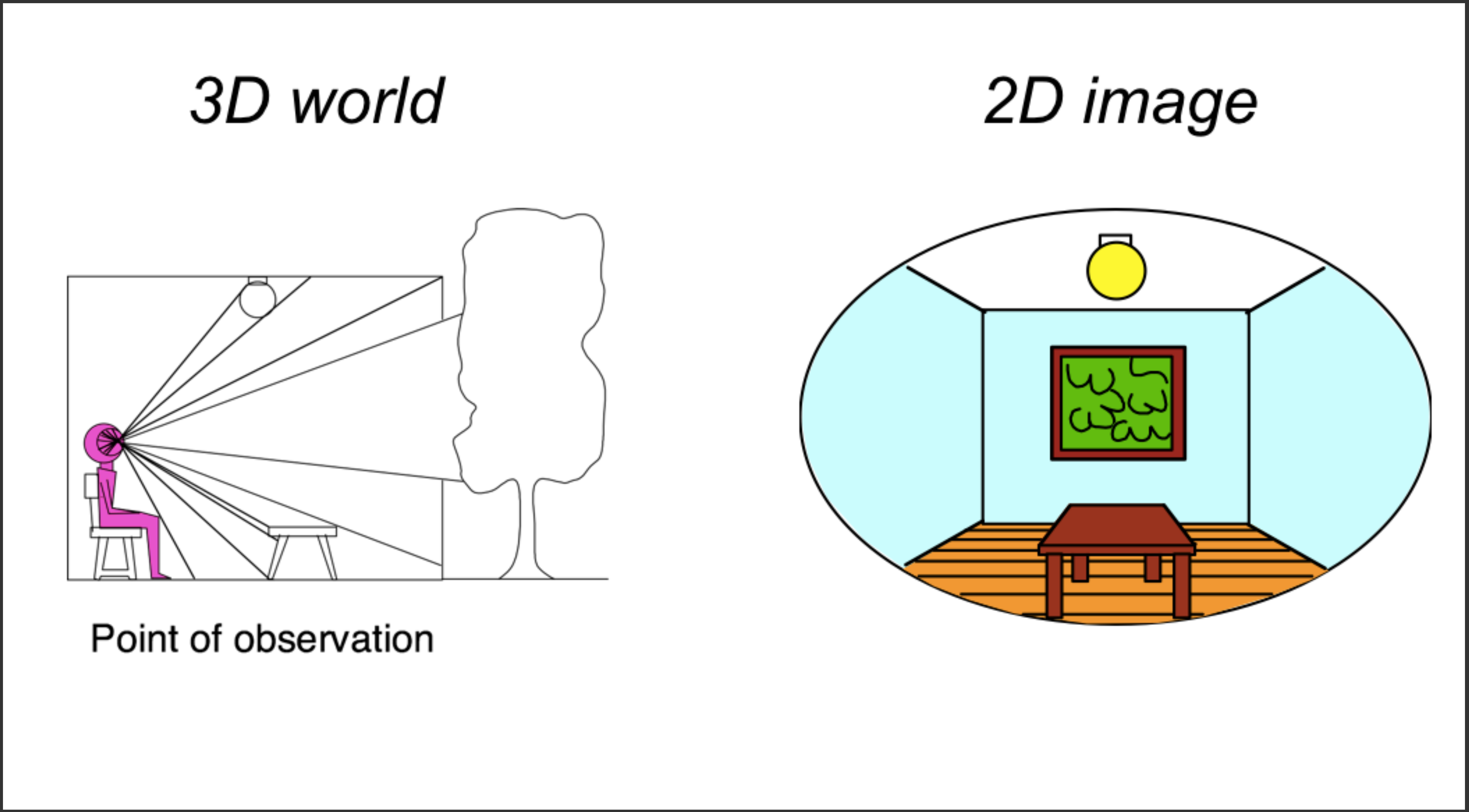
Lecture 26:

Intro to Virtual Reality

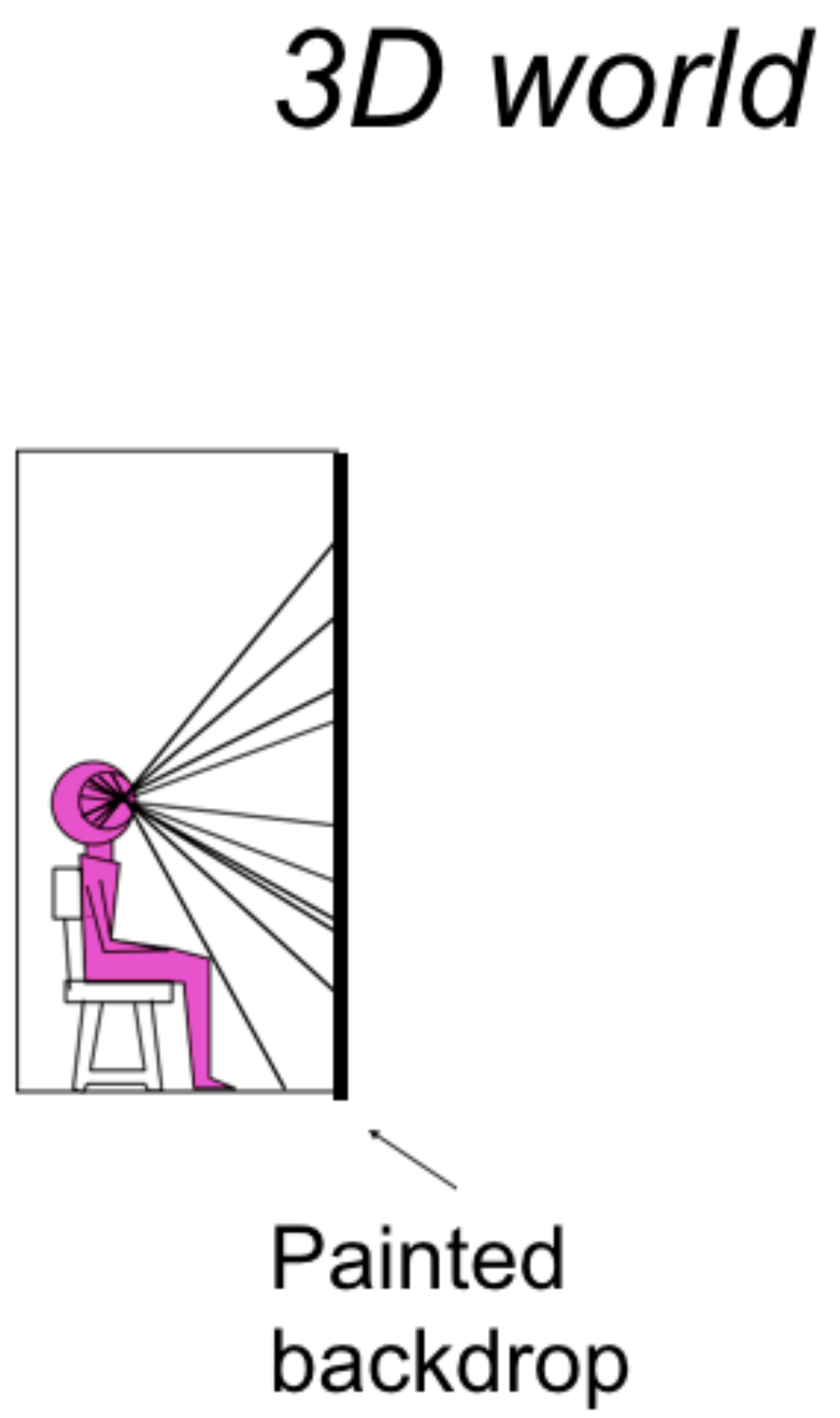
Computer Graphics and Imaging
UC Berkeley CS184/284A

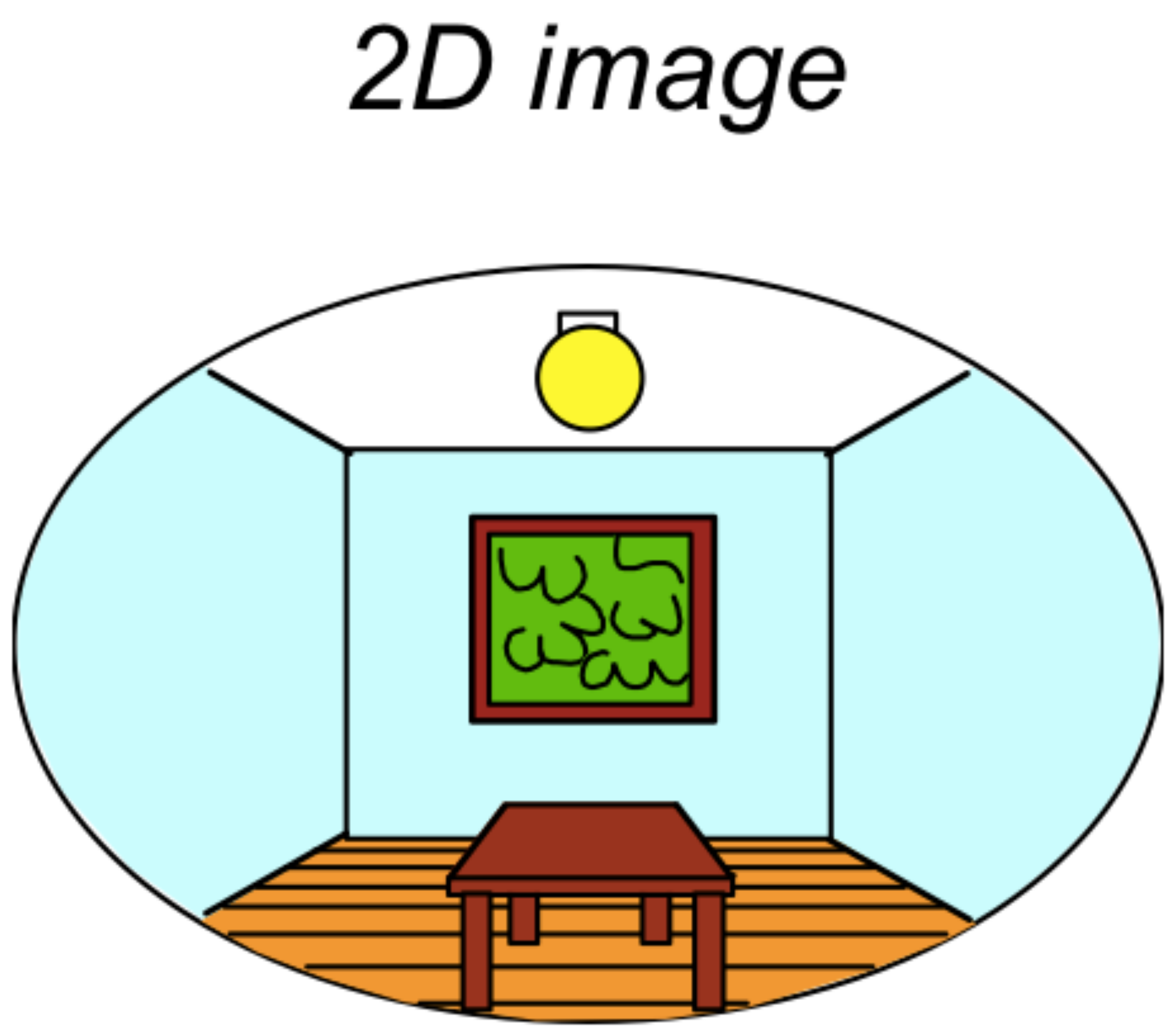
What Do We See?



Slide credit: Alyosha Efros

What Do We See?





Slide credit: Alyosha Efros

The Plenoptic Function

Figure by Leonard McMillan

Q: What is the set of all things that we can ever see?

A: The Plenoptic Function (Adelson & Bergen)

Let's start with a stationary person and try to parameterize <u>everything</u> that person can see…

Slide credit: Alyosha Efros

Grayscale Snapshot

$$P(\theta, \phi)$$

is intensity of light

- Seen from a single view point
- At a single time
- Averaged over the wavelengths of the visible spectrum

(can also do *P(x,y)*, but spherical coordinate are nicer)

Slide credit: Alyosha Efros

Color Snapshot

$$P(\theta, \phi, \lambda)$$

is intensity of light

- Seen from a single view point
- At a single time
- As a function of wavelength

Slide credit: Alyosha Efros

A Movie

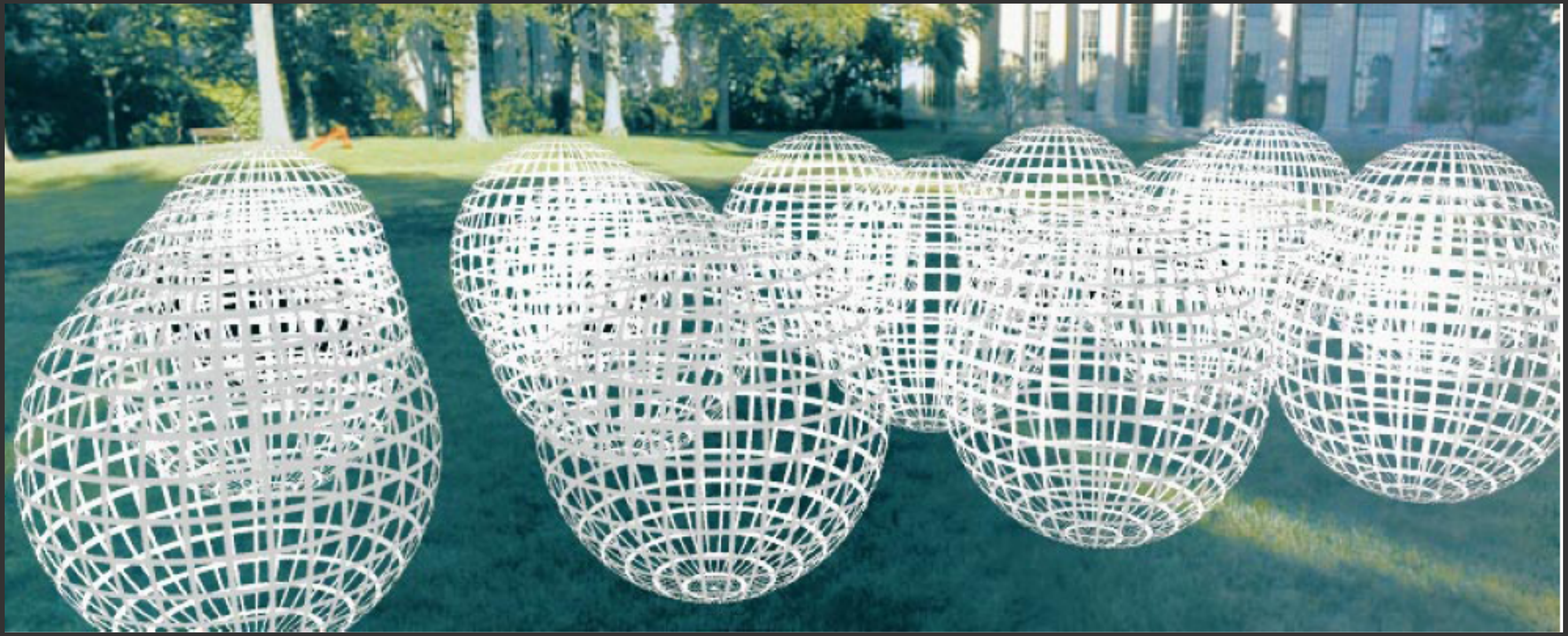
$$P(\theta, \phi, \lambda, t)$$

is intensity of light

- Seen from a single view point
- Over time
- As a function of wavelength

Slide credit: Alyosha Efros

Free Viewpoint Movie



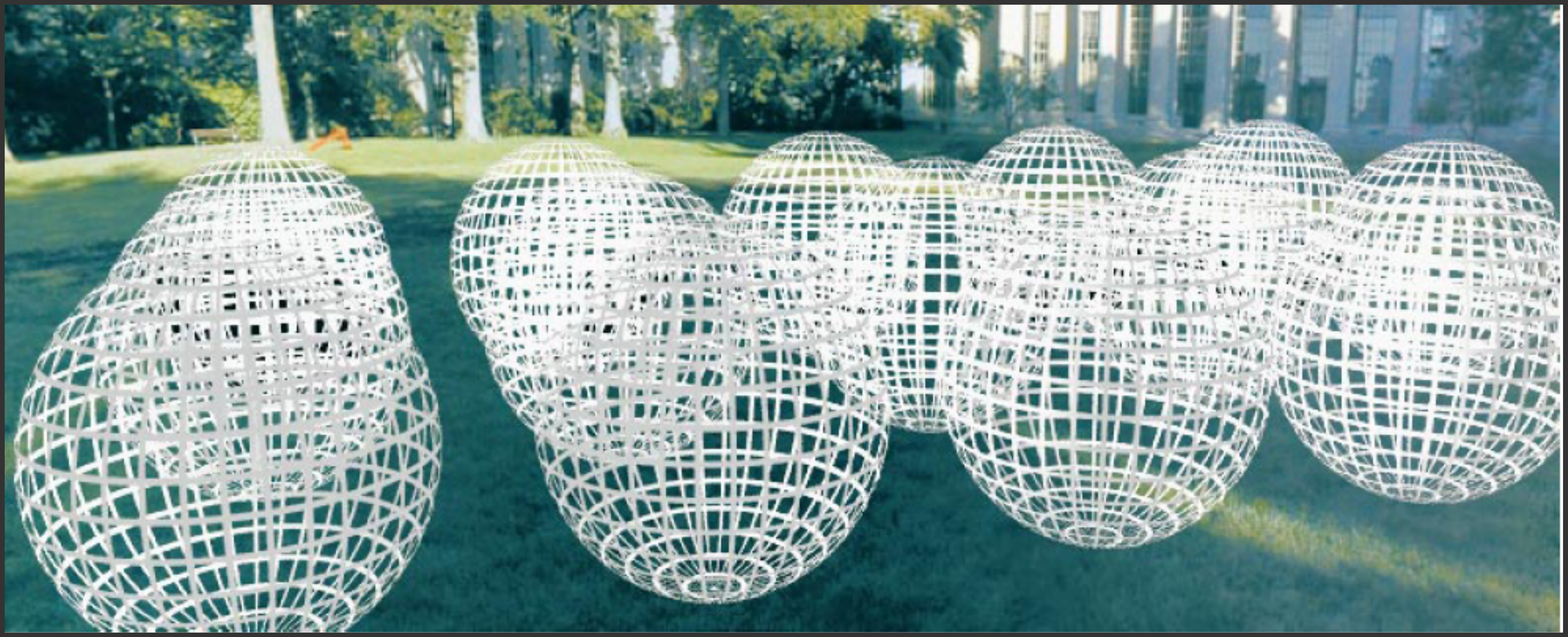
$$P(\theta, \phi, \lambda, t, V_x, V_y, V_z)$$

is intensity of light

- Seen from ANY viewpoint
- Over time
- As a function of wavelength

Slide credit: Alyosha Efros

The Plenoptic Function

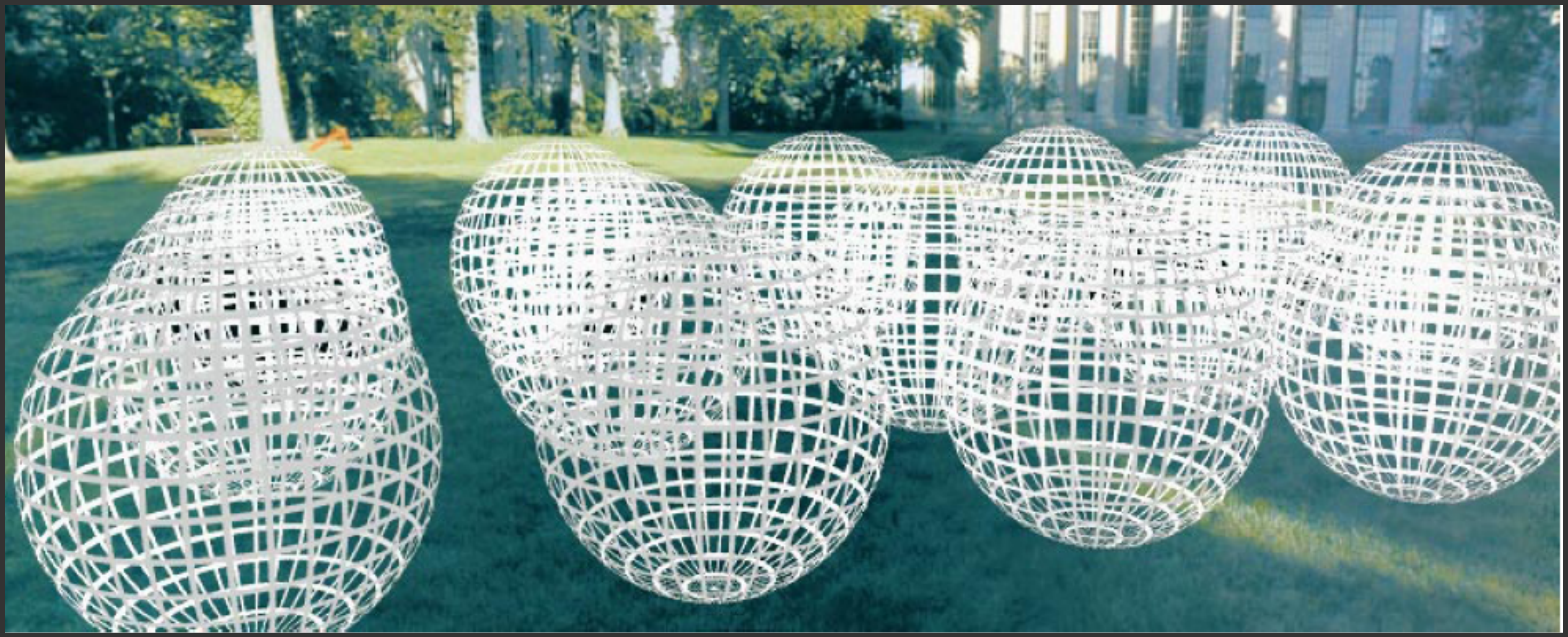


$$P(\theta, \phi, \lambda, t, V_x, V_y, V_z)$$

- Can reconstruct every possible view, at every moment, from every position, at every wavelength
- Contains every photograph, every movie, everything that anyone has ever seen! it completely captures our visual reality! Not bad for a function…

Slide credit: Alyosha Efros

The 5D Plenoptic Function



$$P(\theta, \phi, V_x, V_y, V_z)$$

- Ignore time and wavelength
- Focus just on spatial structure of light

Slide credit: Alyosha Efros

On Simulating the Visual Experience

Just feed the eyes the right data

- No one will know the difference!

Philosophy:

- Ancient question: “Does the world really exist?”

Physics:

- “Slowglass" might be possible?

Computer Science:

- Virtual Reality

Ivan Sutherland's "Sword of Damocles", 1968

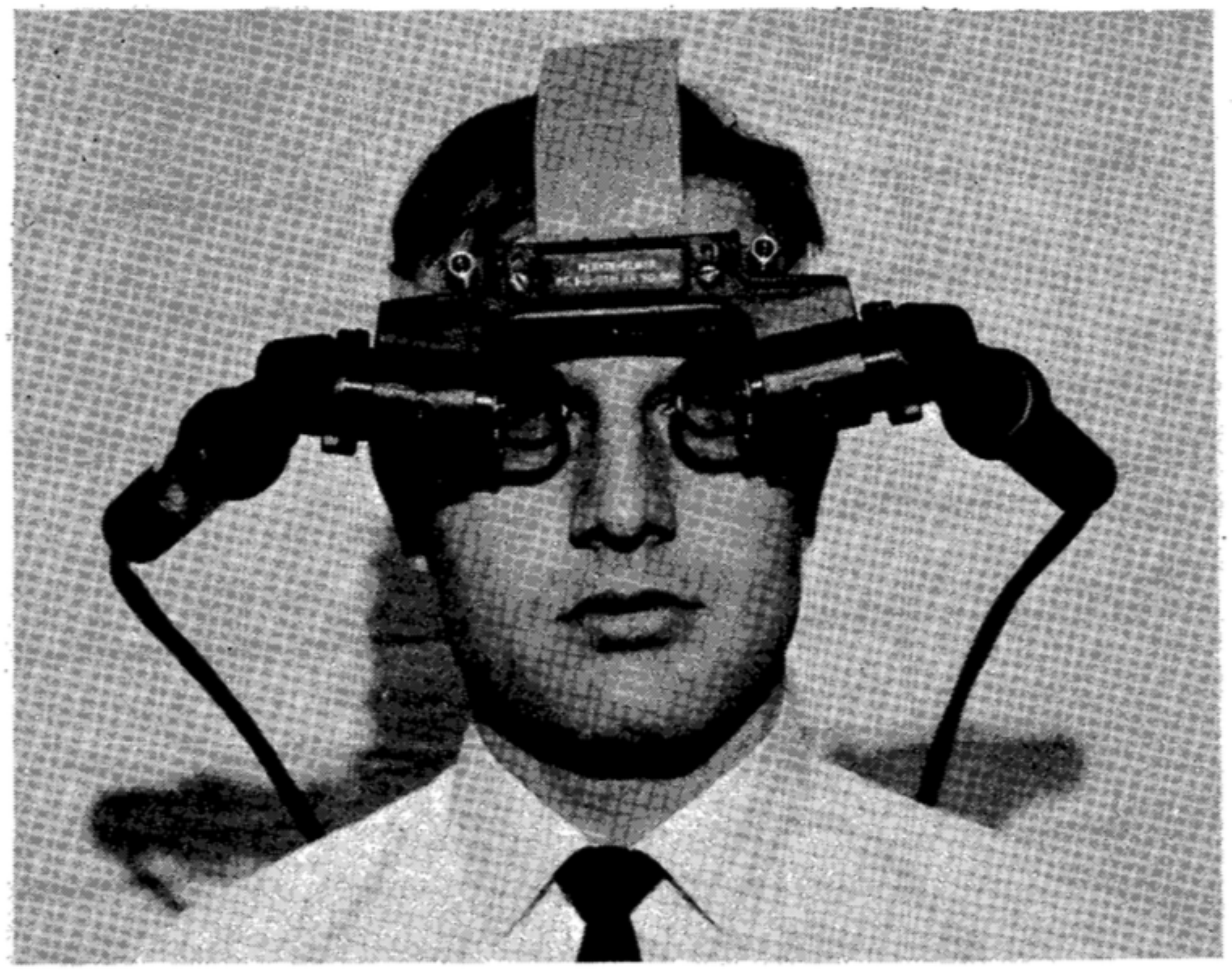


FIGURE 2—The head-mounted display optics
with miniature CRT's

FIGURE 4—The ultrasonic head position sensor in use

VR Head-Mounted Displays (HMDs)

Oculus Rift

PlayStation VR

HTC Vive

Google Cardboard

Virtual Reality (VR) vs Augmented Reality (AR)

VR = virtual reality

- User is completely immersed in virtual world (sees only light emitted by display)

AR = augmented reality

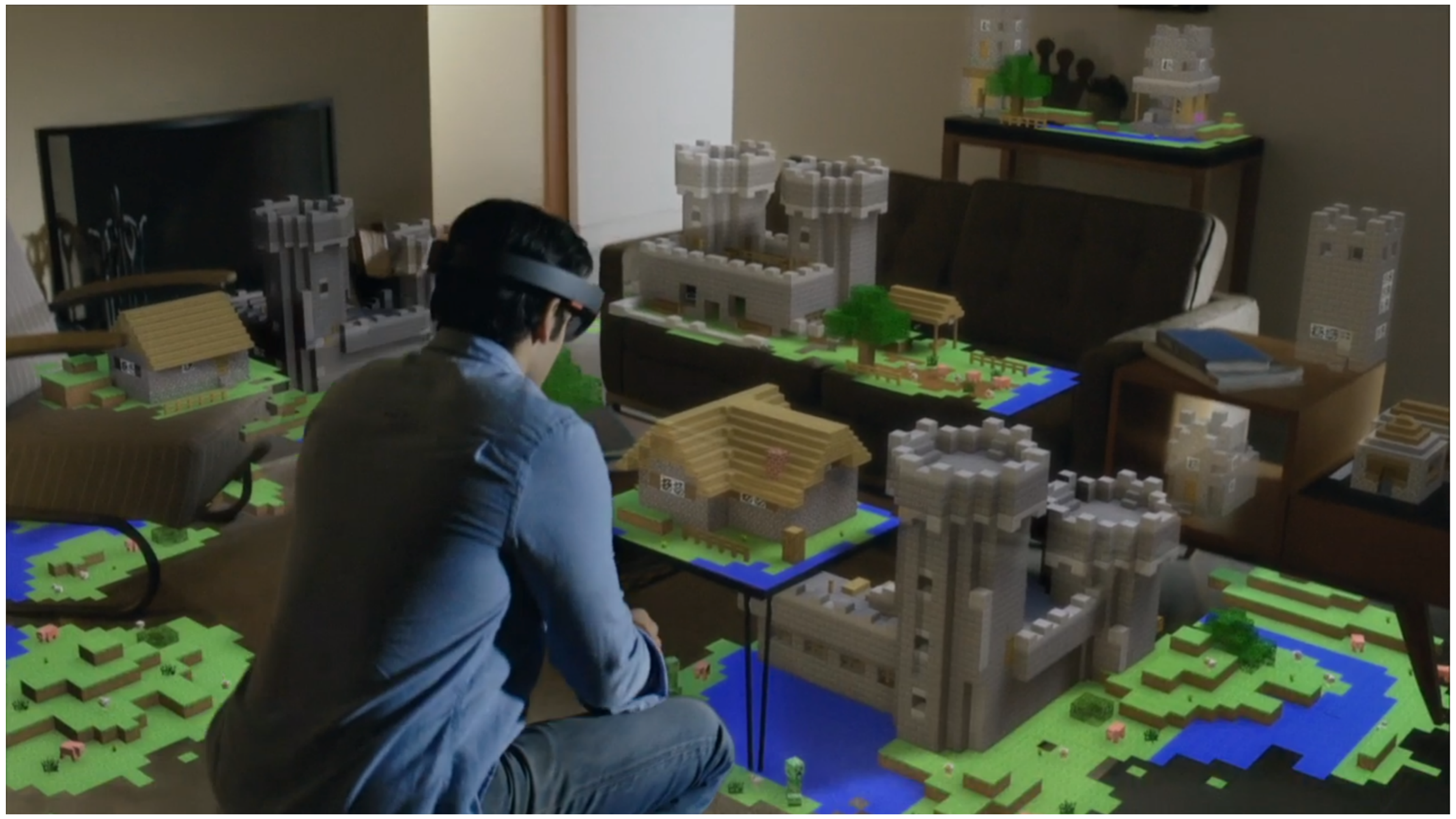
- Display is an overlay that augments user's normal view of the real world (e.g., Terminator)

Image credit: Terminator 2 (naturally)

AR Headsets

Microsoft Hololens

Magic Leap



VR Applications

VR Gaming

Half-Life: Alyx (Valve)

VR Painting

VR Video

VR Video

VR Teleconference / Video Chat

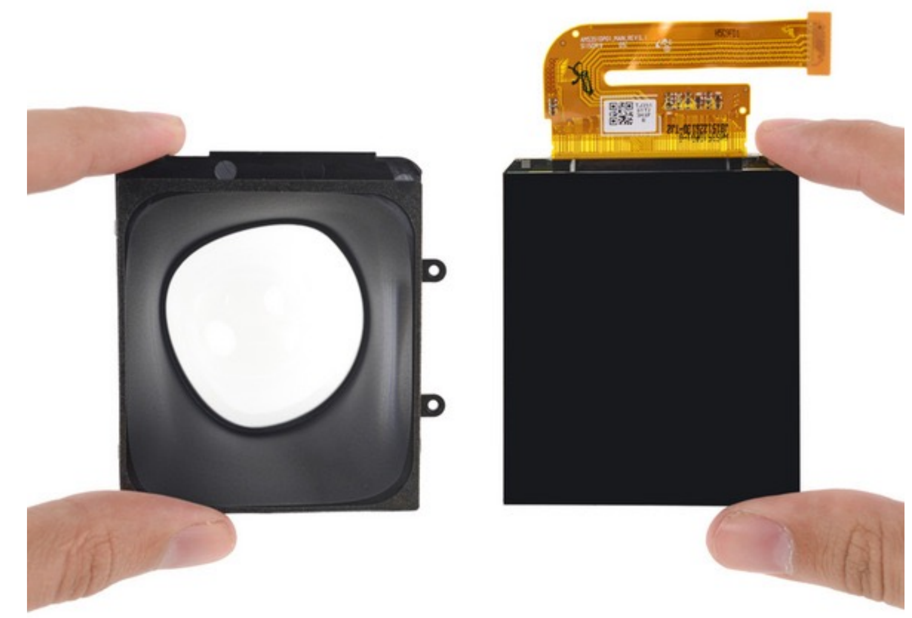
http://vrchat.com/

VR Teleconference / Video Chat

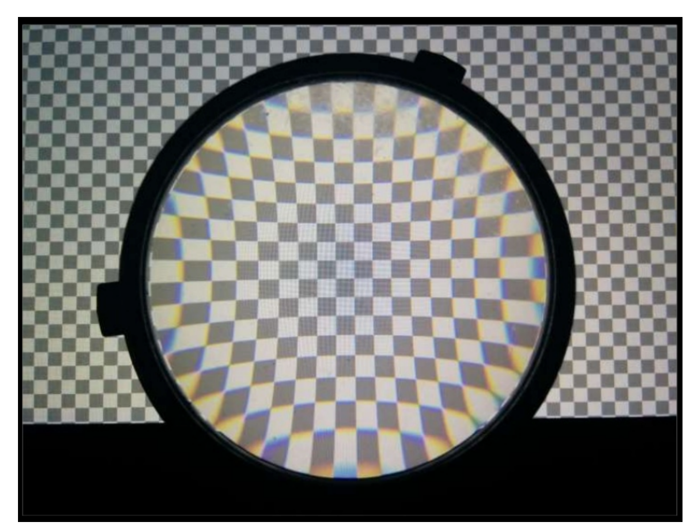
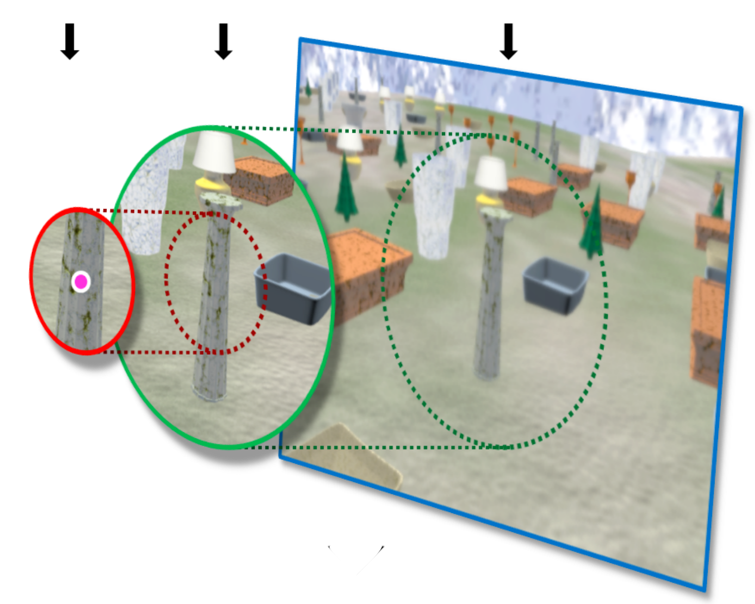
Wei et al. 2019, *VR Facial Animation via Multiview Image Translation*

Overview of VR Topics

- VR Displays

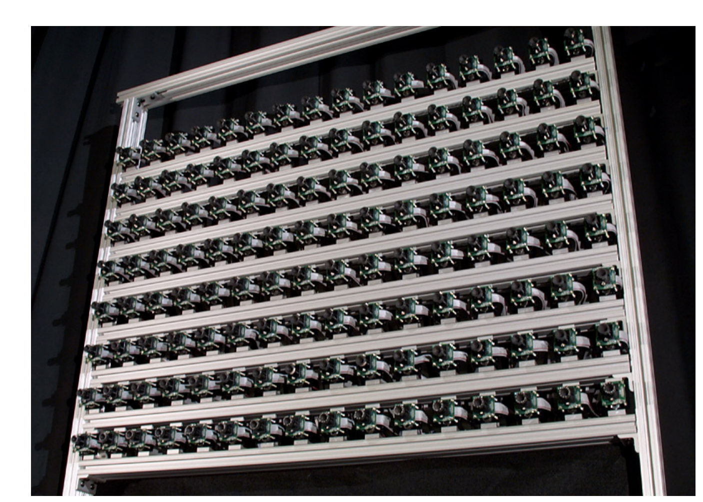




- VR Rendering

- VR Imaging





VR Displays

Field of View

Regular 2D panel displays have windowed FOV

- User orients themselves to the physical window of the display

VR/AR displays provide 360 degree FOV

- Displays attached to head
- Head orientation is tracked physically
- Rendered view synchronized to head orientation in realtime (much more on this later)

3D Visual Cues

Panel displays give 3D cues from monocular rendering

- Occlusion, perspective, shading, focus blur, …
  - Uses z-buffer, 4x4 matrices, lighting calculation, lens calculations…

VR/AR displays add further 3D cues

- Stereo: different perspective view in left/right eyes
  - Physically send different images into each eye
- Parallax (user-motion): different views as user moves
  - Uses head-tracking technology coupled to perspective rendering

VR Headset Components

Google Cardboard

Use mobile phone display inside inexpensive headset with lenses

- Use phone's camera and gyro for tracking view direction
- Stereo 360 degree experience, no head-motion parallax

Image credits: slashgear.com, Google

Oculus Rift

Oculus Rift headset has most documentation of current systems, so will use for this explanation.

Oculus Rift

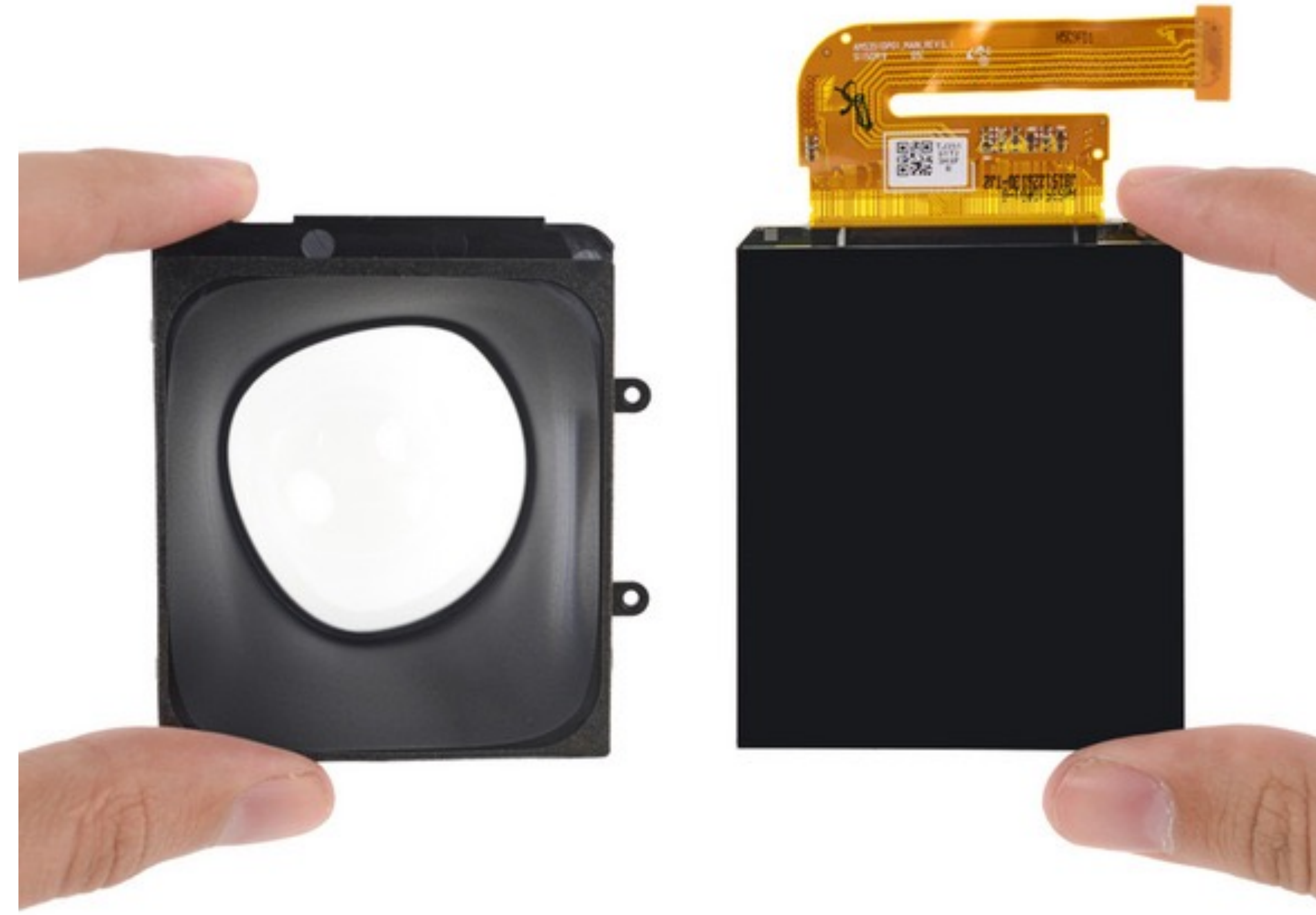


Image credit: ifixit.com https://www.ifixit.com/Teardown/Oculus+Rift+CV1+Teardown/60612

Oculus Rift

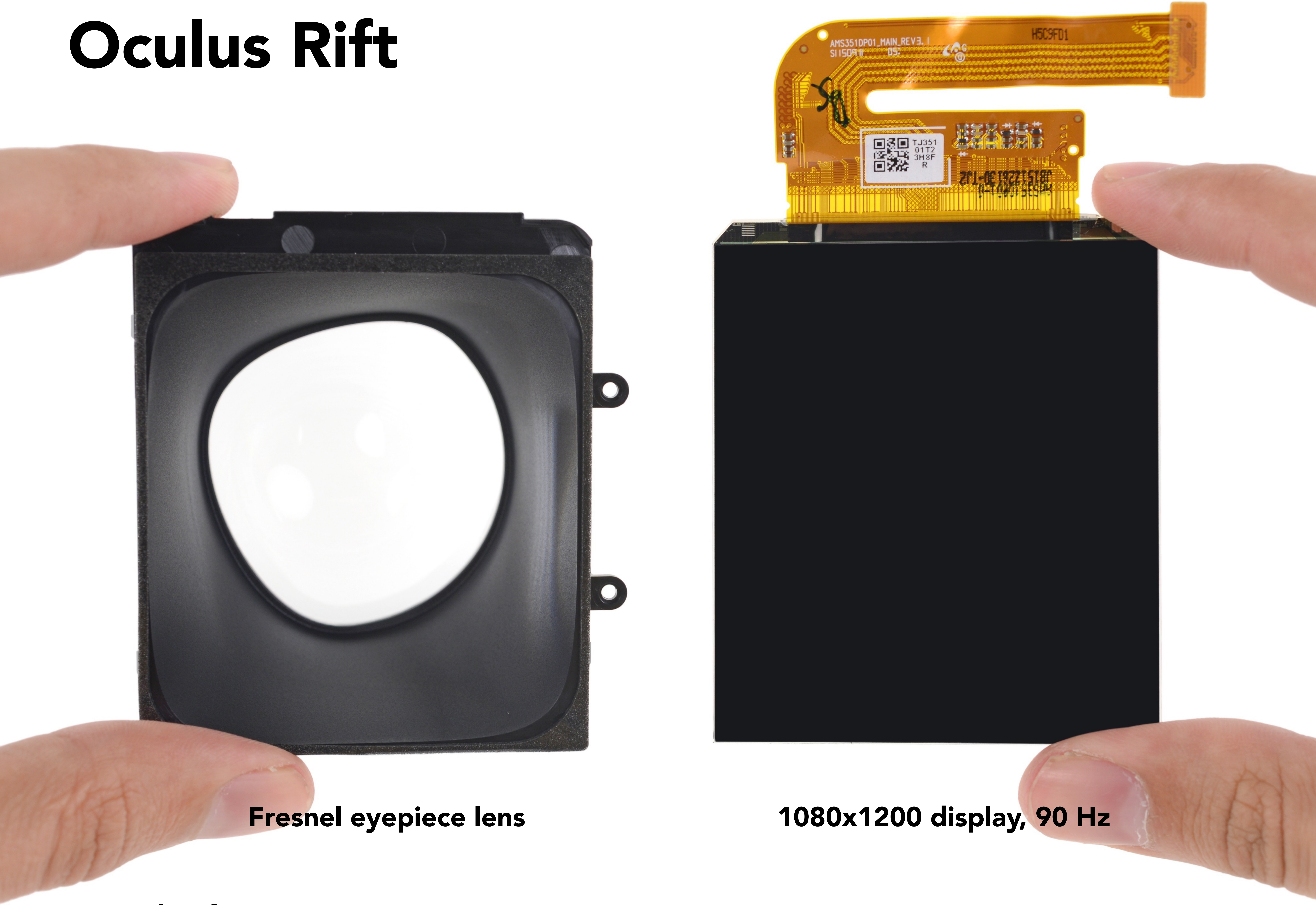
Intra-ocular distance adjustment

Image credit: ifixit.com

Oculus Rift

Image credit: ifixit.com

Oculus Rift

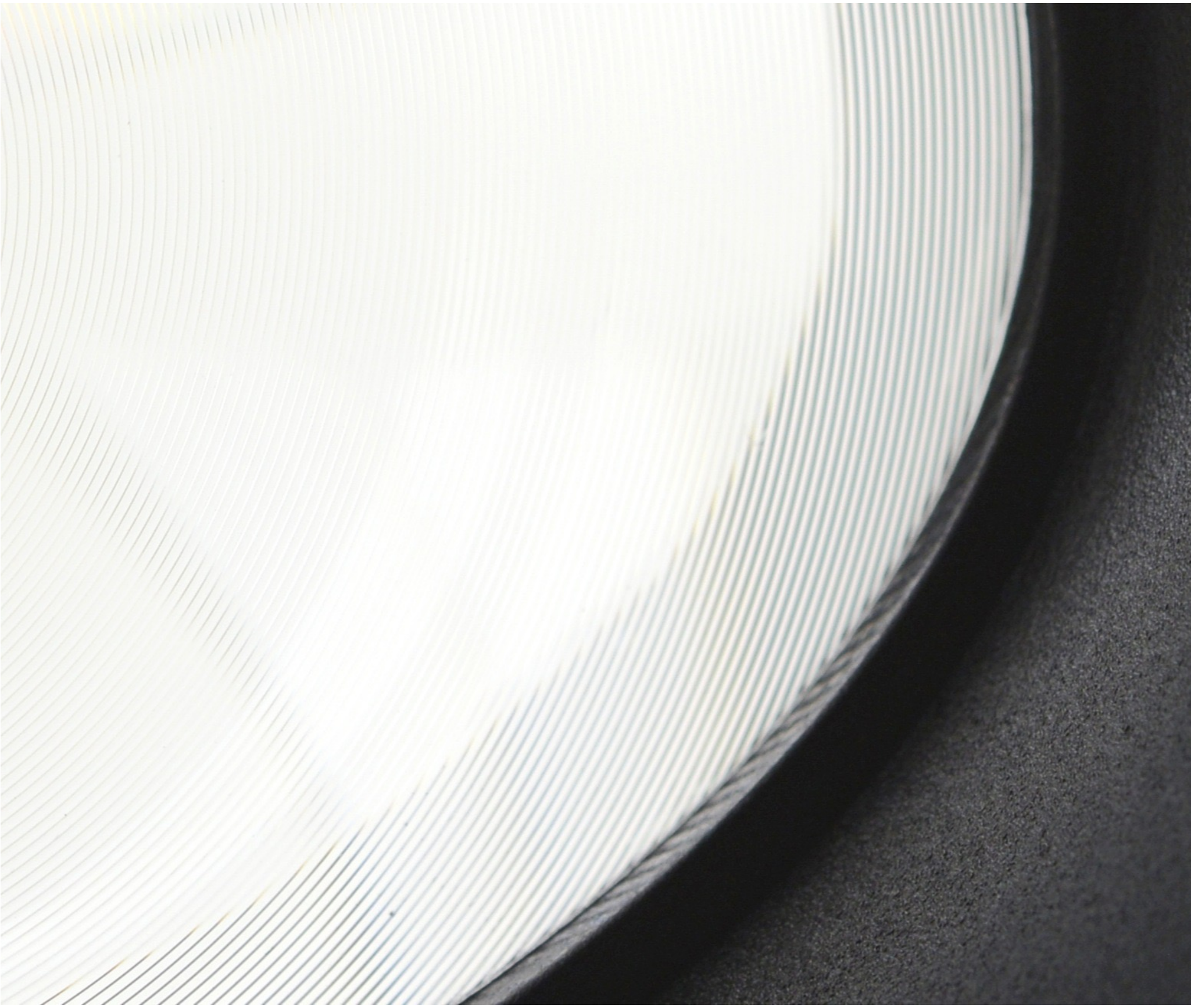


Fresnel eyepiece lens

1080x1200 display, 90 Hz

Image credit: ifixit.com

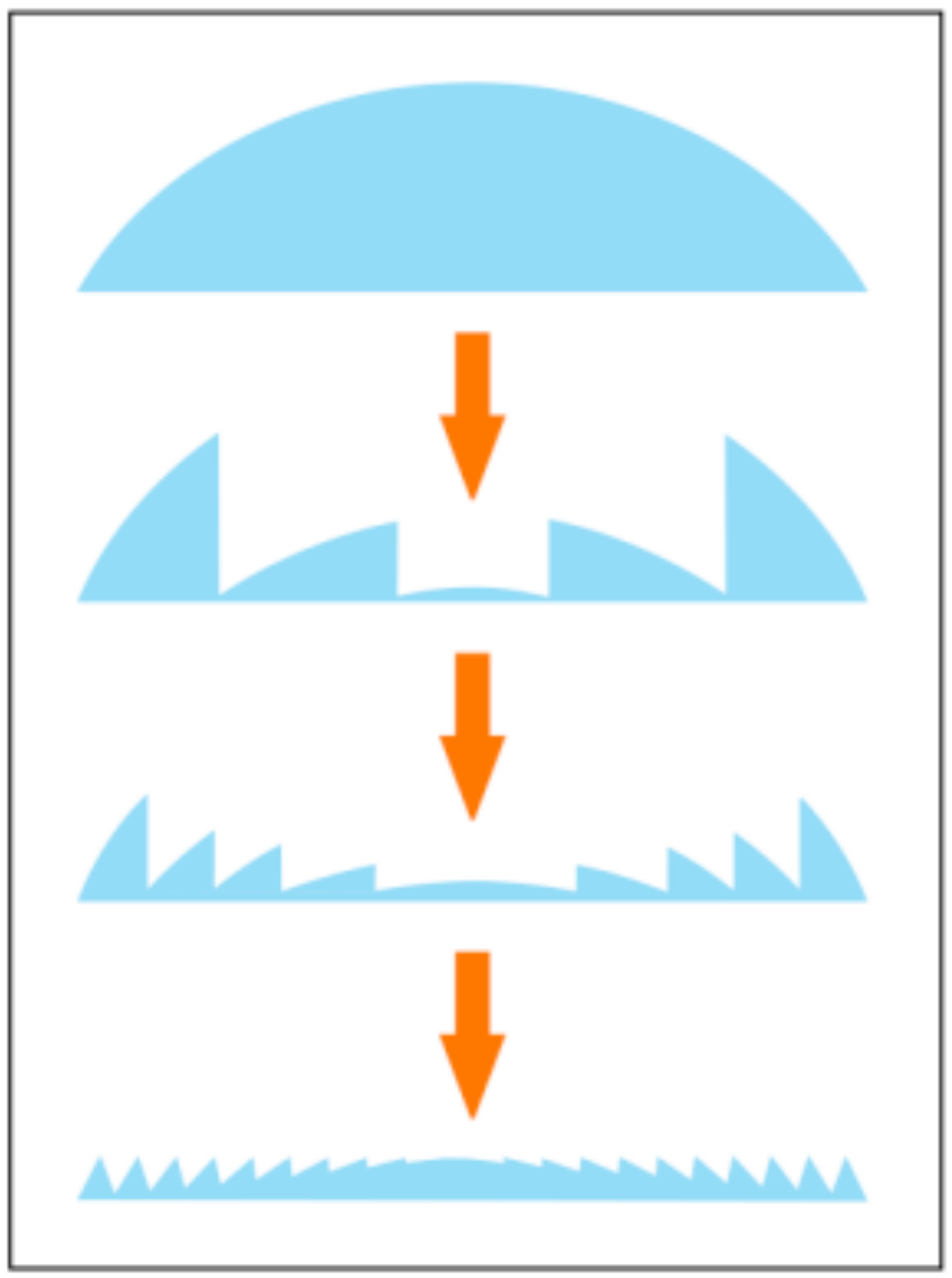
Oculus Rift Lenses



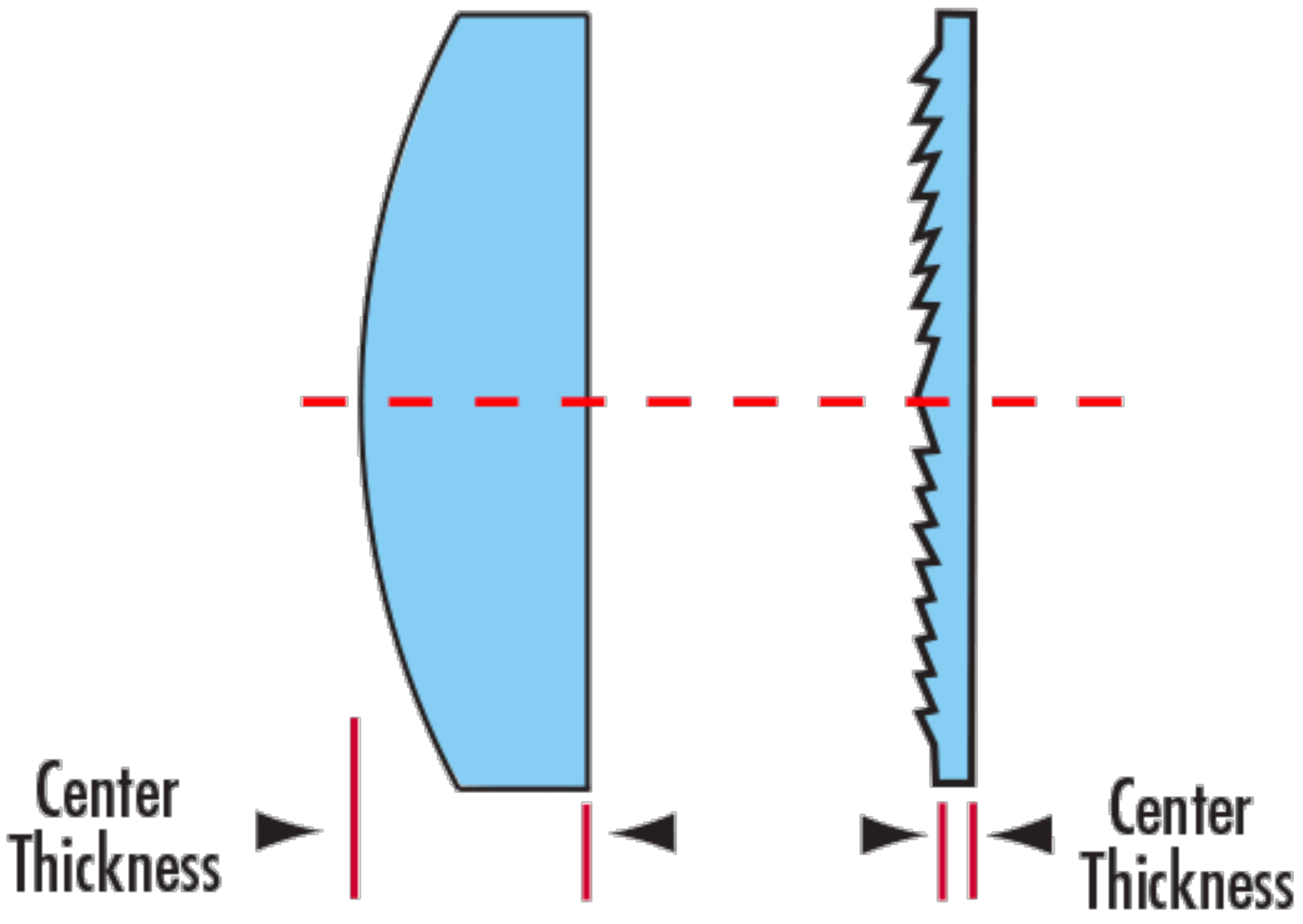
Fresnel eyepiece lens

Image credit: ifixit.com

Fresnel lens



Curvature is what matters for lens, not absolute thickness

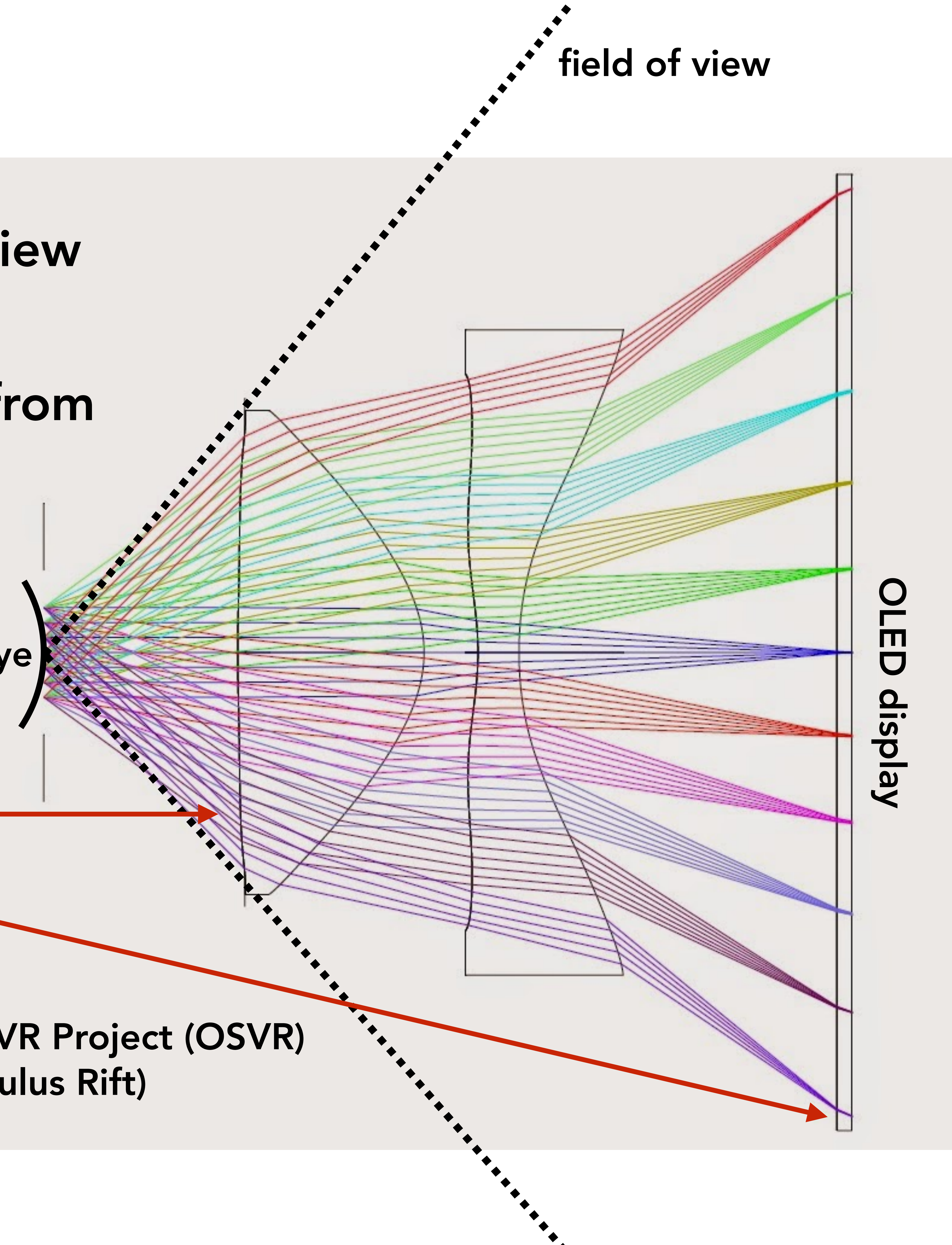


Can shrink lens down, trading off space vs. image quality

Role of Optics

1. Create wide field of view
2. Place focal plane at several meters away from eye (close to infinity)

Note: parallel lines reaching eye converge to a single point on display (eye accommodates to plane near infinity)

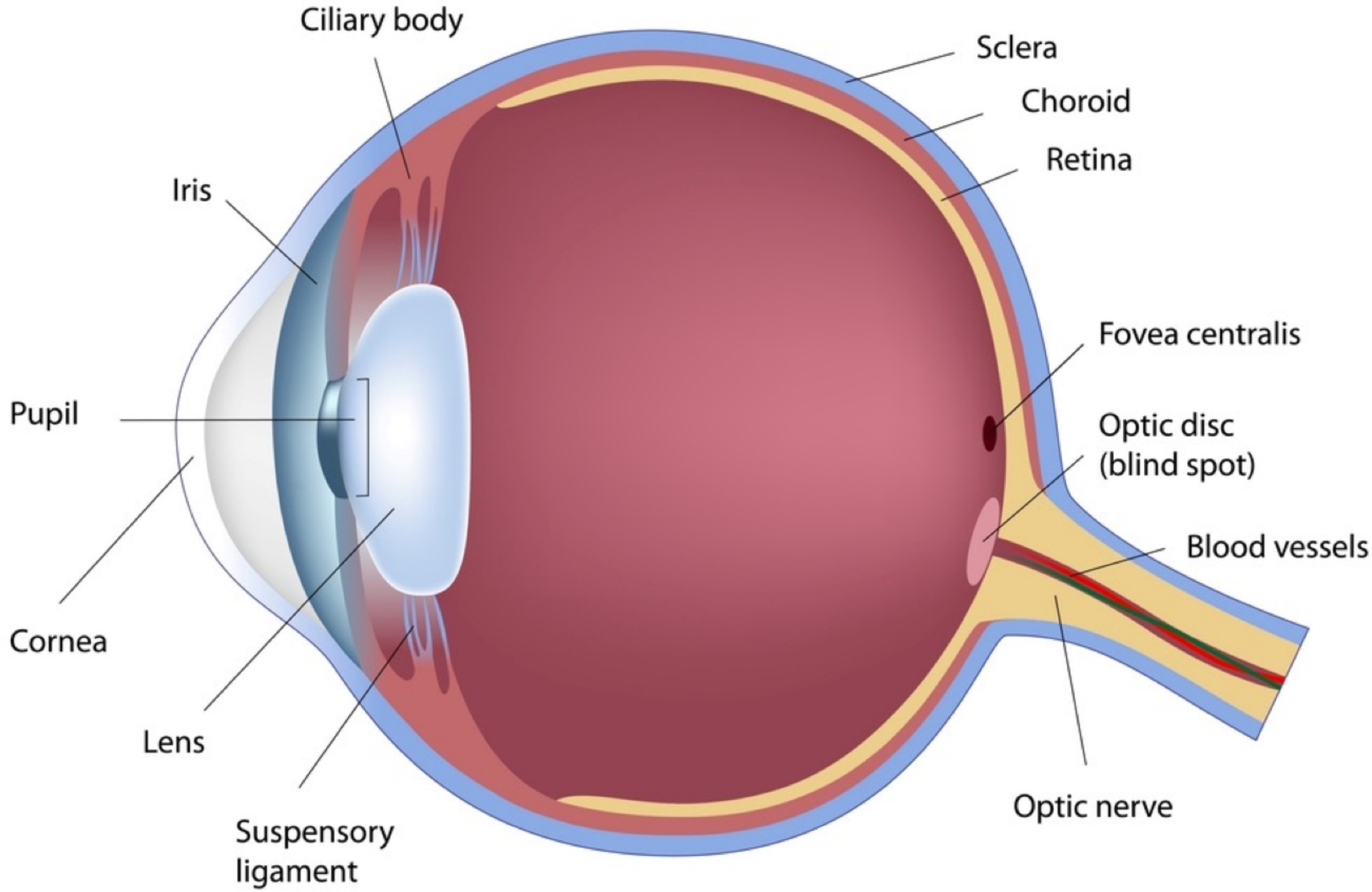


Lens diagram from Open Source VR Project (OSVR)
(Not the lens system from the Oculus Rift)
http://www.osvr.org/

Display Requirements Derive From Human Perception

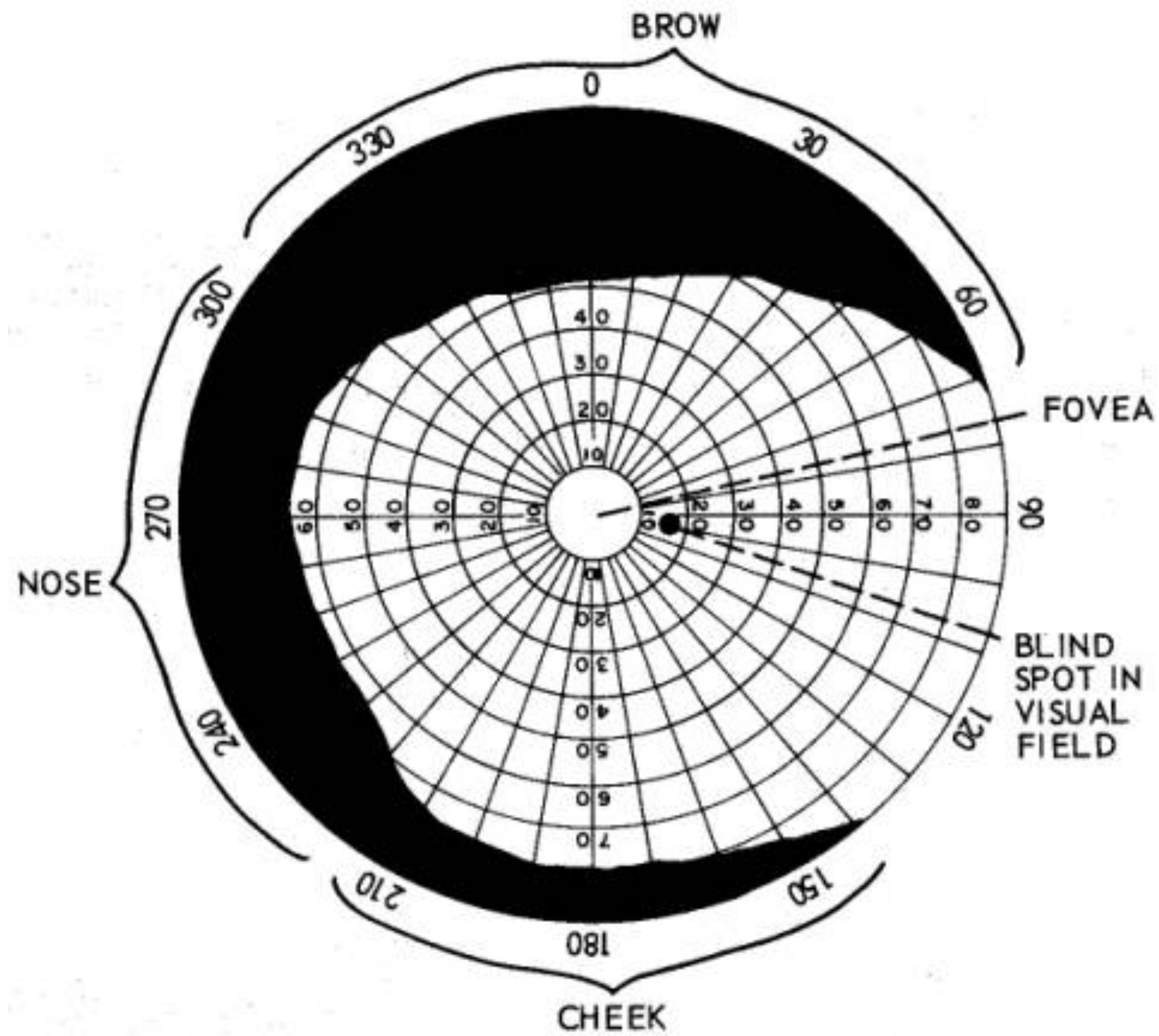
Example: Field of View & Resolution

Anatomy of The Human Eye

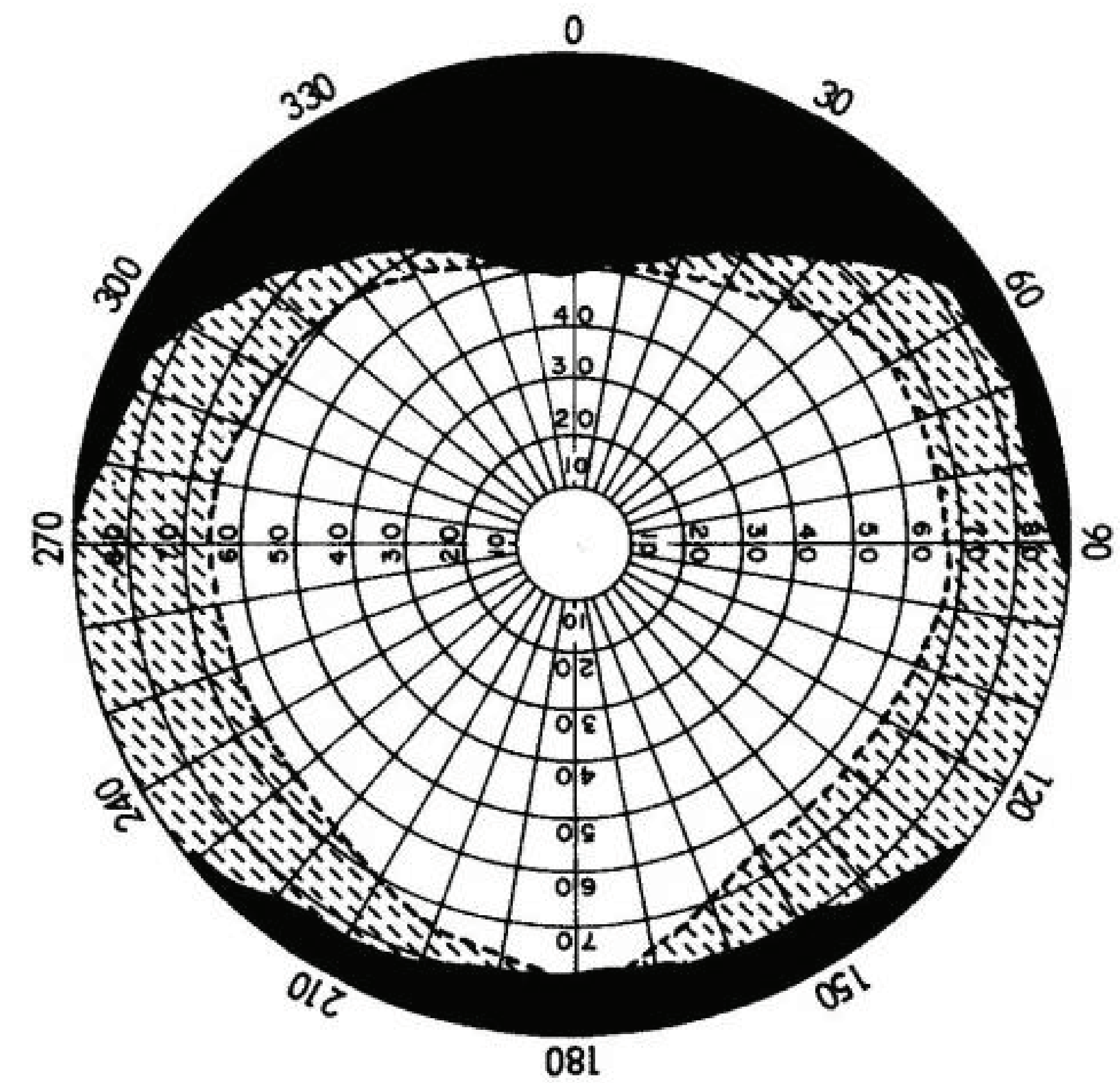


Human Visual Field of View

Ruch & Fulton, 1960



monocular visual field



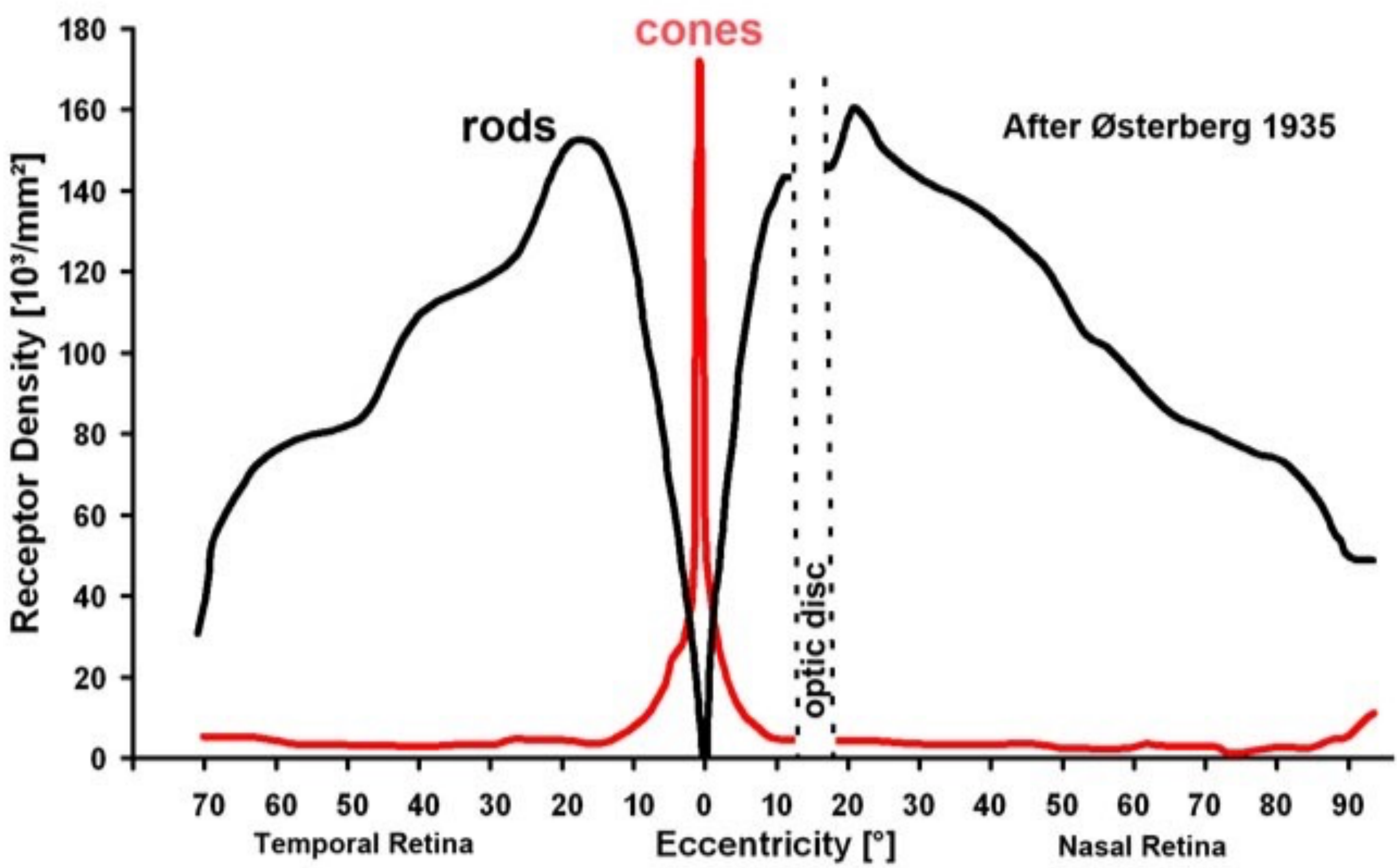
binocular visual field

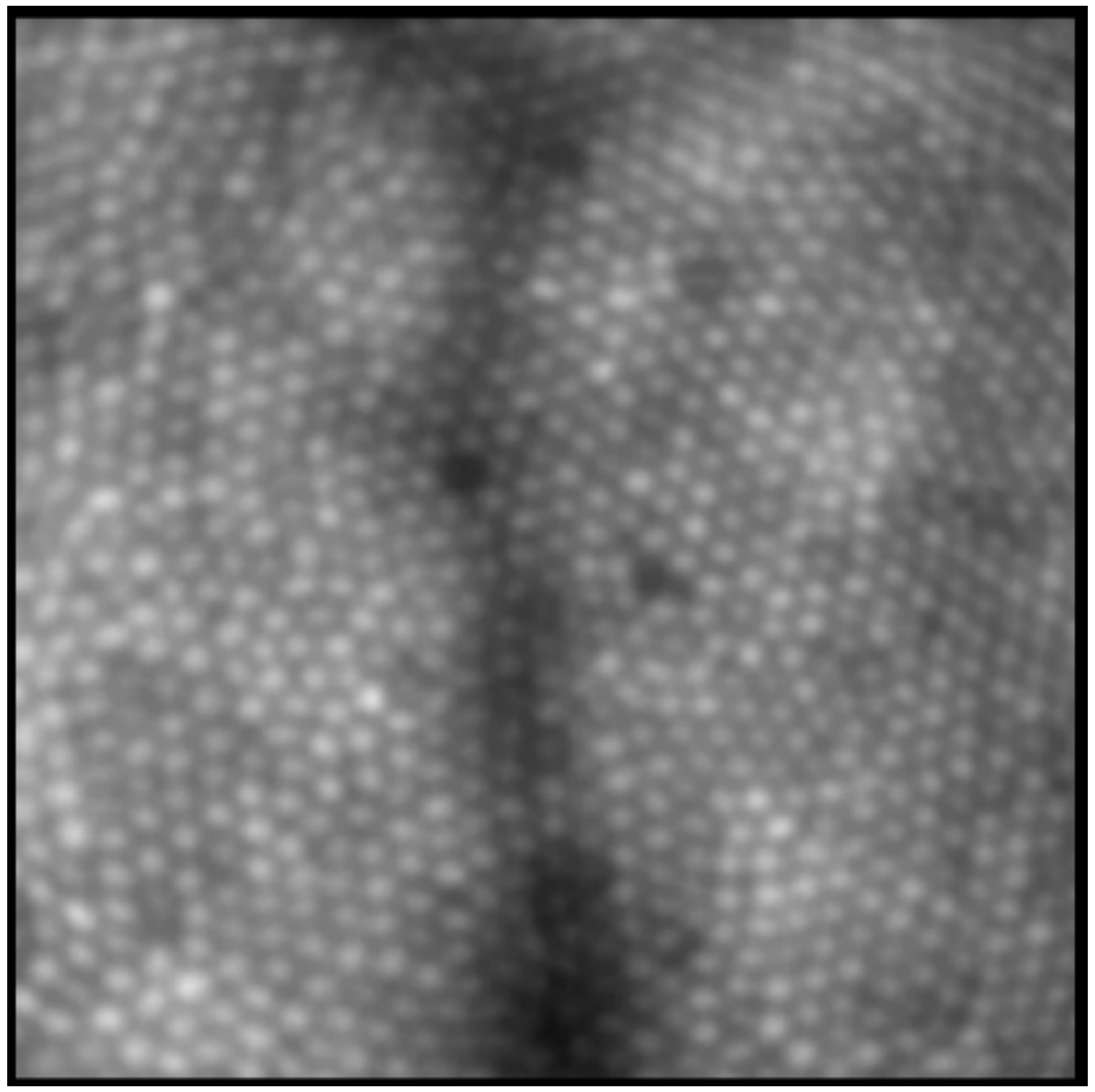
Human: ~160° view of field per eye (~200° overall)

(Note: does not account for eye's ability to rotate in socket)

Slide credit: Gordon Wetzstein

Spatial Resolution of Rods and Cones in the Retina





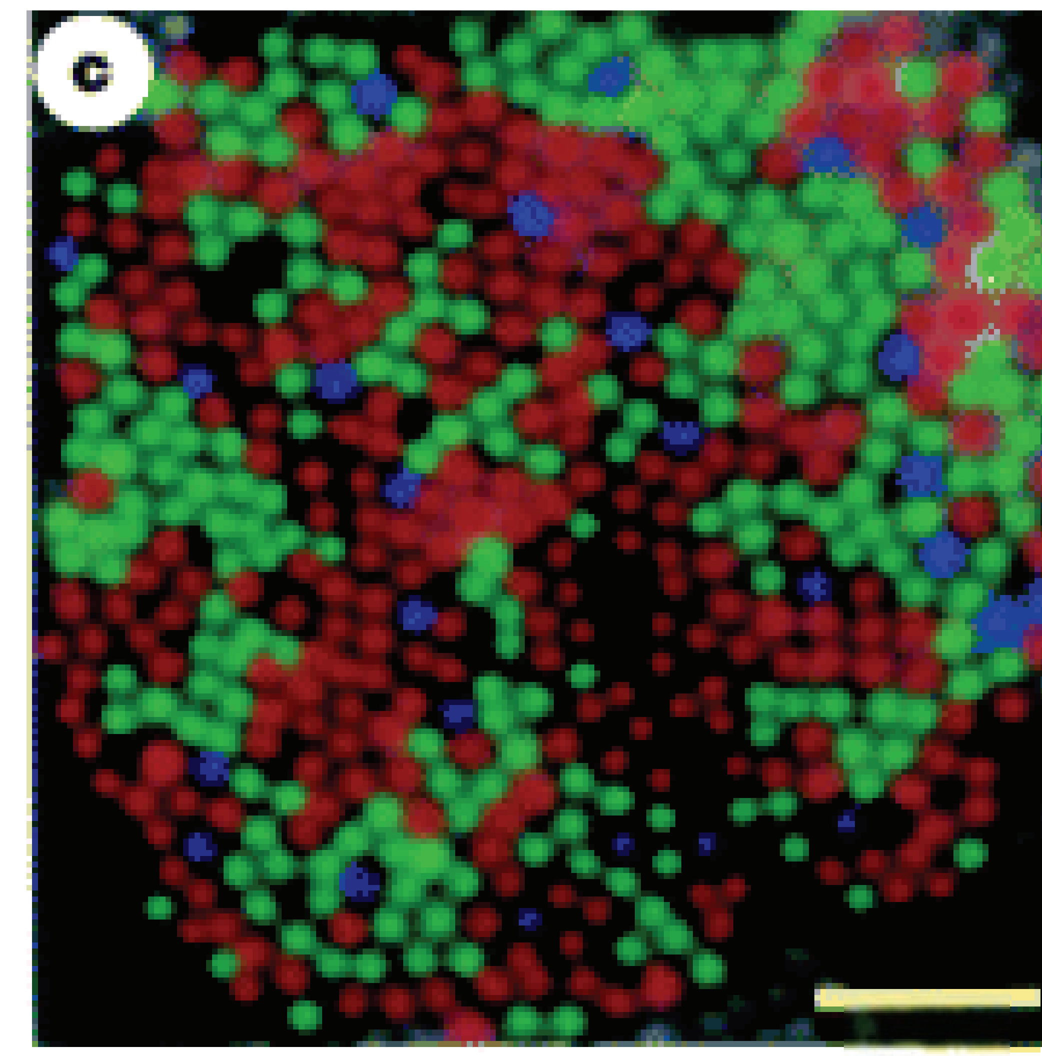
[Roorda 1999]

- Highest density of cones in fovea (and no rods there)
- "Blind spot" at the optic disc, where optic nerve exits eye

Visual Acuity

Roorda & Williams, 1999, Nature

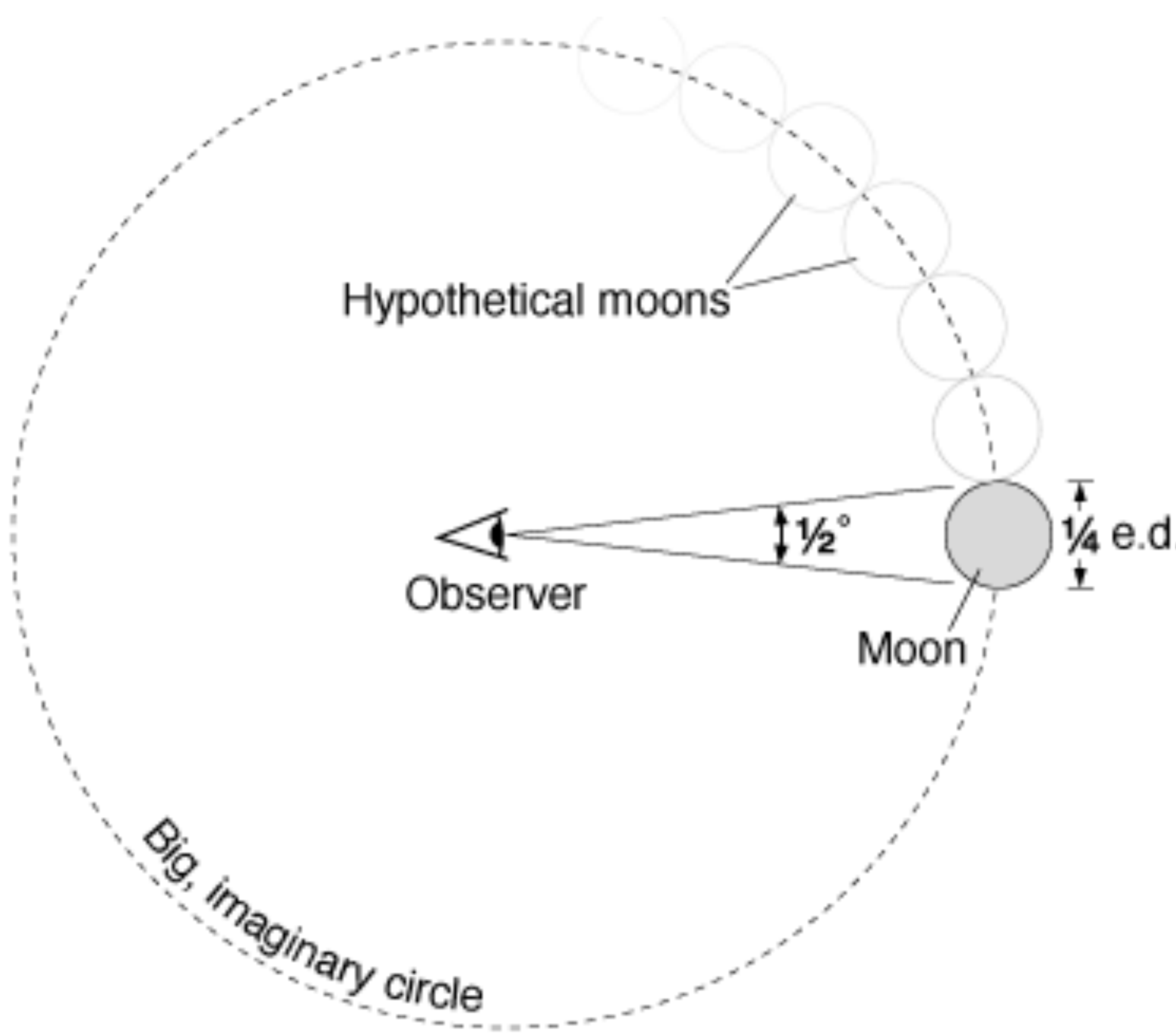
each photorecepter
~ <u>1 arc min</u> (1/60 of a degree)



5 arcmin visual angle

Slide credit: Gordon Wetzstein

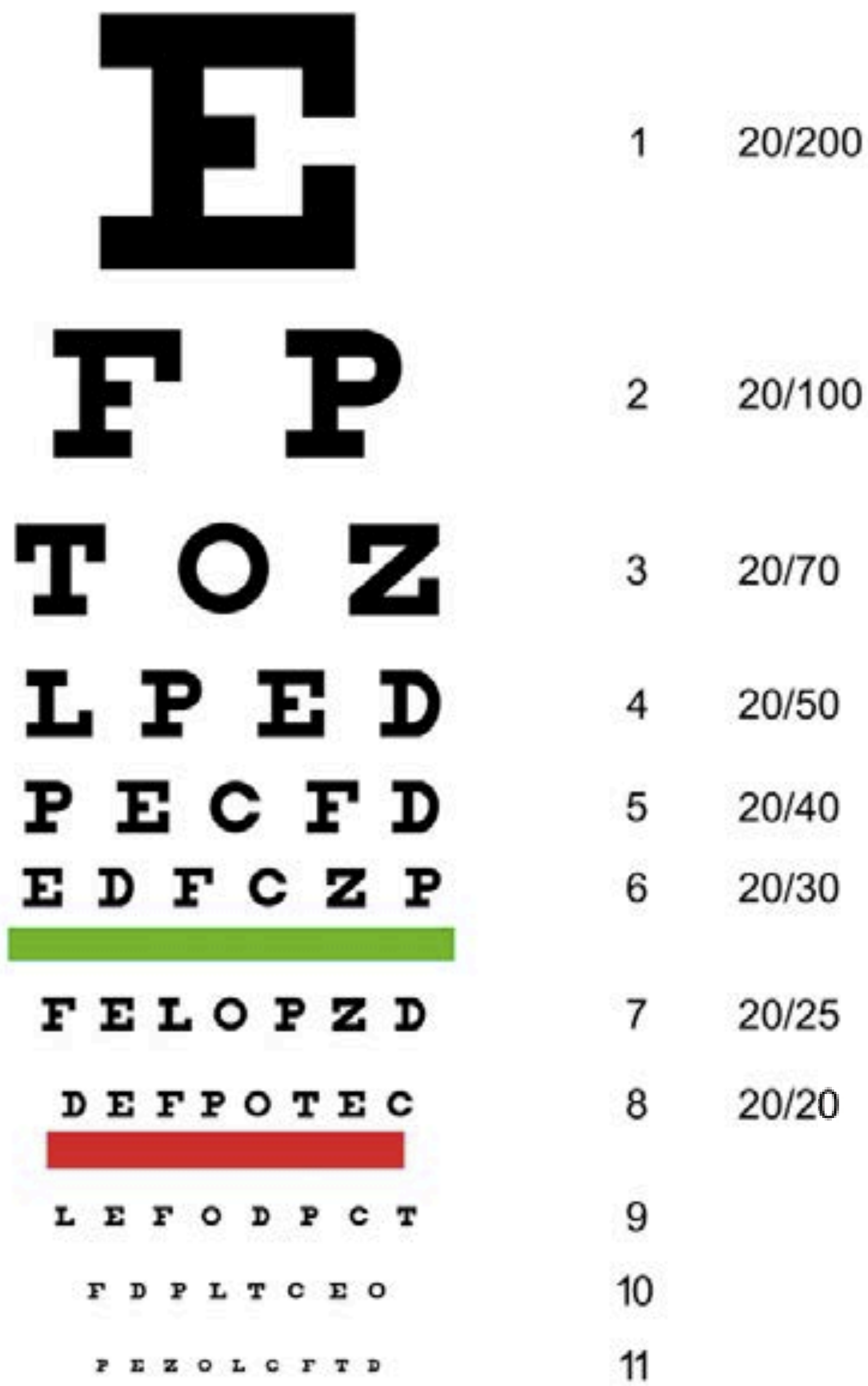
Moon is 1/2 degree (30 arc mins)



Caption: "The half-degree angle has been greatly exaggerated"

Visual Acuity

Snellen chart



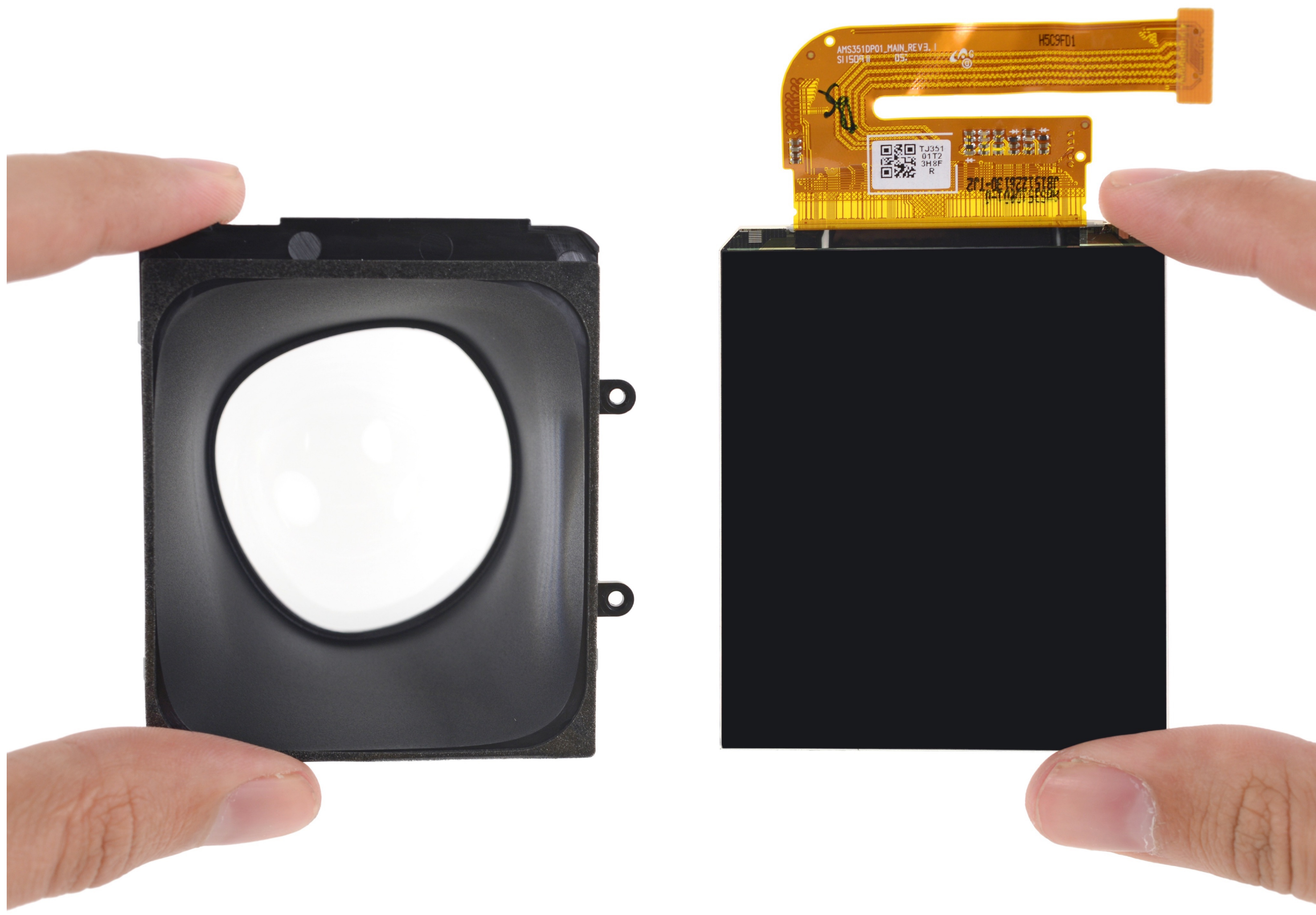
characters are 5 arc min, need to resolve 1 arc min to read

Slide credit: Gordon Wetzstein

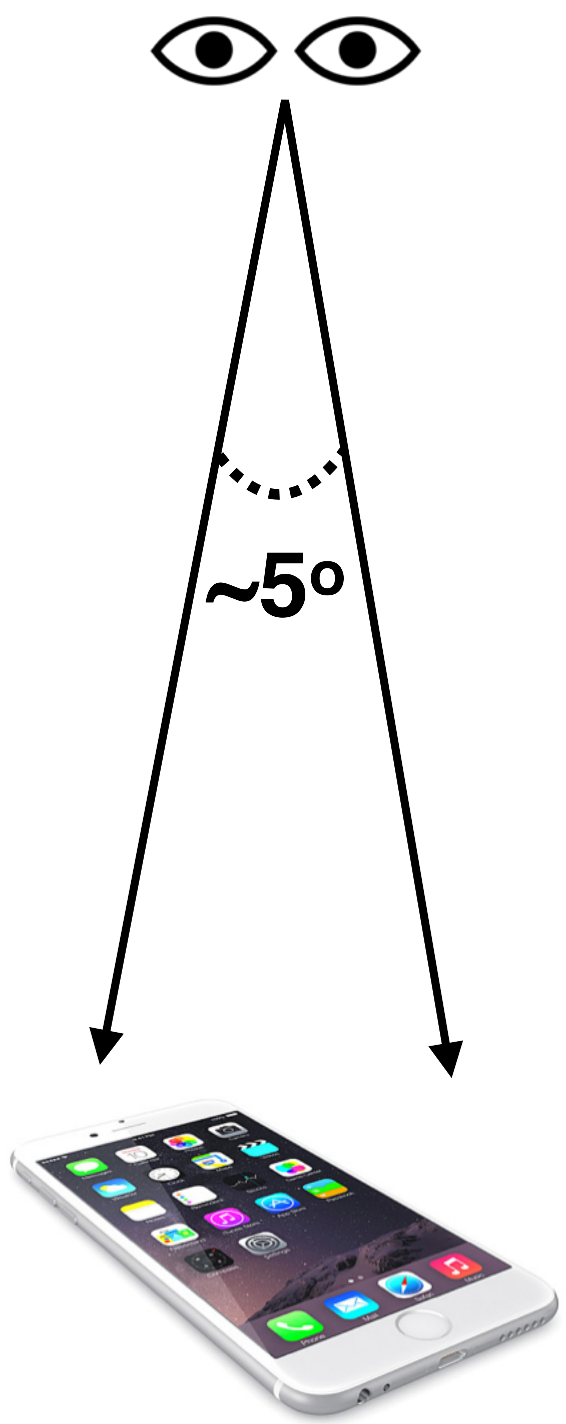
Current VR Headset Field of View and Resolution

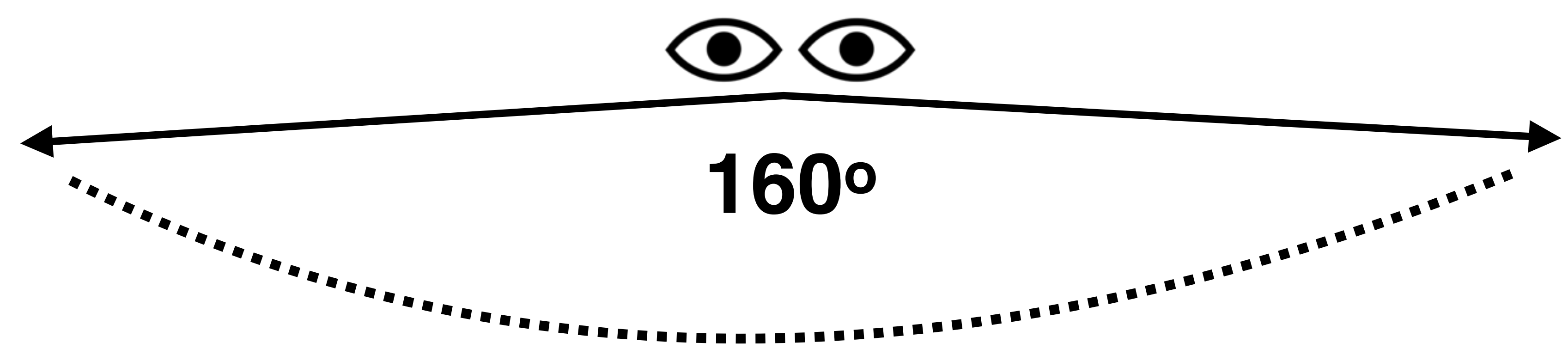
Example: Oculus Rift

- Field of view: approximately 100° per eye
- Resolution: 1080 x 1200 pixel display
- About 10 pixels per degree (as opposed to ~60 samples for 20/20 vision)
- Newer HP/Valve "Reverb G2" headset is 2160 x 2160 pixels!



A VR Display at Human Visual Acuity





Human: ~160° view of field per eye (~200° overall)
(Note: does not account for eye's ability to rotate in socket)

Future "retina" VR display:
57 ppd covering 200°
= 11K x 11K display per eye
= 220 MPixel

iPhone 6: 4.7 in "retina" display:
1.3 MPixel
326 ppi → 57 ppd

Strongly suggests need for eye tracking and foveated rendering (eye can only perceive detail in 5° region about gaze point)

Eye icons designed by SuperAtic LABS from the thenounproject.com

Display Requirements Derive From Human Perception

Example: Binocular Stereo and Eye Focus ("Accommodation")

Two Eyes: Two Views

Charles Wheatstone stereoscope, 1838

Recall: Current VR HMD Optical Design

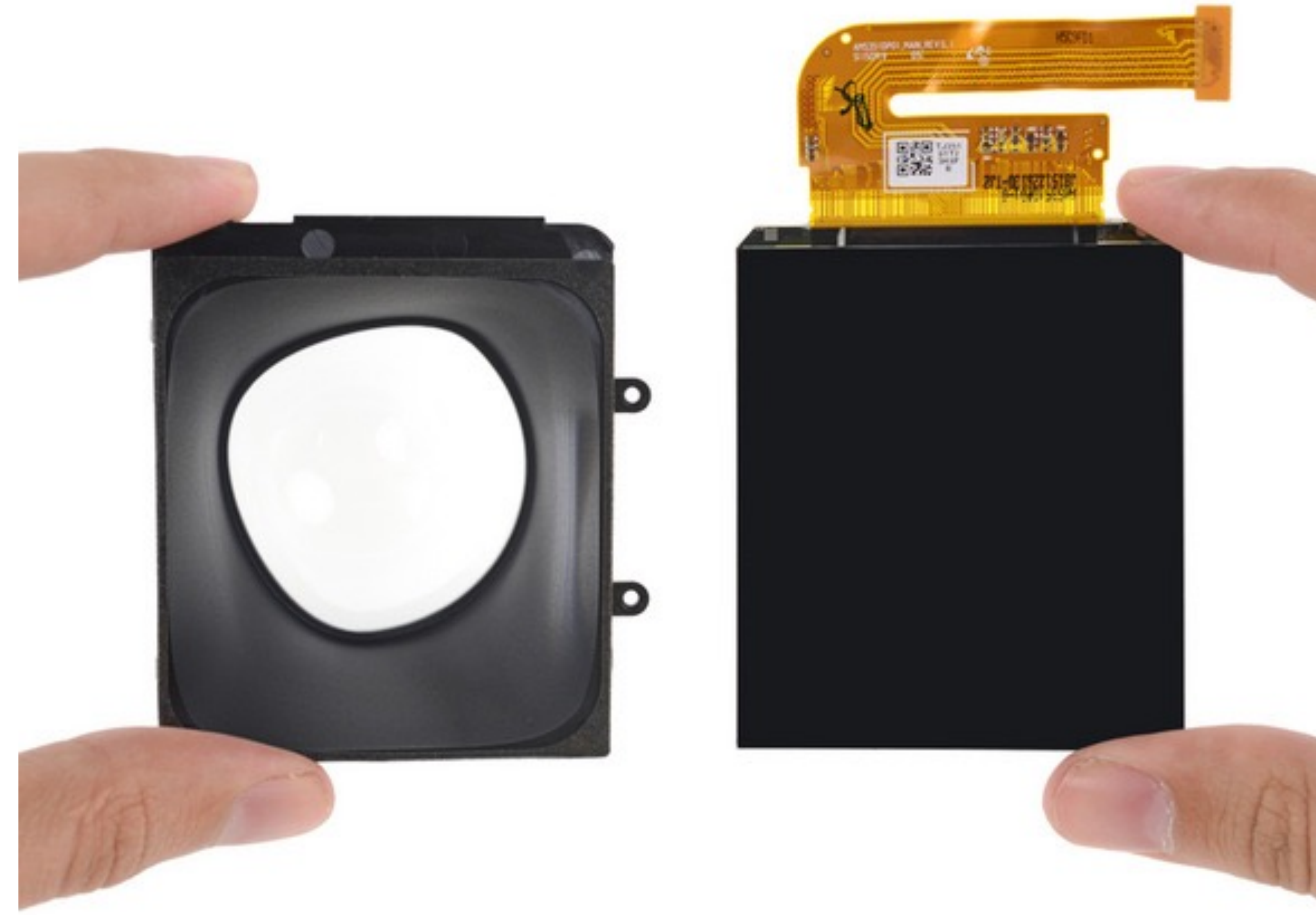
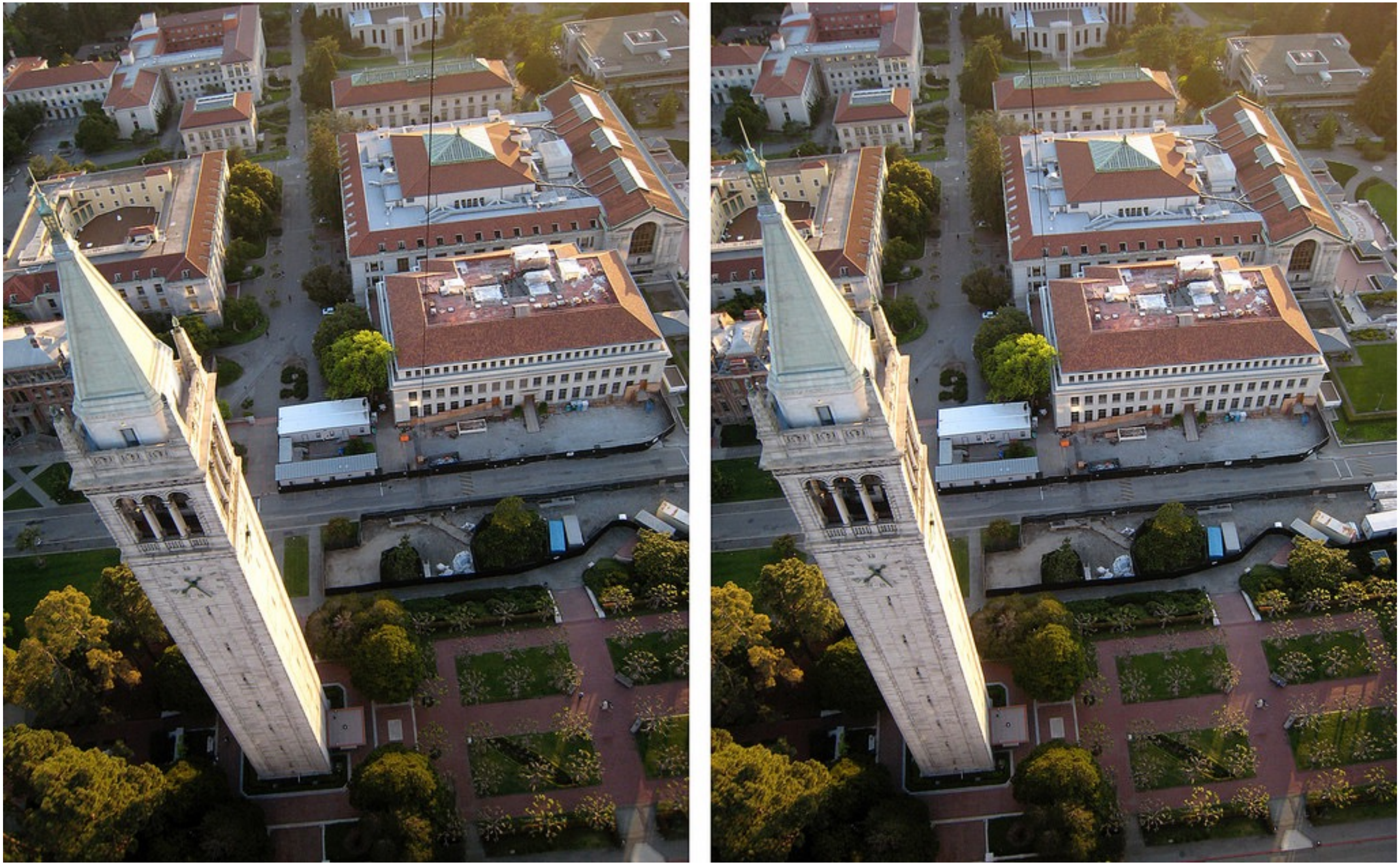
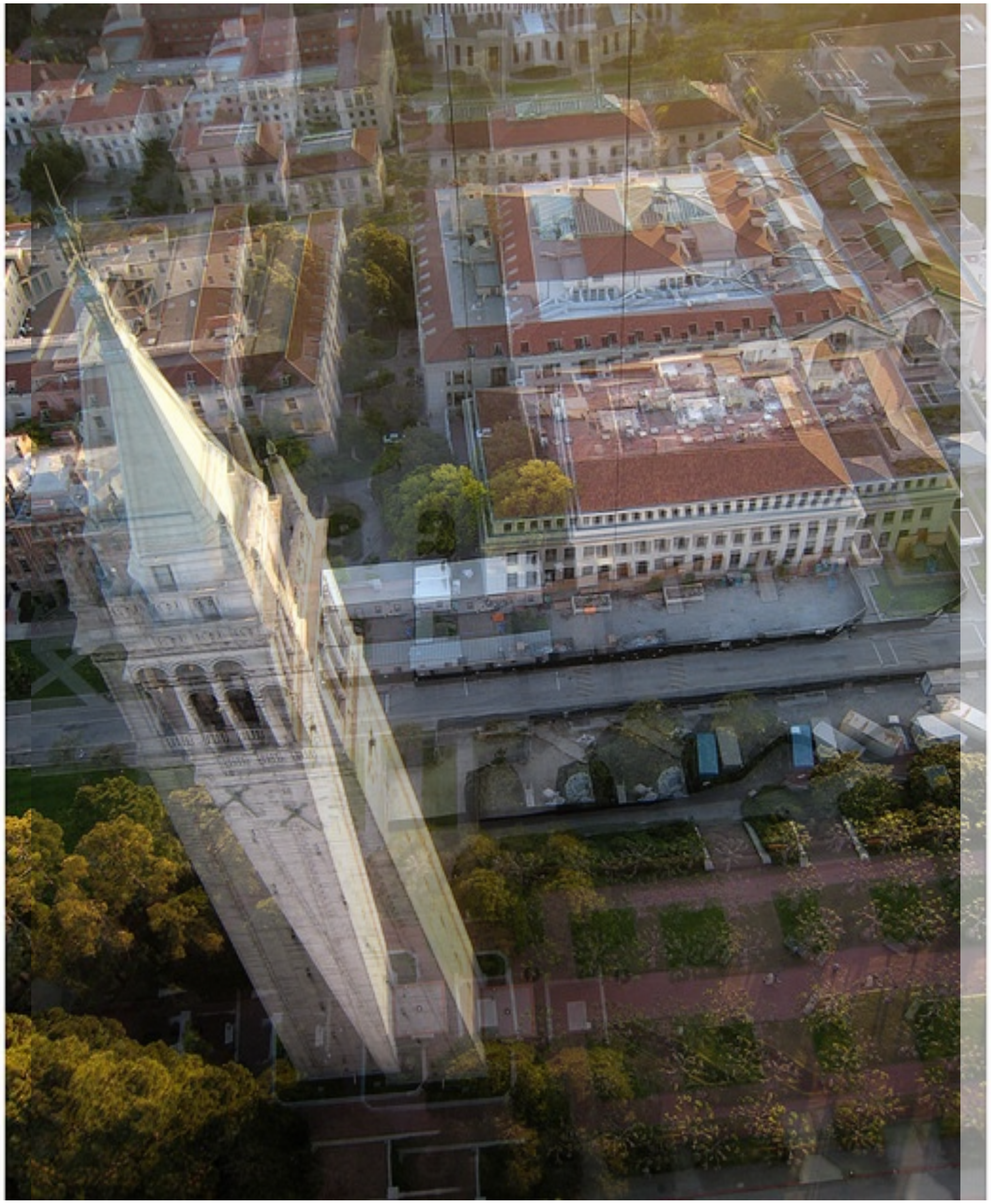


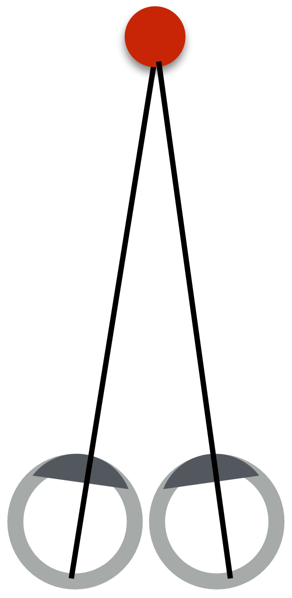
Image credit: ifixit.com https://www.ifixit.com/Teardown/Oculus+Rift+CV1+Teardown/60612

Stereo Vergence

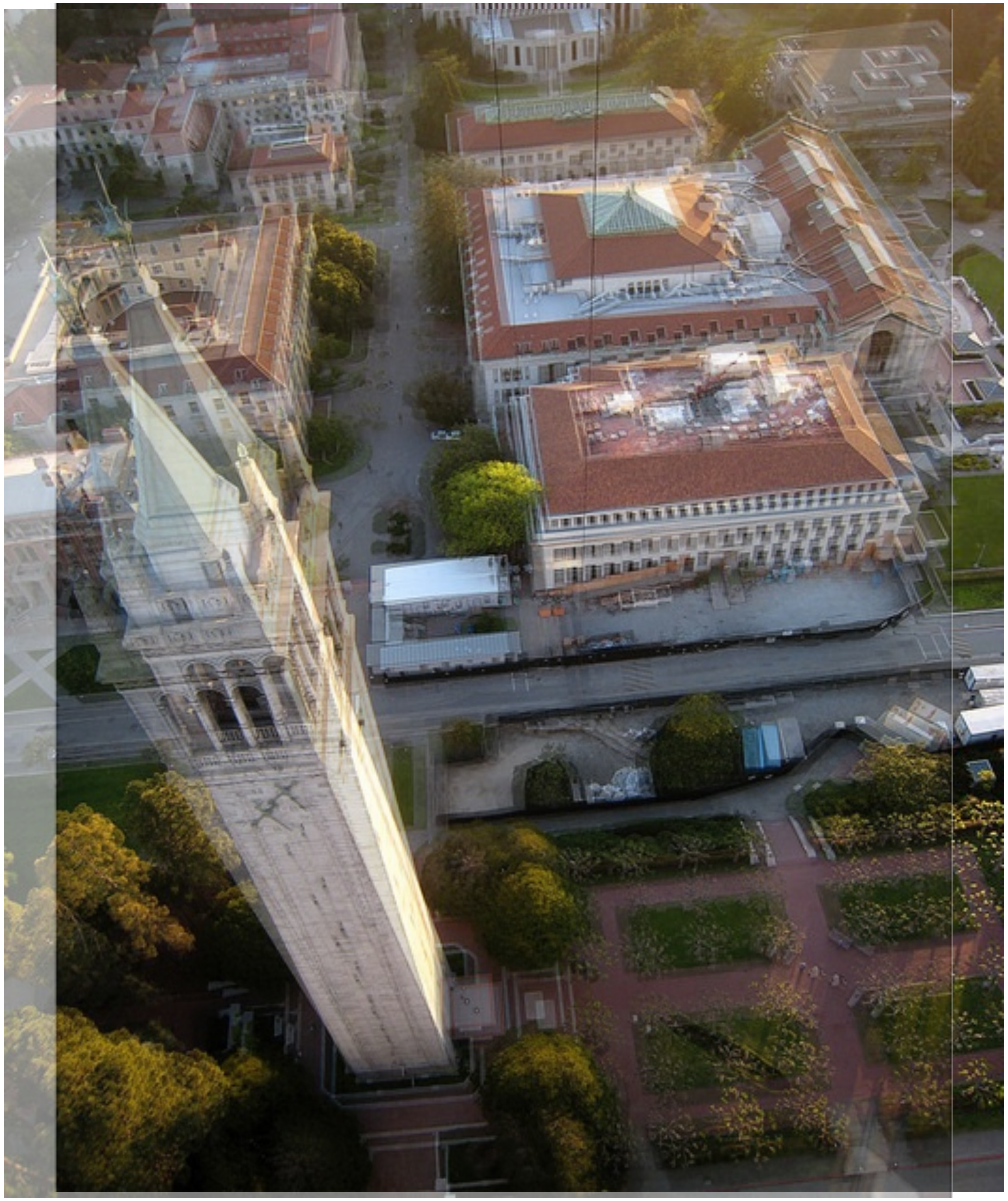


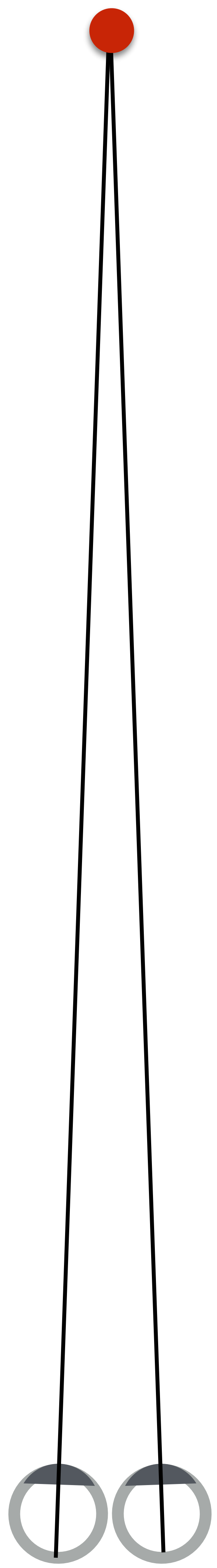
Stereo Vergence





Stereo Vergence

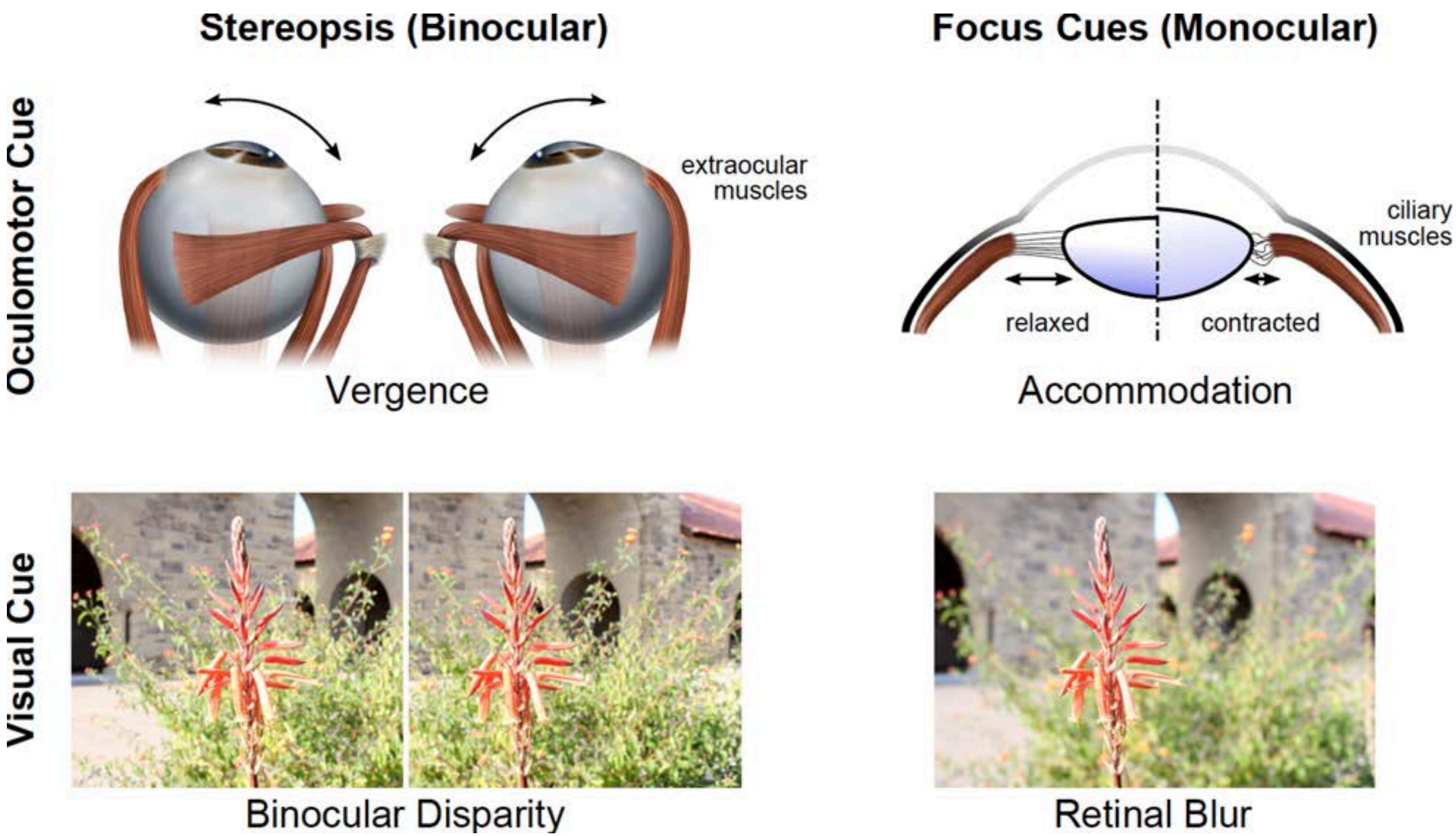




Stereo

- Passive (no tracking of eyes)
- Present each eye with perspective view corresponding to that eye's location relative to the other eye
- Eyes will con(verge) by rotating physically in sockets in order to bring closer and further objects into physical alignment on retina

Human Eye Muscles and Optical Controls



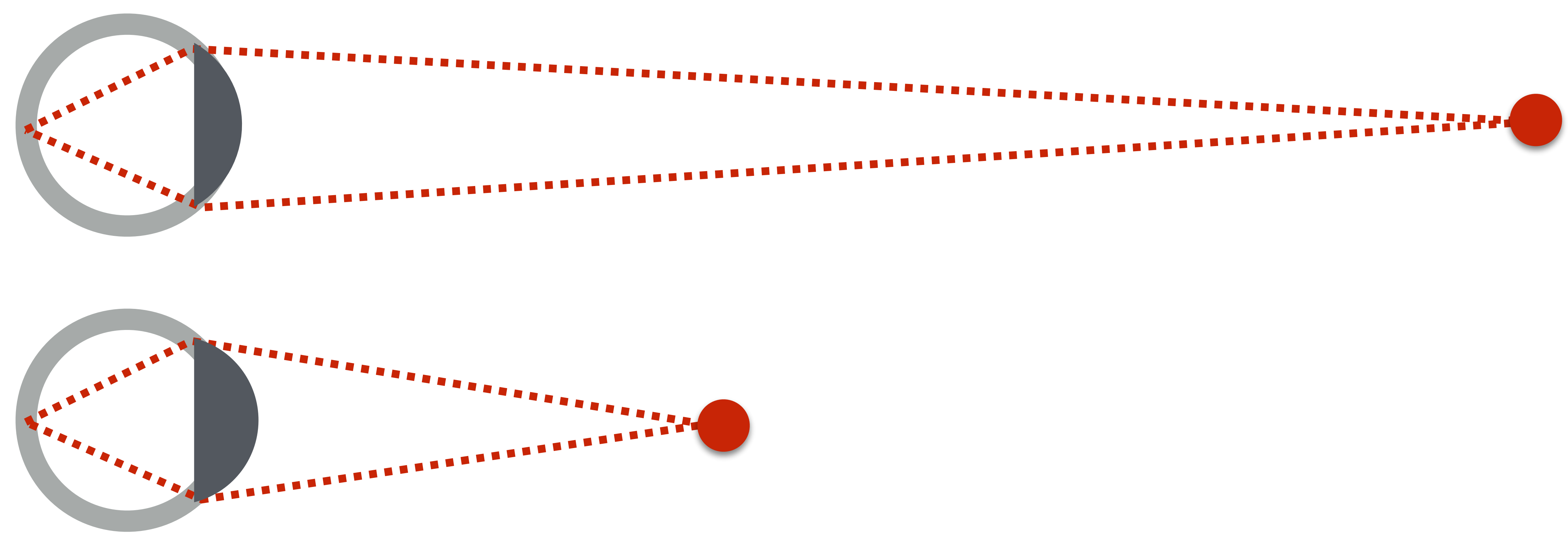
Slide credit: Gordon Wetzstein

Accommodation and Vergence

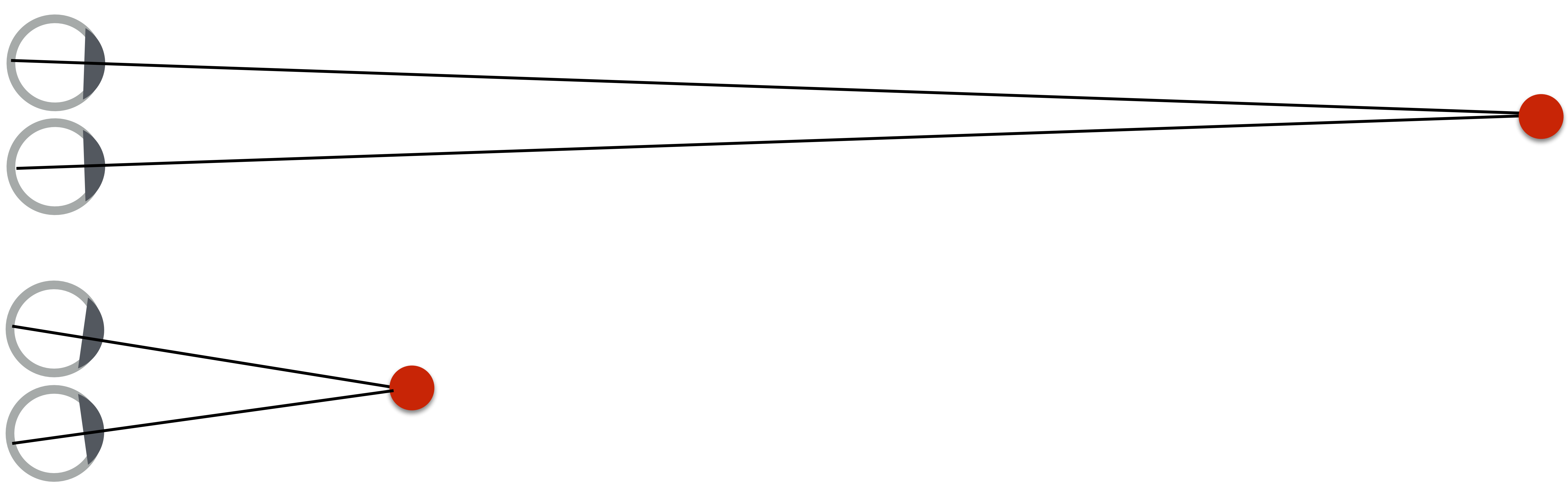
Accommodation: changing the optical power of the eye (lens) to focus at different distances

Eye accommodated to focus on a distant object

Eye accommodated to focus on a nearby object



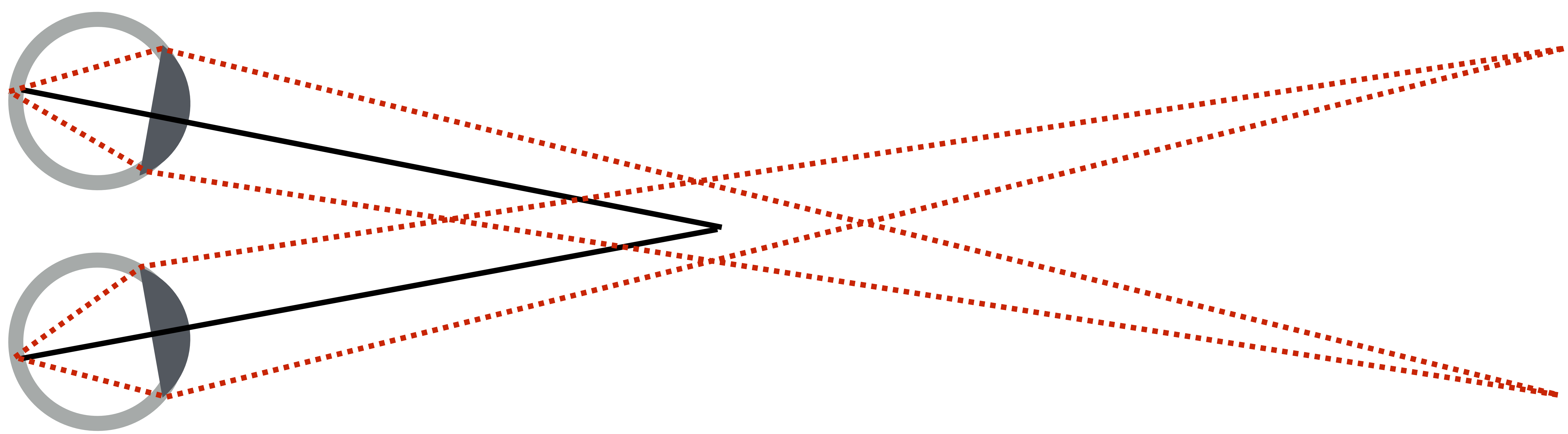
Vergence: rotation of the eye in its socket to ensure projection of object is centered on the retina



Accommodation – Vergence Conflict

Given design of current VR displays, consider what happens when objects are up-close to eye in virtual scene

- Eyes must remain accommodated to far distance (otherwise image on screen won't be in focus)
- But eyes must converge in attempt to fuse stereoscopic images of object up close
- Brain receives conflicting depth clues… (discomfort, fatigue, nausea)

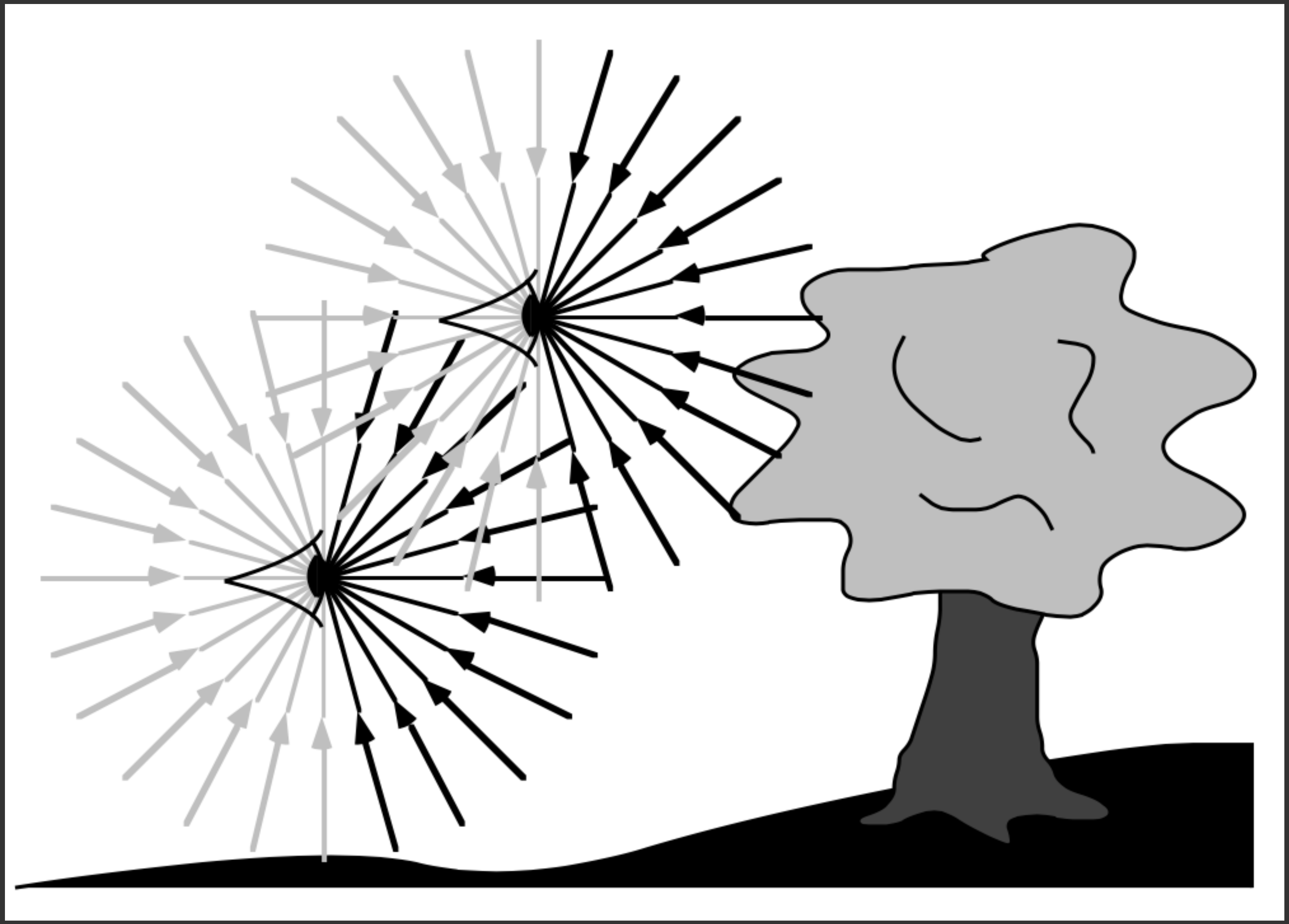


This problem stems from nature of display design. If you could just make a display that emits the light field that would be produced by a virtual scene, then you could avoid the accommodation - vergence conflict…

Display Requirements Derive From Human Perception

Example: Parallax from Head Motion

The 5D Plenoptic Function



$$P(x, y, z, \theta, \phi)$$

3D Position

2D Direction

[Adelson, Bergen 1991]

Google Cardboard: Tracking Using Headset Camera

Tracking uses gyro / rear-facing camera to estimate user's viewpoint

- 2D rotation tracking generally works well
- 3D positional tracking a challenge in general environments

Environment-Supported Vision-Based Tracking?

Image credit: gizmodo.com

Early VR test room at Valve, with markers positioned throughout environment

Oculus Rift IR LED Tracking System

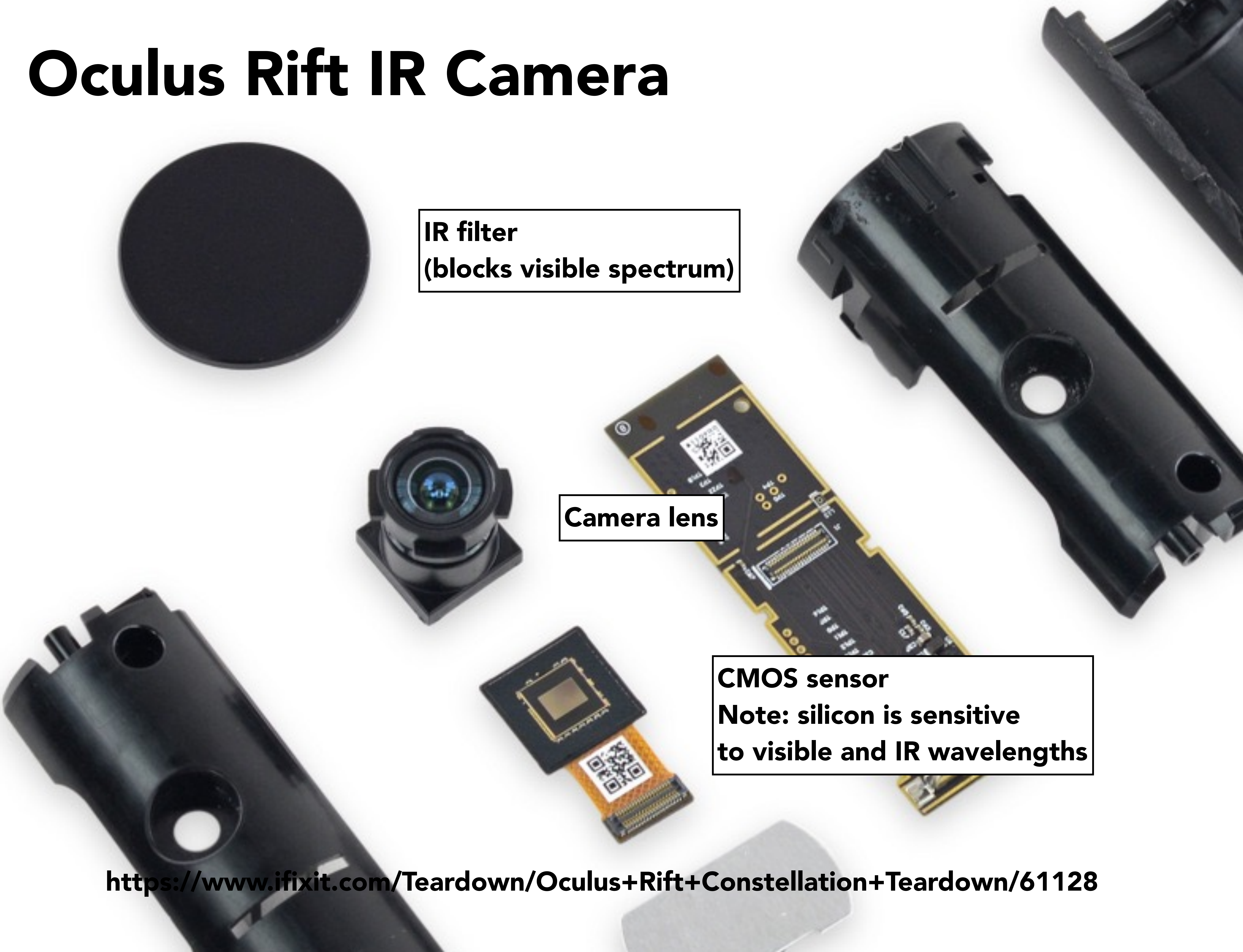
Oculus Rift + IR LED sensor

Oculus Rift IR LED Tracking Hardware

Photo taken with IR-sensitive camera

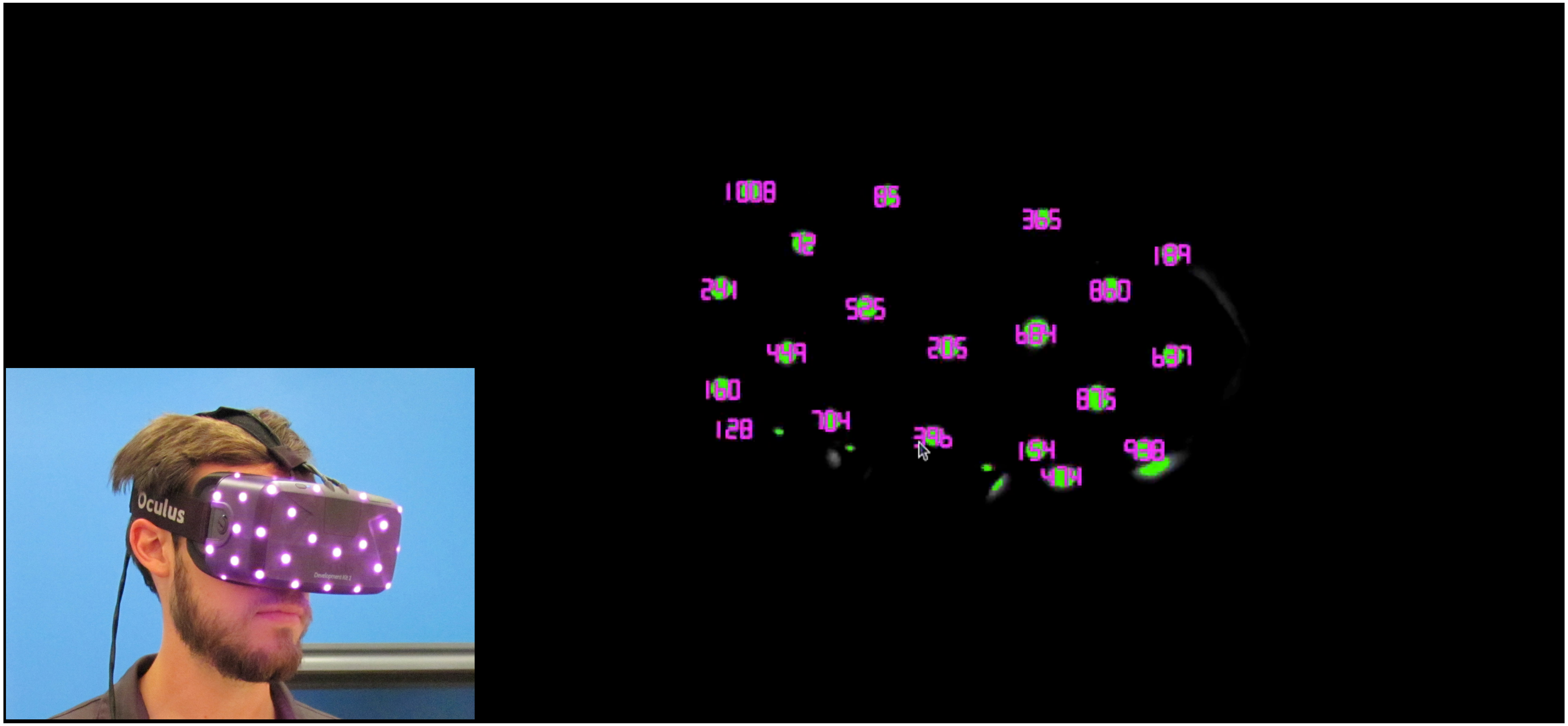
https://www.ifixit.com/Teardown/Oculus+Rift+Constellation+Teardown/61128

Oculus Rift IR Camera



https://www.ifixit.com/Teardown/Oculus+Rift+Constellation+Teardown/61128

Oculus Rift Uses Active Marker Motion Capture



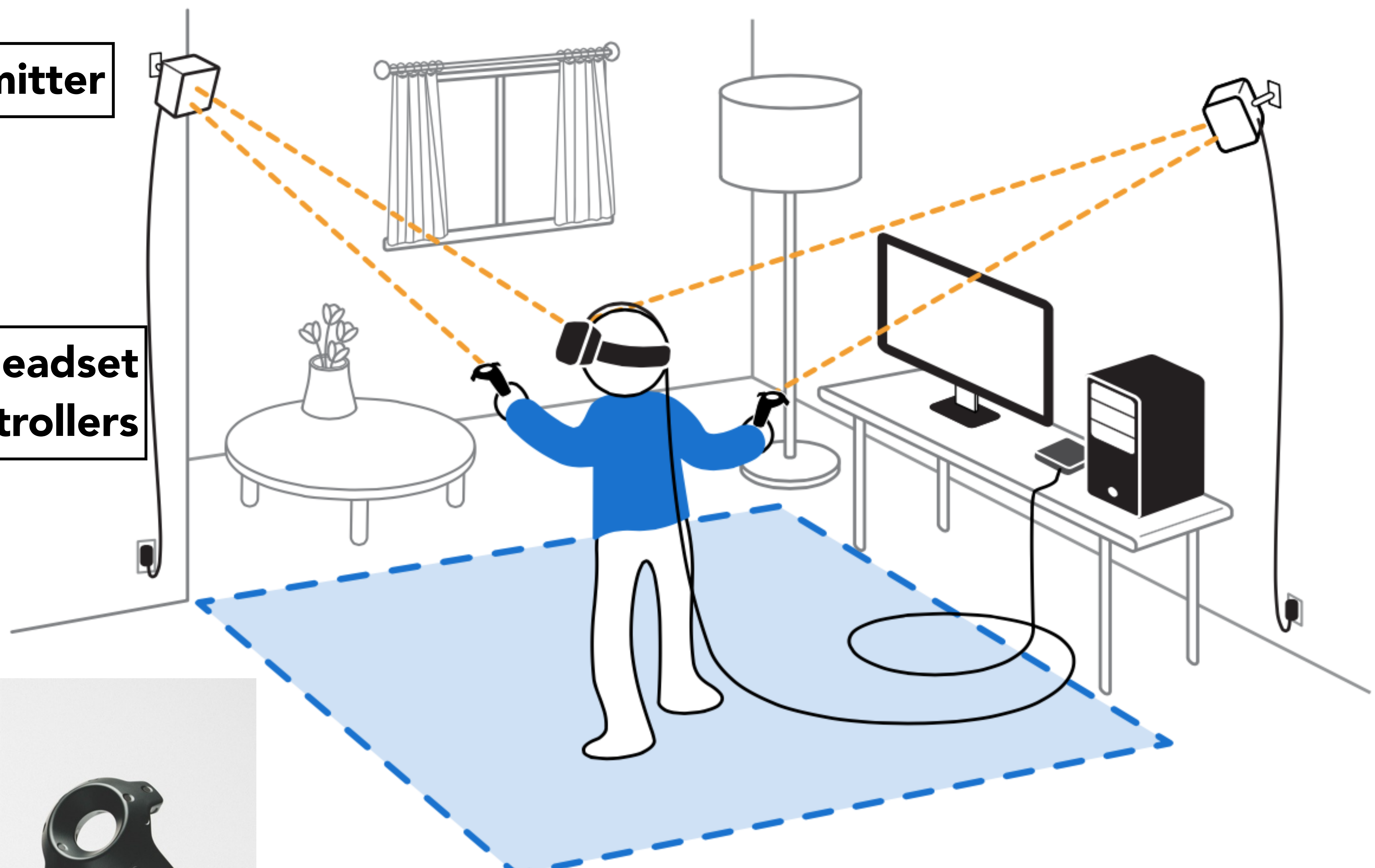
Credit: Oliver Kreylos, https://www.youtube.com/watch?v=O7Dt9Im34OI

- Motion capture: unknown shape, multiple cameras
- VR head tracking: known shape, single camera

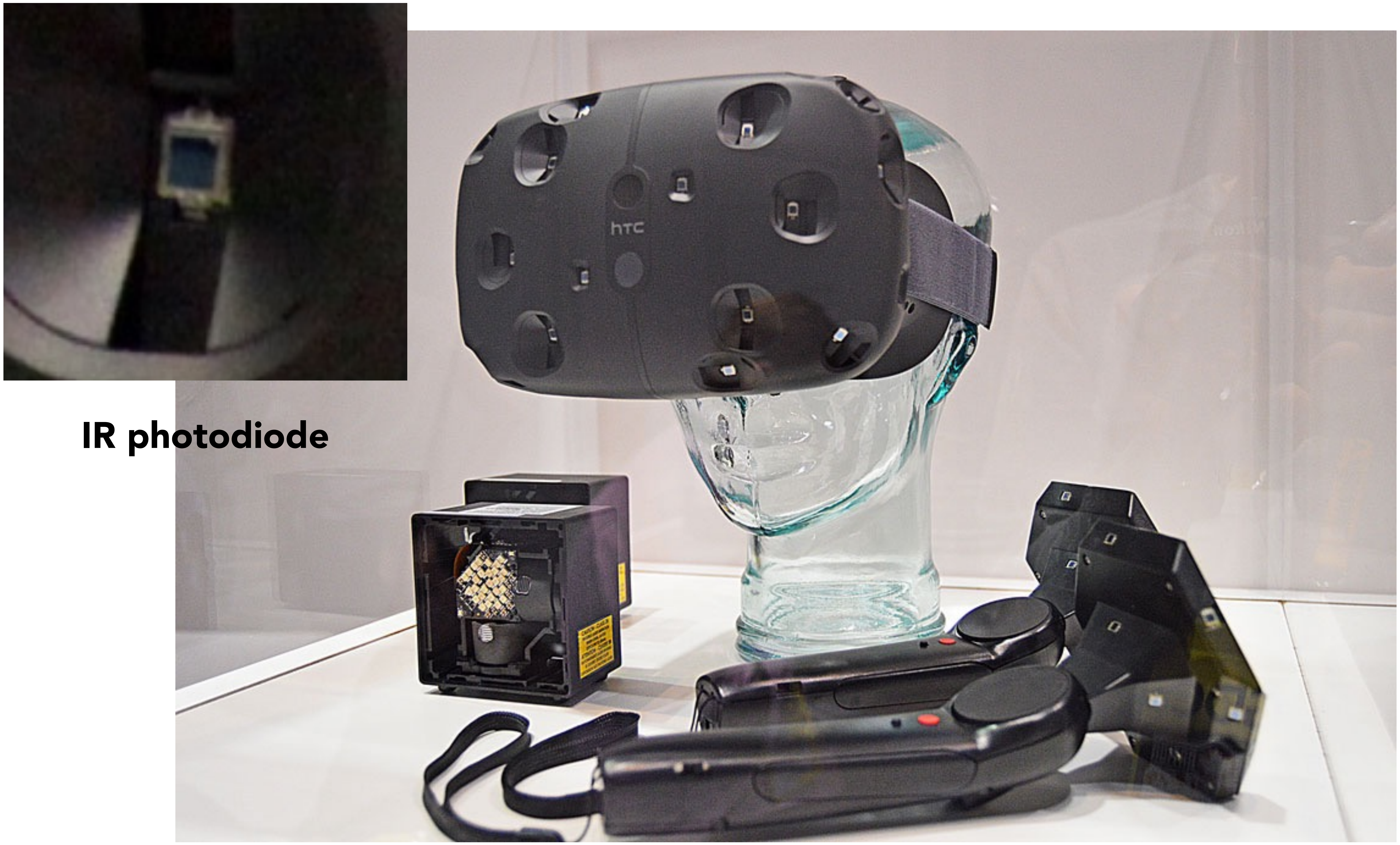
HTC Vive Tracking System ("Lighthouse")

Structured light transmitter

Photodiode arrays on headset and hand-held controllers



Vive Headset & Controllers Have Array of IR Photodiodes

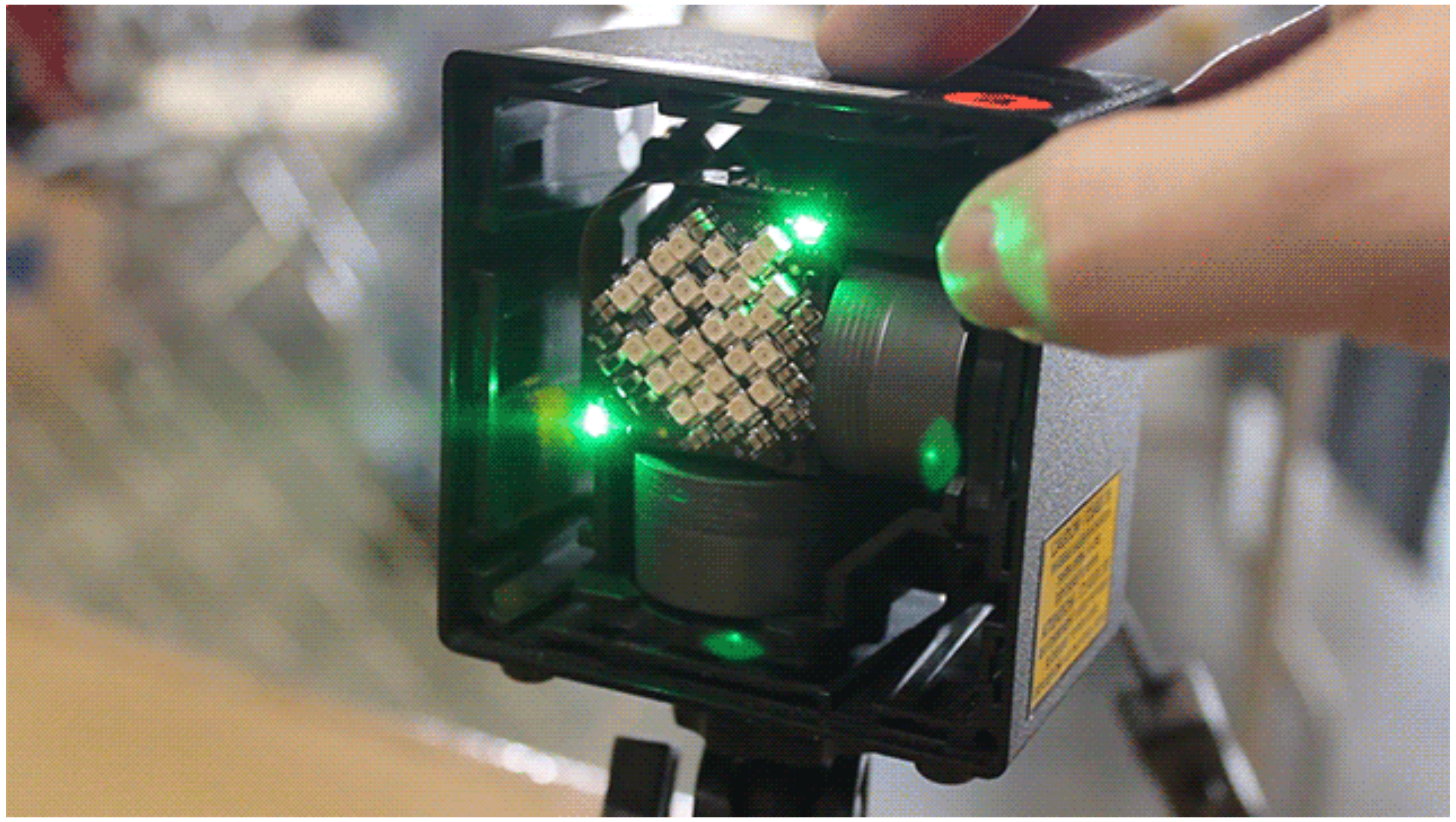


IR photodiode

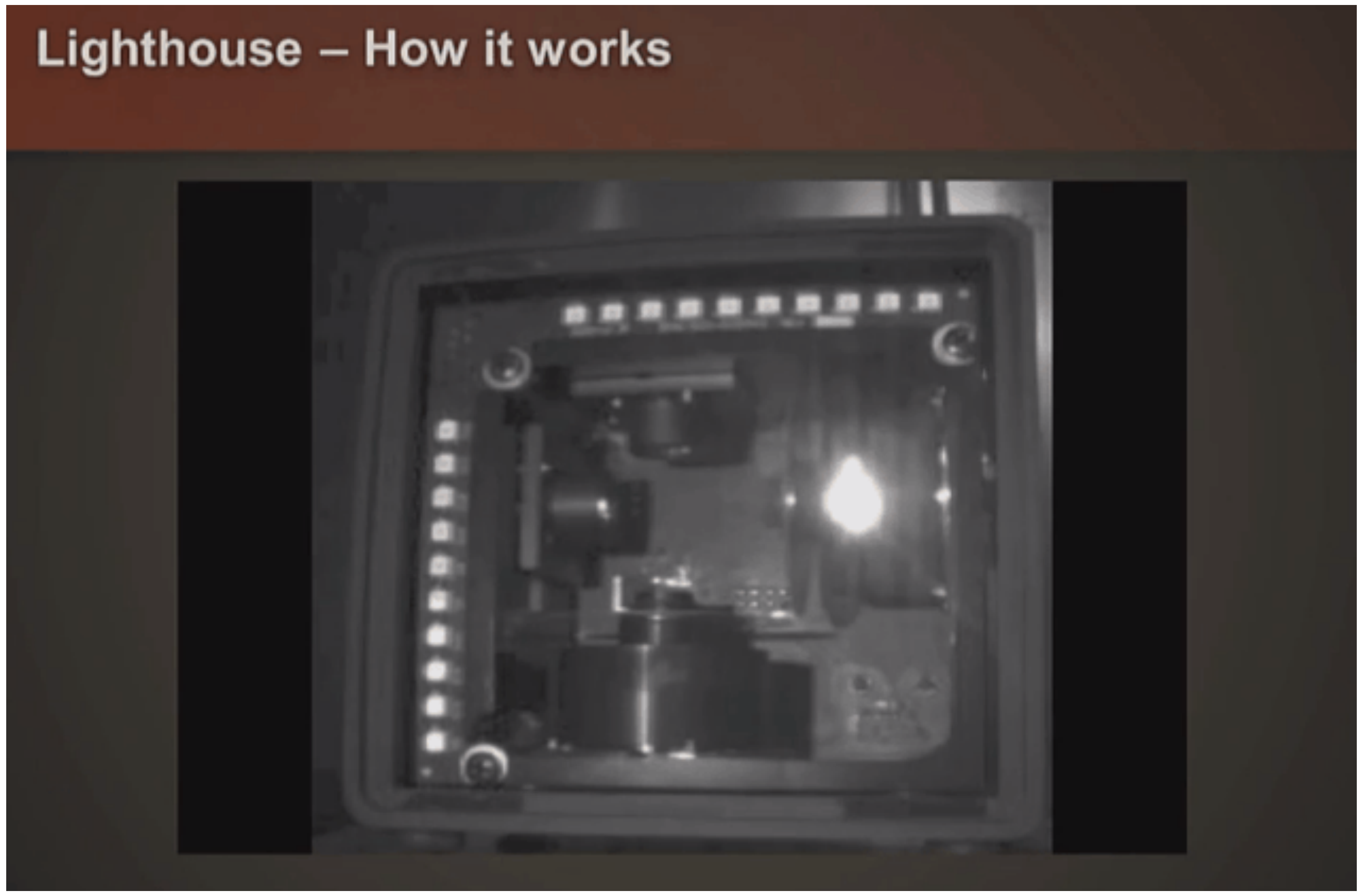
Image credit: uploadvr.com

(Prototype) Headset and controller are covered with IR photodiodes

HTC Vive Structured Light Emitter ("Lighthouse")



Light emitter contains array of LEDs (white) and two spinning wheels with lasers



Sequence of LED flash and laser sweeps provide structured lighting throughout room

Credit: Gizmodo: http://gizmodo.com/this-is-how-valve-s-amazing-lighthouse-tracking-technol-1705356768

HTC Vive Tracking System

For each frame, lighthouse does the following:

- LED pulse, followed by horizontal laser sweep
- LED pulse, followed by vertical laser sweep

Each photodiode on headset measures time offset between pulse and laser arrival

- Determines the x and y offset in the lighthouse's field of view
- In effect, obtain an image containing the 2D location of each photodiode in the world
  - (Can think of the lighthouse as a virtual "camera")

HTC Vive Tracking System ("Lighthouse")

Credit: rvdm88 / youtube. https://www.youtube.com/watch?v=J54dotTt7k0

Tracking Summary

Looked at three tracking methods

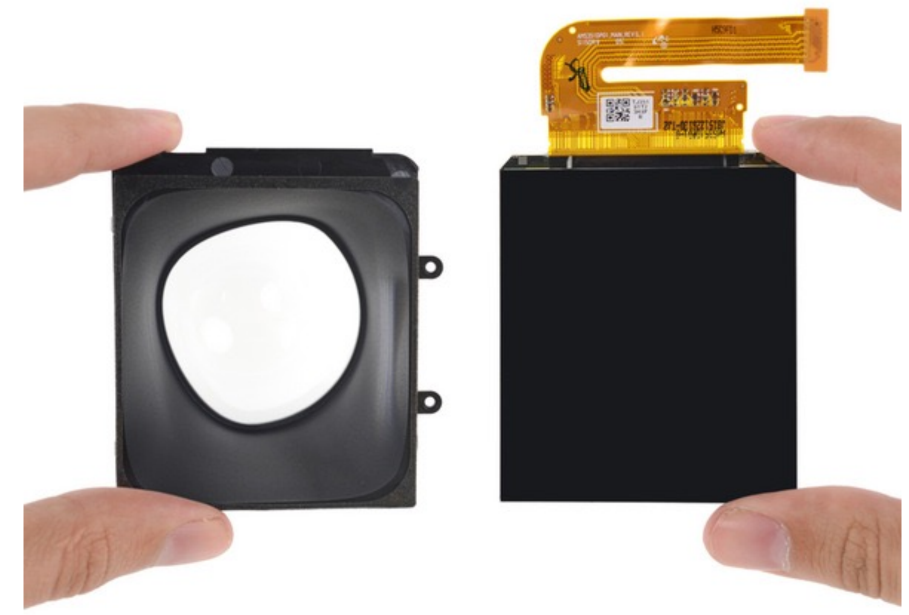
- Camera on headset + computer vision + gyro
- External camera + marker array on headset
- External structured light + sensor array on headset

3D tracking + depth sensing an active research area

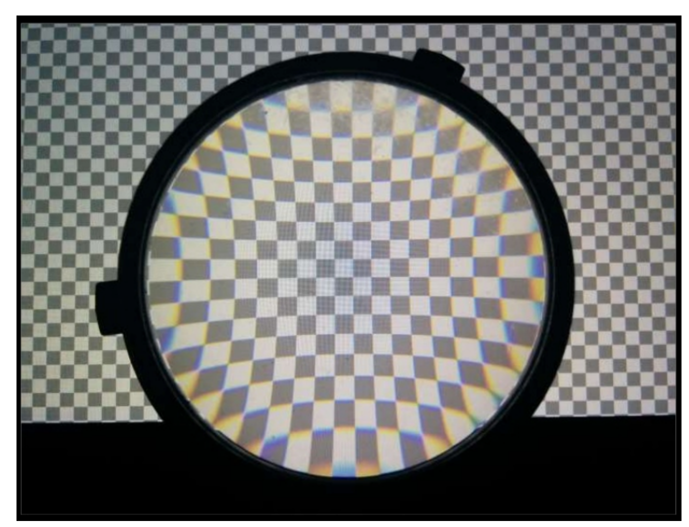
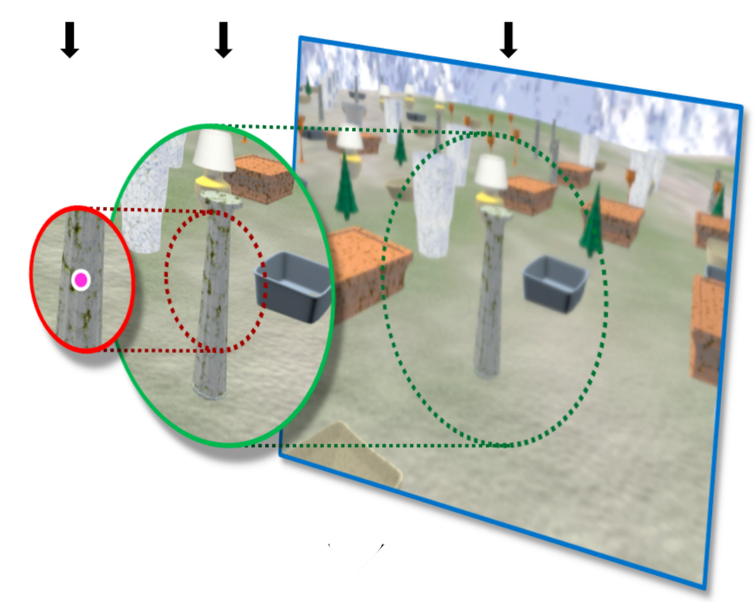
- SLAM, PTAM, DTAM…
- Microsoft Hololens, Magic Leap, Google Tango, Intel Realsense, …

Overview of VR Topics

- VR Displays

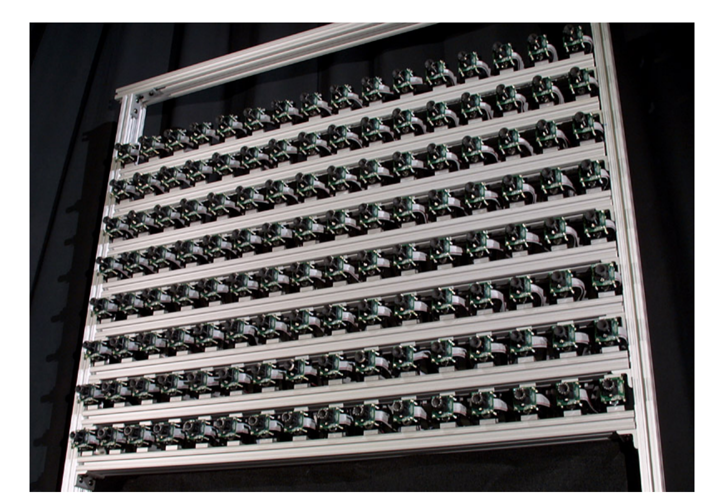




- VR Rendering

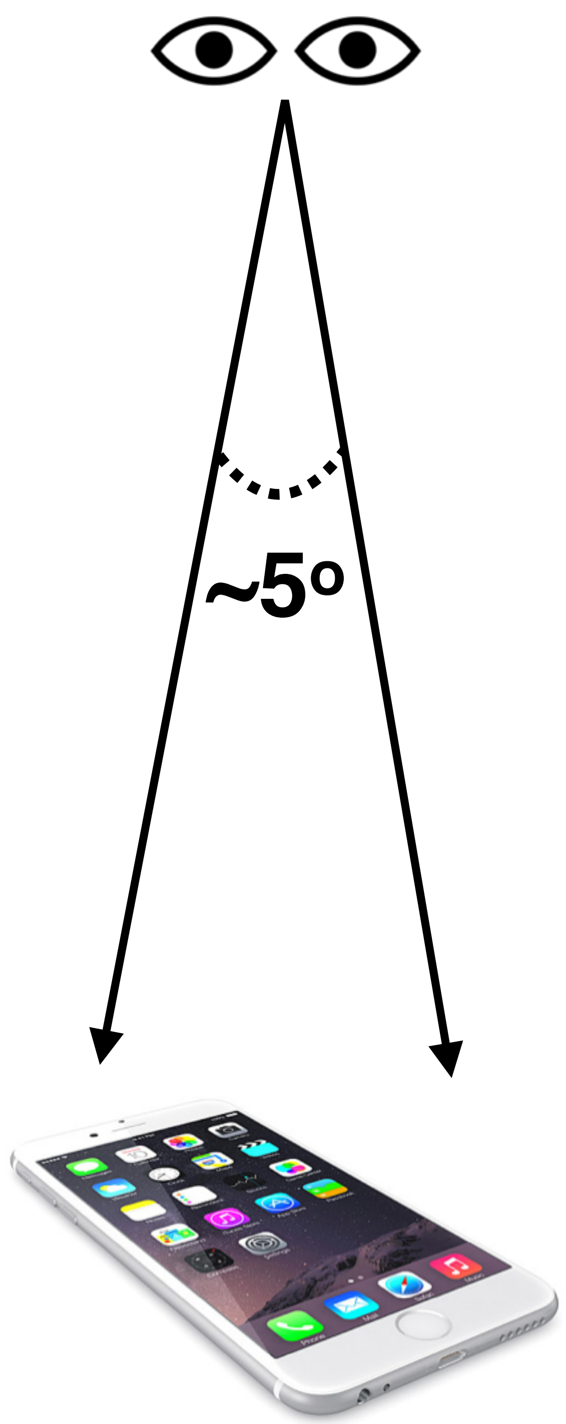
- VR Imaging



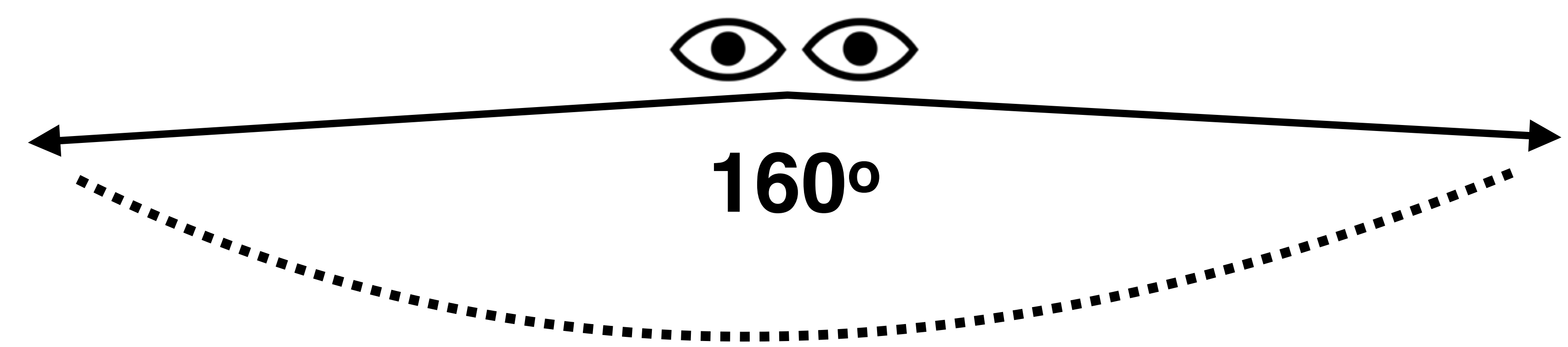


Rendering Latency in VR

Resolution Requirements in VR Are Very High



iPhone 6: 4.7 in "retina" display:
1.3 MPixel
326 ppi → 57 ppd



Human: ~160° view of field per eye (~200° overall)
(Note: does not account for eye's ability to rotate in socket)

Future "retina" VR display:
57 ppd covering 200°
= 11K x 11K display per eye
= 220 MPixel

Eyes designed by SuperAtic LABS from the thenounproject.com

Latency Requirements in VR Are Challenging

The goal of a VR graphics system is to achieve "presence", tricking the brain into thinking what it is seeing is real

Achieving presence requires an exceptionally low-latency system

- What you see must change when you move your head!
- End-to-end latency: time from moving your head to the time new photons hit your eyes
  - Measure user's head movement
  - Update scene/camera position
  - Render new image
  - Transfer image to headset, then transfer to display in headset
  - Actually emit light from display (photons hit user's eyes)
- Latency goal of VR: 10-25 ms
  - Requires exceptionally low-latency head tracking
  - Requires exceptionally low-latency rendering and display

Thought Experiment: Effect of Latency

Consider 1,000 x 1,000 display spanning 100° field of view

- 10 pixels per degree

Assume:

- You move your head 90° in 1 second (only modest speed)
- End-to-end latency of system is a slow 50 ms (1/20 sec)

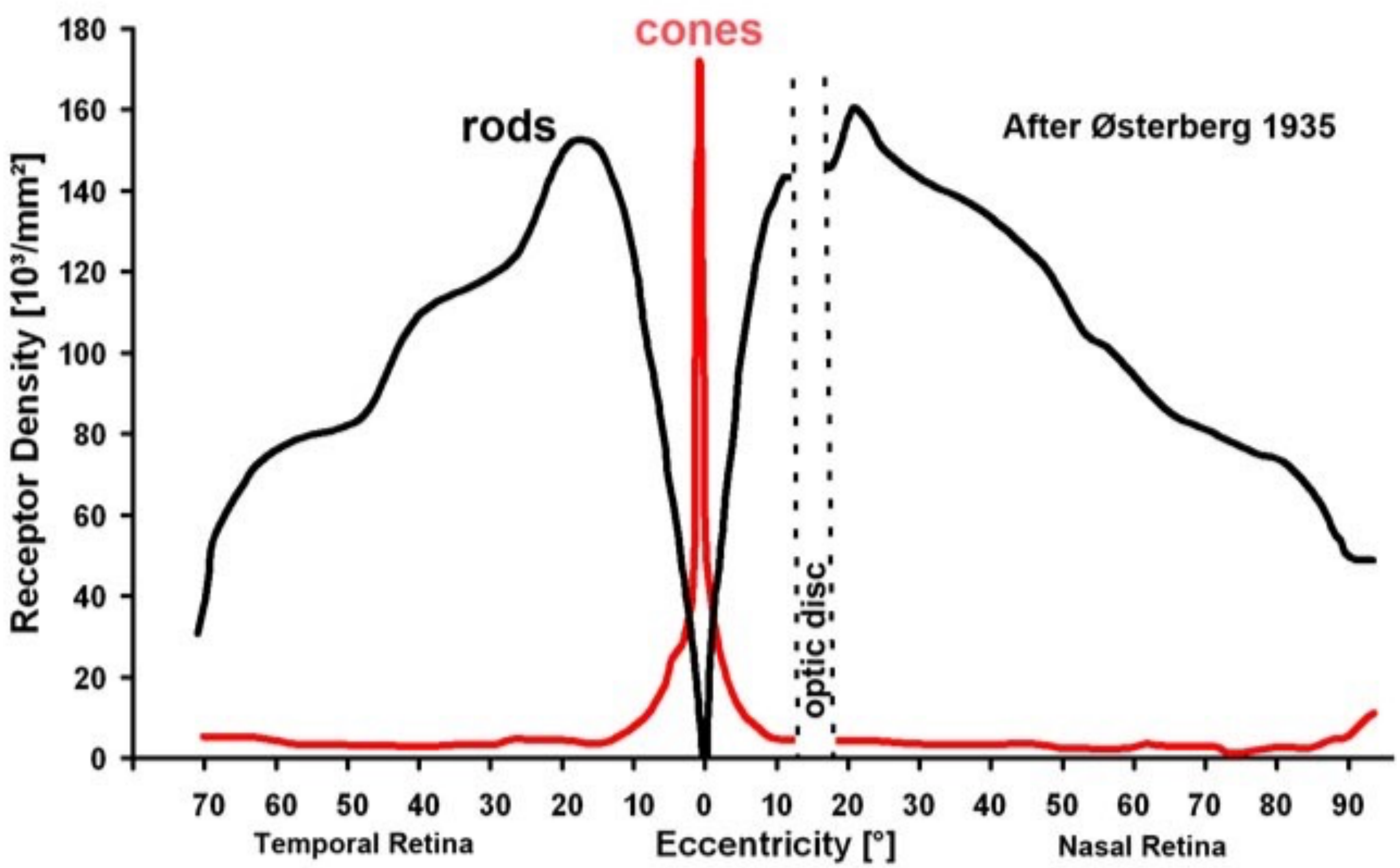
Result:

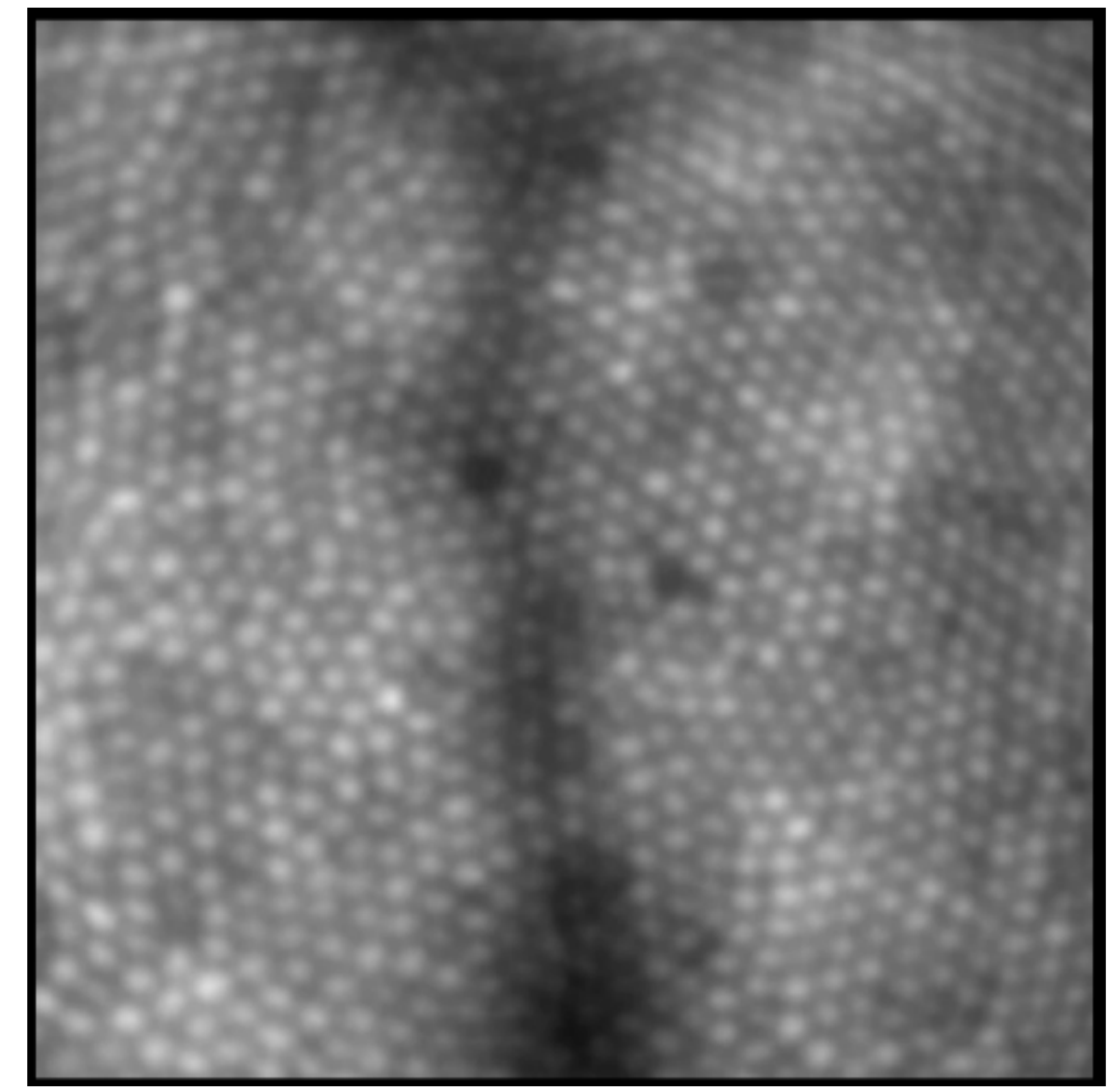
- Displayed pixels are off by 4.5° ~ 45 pixels from where they would be in an ideal system with 0 latency

Example credit: Michael Abrash

Challenge:
Low Latency and High Resolution
Require High Rendering Speed

Recall: Retinal Resolution Falls Away from Fovea



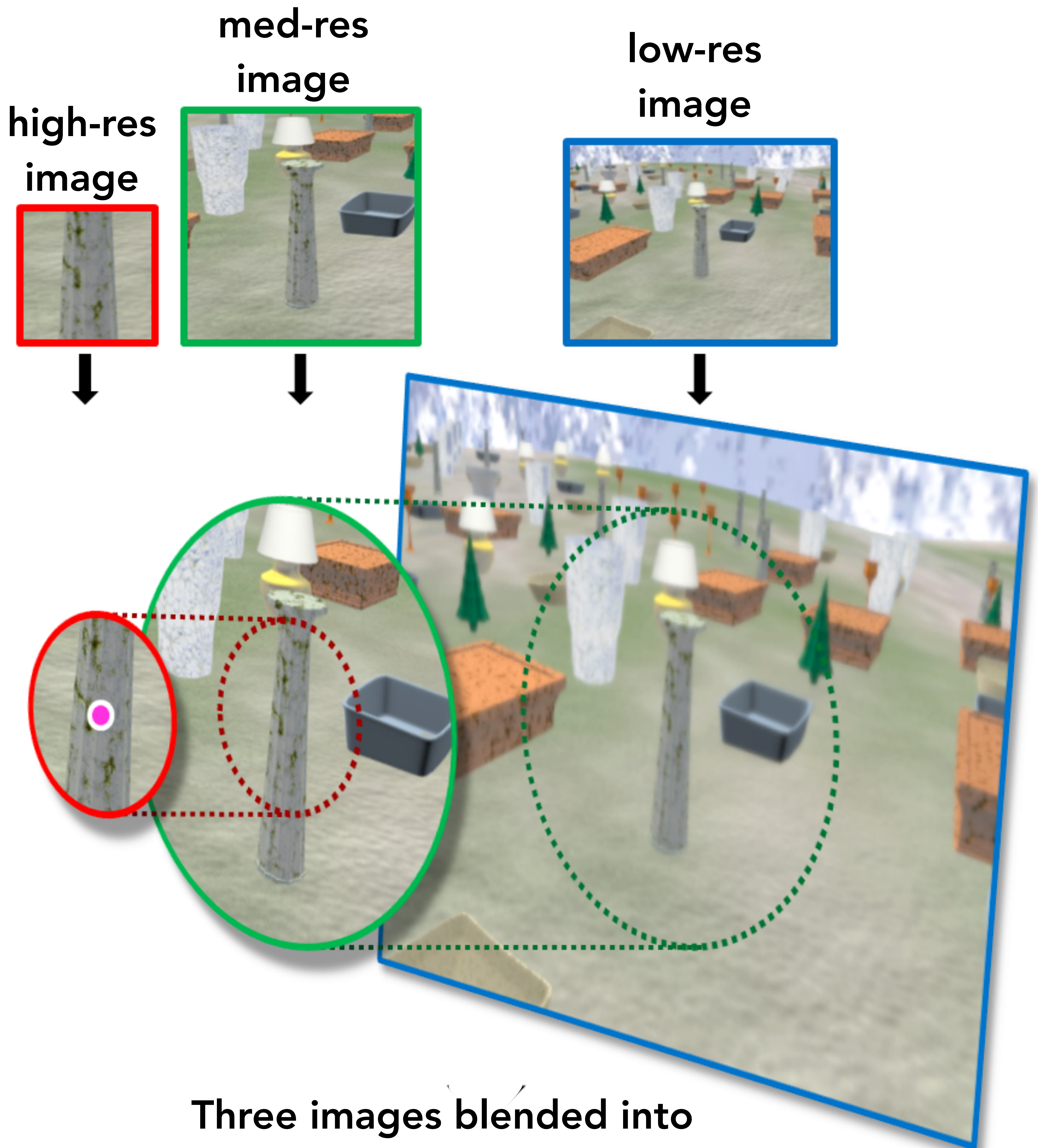


[Roorda 1999]

- Highest density of cones in fovea (and no rods there)
- "Blind spot" at the optic disc, where optic nerve exits eye

Foveated Rendering

Idea: track user's gaze, render with increasingly lower resolution farther away from gaze point

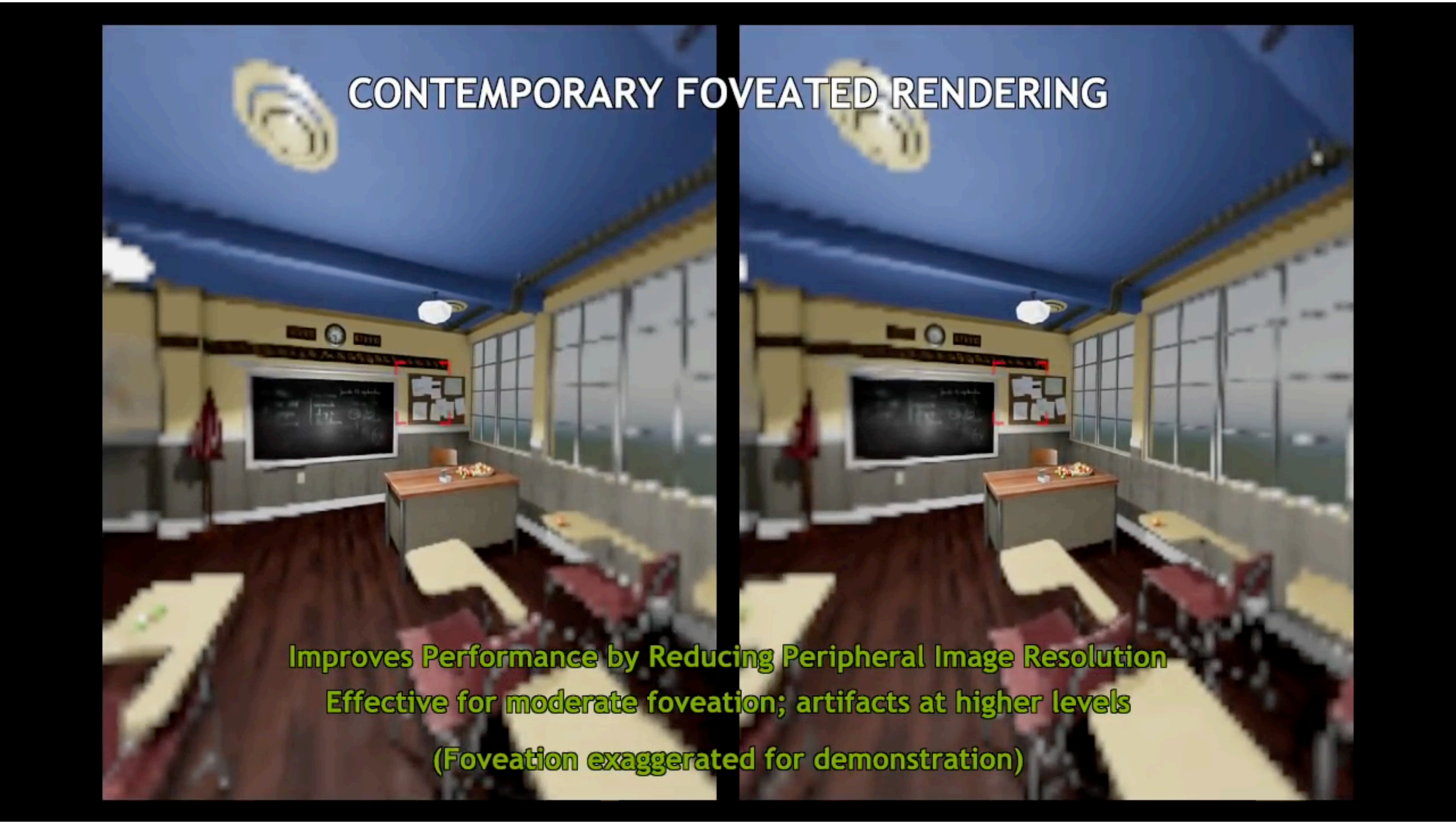


Three images blended into one for display

Foveated Rendering - Perceptual Effects

Patney et al., Towards Foveated Rendering for Gaze-Tracked Virtual Reality
SIGGRAPH Asisa 2016.

Foveated Rendering - Perceptual Effects

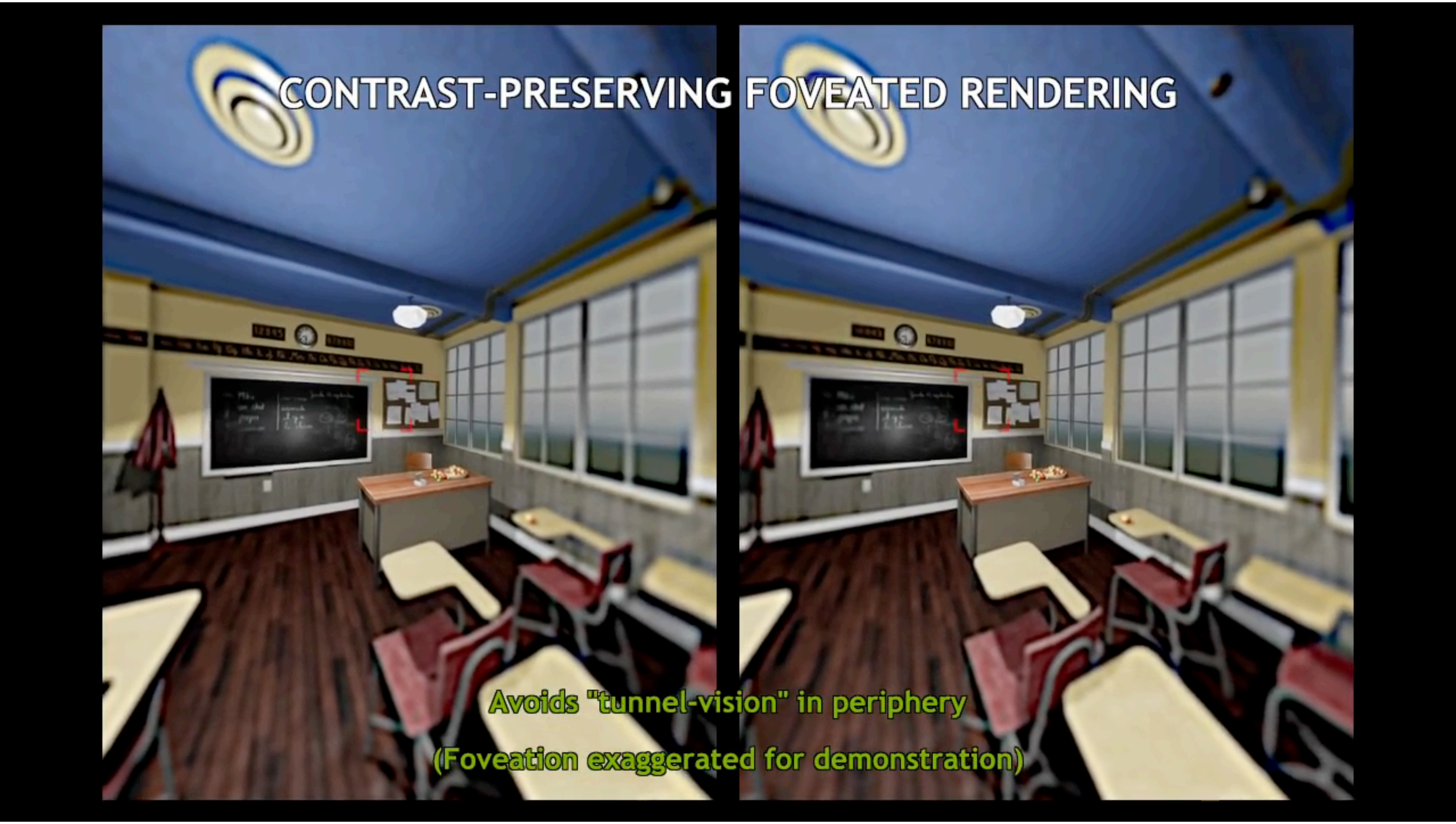


Patney et al., Towards Foveated Rendering for Gaze-Tracked Virtual Reality
SIGGRAPH Asisa 2016.

Foveated Rendering - Perceptual Effects

Patney et al., Towards Foveated Rendering for Gaze-Tracked Virtual Reality
SIGGRAPH Asisa 2016.

Foveated Rendering - Perceptual Effects



Patney et al., Towards Foveated Rendering for Gaze-Tracked Virtual Reality
SIGGRAPH Asisa 2016.

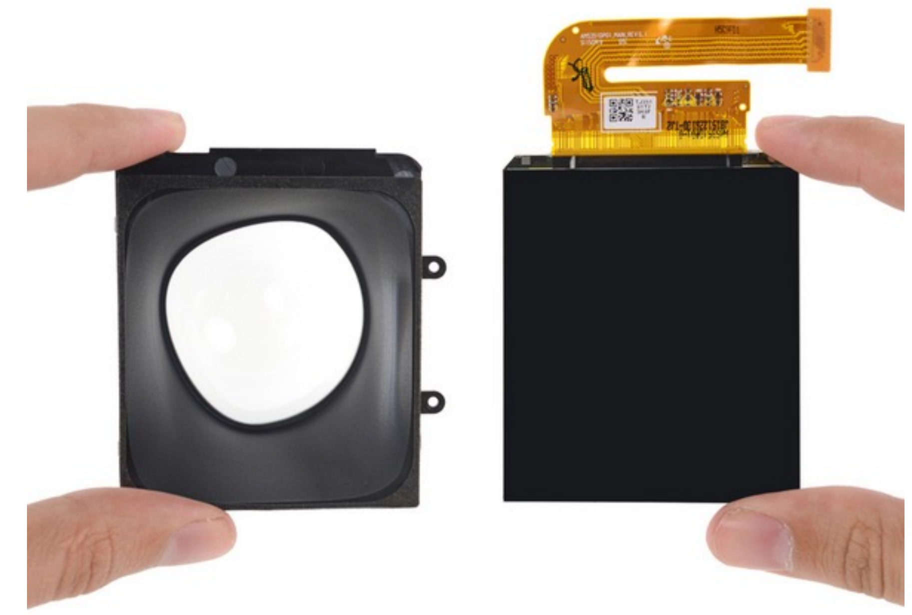
Foveated Rendering

Perceptual considerations:

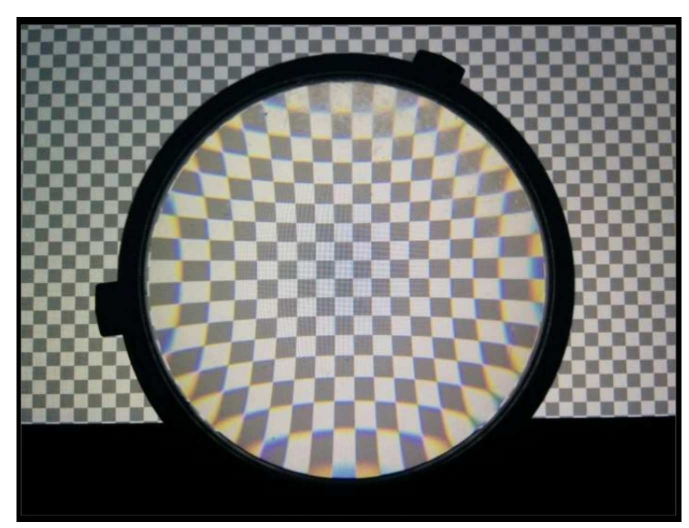
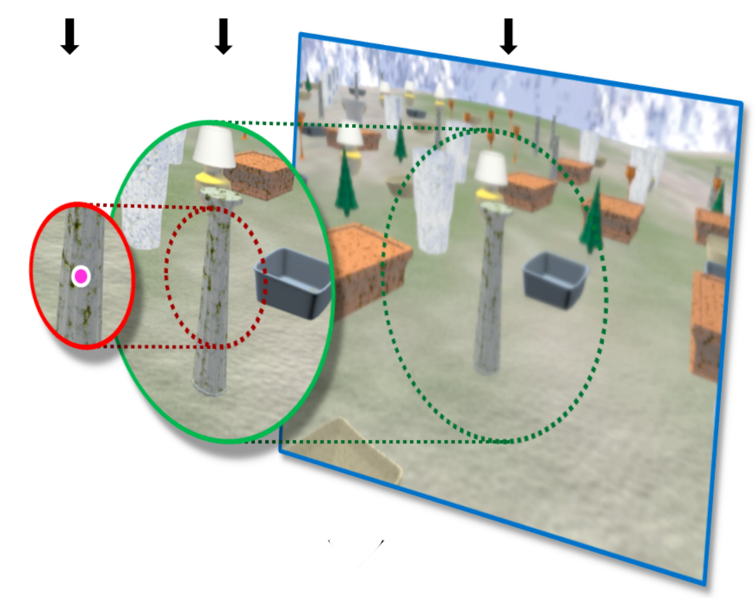
- If we render low resolution in periphery, have to be careful of aliasing / flickering
- If we render with a smooth image blur in the periphery, users experience a “tunnel vision” effect
- Research indicates that we should boost the contrast of low-frequency content in the periphery

Overview of VR Topics

- VR Displays

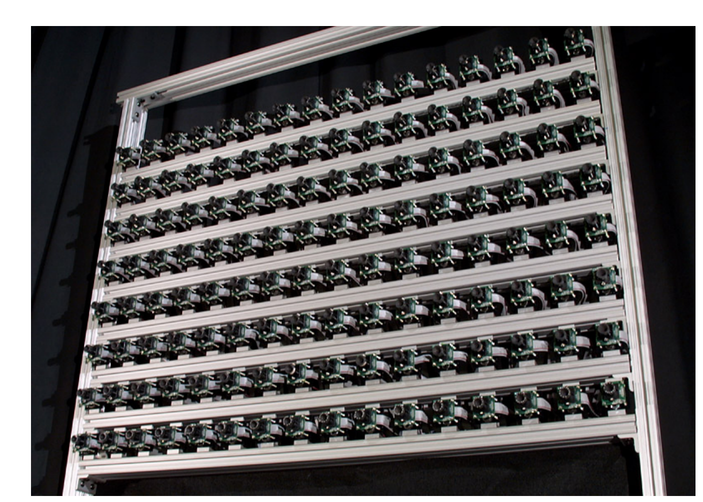




- VR Rendering

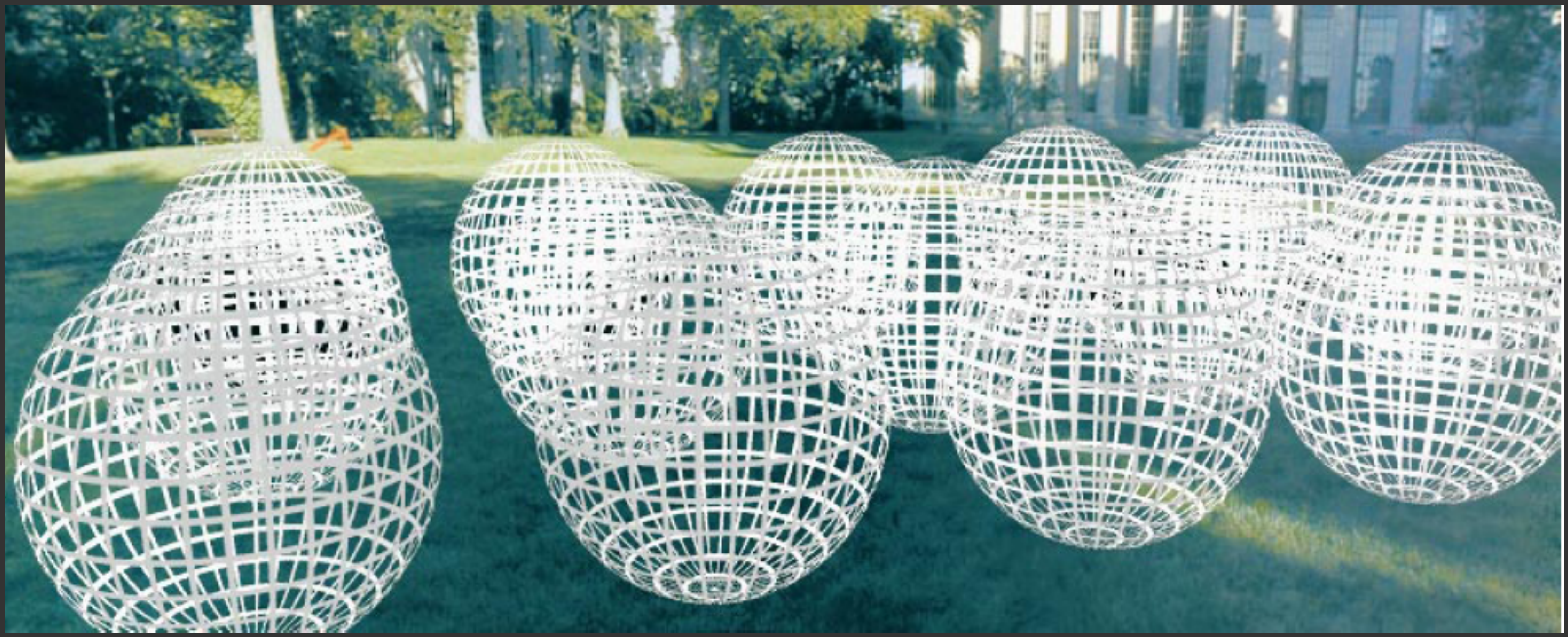
- VR Imaging





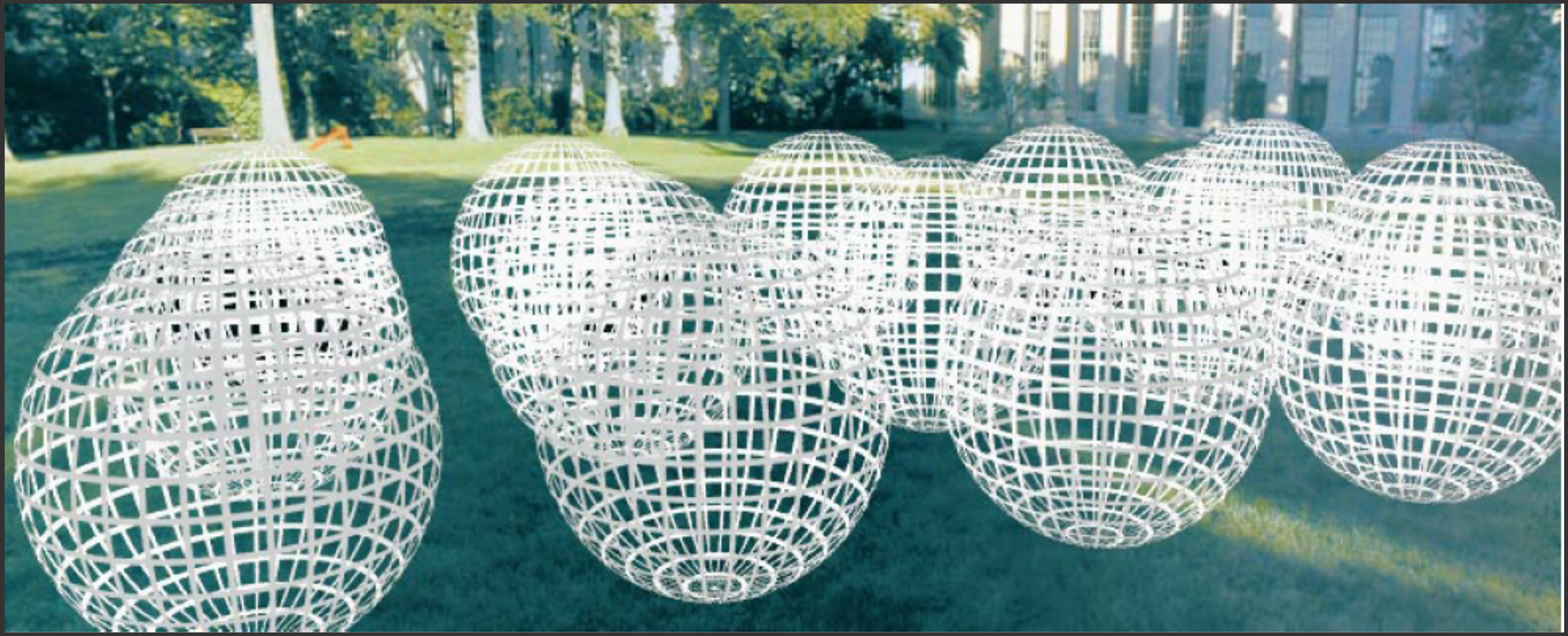
Moving-Viewpoint Imaging
(Full Plenoptic Function?)

The 5D Plenoptic Function



$$P(\theta, \phi, V_x, V_y, V_z)$$

4D Light Field



$$P(\theta, \phi, V_x, V_y) = P(u, v, s, t)$$

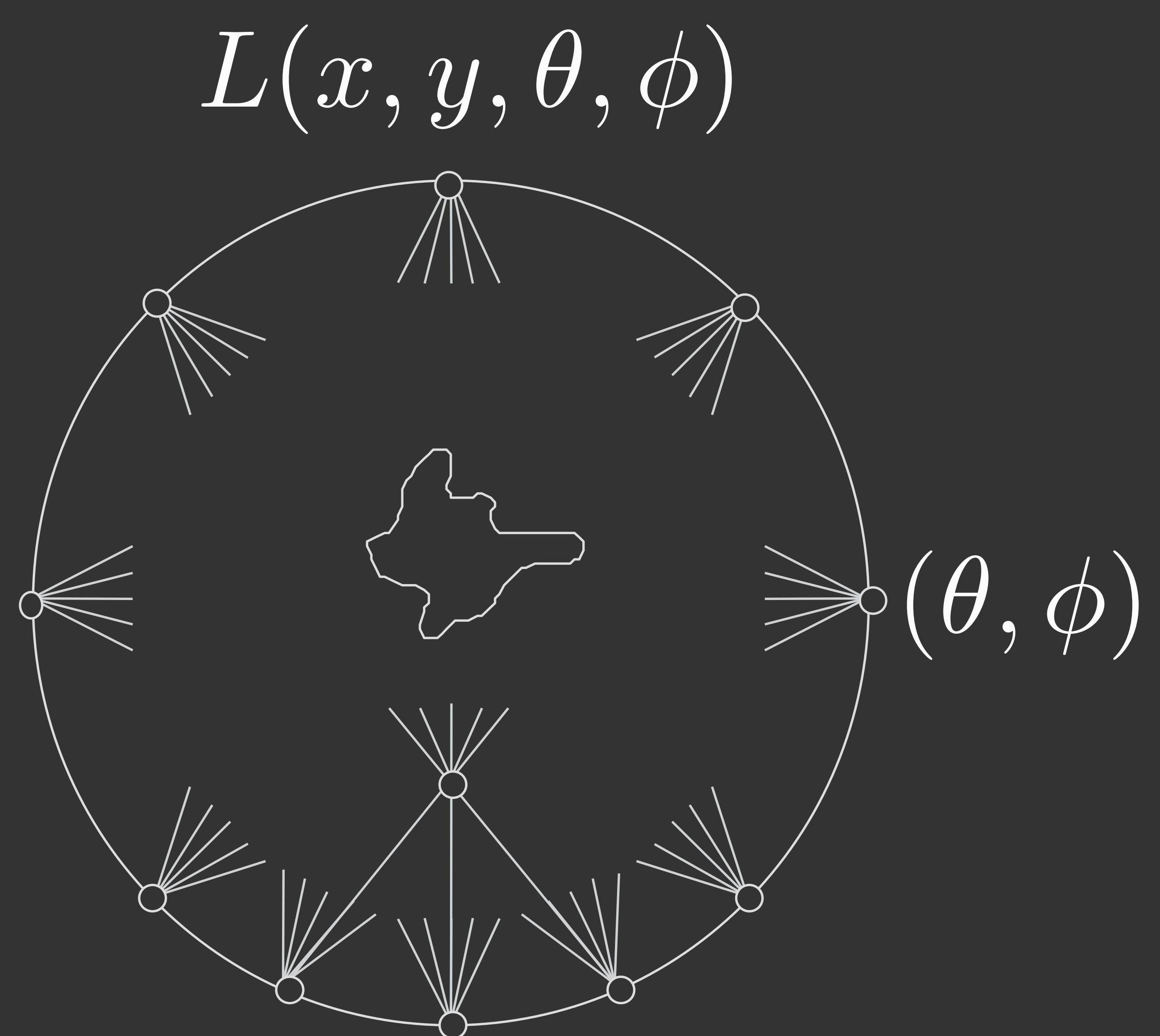
- In a region of free-space, 5D plenoptic function simplifies to 4D because light is constant along a ray

Light Field Capture Robot

Original light field rendering paper

Take photographs of an object from all points on an enclosing sphere

Captures all light leaving an object – like a hologram



[Levoy & Hanrahan 1996] [Gortler et al. 1996]

Slide credit: Pat Hanrahan

Multi-Camera Array ⇒ 4D Light Field

[Wilburn et al. SIGGRAPH 2005]

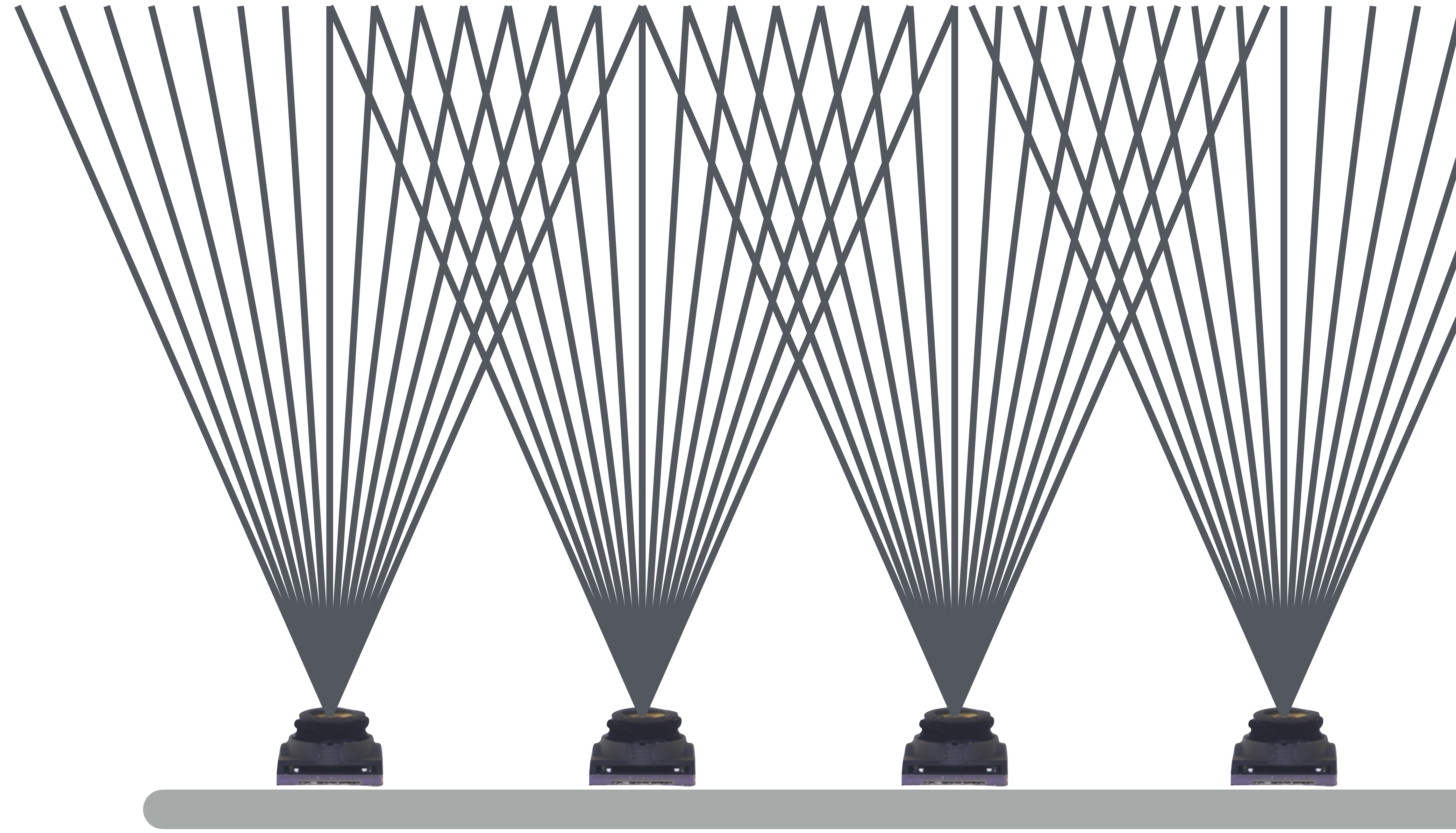
Slide credit: Pat Hanrahan

[Wilburn et al. SIGGRAPH 2005]

Handheld 4D Light Field Camera (Plenoptic Camera)

Lytro Gen-2 Light Field Camera

Handheld Light Field Camera vs Camera Array



Camera array: e.g. 10x10 views distributed across large planar support

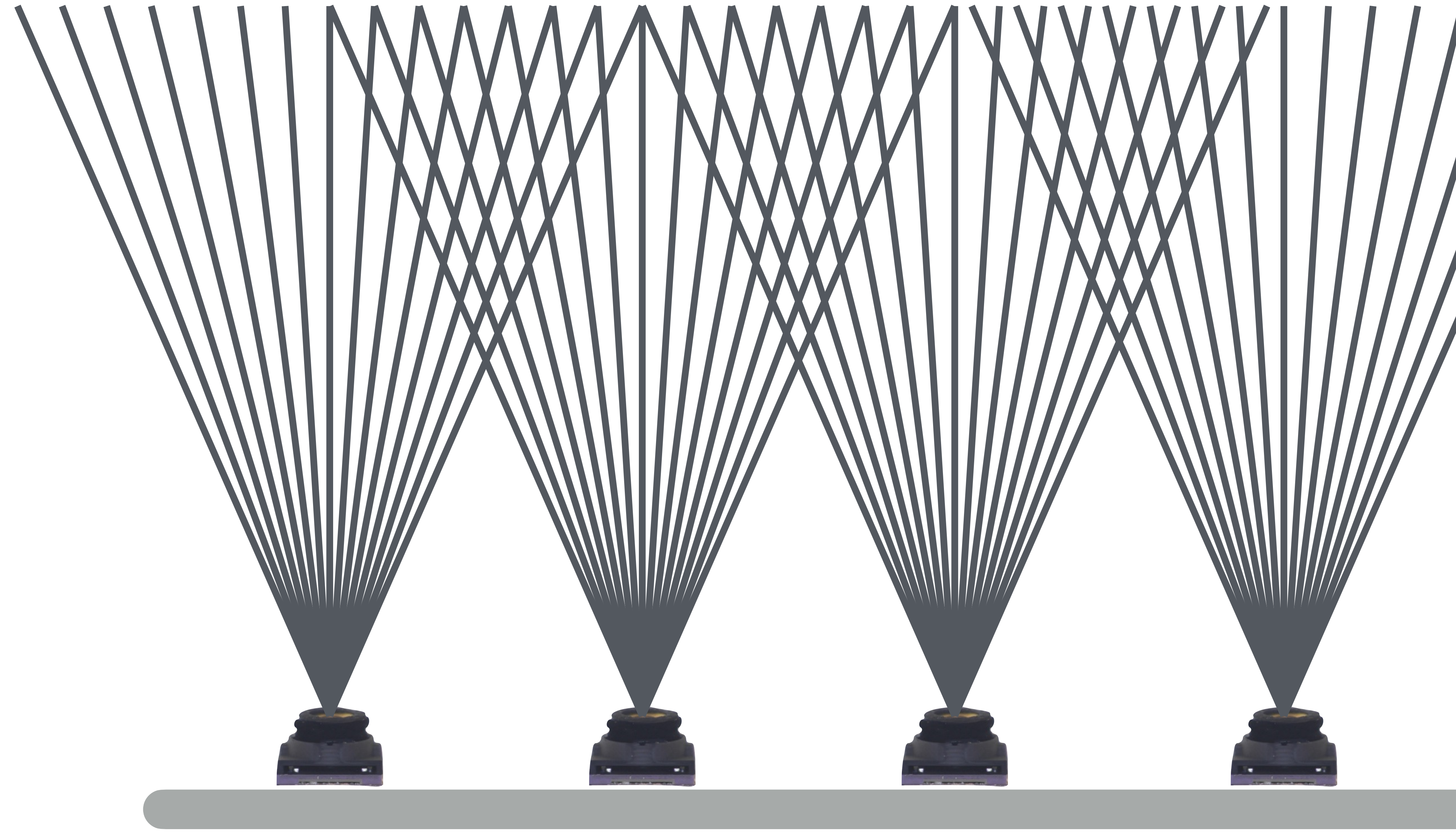
Plenoptic camera: e.g. 14x14 views distributed across small lens pupil
Note: antialiased across views, unlike camera array

The Intimacy of VR Graphics

Google's Tilt Brush on HTC Vive


A Challenge: Intimate Proximity in VR Imaging

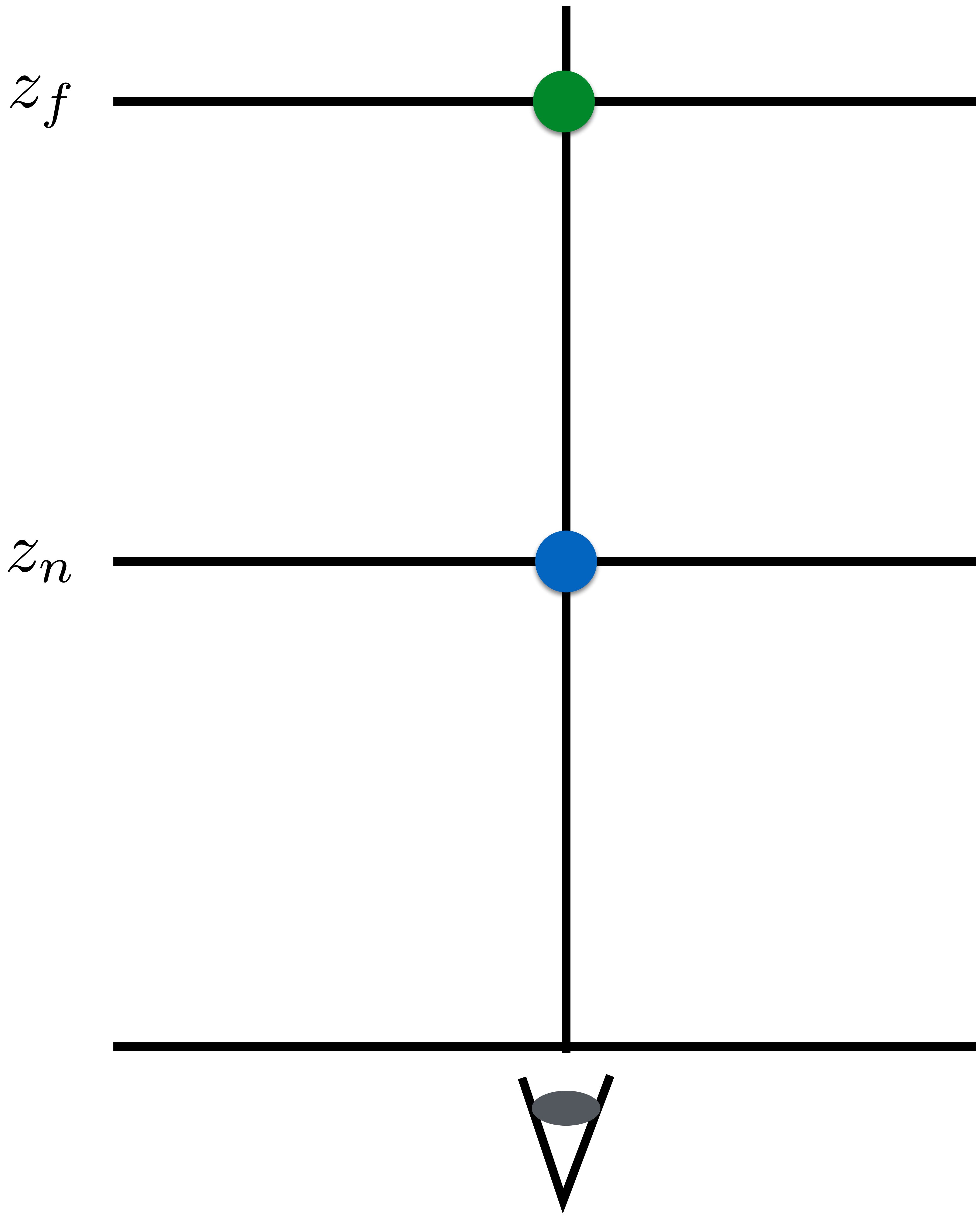
How Dense Are Camera Views Today?



Multi-camera arrays:	50 - 100 views
Plenoptic cameras:	100 - 200 views


How Dense Must Cameras Views Be?

How Dense Must Camera Views Be?

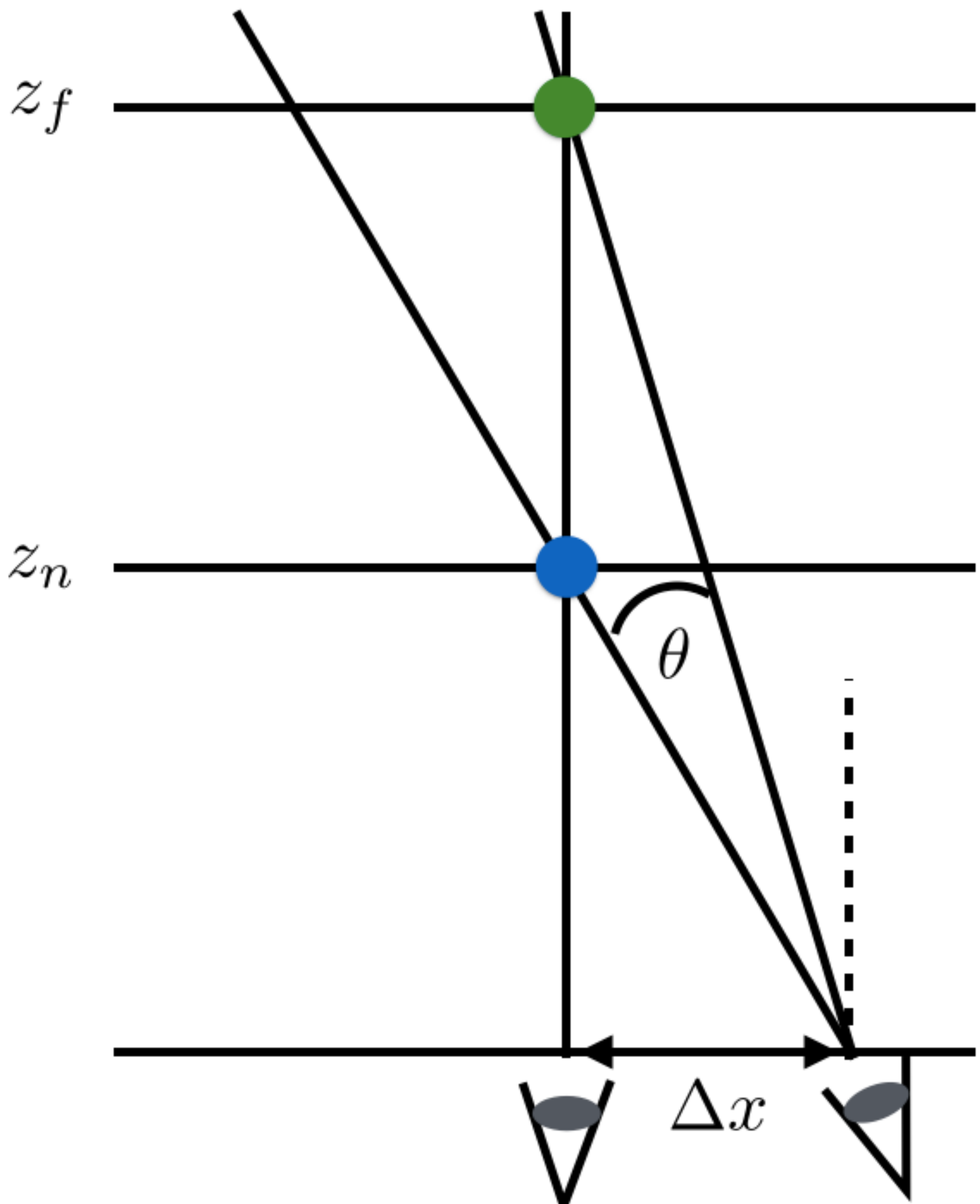


Child in lap, front to back of head

$$z_n = 0.3\text{m}$$

$$z_f = 0.6\text{m}$$

How Dense Must Camera Views Be?



20/20 vision: $\theta \approx (1/60)^\circ$

Curent HMDs: $\theta \approx (1/10)^\circ$

How Dense Must Camera Views Be?

Solving for minimum lateral motion:

$$\Delta x = \frac{(z_f - z_n) - \sqrt{(z_f - z_n)^2 - 4\tan^2\theta z_n z_f}}{2\tan\theta}$$

Child in lap, front to back of head

$z_n = 0.3\text{m}$

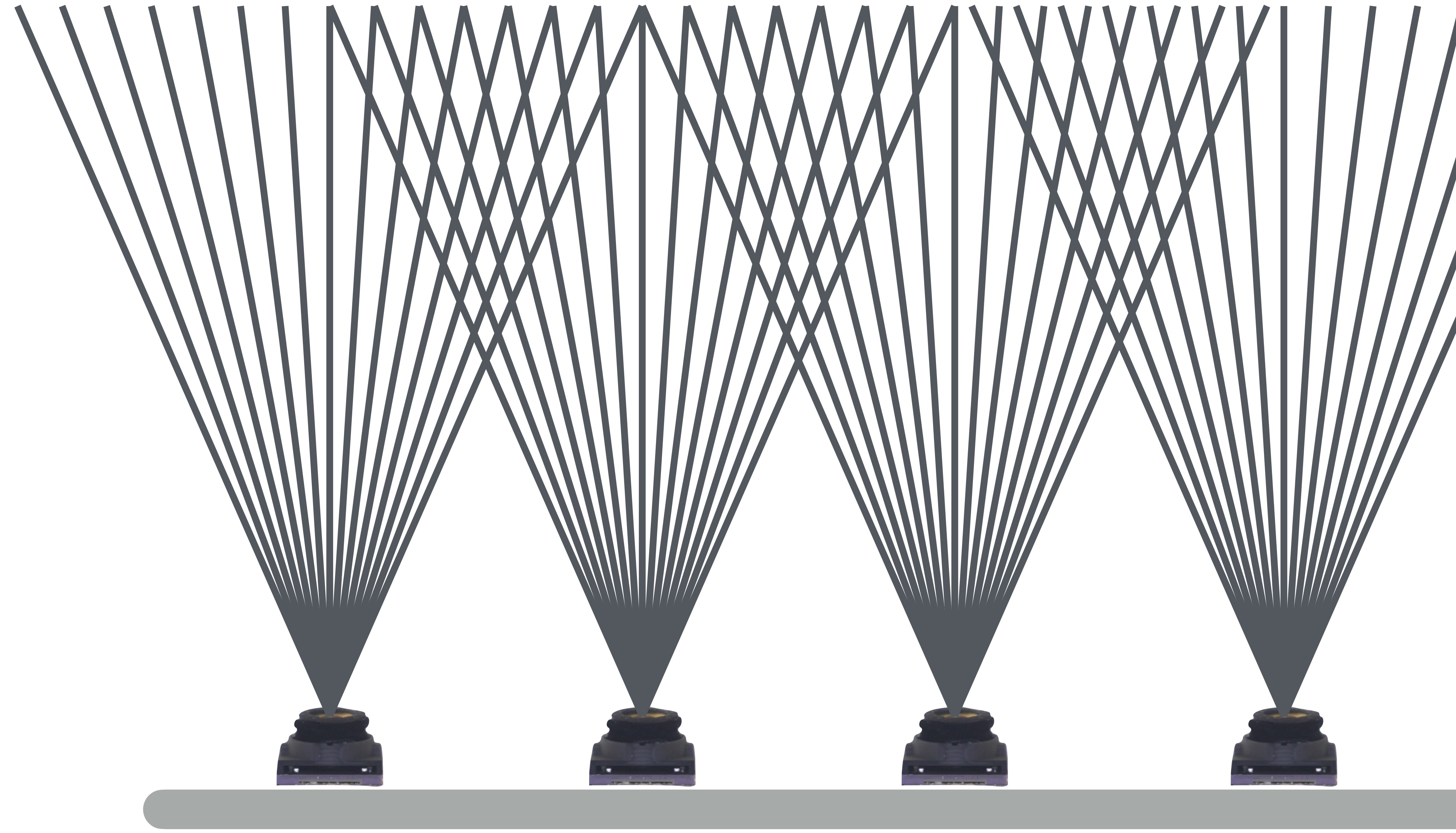
$z_f = 0.6\text{m}$

20/20 vision: $\theta \approx (1/60)^\circ \implies \Delta x \approx (1/1719)\text{ft}$

Curent HMDs: $\theta \approx (1/10)^\circ \implies \Delta x \approx (1/286)/\text{ft}$

20/20 vision:	millions of views per square foot
Current HMDs:	a hundred thousand views per square foot

How Dense Are Camera Views Today?



Multi-camera arrays: 50 - 100 views
Plenoptic cameras: 100 - 200 views

One of our group's research areas

MILDENHALL & SRINIVASAN ET AL 2019, *LOCAL LIGHT FIELD FUSION*

Input Images

MILDENHALL & SRINIVASAN ET AL 2019, *LOCAL LIGHT FIELD FUSION*

Output Novel Views

MILDENHALL & SRINIVASAN ET AL 2019, *LOCAL LIGHT FIELD FUSION*

Google VR Camera Rig

Paul Debevec, Google

Google VR Camera Rig

Paul Debevec, Google

Active Area of Research

Applying machine learning to intelligently up-sample from tens of camera views to the very high sampling rates required for Nyquist-sampled VR rendering.

Things to Remember

VR presents many new graphics challenges!

Displays

- Head-pose tracking with high accuracy and low latency

Rendering

- Low-latency, high resolution & frame-rate, wide field of view, …

Imaging

- 360 spherical, stereo, light field

Acknowledgments

Original slides made by Ren Ng

Thanks to Kayvon Fatahalian, Alyosha Efros, Brian Wandell and Pratul Srinivasan for lecture resources and slides!