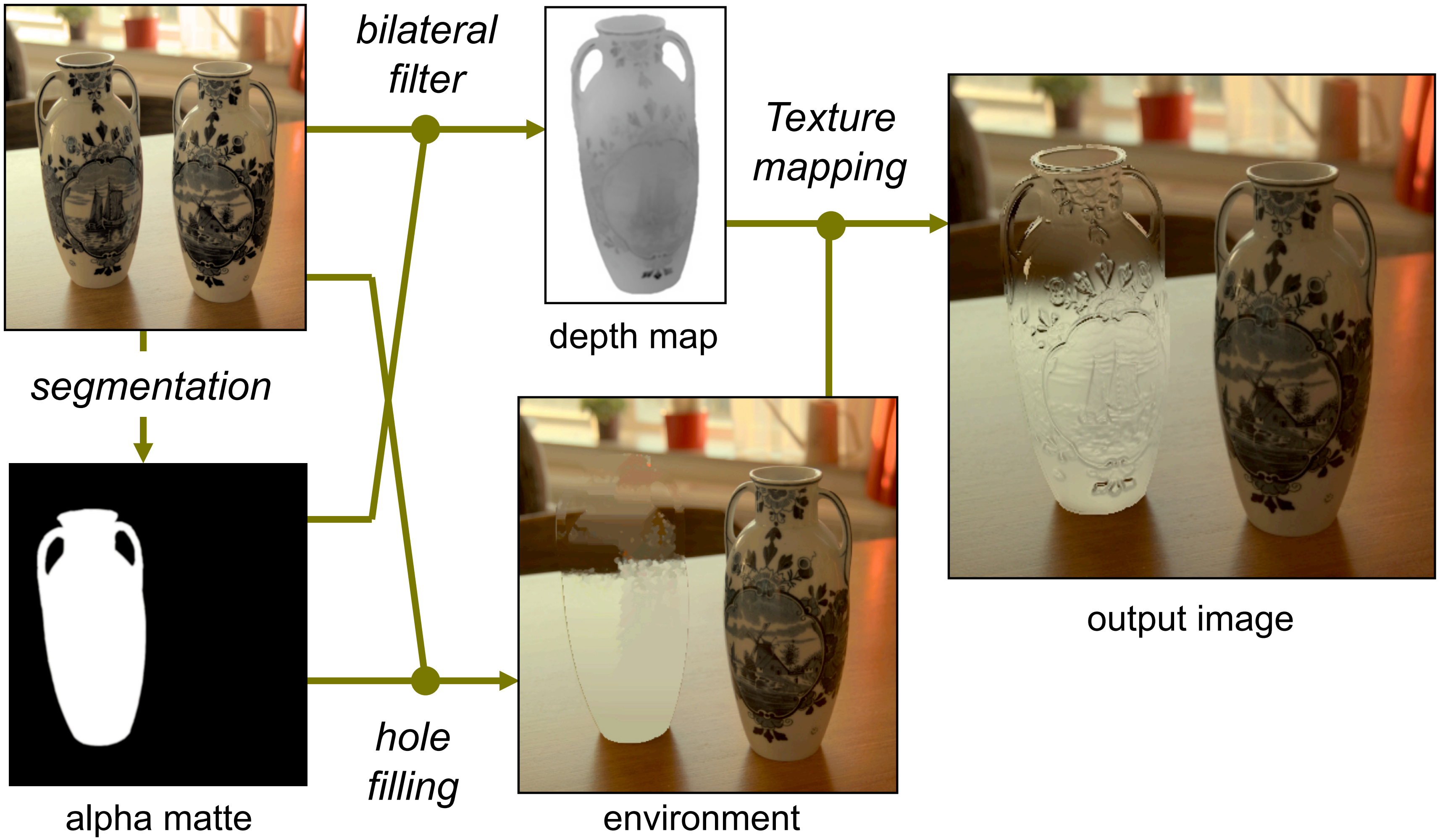
transparency

new BRDF

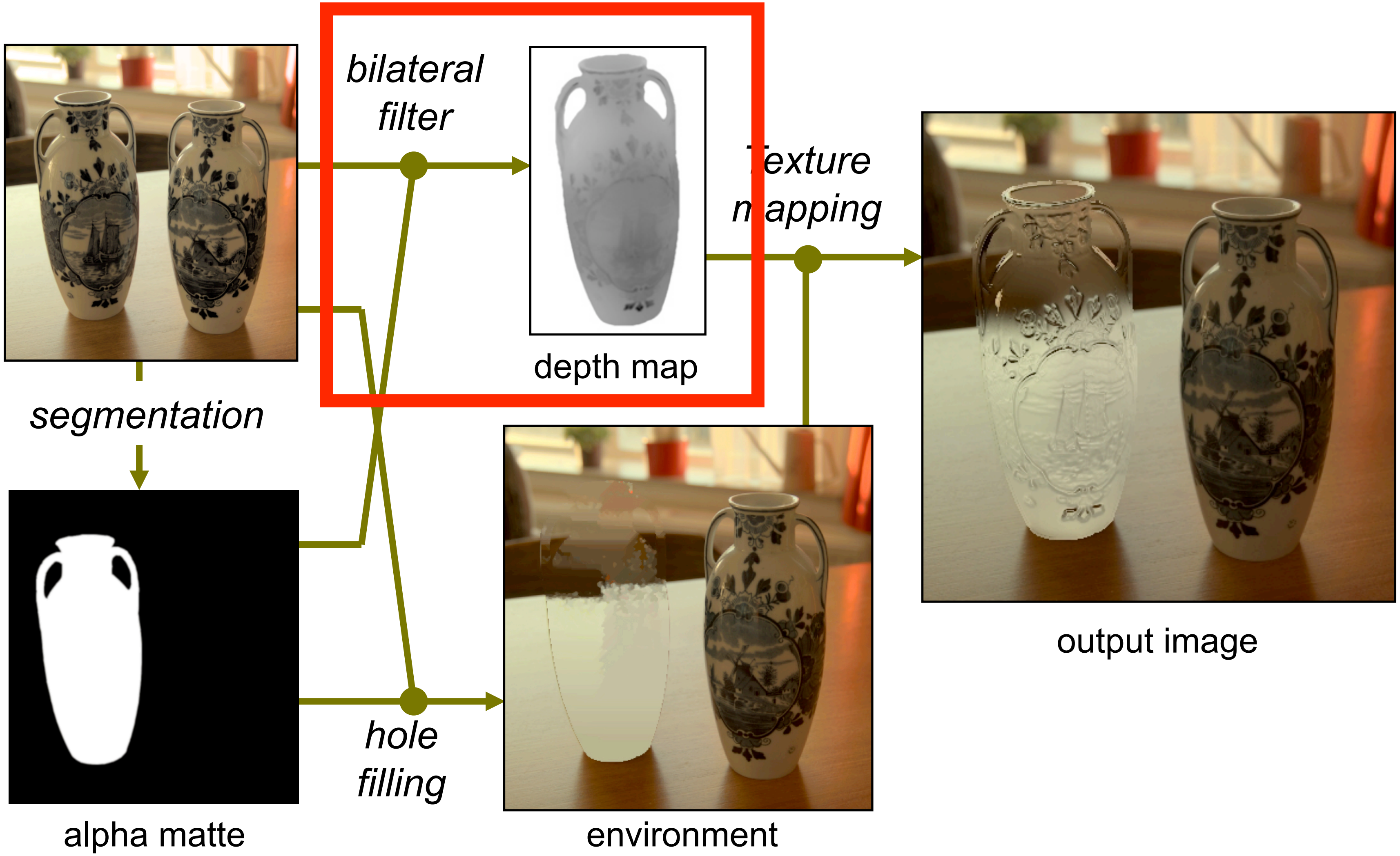
Presented at SIGGRAPH2006



Processing Pipeline

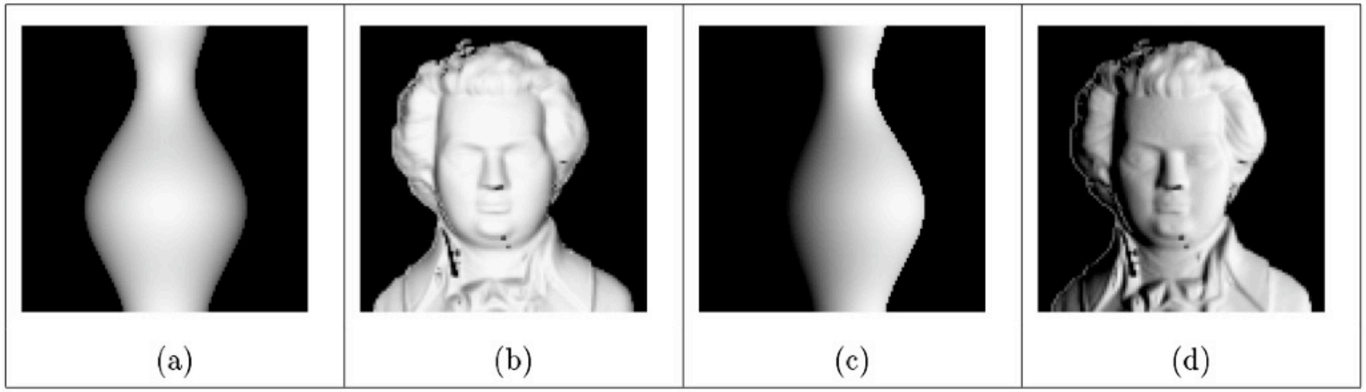


Processing Pipeline

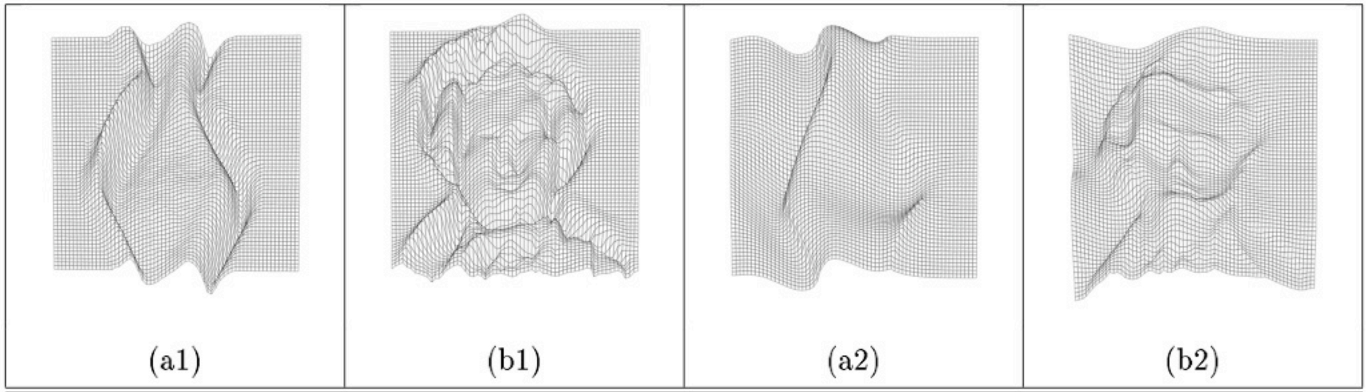


How **not** to do shape-from-shading

Try using the state-of-the-art algorithms and you will generally be disappointed!



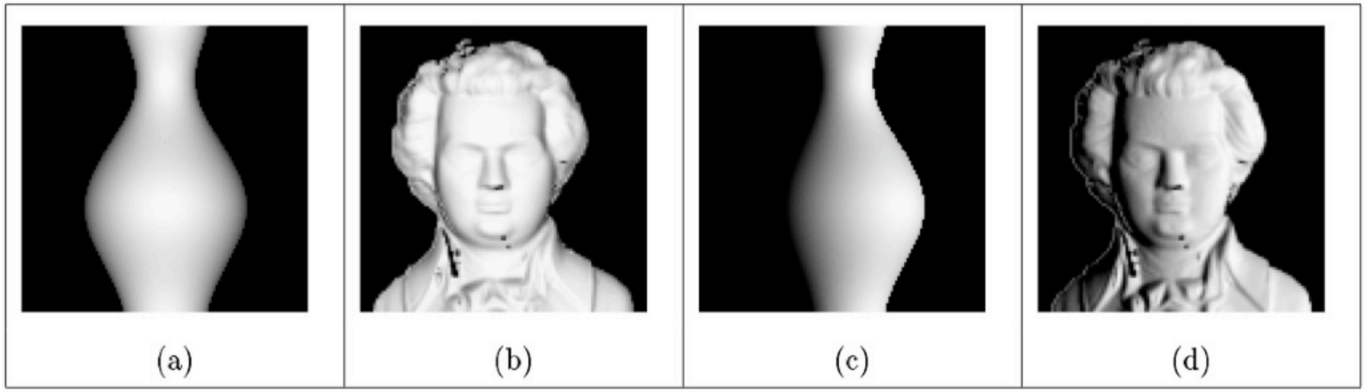
input



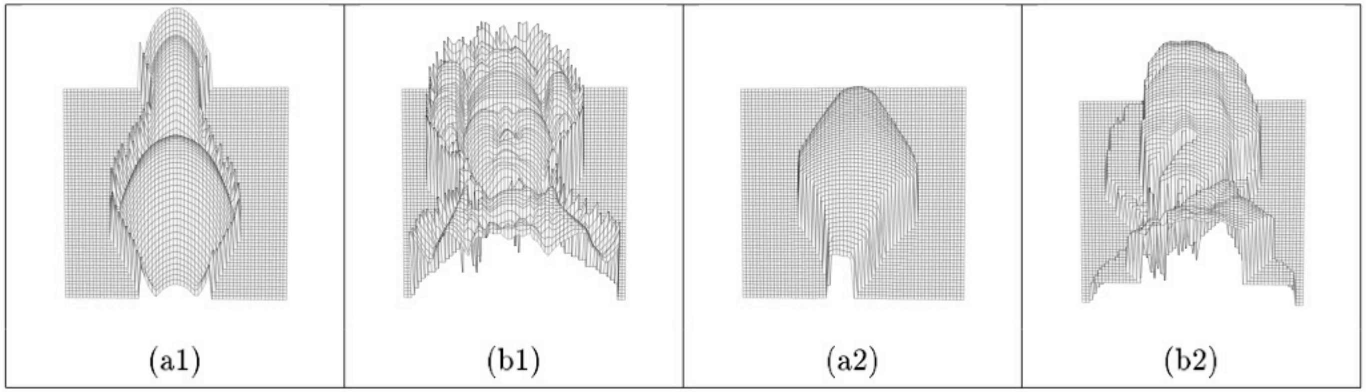
reconstruction

How **not** to do shape-from-shading

Try using the state-of-the-art algorithms and you will generally be disappointed!



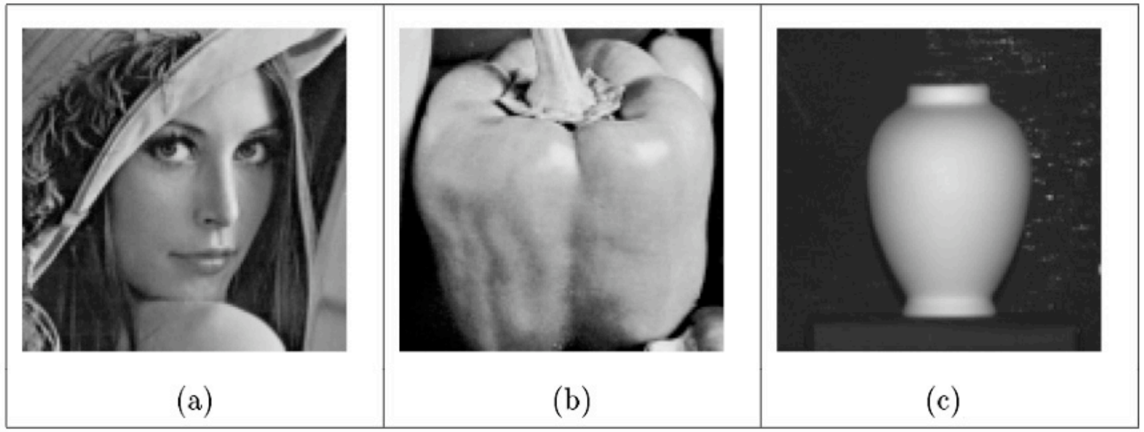
input



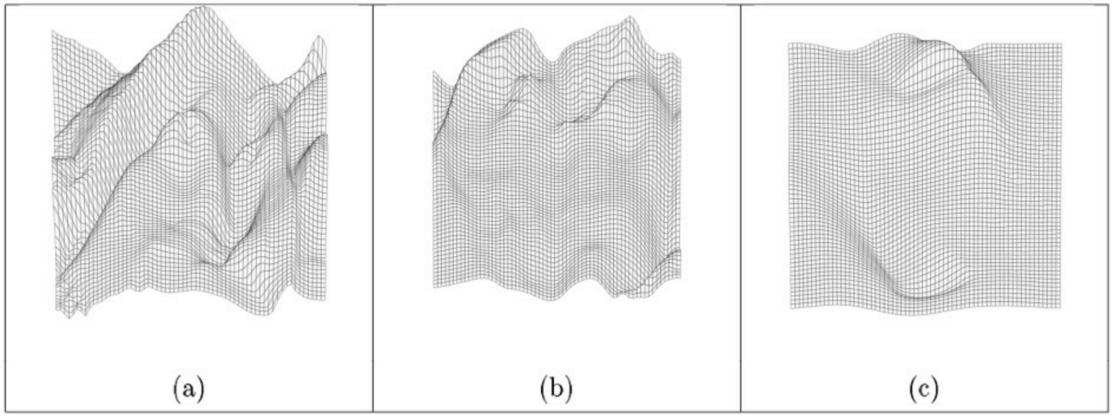
reconstruction

How **not** to do shape-from-shading

Try using the state-of-the-art algorithms and you will generally be disappointed!



input



reconstruction

We use a simple but suprisingly effective heuristic:

Dark is Deep

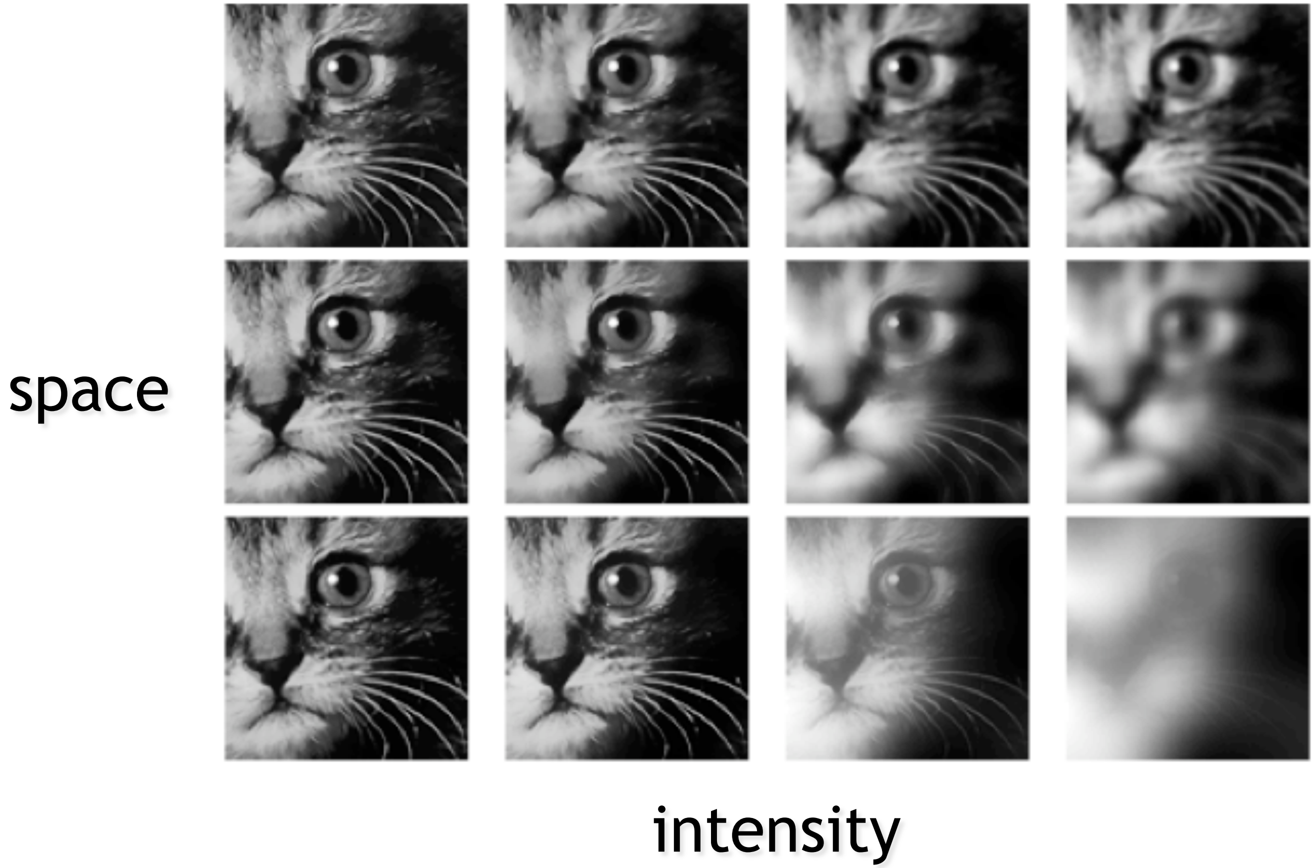
In other words …

$$z(x, y) = L(x, y)$$

input

initial depth estimate

Bilateral Filter



- The ‘recovered depths’ are conditioned using a **bilateral filter** (Tomasi & Manduchi, 1998; Durand & Dorsey, 2002).
- Simple non-linear **edge-preserving filter** with kernels in space and intensity domains.

Bilateral Filter:
3 main functions

- 1. De-noising depth-map
    - **Intuition**: depths are generally smoother than intensities in the real world.

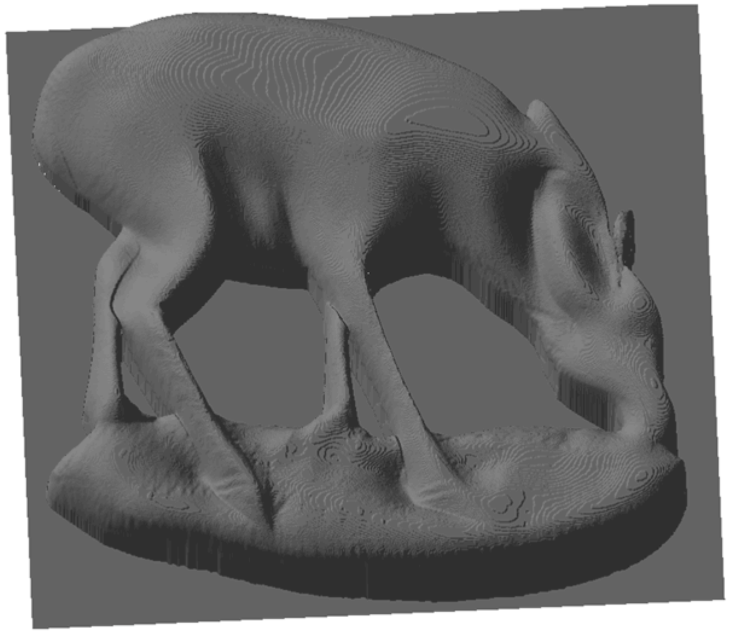
- 2. Selectively enhance or remove textures for embossed effect

Bilateral Filter:
3 main functions

- 3. **Shape-from-silhouette**, like level-sets shape 'inflation' (e.g. Williams, 1998)
  - **Intuition**: values outside object are set to zero, so blurring across boundary makes recovered depths smooth and convex.

Forgiving case

- Diffuse surface reflectance leads to clear shading pattern
- Silhouette provides good constraints



original **reconstructed depths**

Difficult case

- Strong highlights create large spurious depth peaks
- Silhouette is relatively uninformative

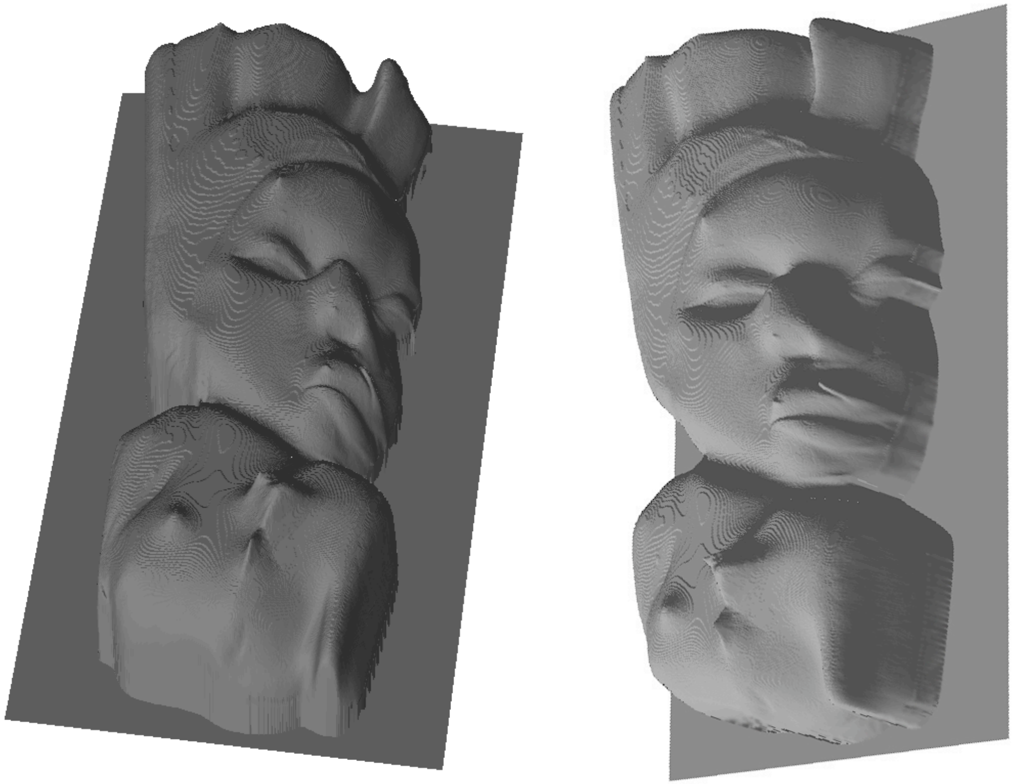
original

reconstructed depths

Light from the side

- Shadows and intensity gradient leads to substantial distortions of the face

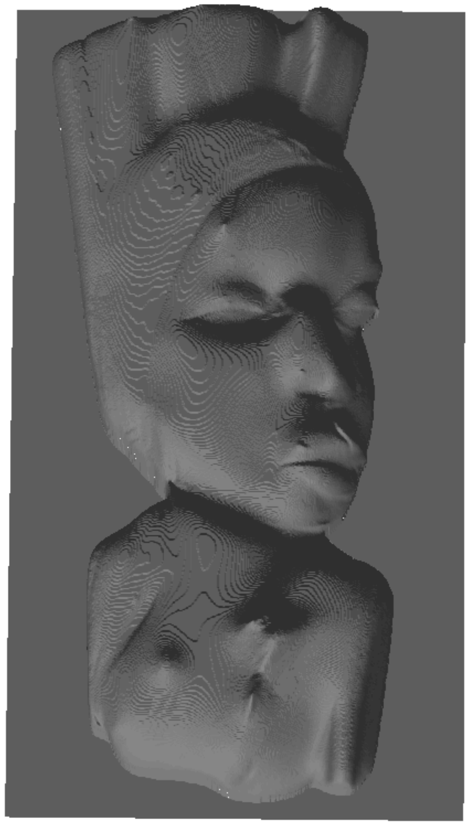
original



reconstructed depths

Importance of viewpoint

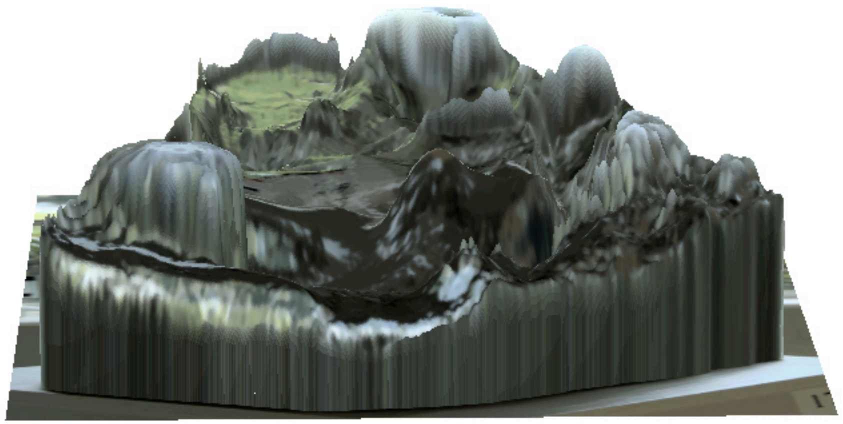
- Substantial errors in depth reconstruction are not visible in transformed image



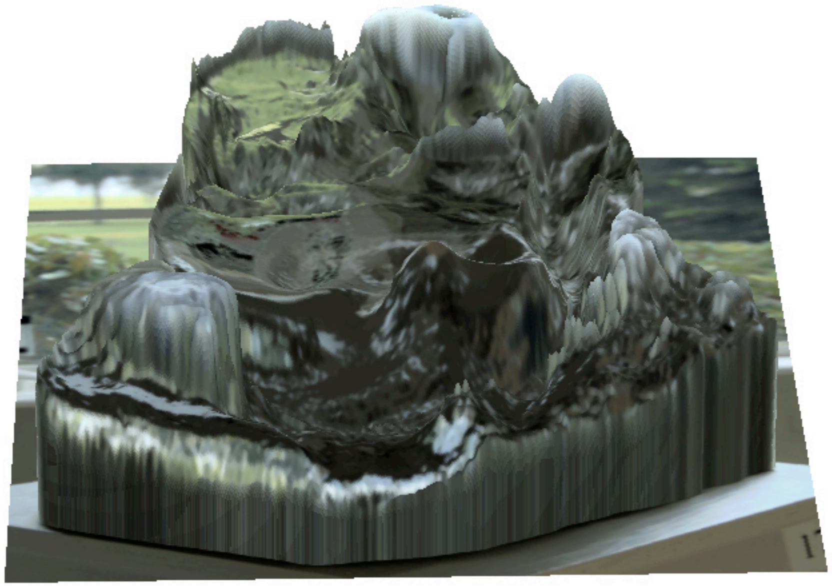
correct viewpoint

transformed image

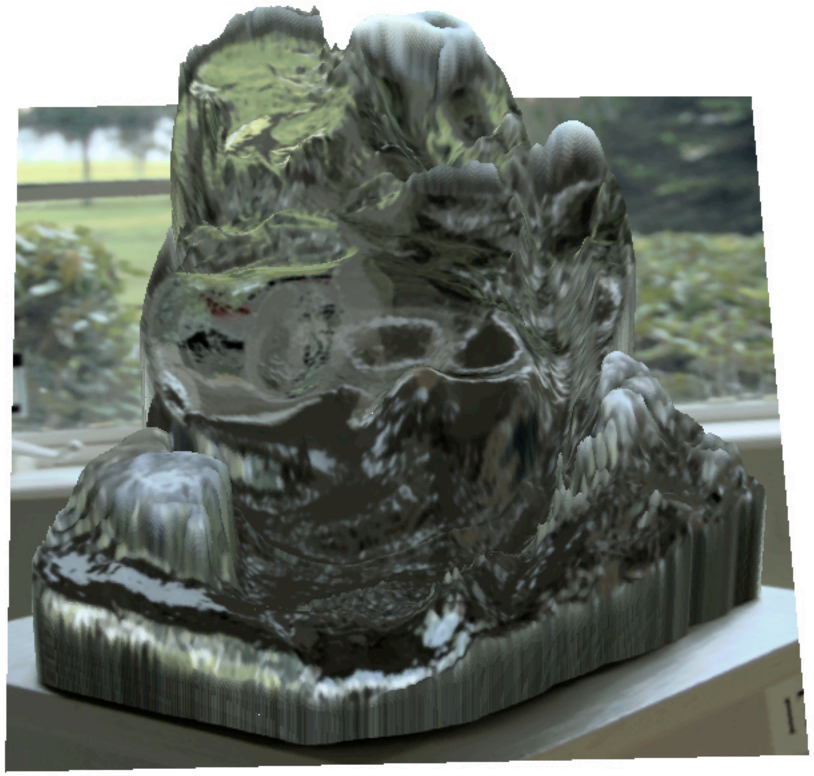
Importance of viewpoint



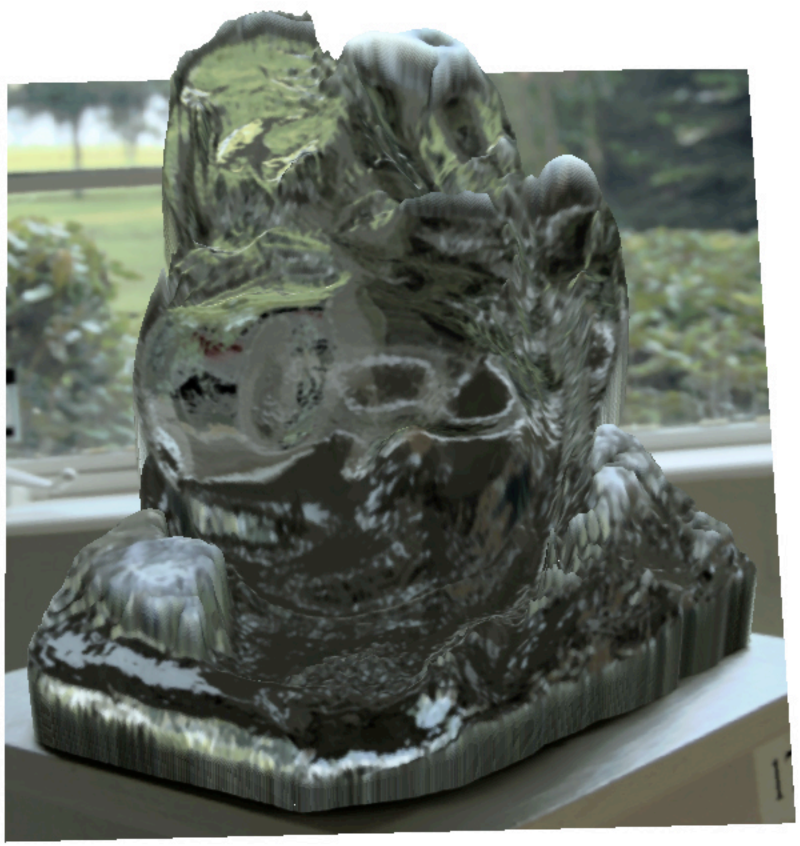
Importance of viewpoint



Importance of viewpoint



Importance of viewpoint

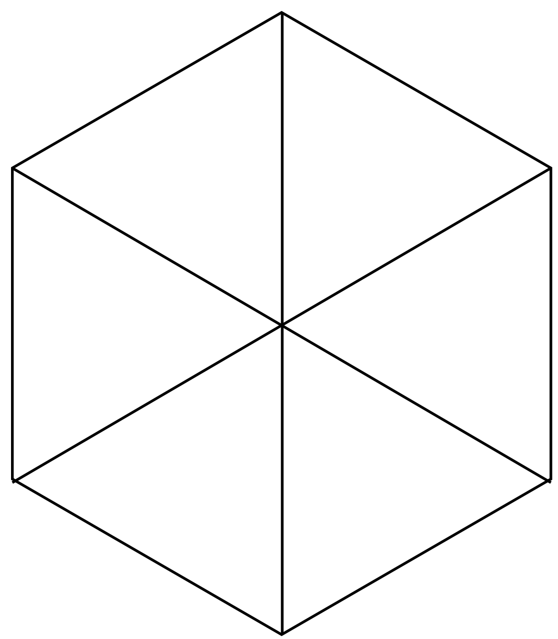


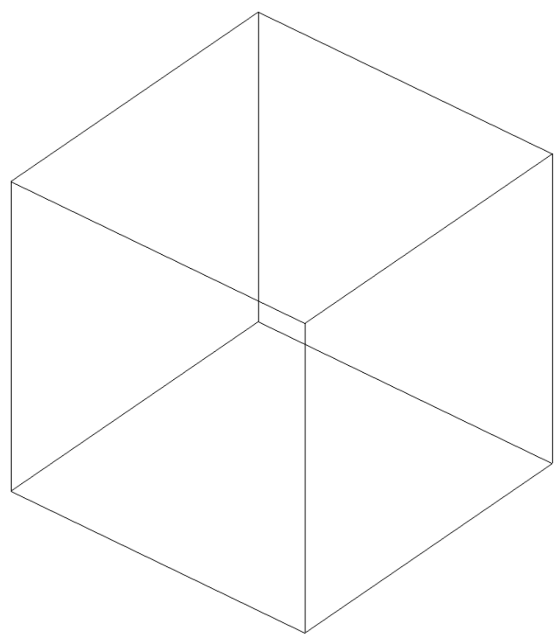
Importance of viewpoint

Importance of viewpoint

Why does it work ?

- Generic viewpoint assumption (Koenderink & van Doorn, 1979; Binford, 1981; Freeman, 1994)





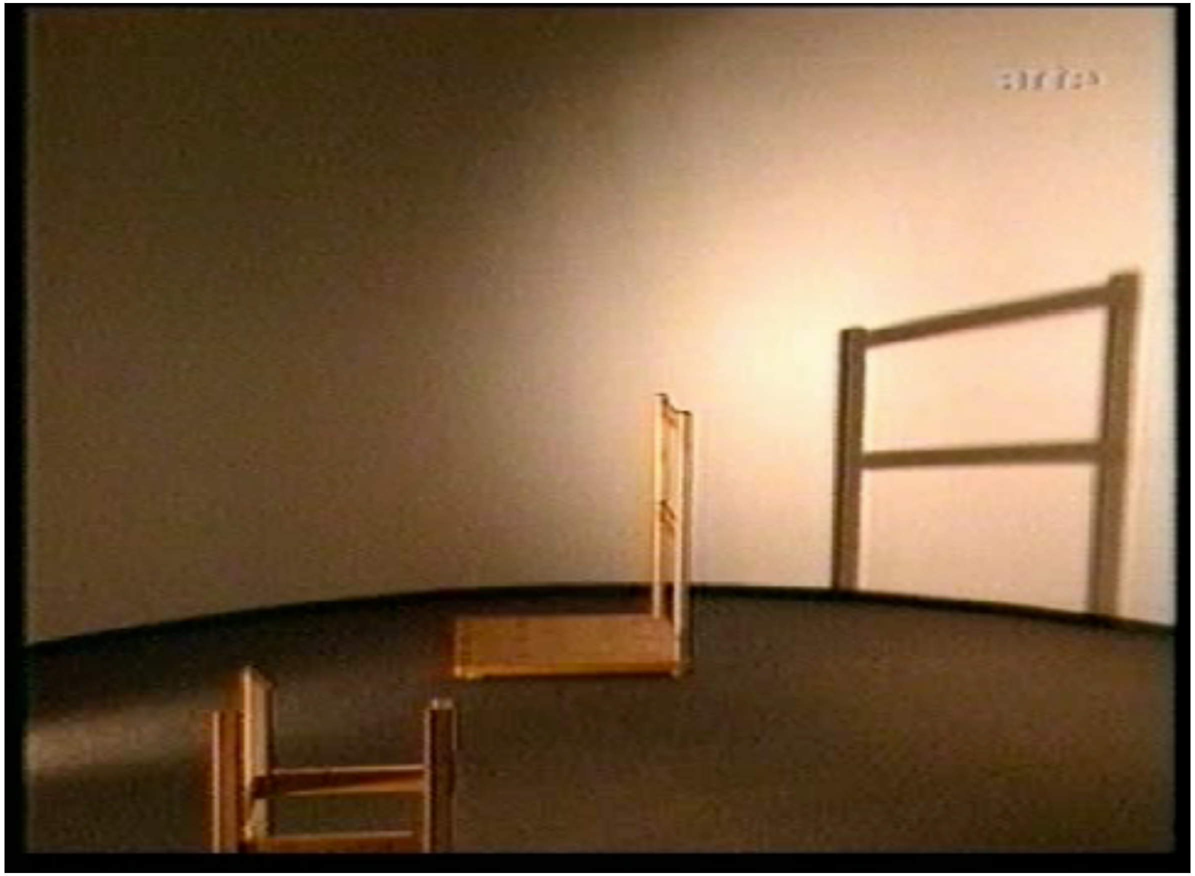
Piano-Illusion

Shigeo Fukuda

Wrong assumptions

- wrong assumptions can lead to perceptual illusions (Beuchet Chair)
- the brain assumes that parts in close proximity belong together
- this assumption can be wrong in rare cases
- an “accidental view” leads to the wrong 3D interpretation


Proximity assumption works in most cases

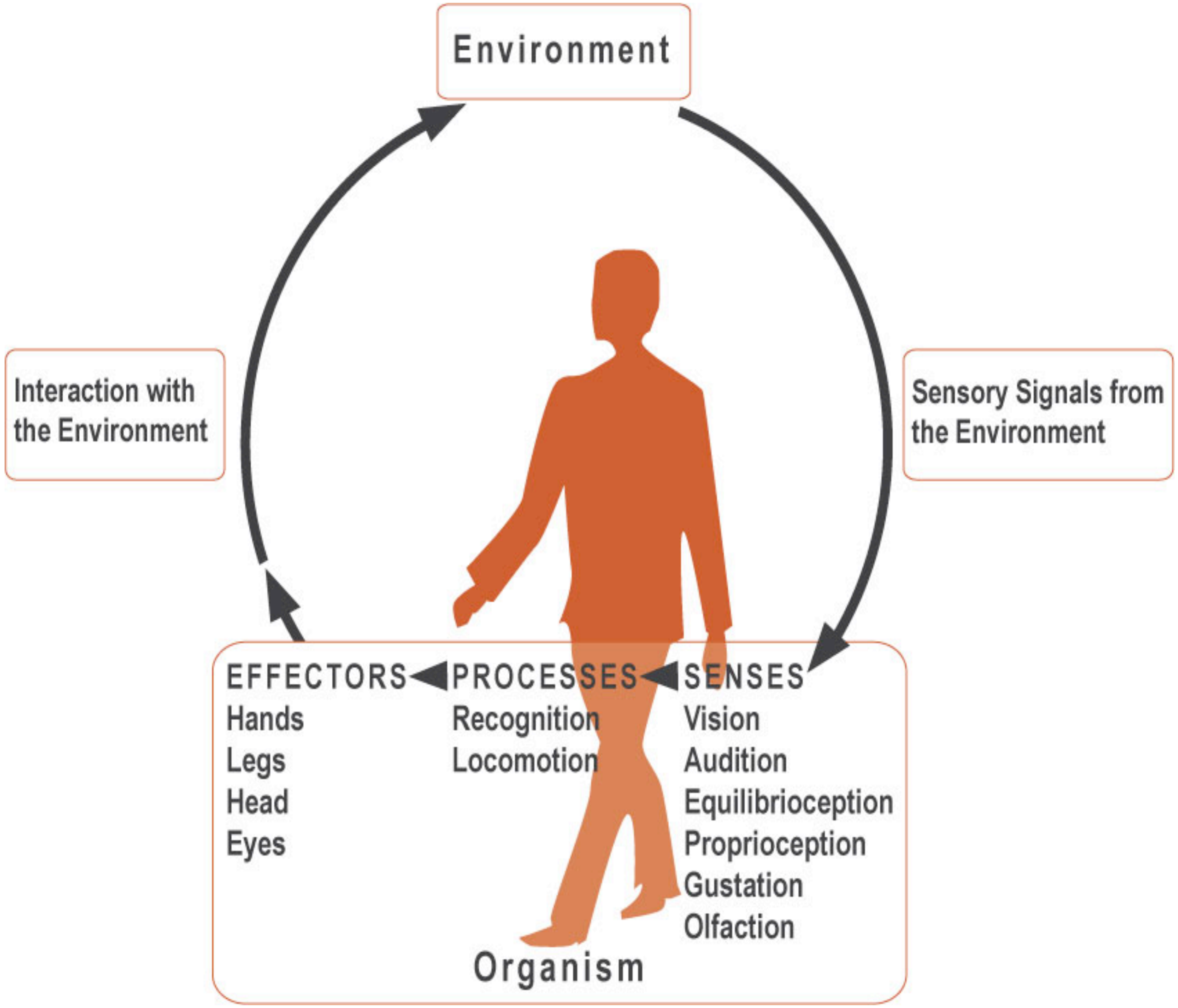


- 2D images usually are sufficient
  for the correct interpretation of a scene
- only from a single viewpoint (accidental view)
  the *proximity assumption* leads to the wrong
  conclusion

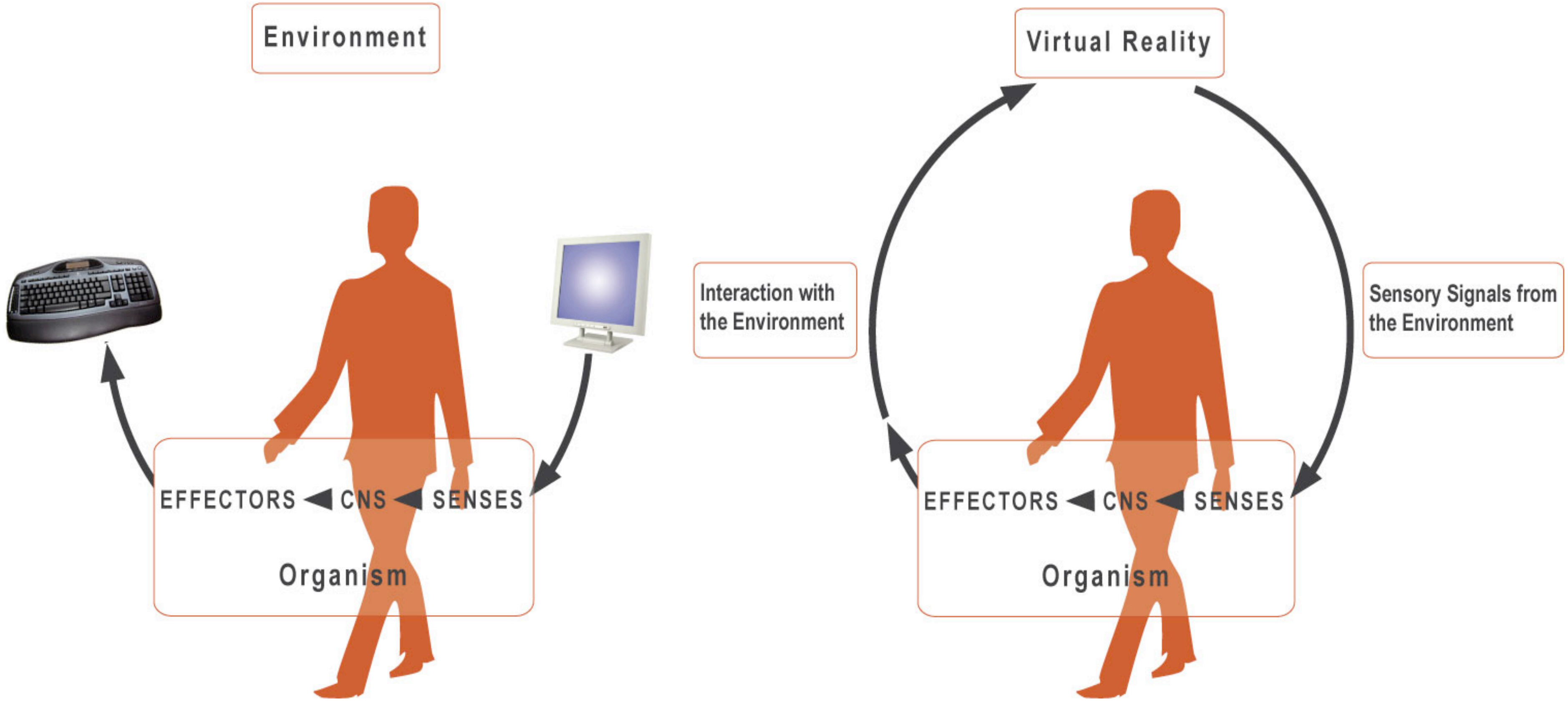
Dwarfs and Giants

- high level interpretation (size, shadows) is ignored
- occlusions can solve the perceptual puzzle

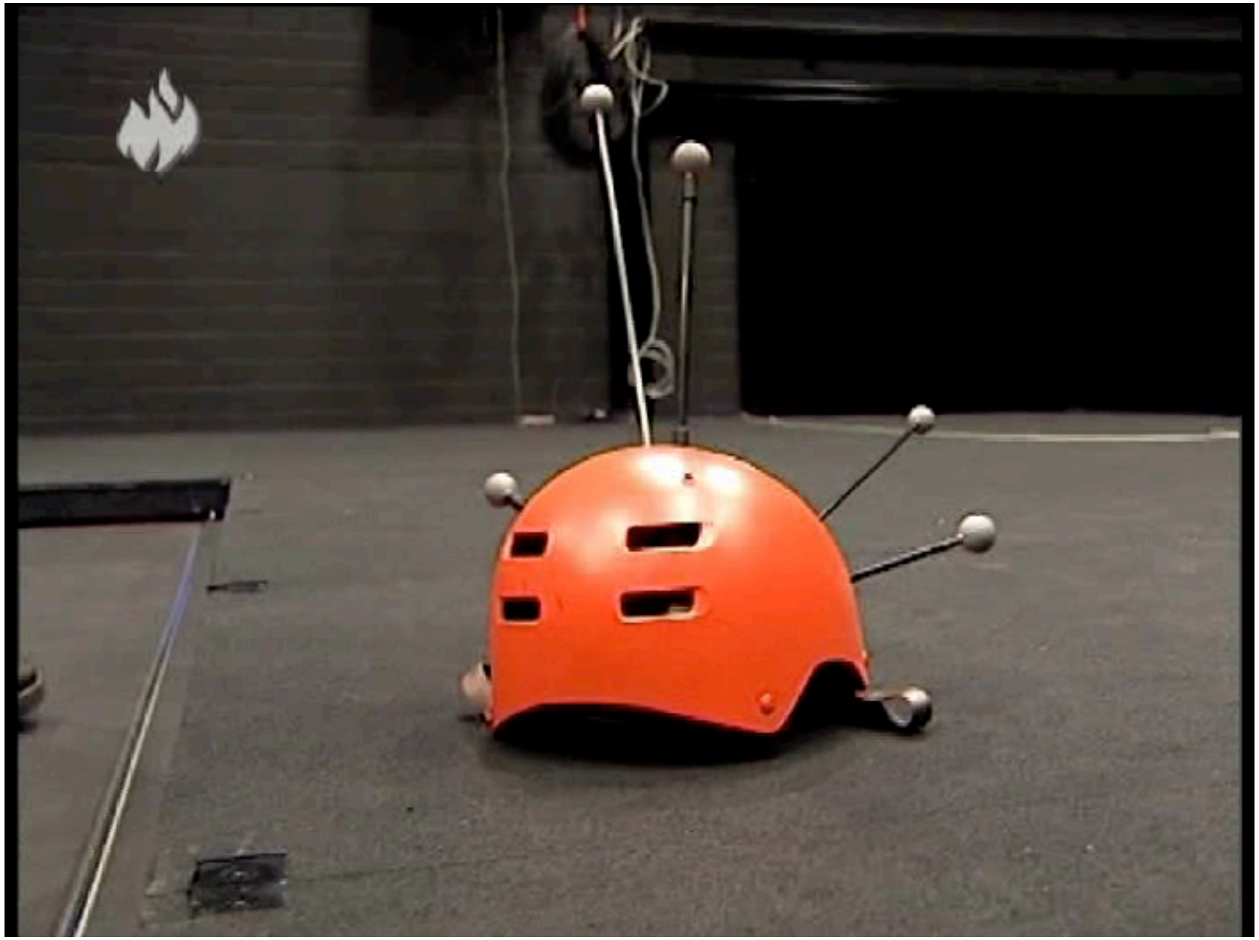
Perception Action Cycle



VR Psychophysics closed loop

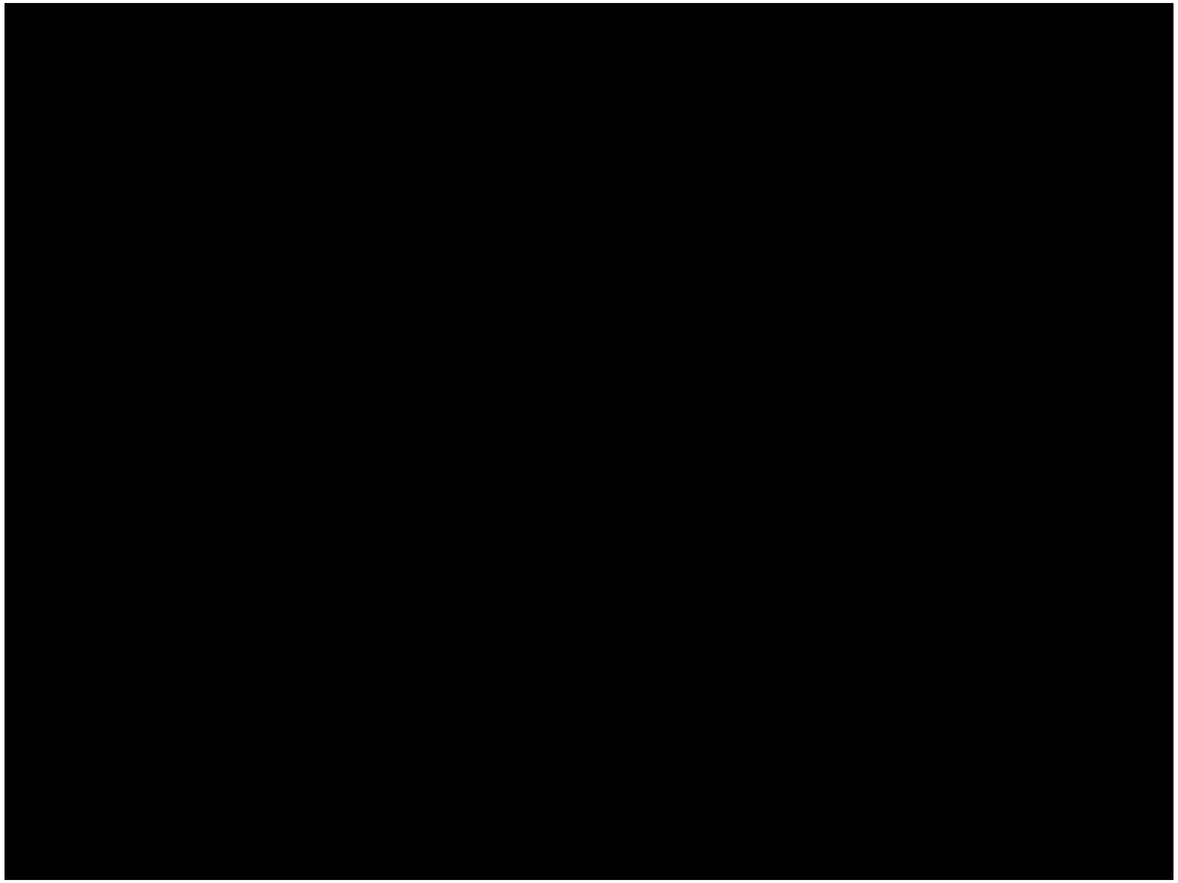


Omni-Directional Treadmill



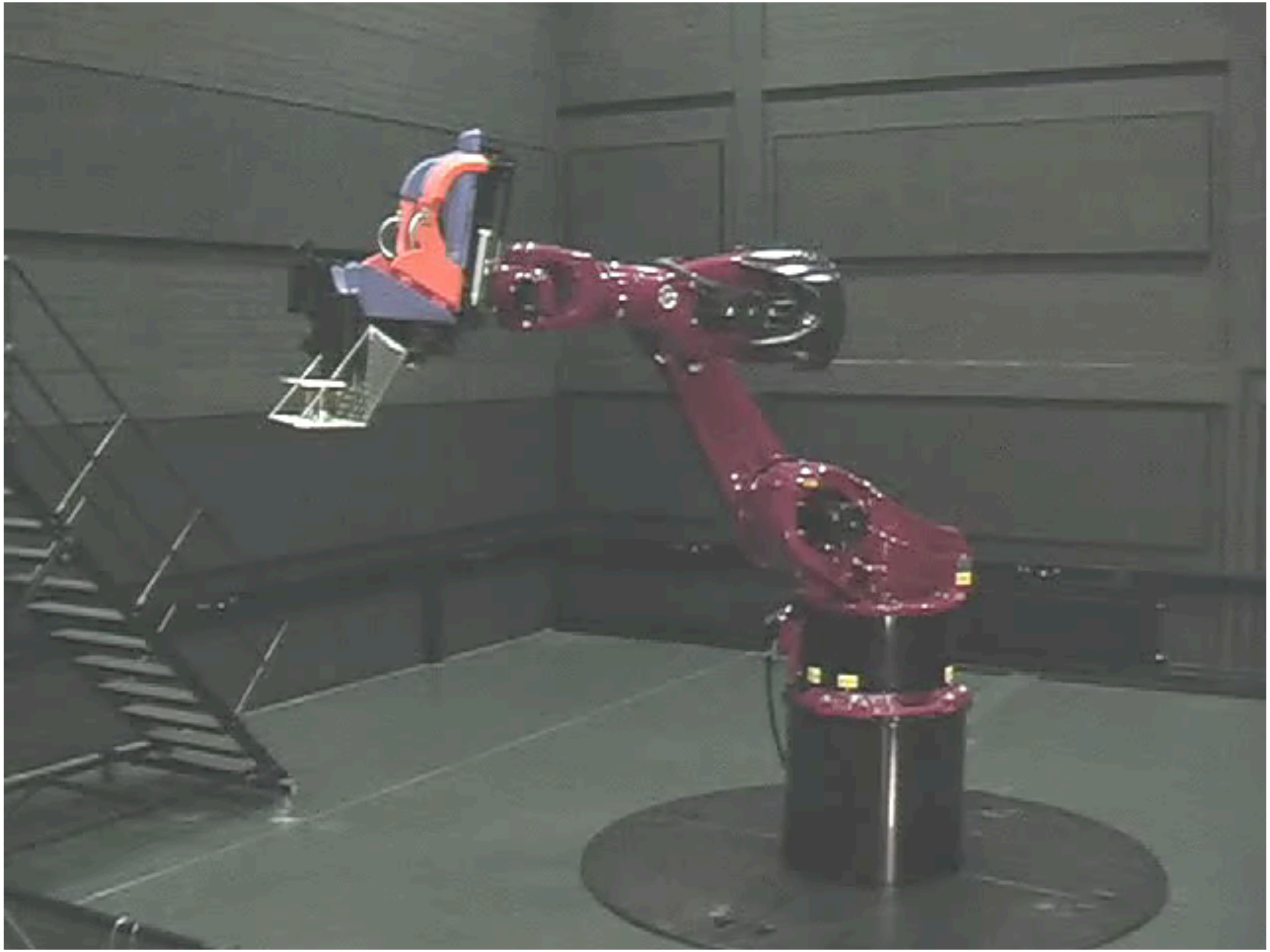
- Unconstrained walking in all directions (2D), creating a truly immersive locomotion interface for VR

Omni-Directional Treadmill

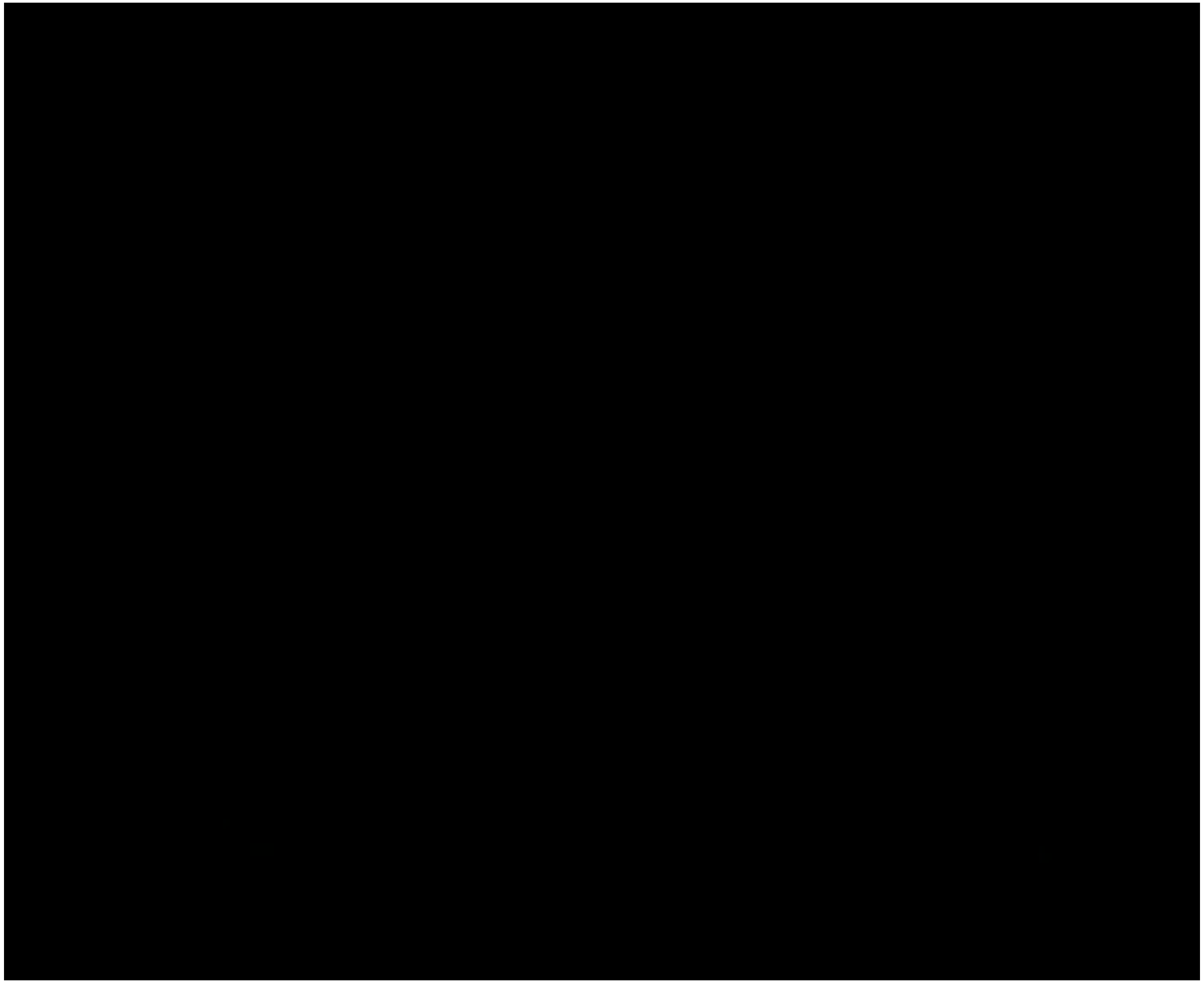


- Unconstrained walking in all directions (2D), creating a truly immersive locomotion interface for VR

MPI Motion Simulator

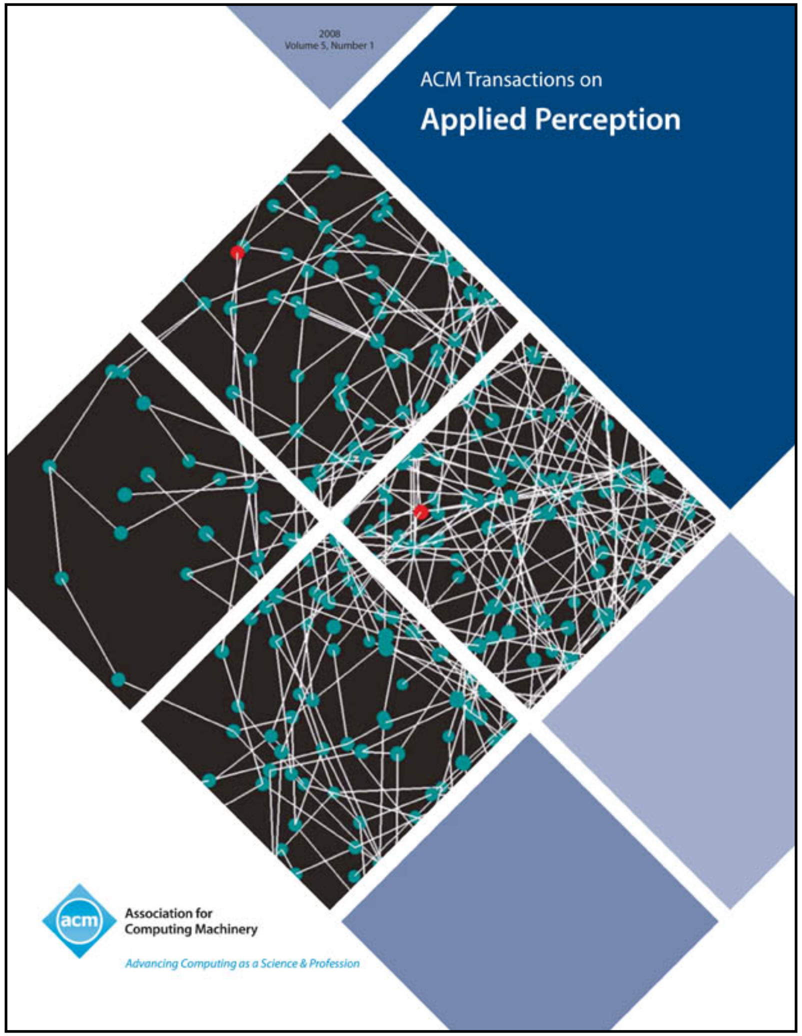


Helicopter Simulator

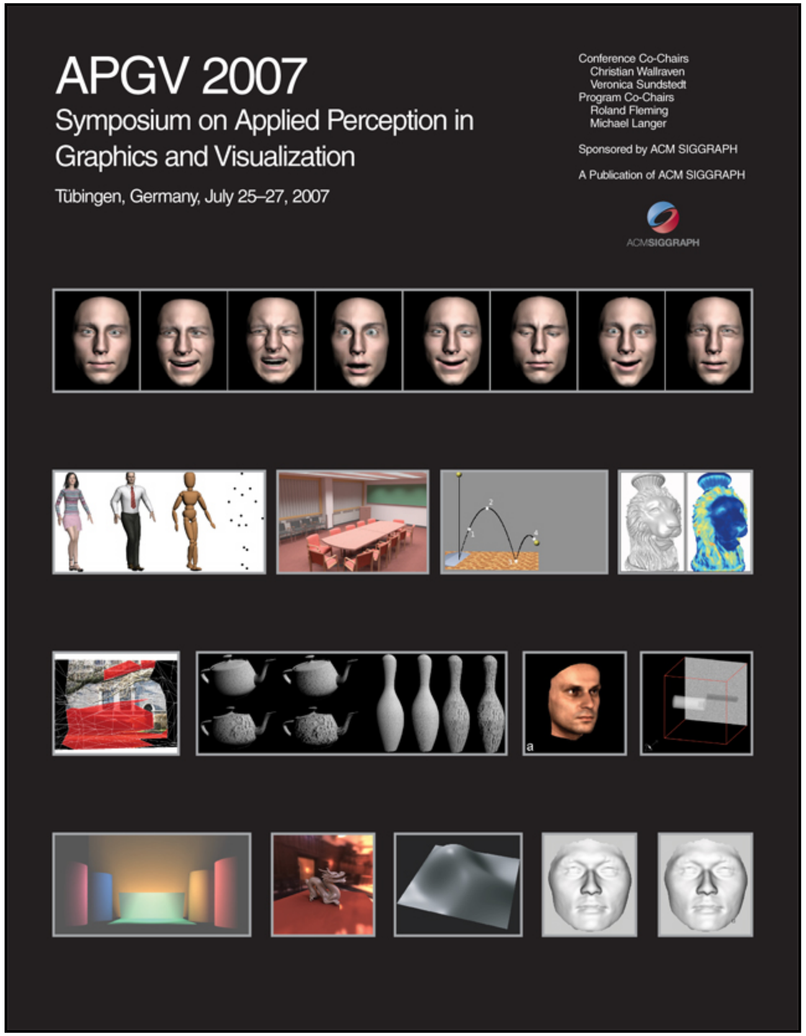


Forums for Perceptual Graphics Research

ACM Trans. on Applied Perception



Symposium on Applied Perception in Graphics and Visualization



Thank You