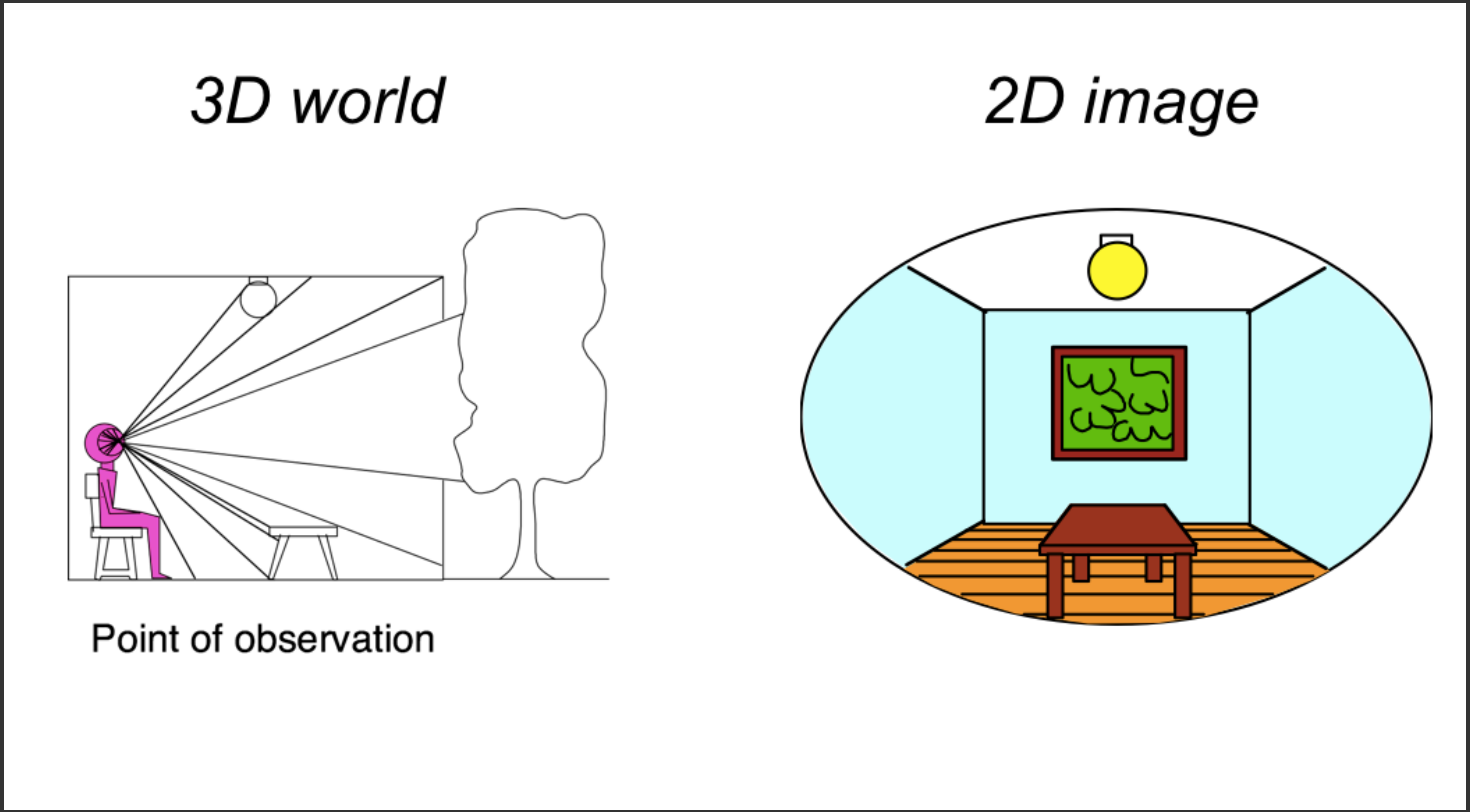
Lecture 26:

Intro to Virtual Reality

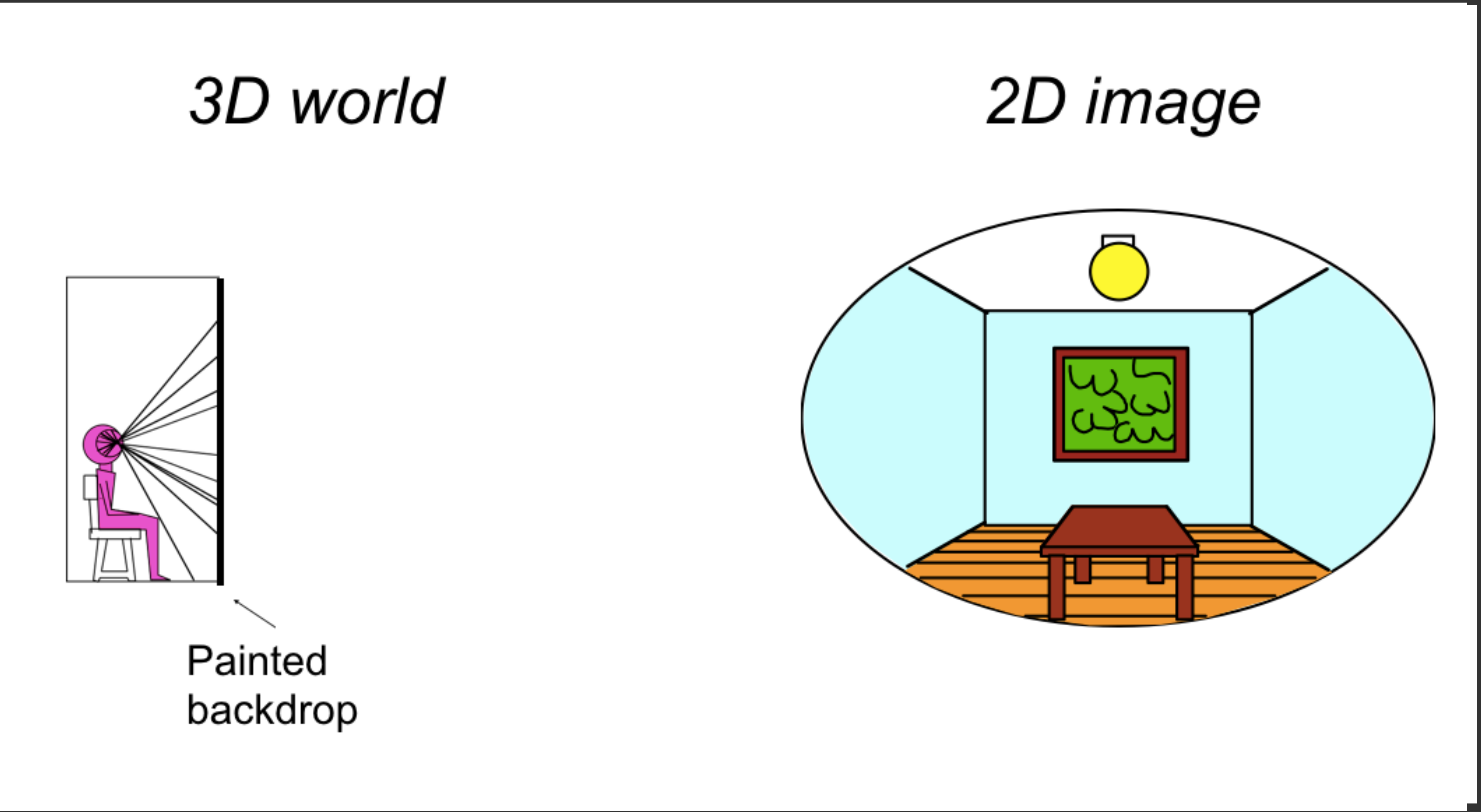
Computer Graphics and Imaging
UC Berkeley CS184/284A

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 everything that person can see…

Slide credit: Alyosha Efros

Grayscale Snapshot

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

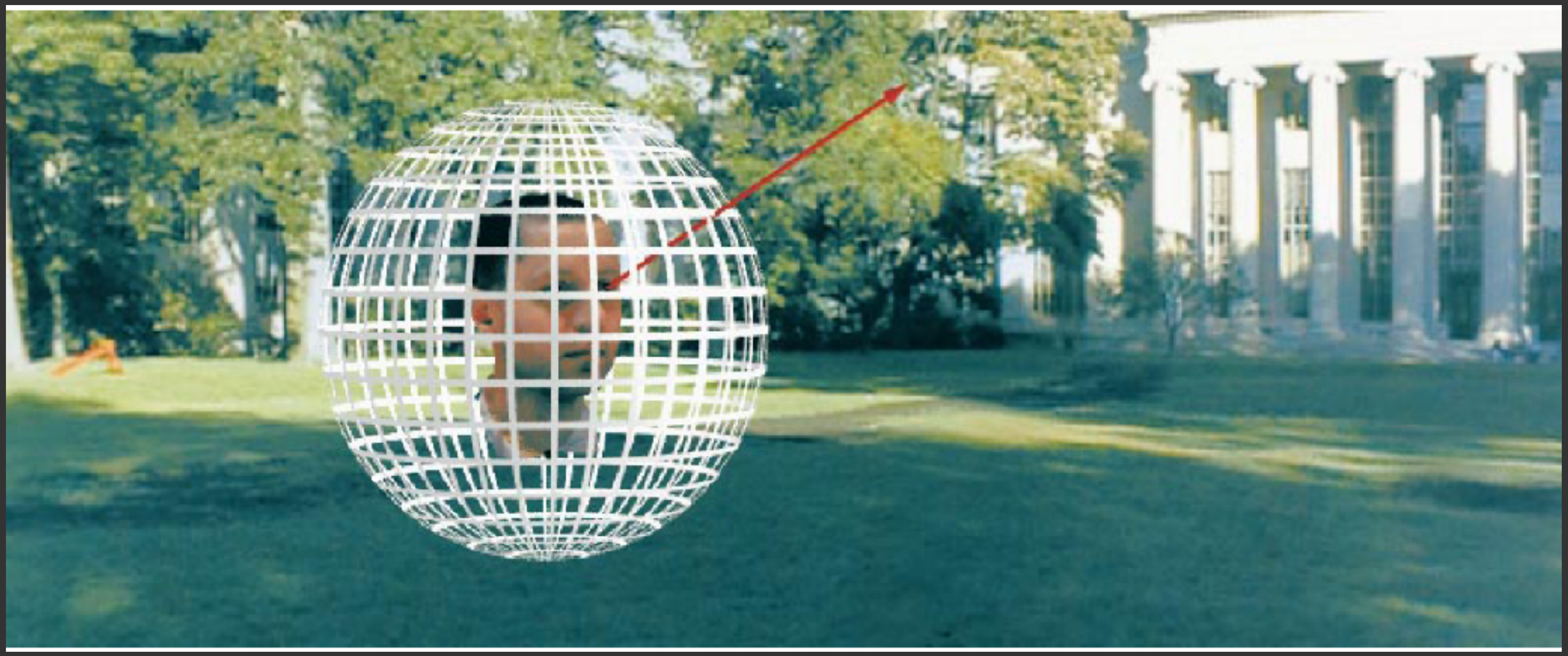
is intensity of light

- 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

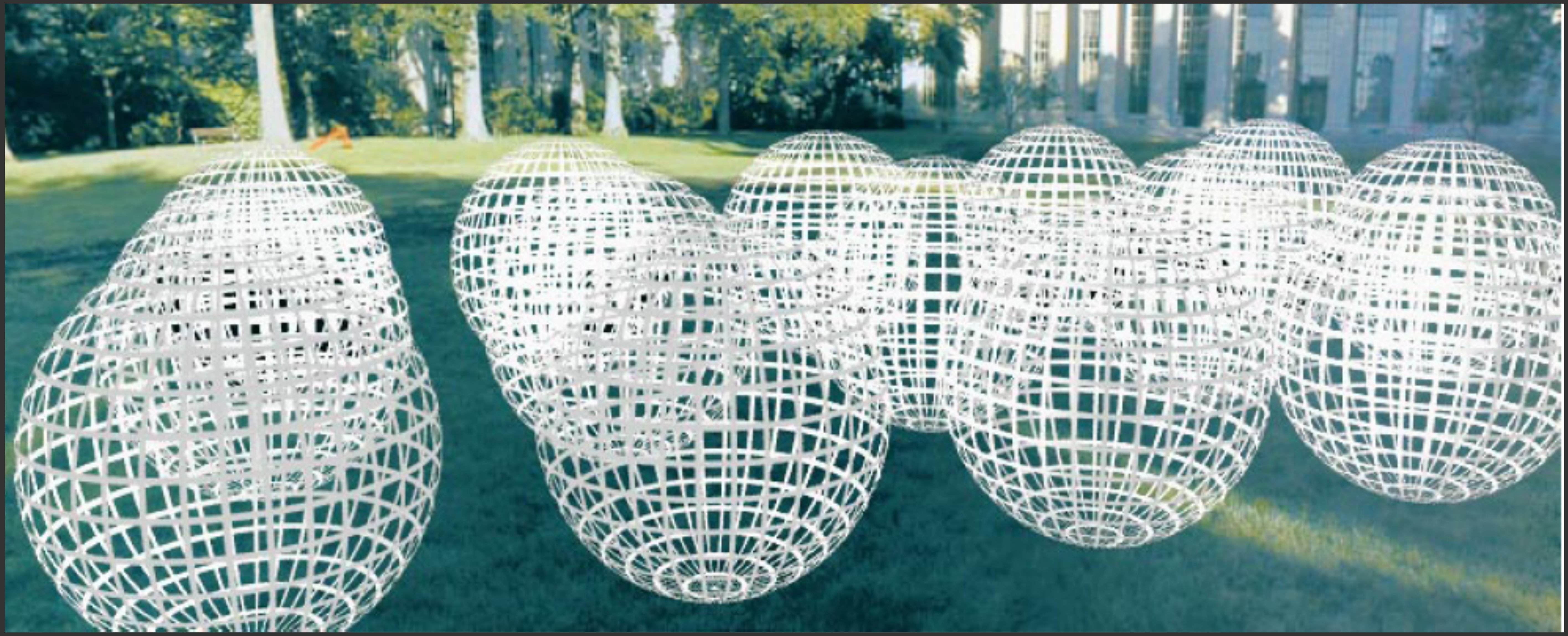
$$P(\theta, \phi, \lambda, t)$$

is intensity of light

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

Slide credit: Alyosha Efros

Free Viewpoint Movie



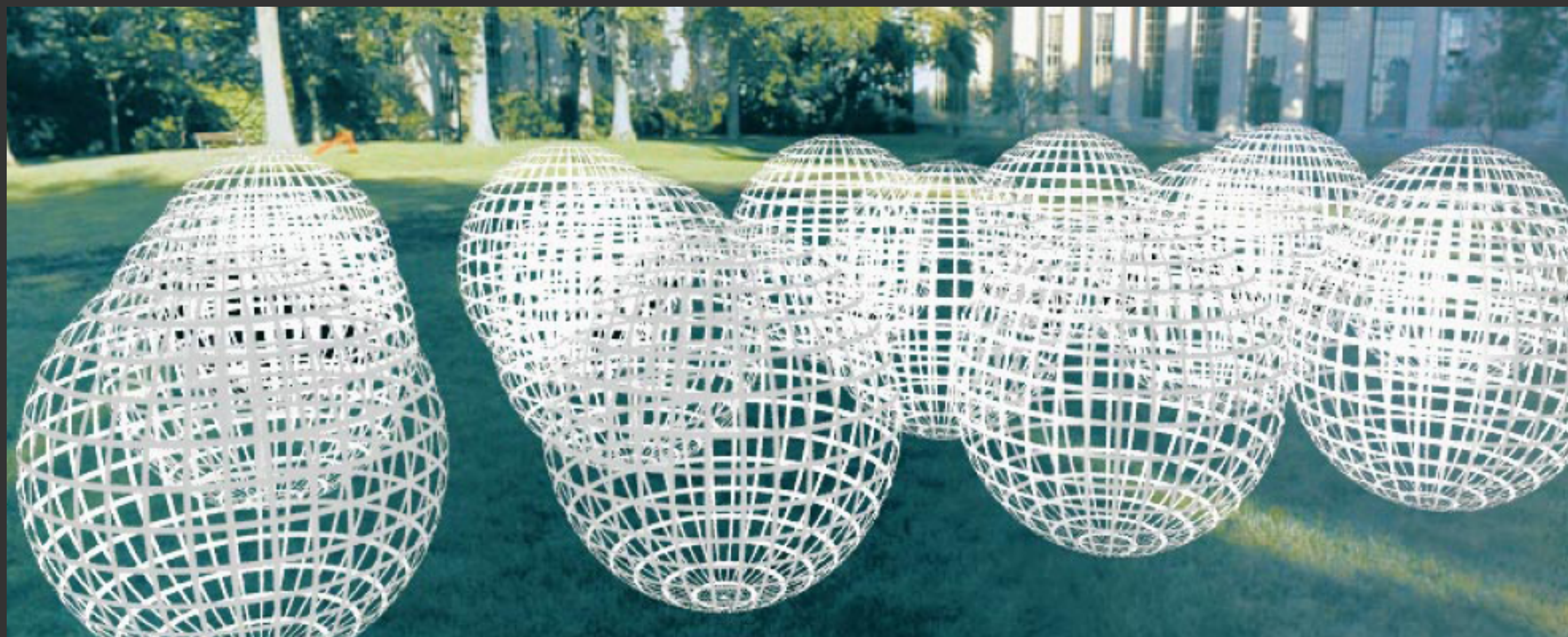
$$P(\theta, \phi, \lambda, t, V_x, V_y, V_z)$$

is intensity of light

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

Slide credit: Alyosha Efros

The Plenoptic Function

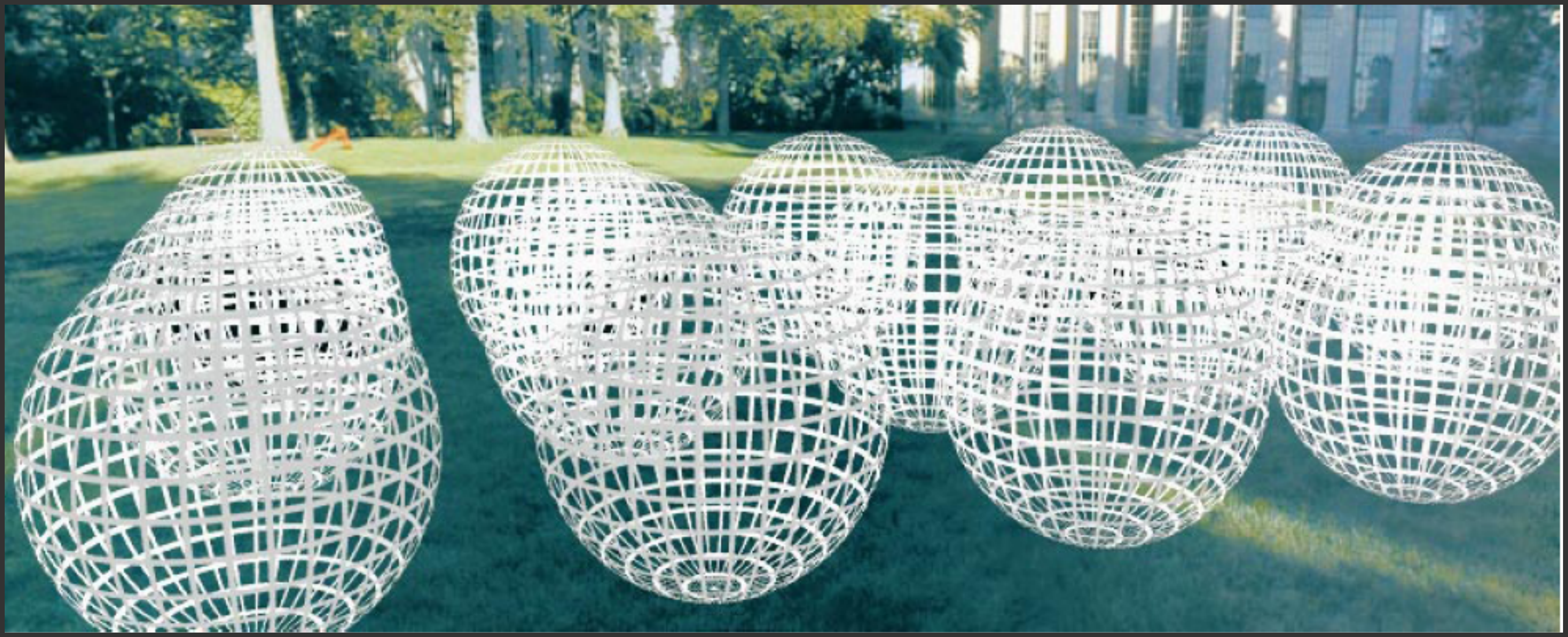


$$P(\theta, \phi, \lambda, t, V_x, V_y, V_z)$$

- 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

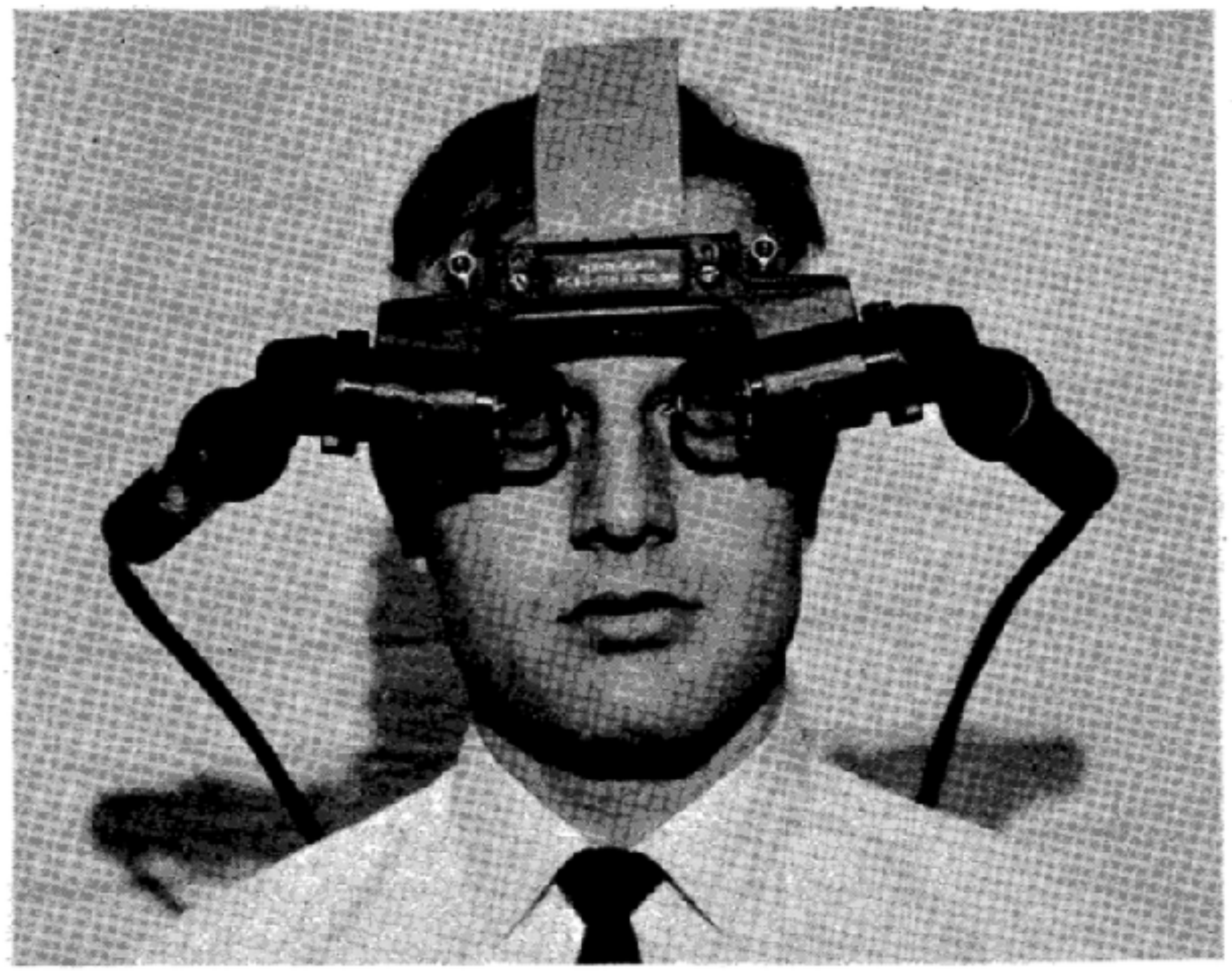


FIGURE 2—The head-mounted display optics with miniature CRT's

FIGURE 4—The ultrasonic head position sensor in use

VR Head-Mounted Displays (HMDs)

Oculus Rift

Sony Morpheus

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

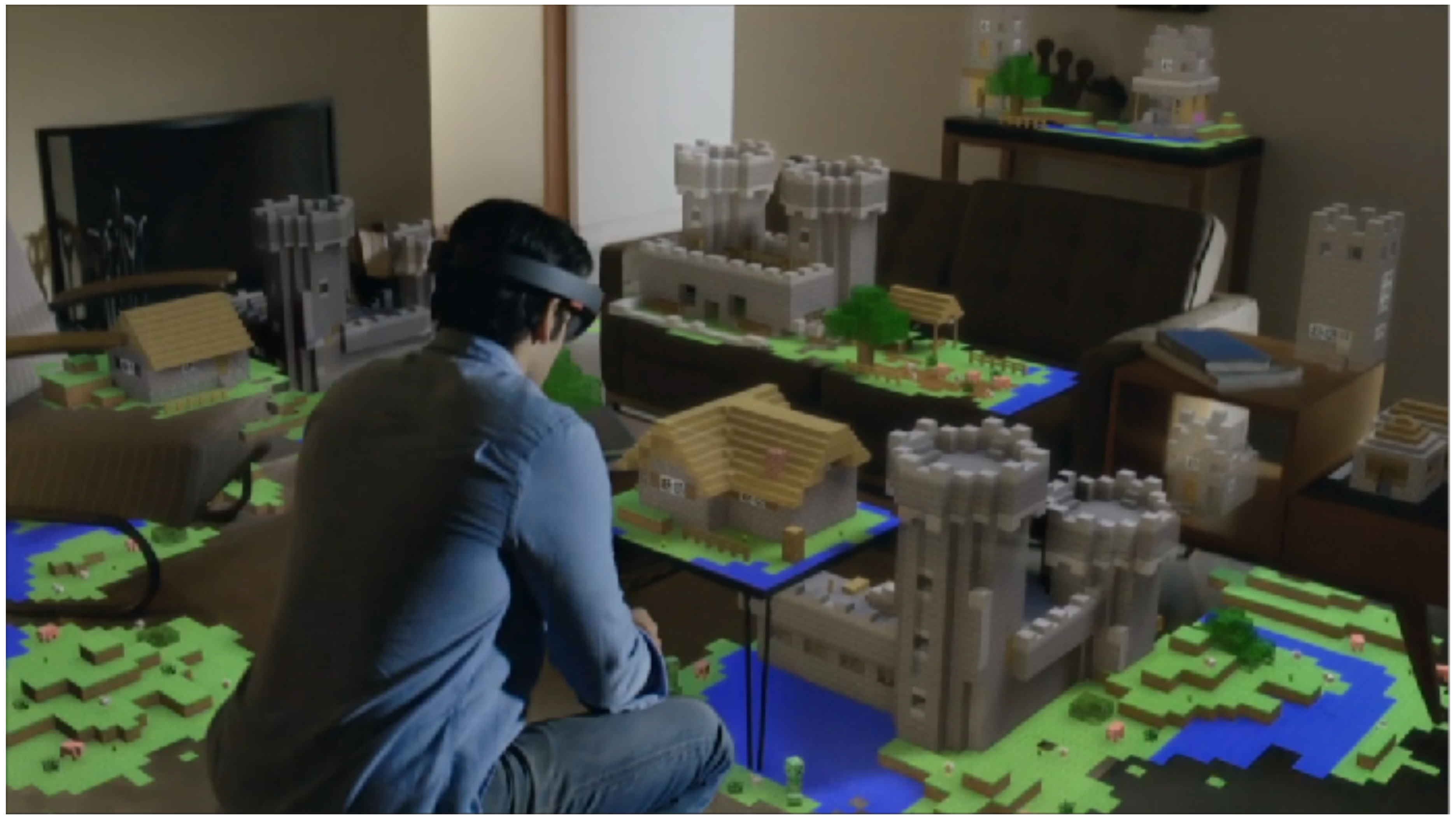
- Display is an overlay that augments user's normal view of the real world (e.g., Terminator)

Image credit: Terminator 2 (naturally)

AR Headsets

Microsoft Hololens

Magic Leap



VR Applications

VR Gaming

Bullet Train Demo (Epic)

VR Painting

VR Video

VR Video

VR Teleconference / Video Chat

http://vrchat.com/

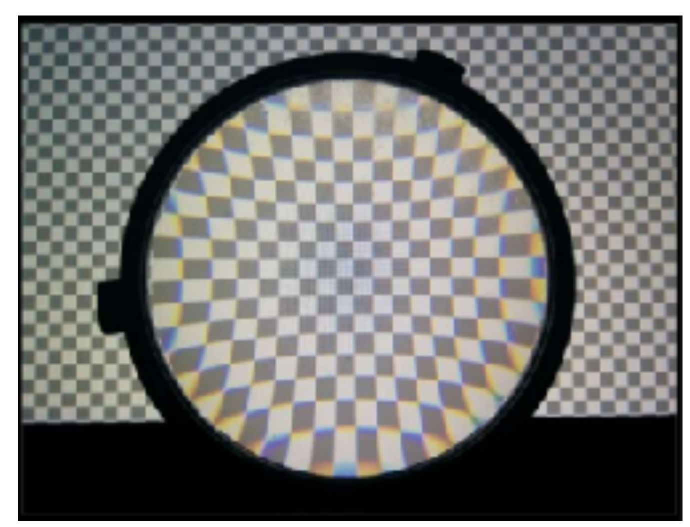
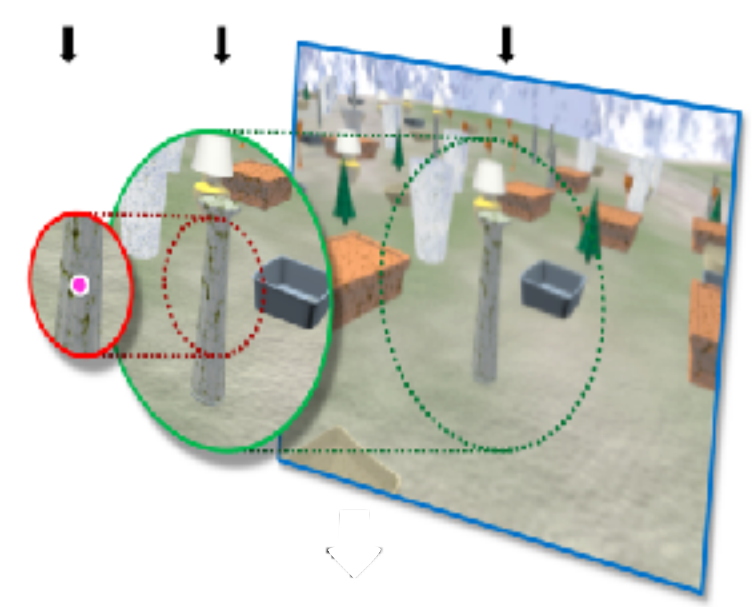
Overview of VR Topics

Areas we will discuss over next few lectures

- 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

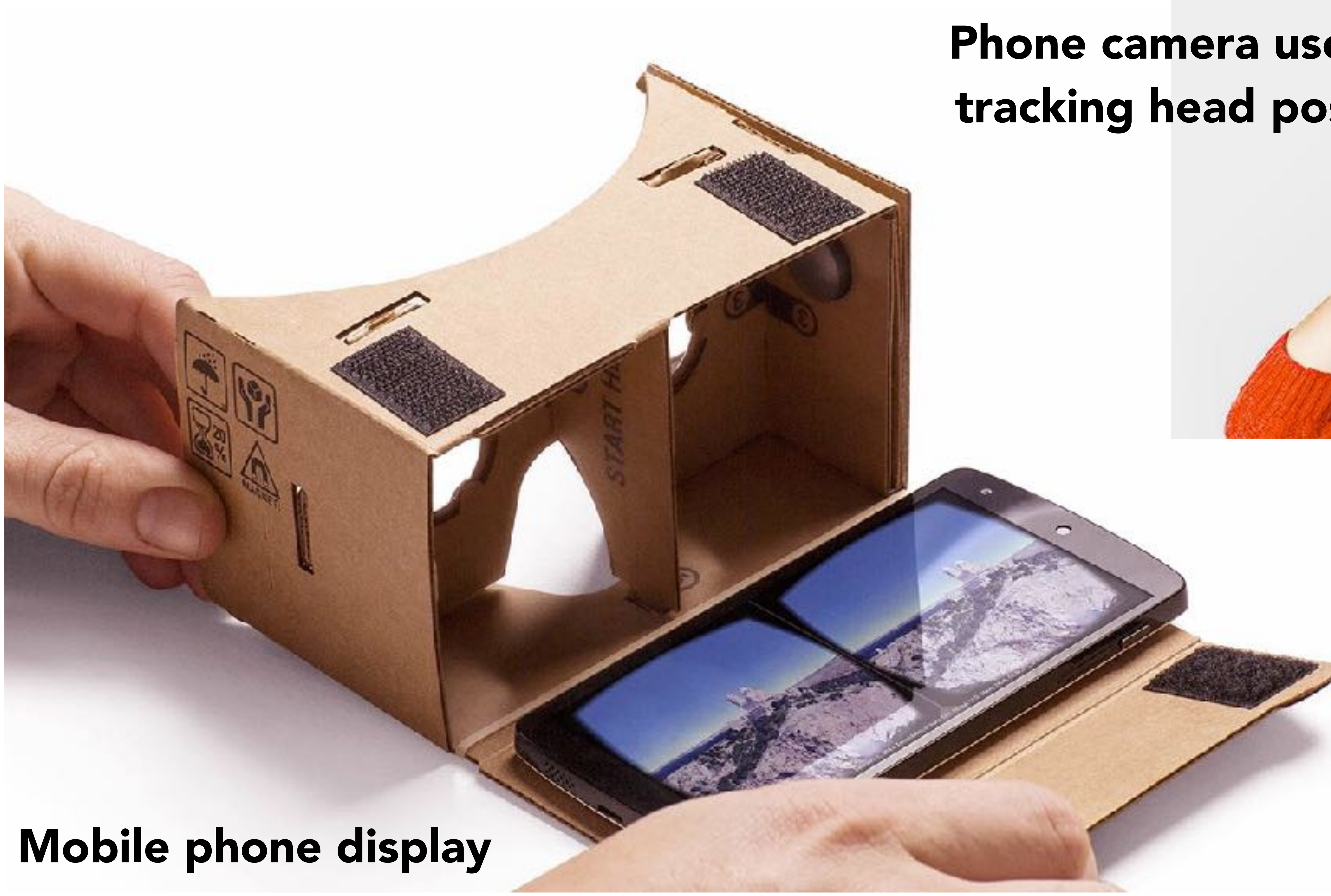


Image credits: slashgear.com, Google

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

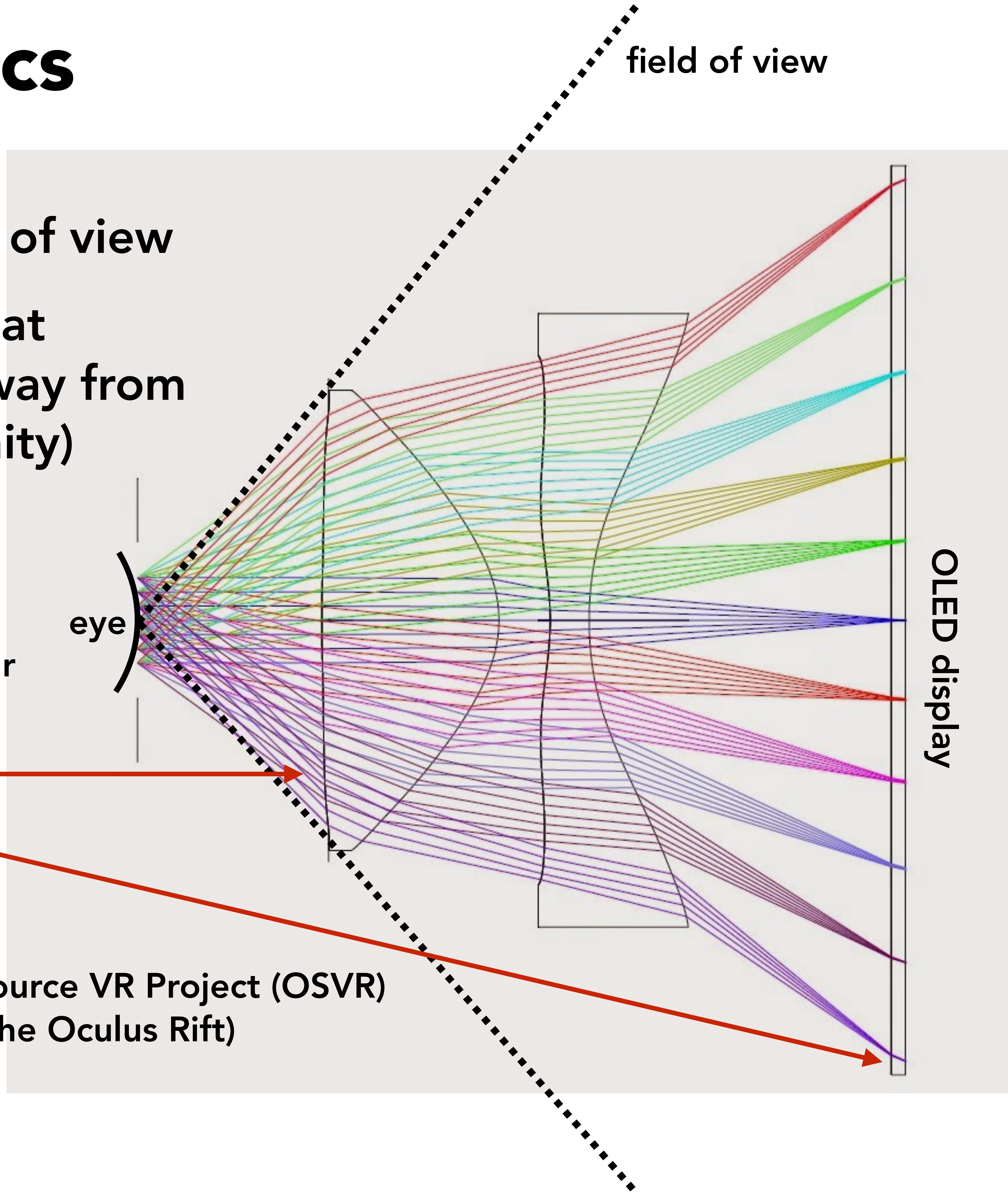
Fresnel eyepiece lens

Image credit: ifixit.com

Role of Optics

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

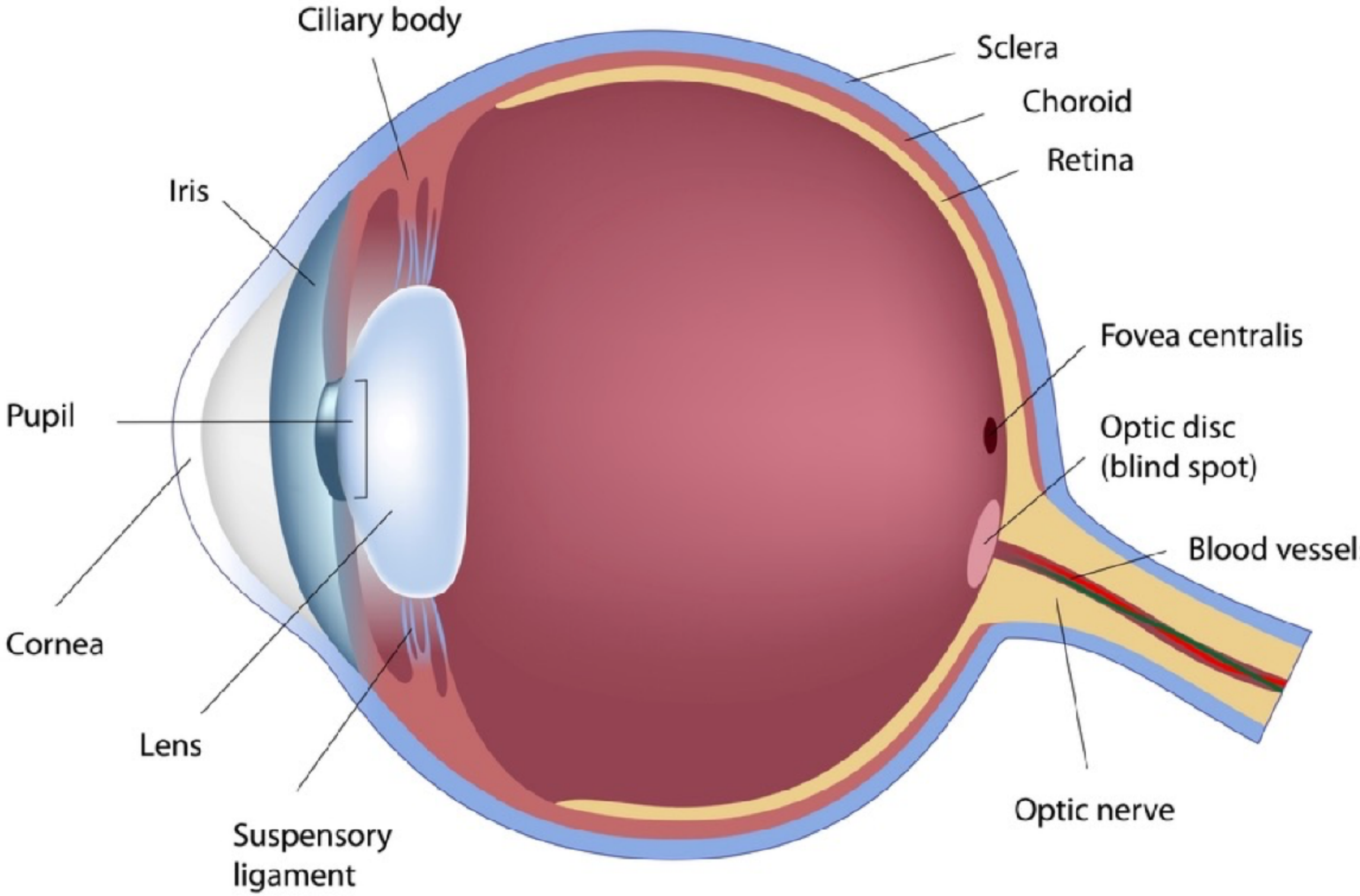
Note: parallel lines reaching eye converge to a single point on display (eye accommodates to plane near infinity)



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

Anatomy of The Human Eye



Display Requirements Derive From Human Perception

Example 1: Color

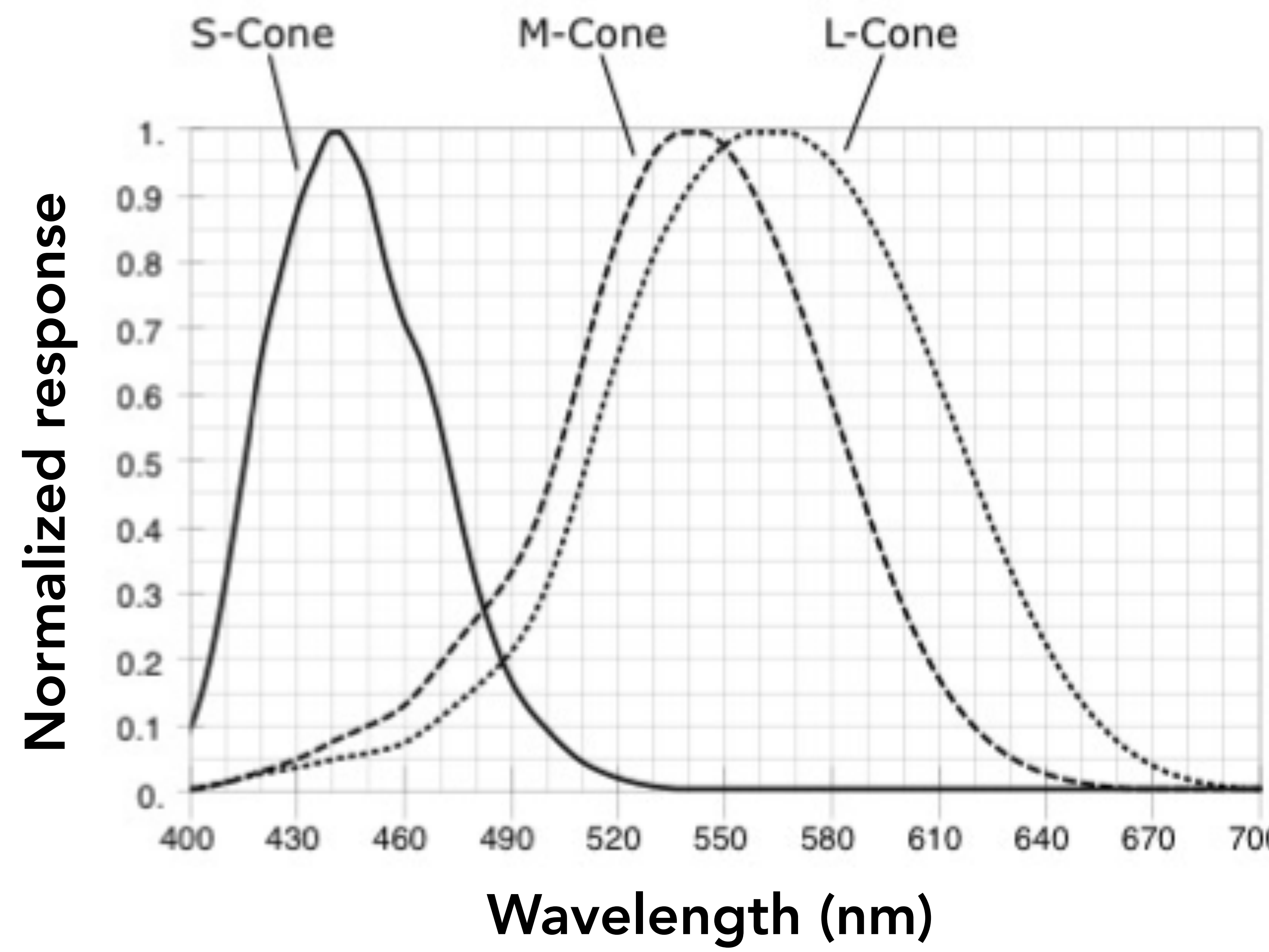
Recall: Spectral Response of Human Cone Cells

Three types of cones: S, M, and L cones (corresponding to peak response at short, medium, and long wavelengths)

$$S = \int r_S(\lambda) s(\lambda) \, d\lambda$$

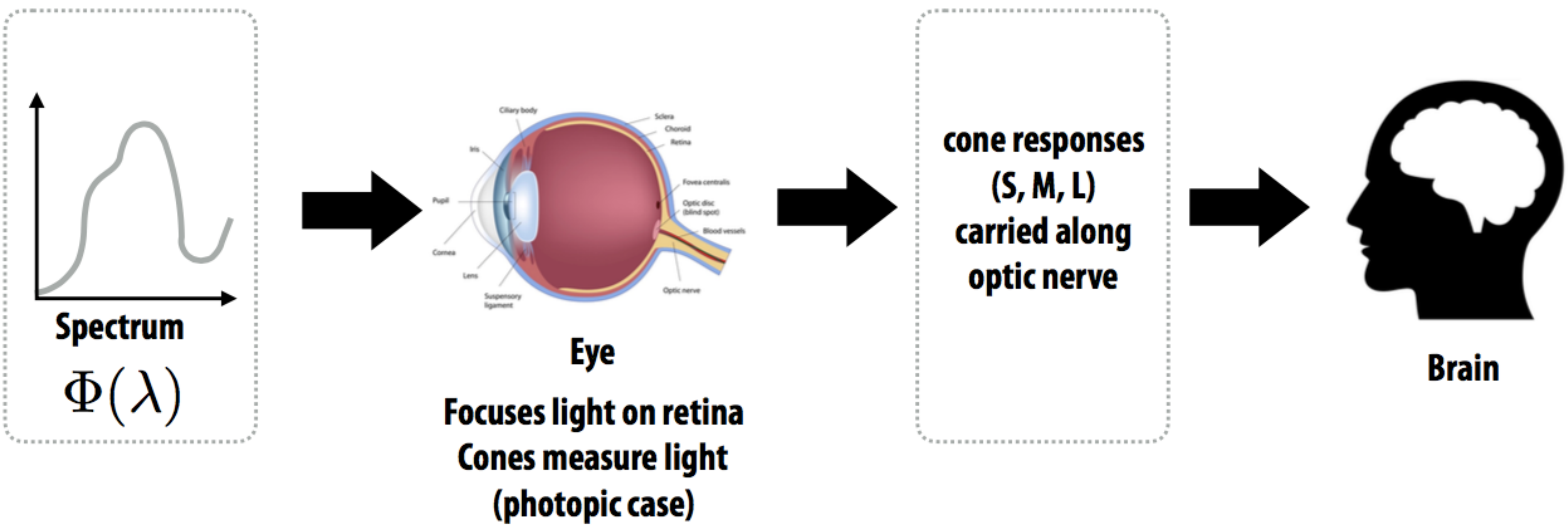
$$M = \int r_M(\lambda) s(\lambda) \, d\lambda$$

$$L = \int r_L(\lambda) s(\lambda) \, d\lambda$$



Recall: The Human Visual Color System

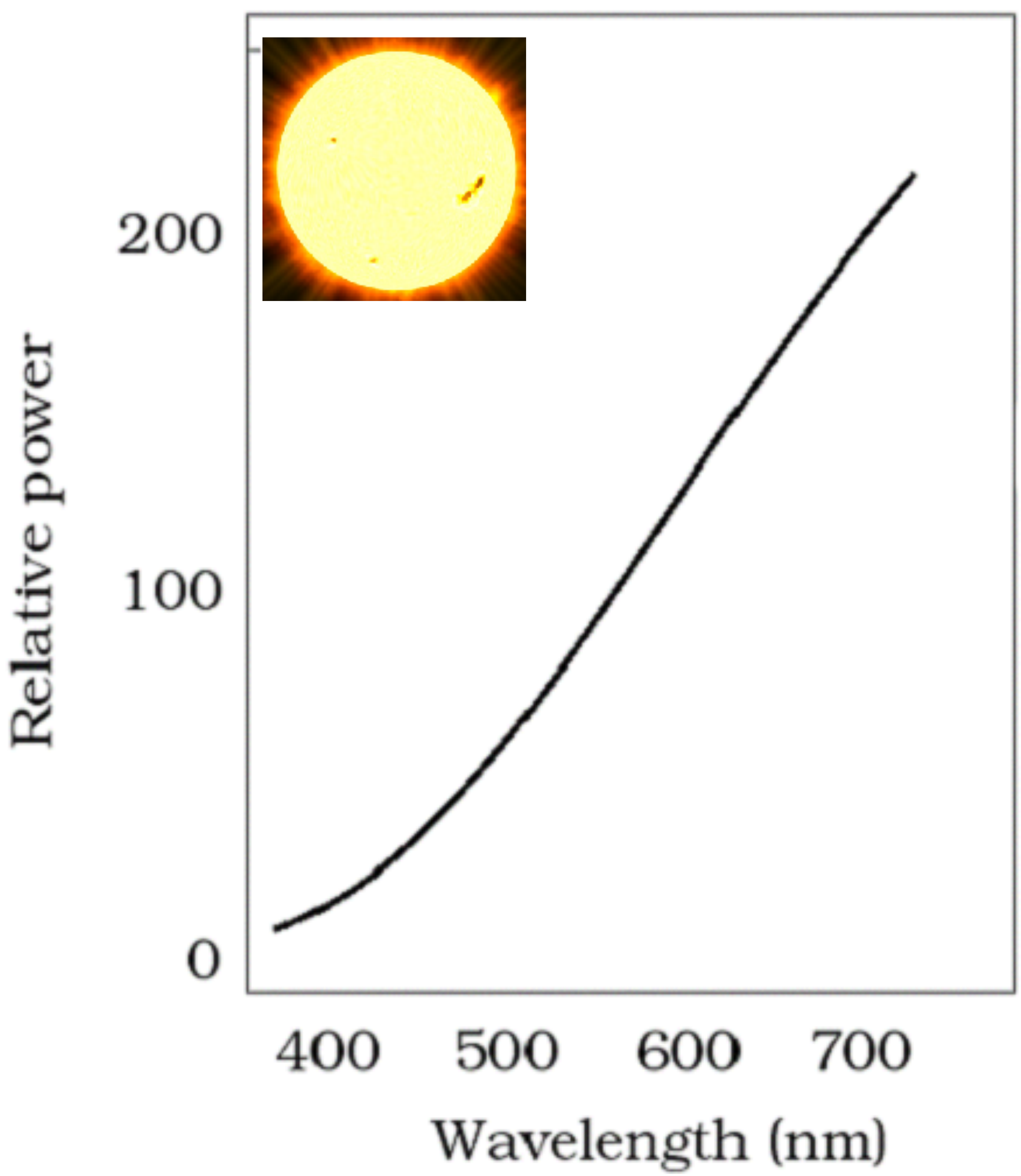
- Human eye does not measure and brain does not receive measure of each wavelength of light
- Rather, the eye measures three response values = (S, M, L).
  - The result of integrating the incoming spectrum against response functions of S, M, L cones

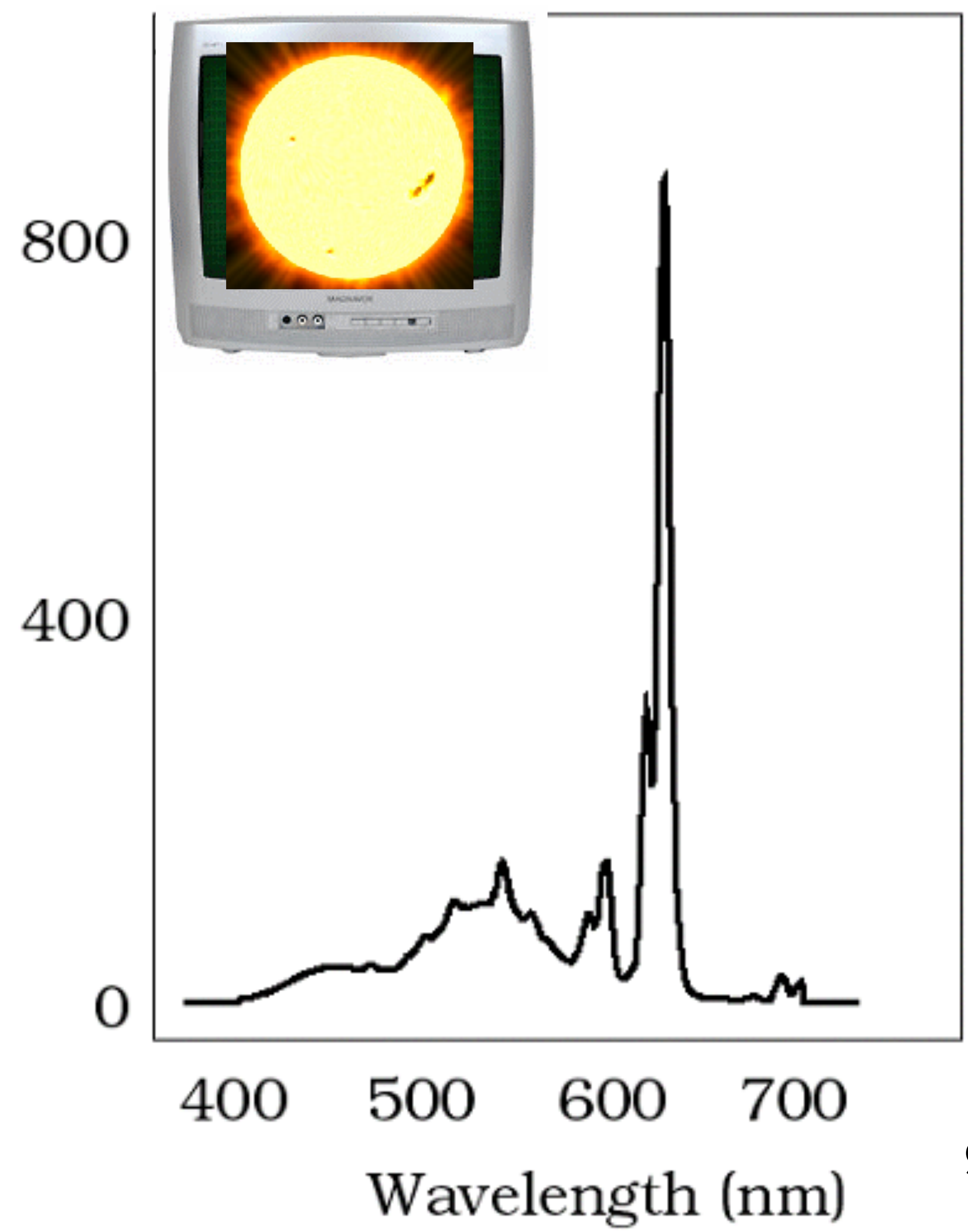


Kayvon Fatahalian

Recall: Metamerism

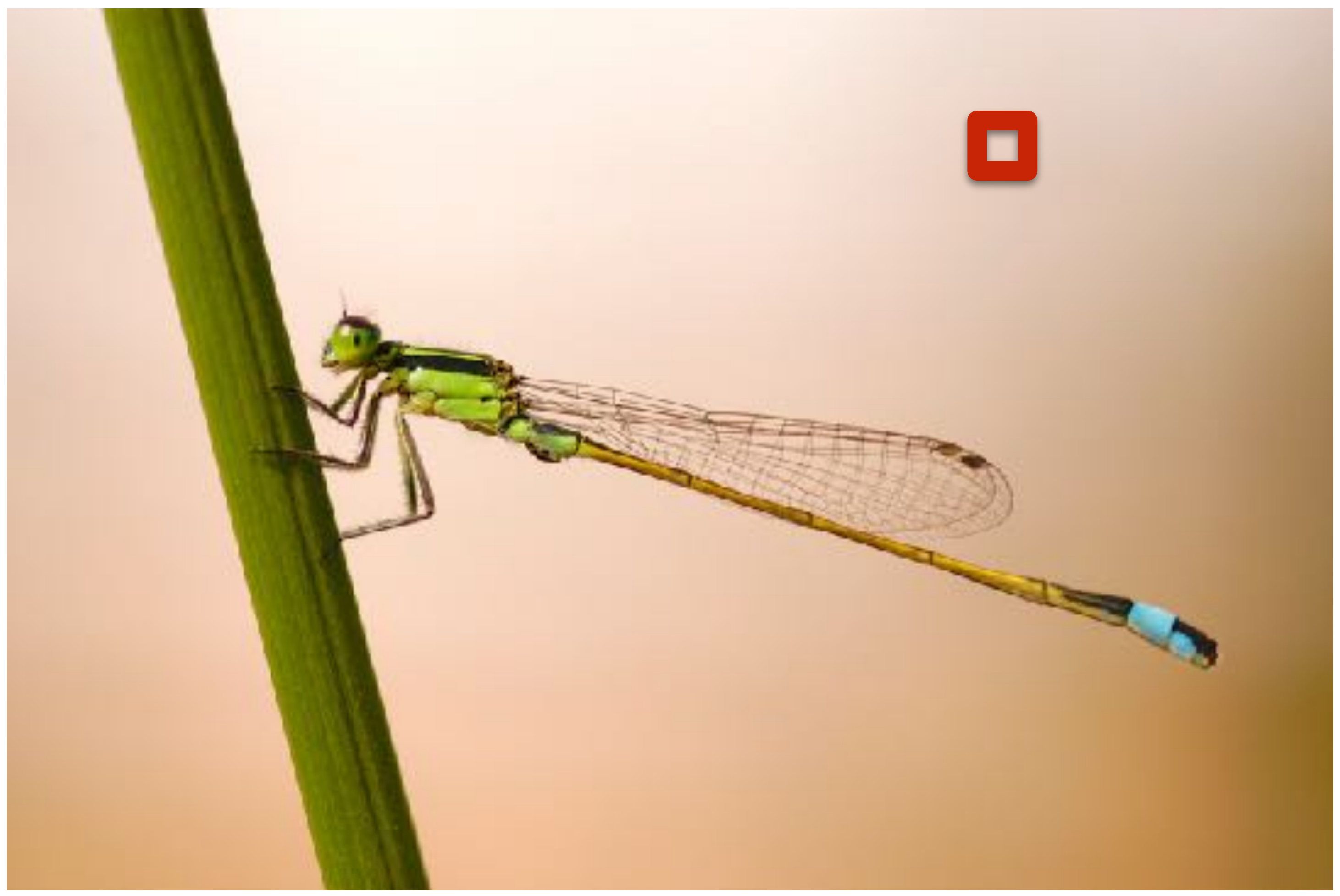
Color matching is an important illusion that is understood quantitatively

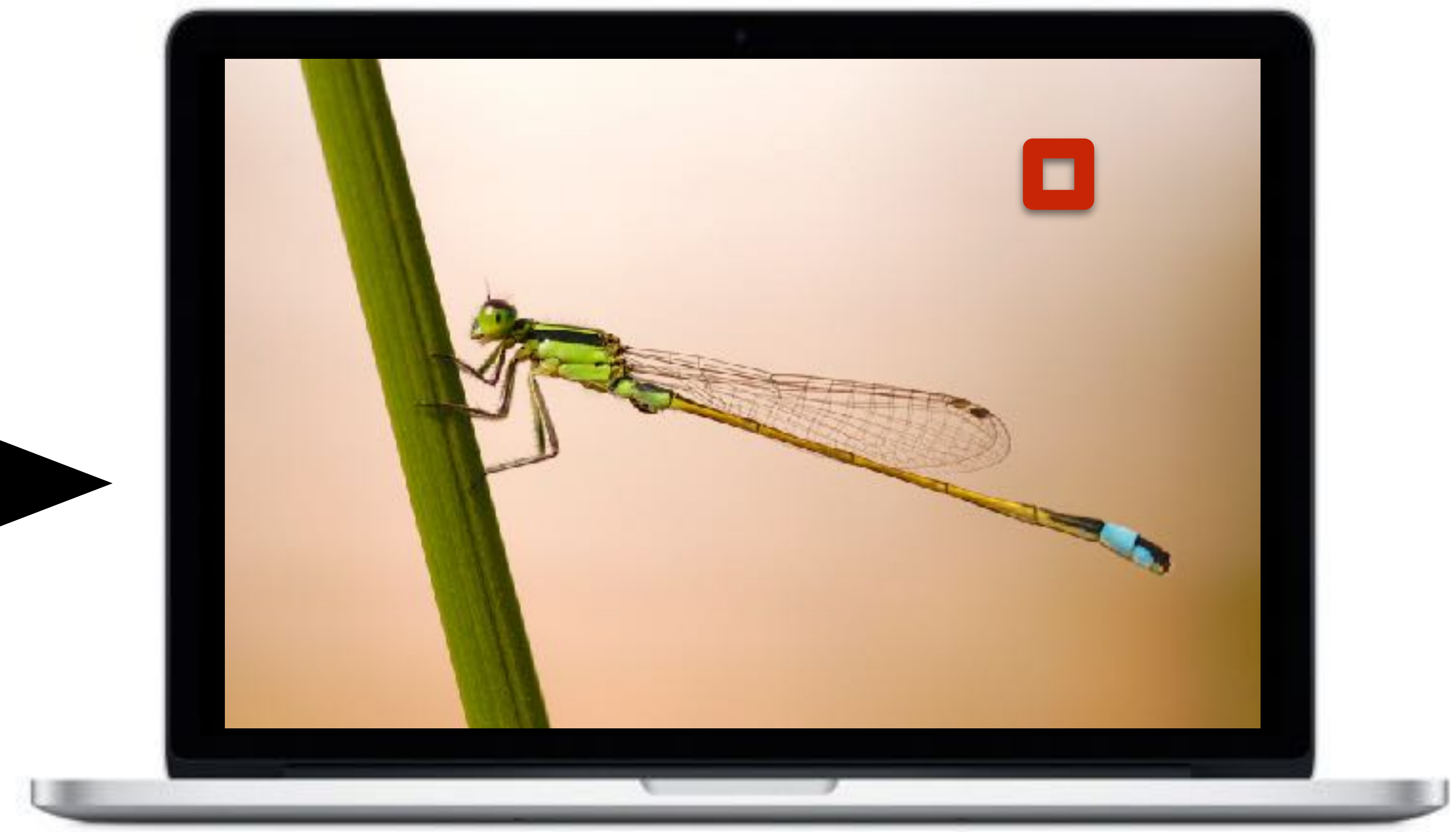




Brian Wandell

Recall: Color Reproduction Problem





Target real spectrum $s(\lambda)$

Display outputs spectrum
$R\, s_R(\lambda) + G\, s_G(\lambda) + B\, s_B(\lambda)$

Goal: at each pixel, choose R, G, B values for display so that the output color matches the appearance of the target color in the real world.

Recall: Color Reproduction as Linear Algebra

Color perceived for display spectra with values R,G,B

$$\begin{bmatrix} S \\ M \\ L \end{bmatrix}_{\text{disp}} = \begin{bmatrix} \text{---} & r_S & \text{---} \\ \text{---} & r_M & \text{---} \\ \text{---} & r_L & \text{---} \end{bmatrix} \begin{bmatrix} | & | & | \\ s_R & s_G & s_B \\ | & | & | \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

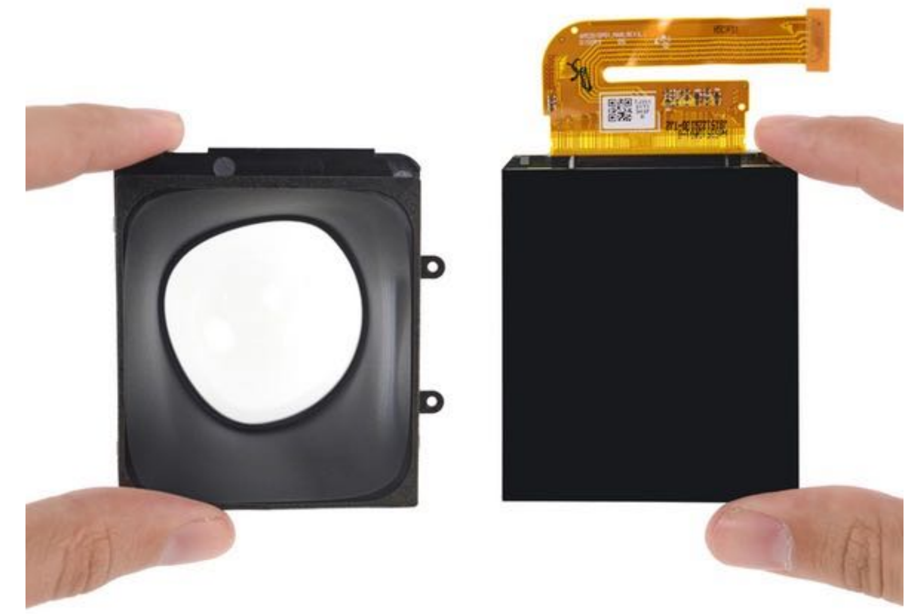
Color perceived for real scene spectra, s

$$\begin{bmatrix} S \\ M \\ L \end{bmatrix}_{\text{real}} = \begin{bmatrix} \text{---} & r_S & \text{---} \\ \text{---} & r_M & \text{---} \\ \text{---} & r_L & \text{---} \end{bmatrix} \begin{bmatrix} | \\ s \\ | \end{bmatrix}$$

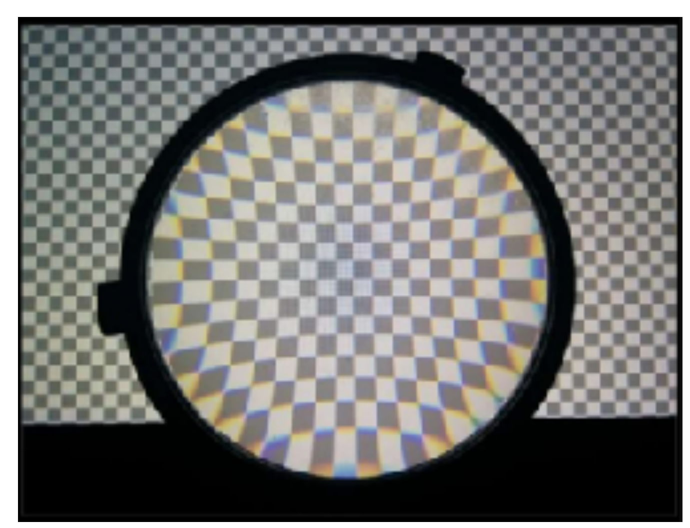
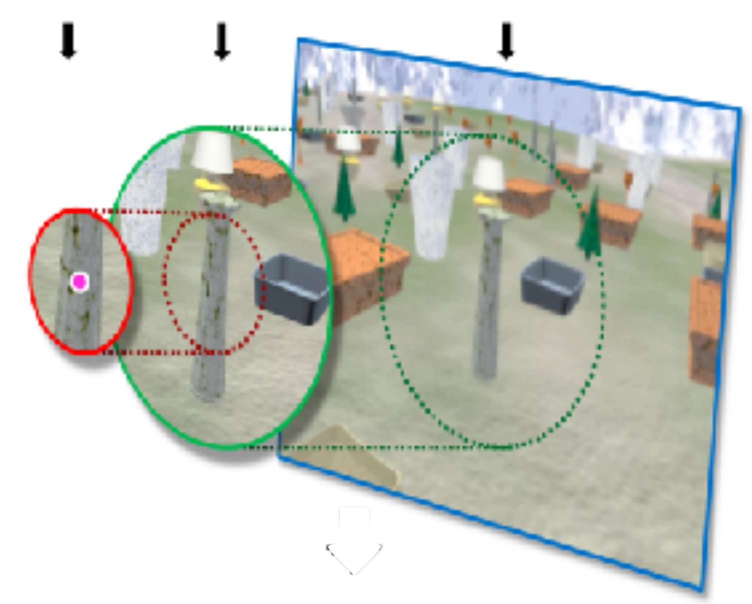
Reproduce color of s? Set these lines equal and solve for R,G,B as a function of s!

Overview of VR Topics

- VR Displays


- VR Rendering

- VR Imaging

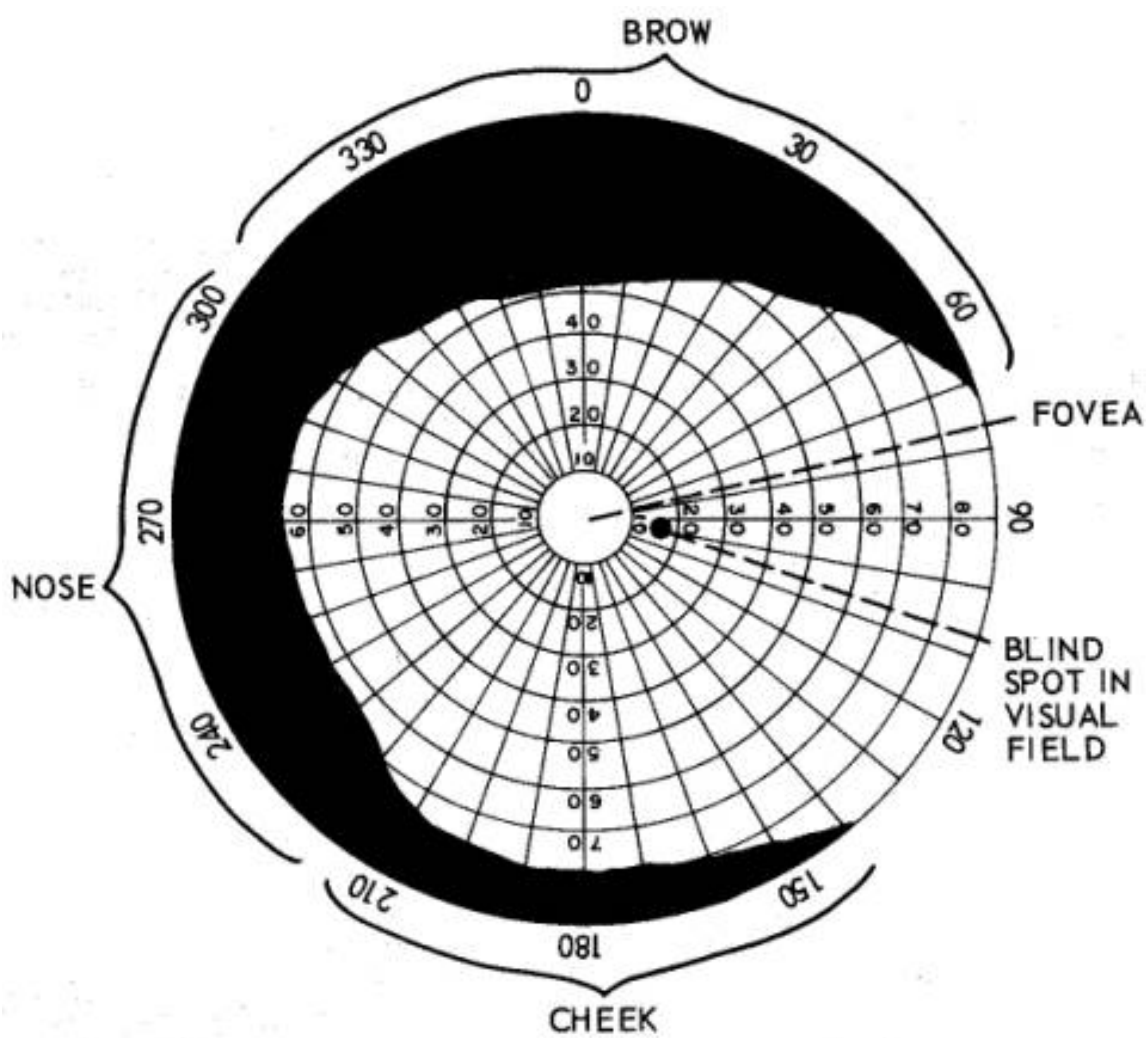


Display Requirements Derive From Human Perception

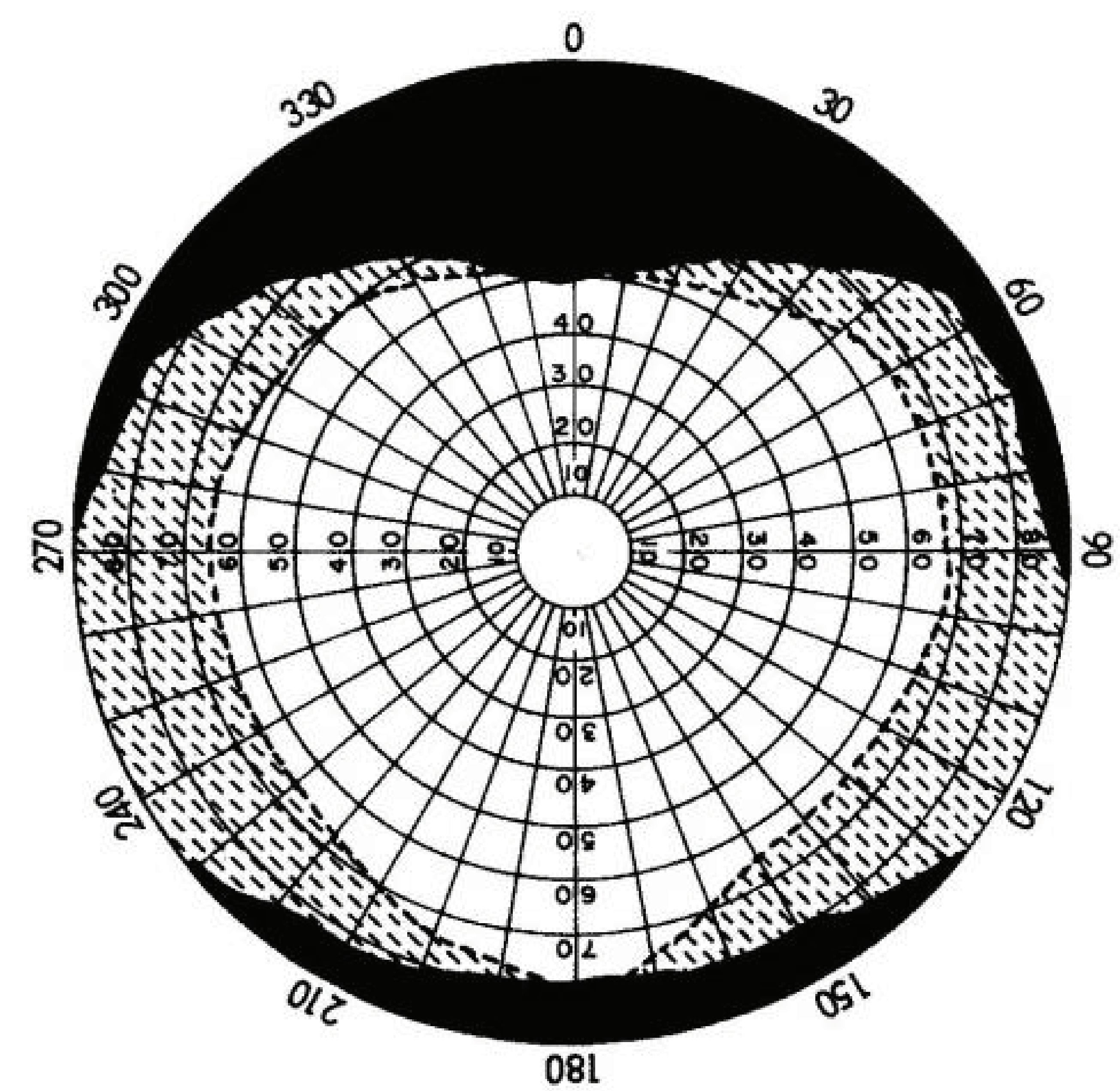
Example 2: Field of View & Resolution

Human Visual Field of View

Ruch & Fulton, 1960



monocular visual field

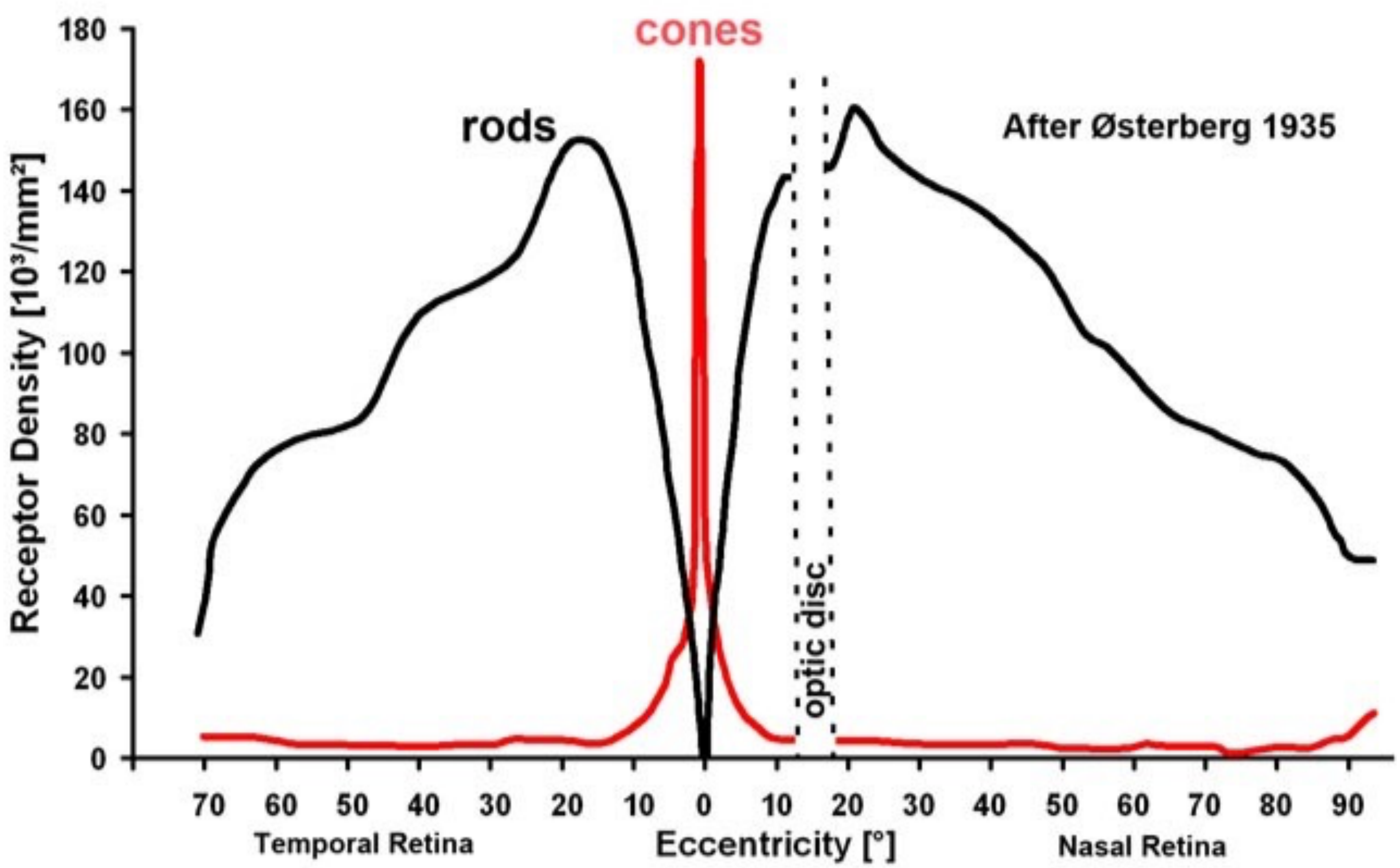


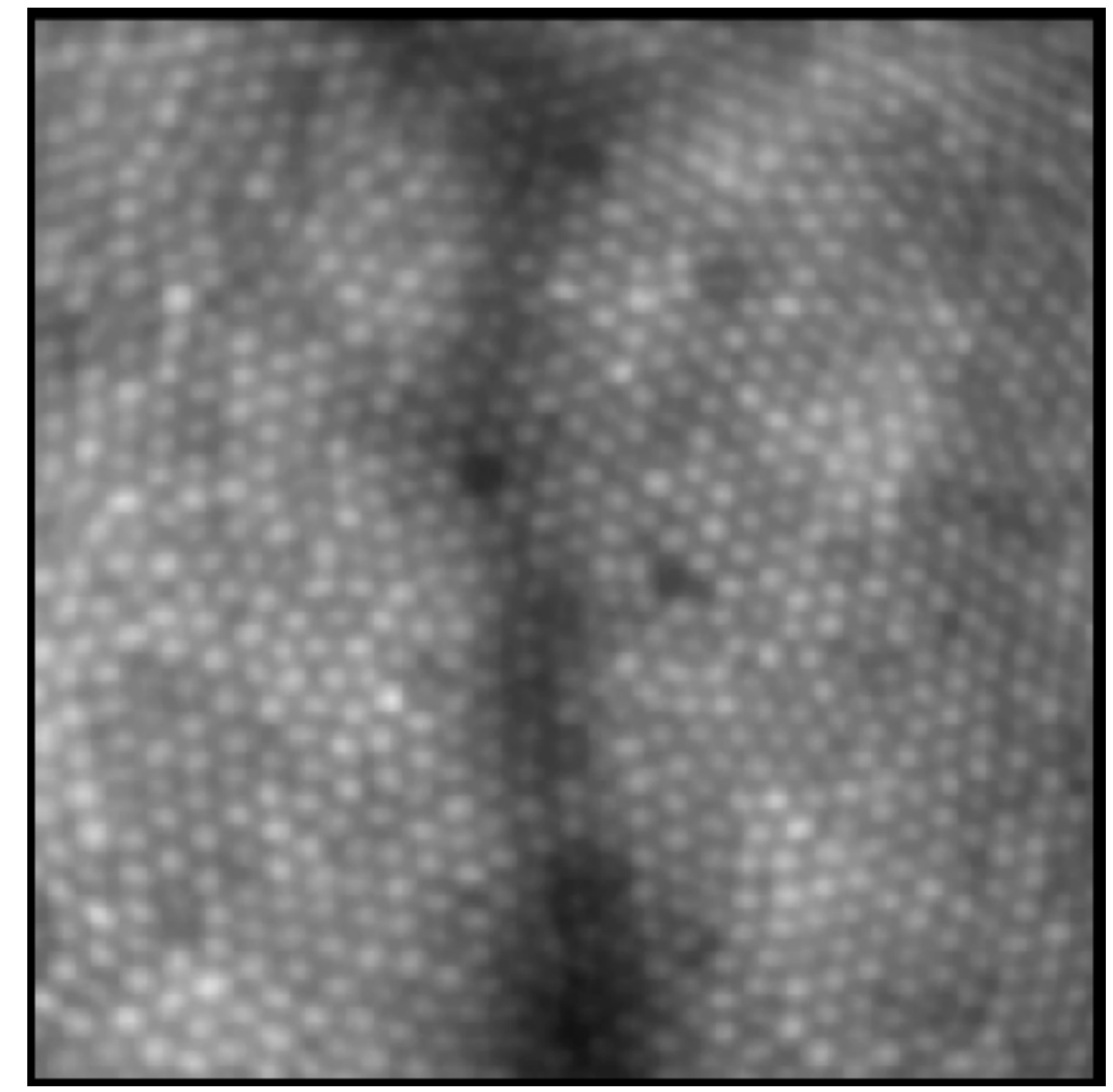
binocular visual field

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





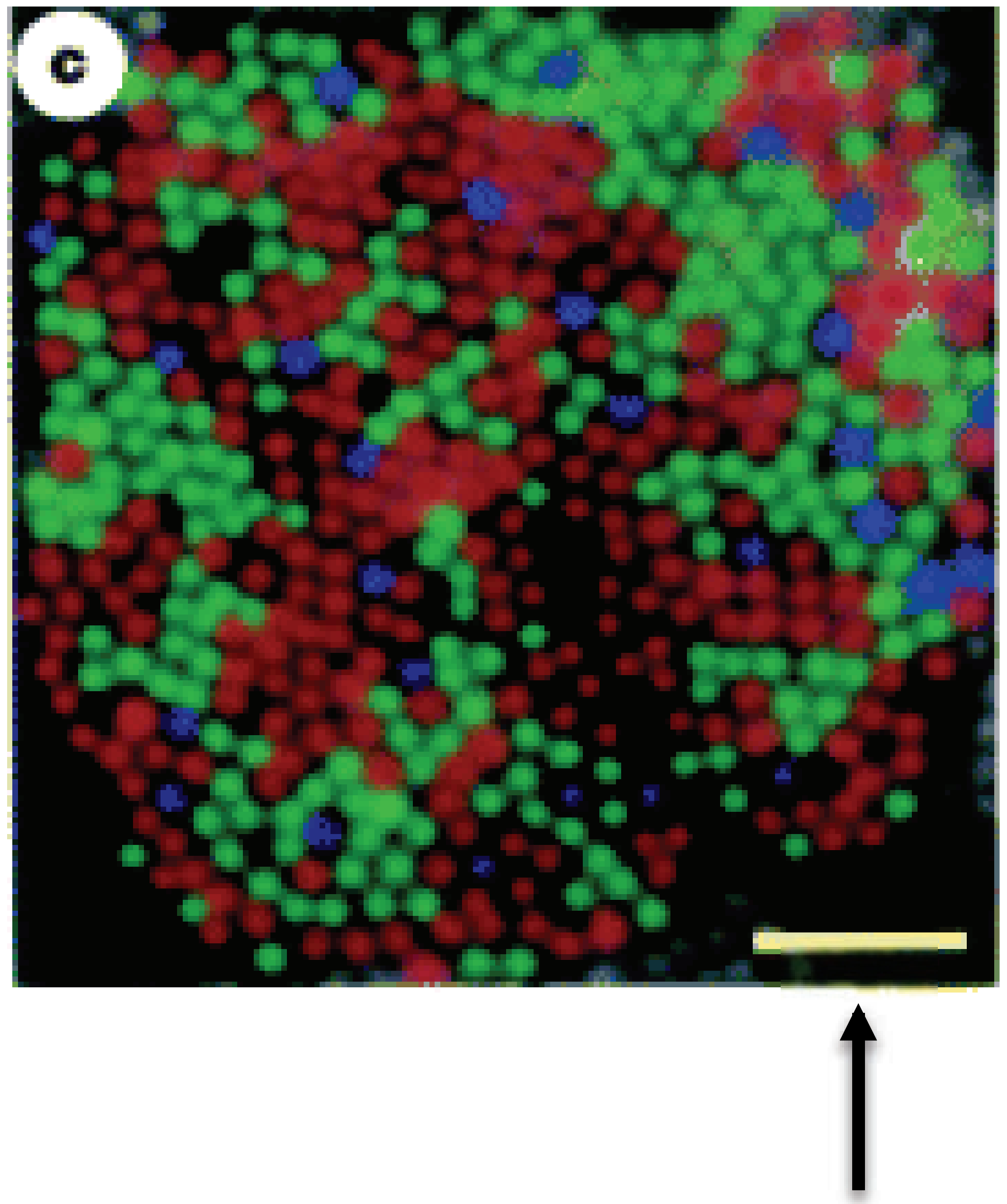
[Roorda 1999]

- 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
~ 1 arc min (1/60 of a degree)

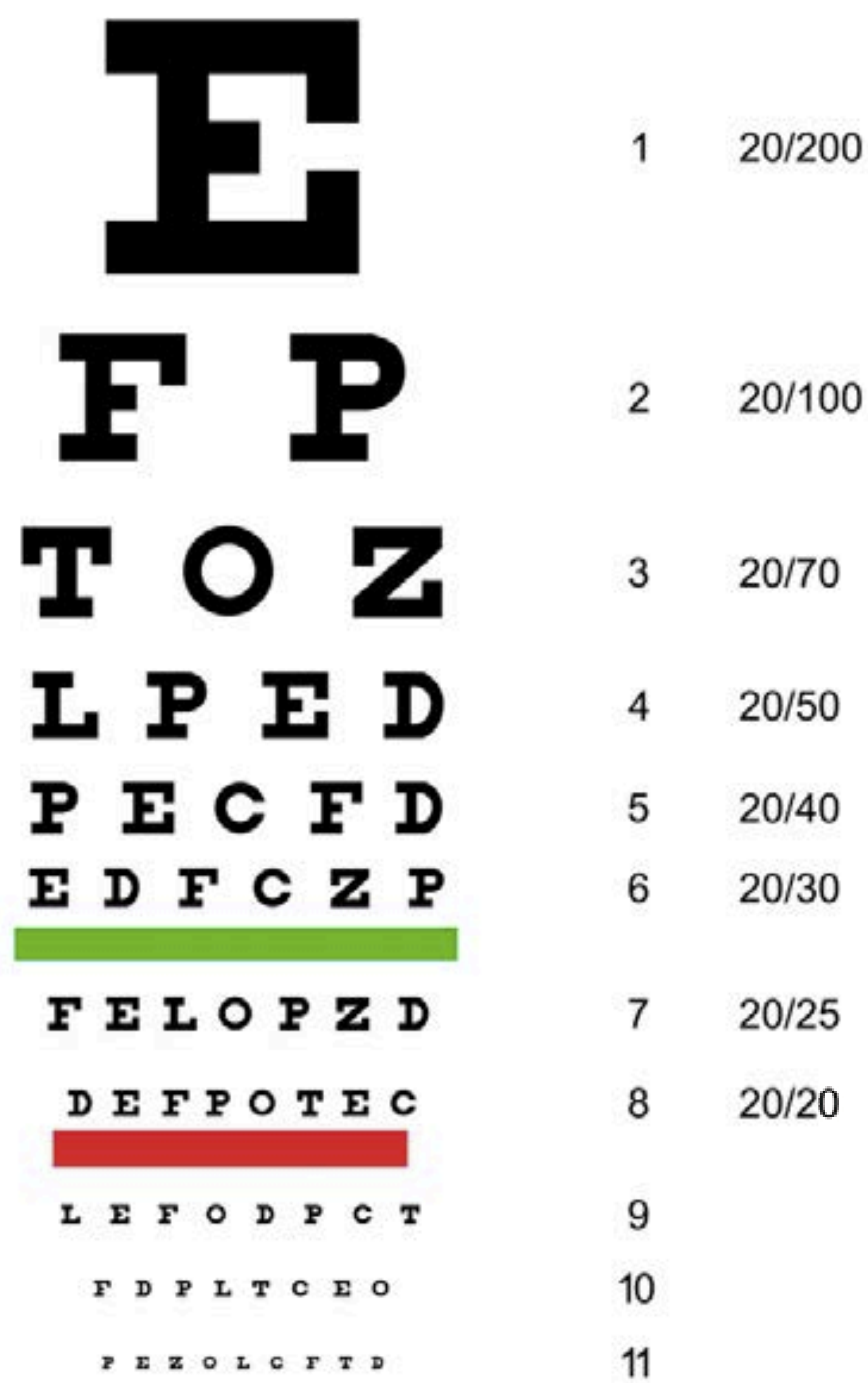


5 arcmin visual angle

Slide credit: Gordon Wetzstein

Visual Acuity

Snellen chart



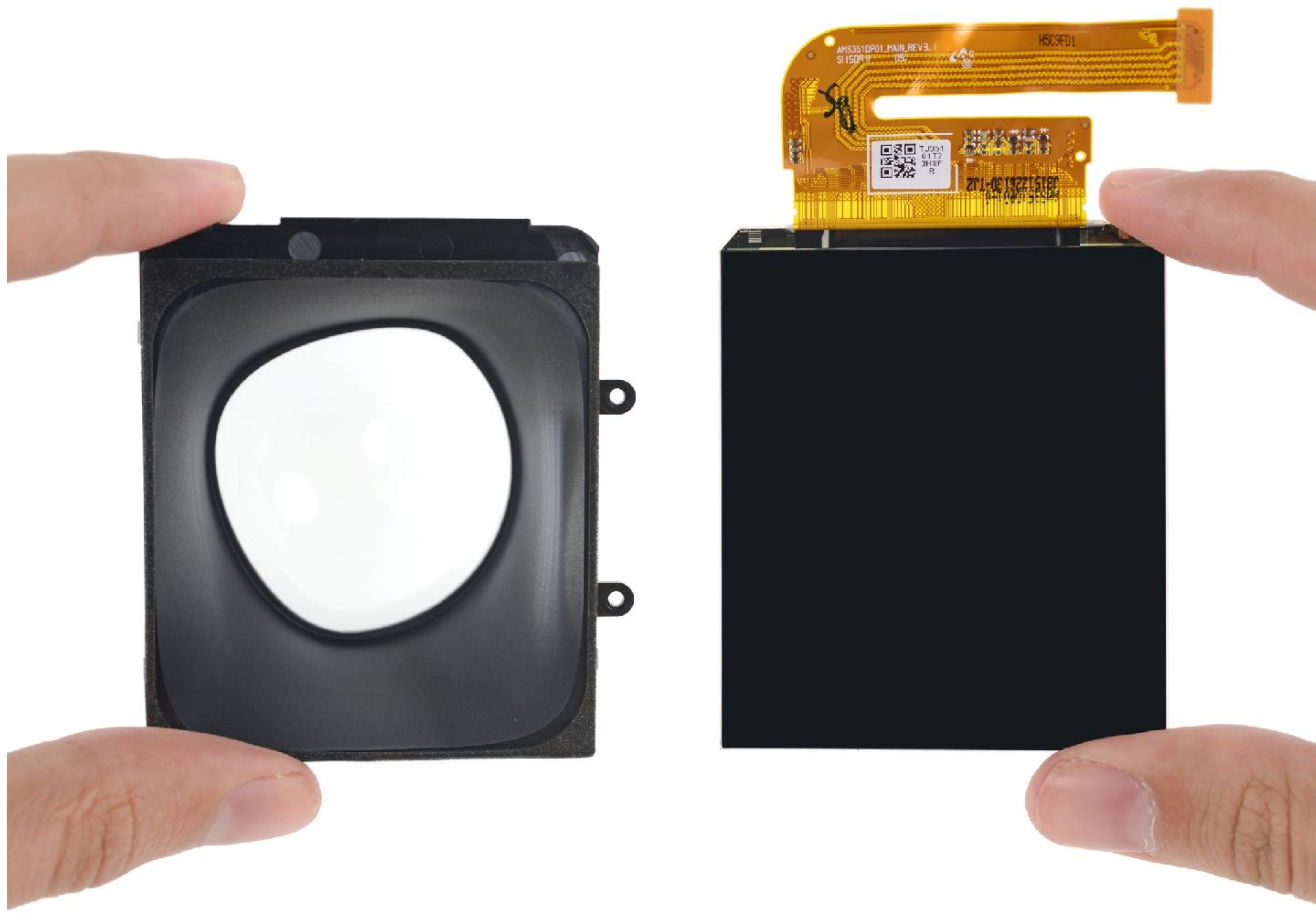
characters are 5 arc min, need to resolve 1 arc min to read

Slide credit: Gordon Wetzstein

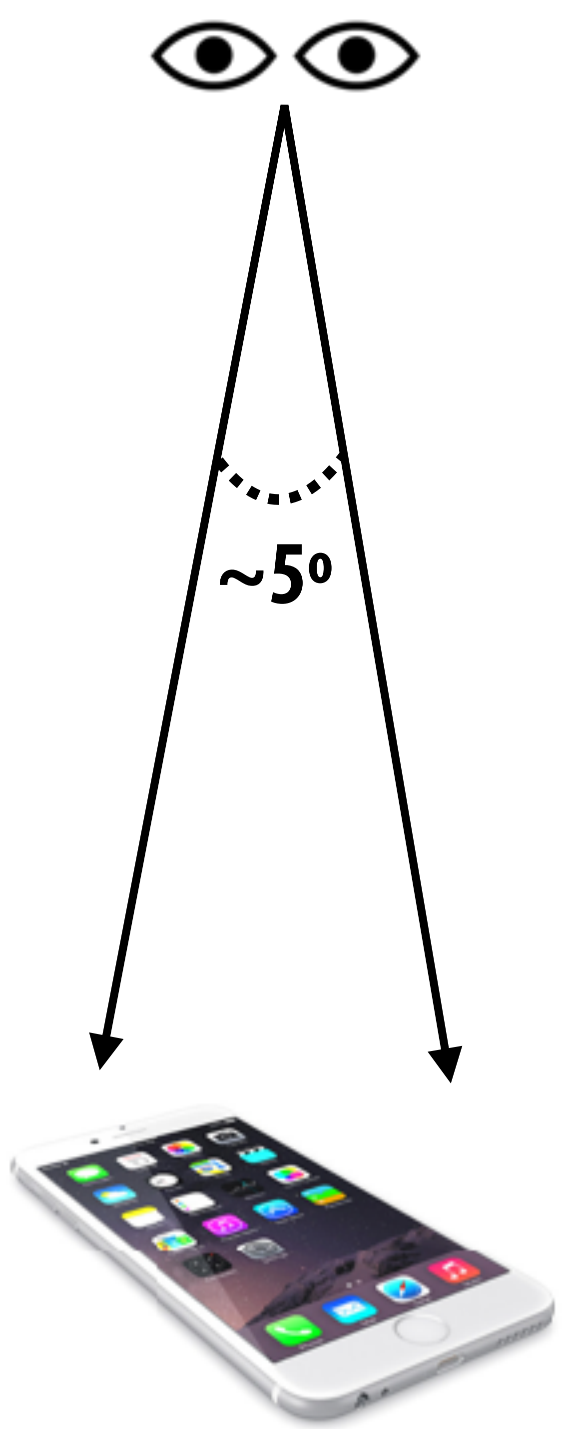
Current VR Headset Field of View and Resolution

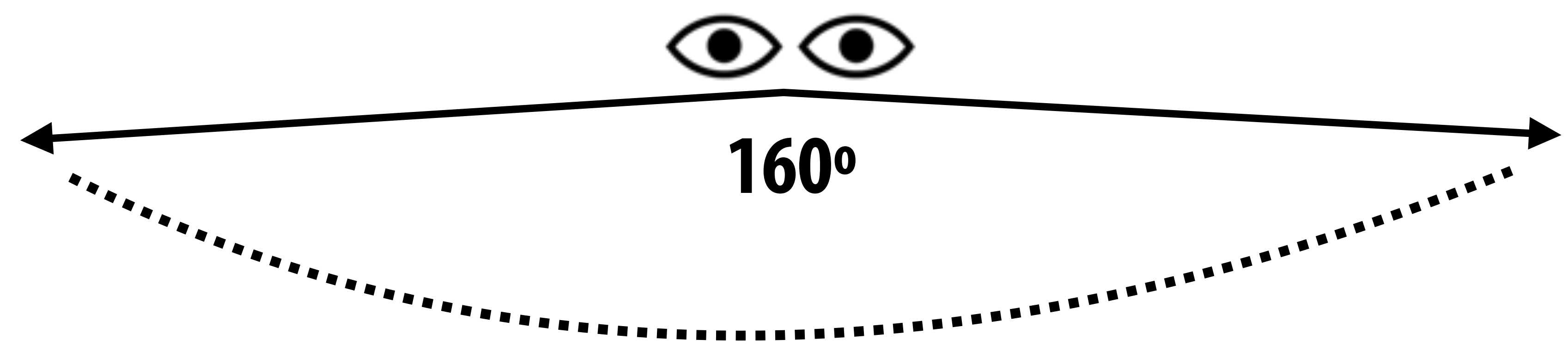
Example: Oculus Rift

- 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)
- [Note: modern headsets of this kind up to 1440x1700 now]



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 3: Binocular Stereo and Eye Focus (“Accommodation”)

Two Eyes: Two Views

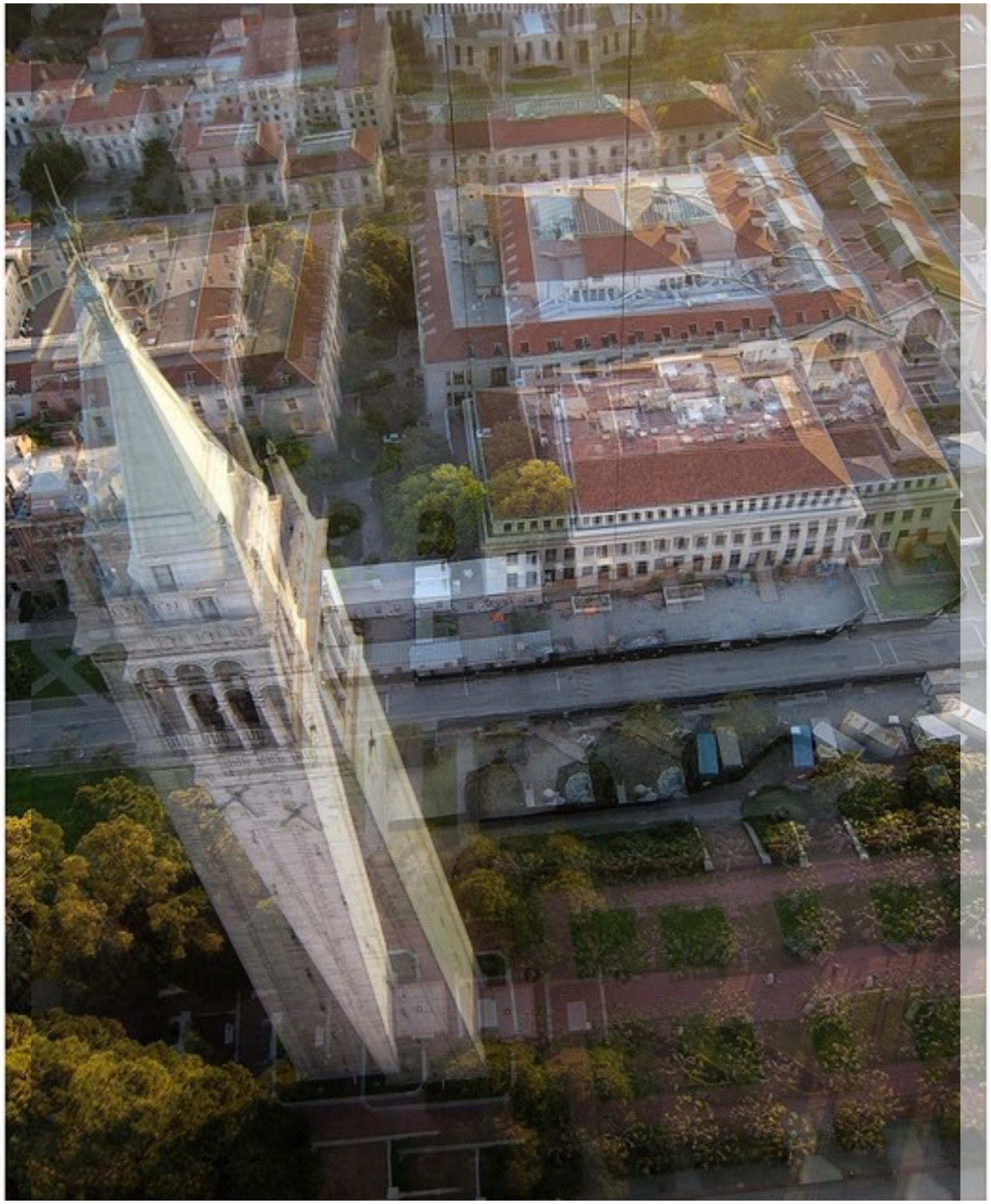
Charles Wheatstone stereoscope, 1838

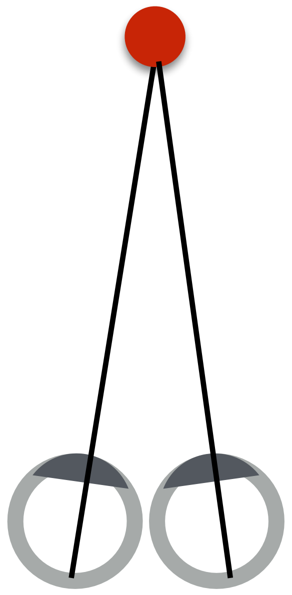
Recall: Current VR HMD Optical Design

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

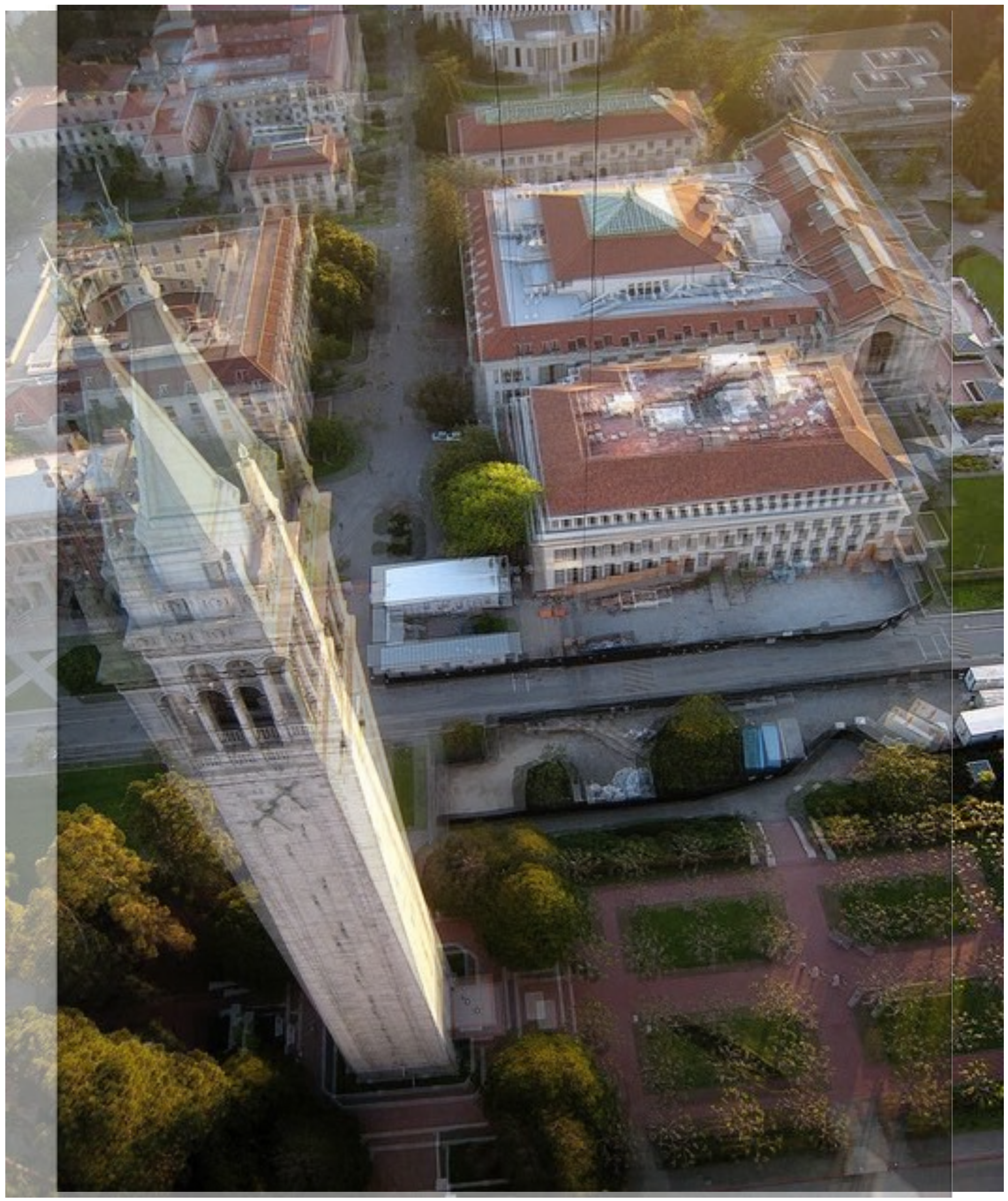
Stereo Vergence

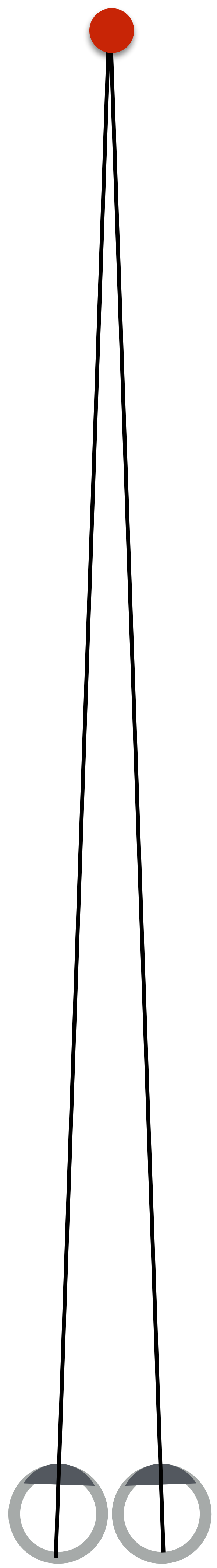
Stereo Vergence





Stereo Vergence

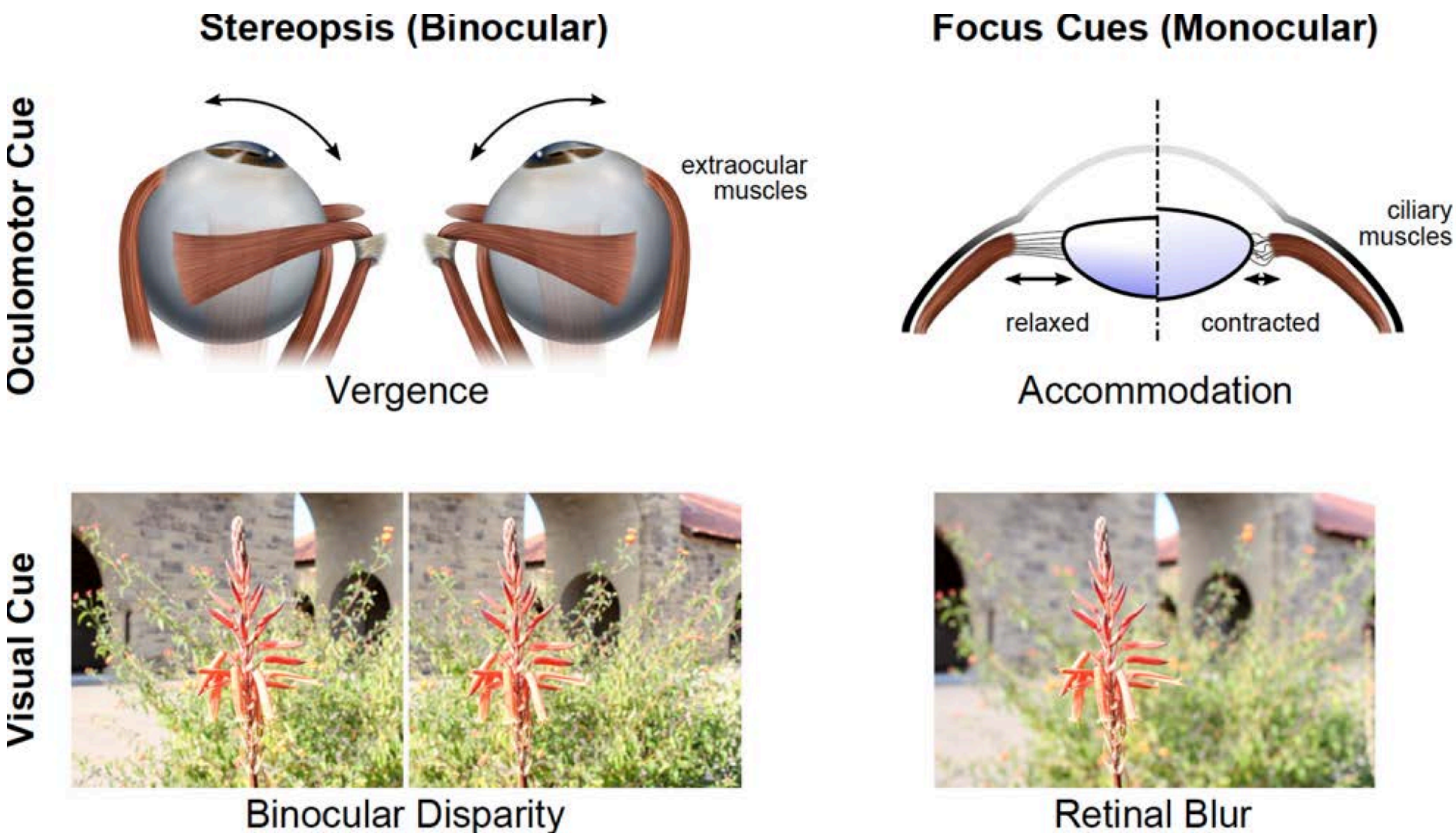




Stereo

- 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



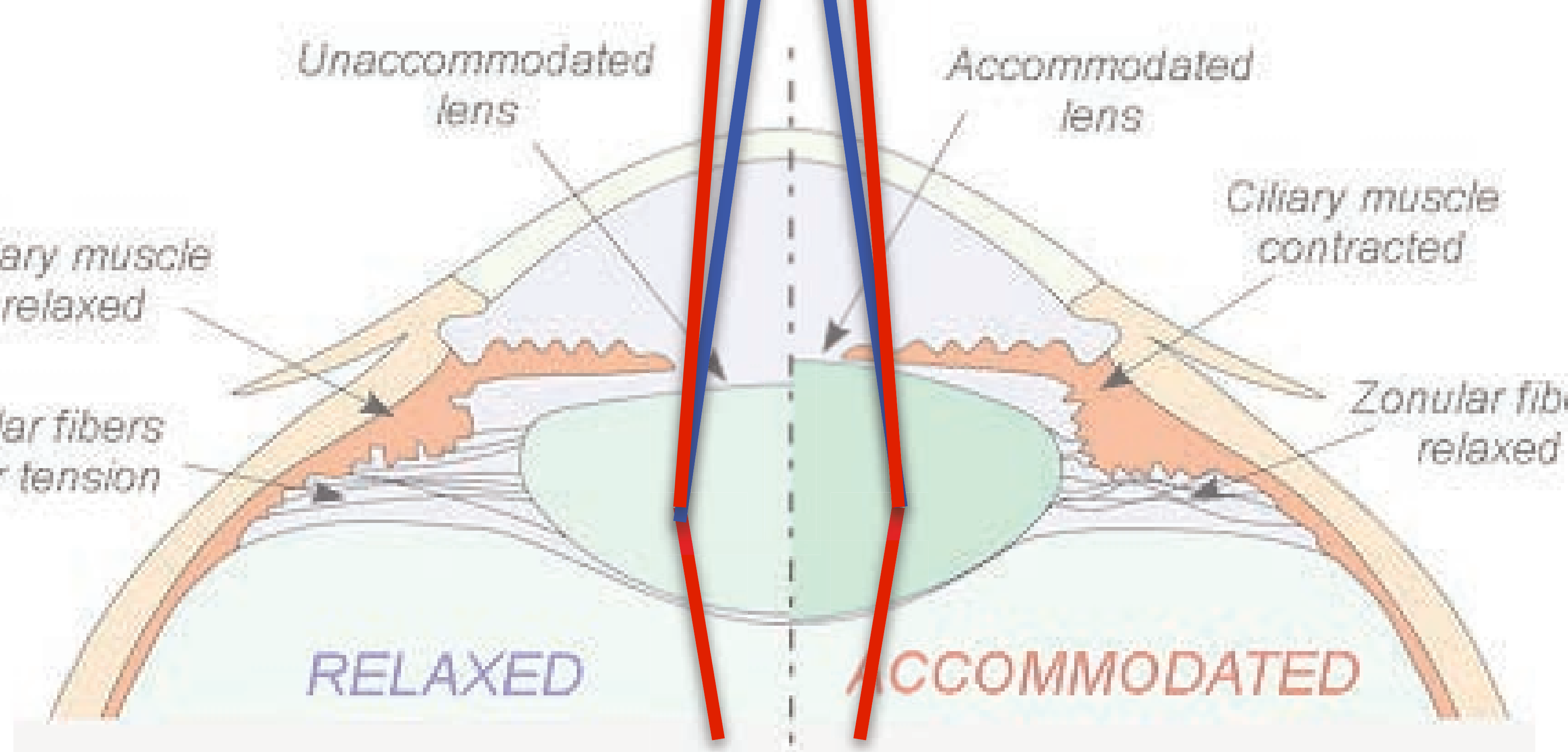
Slide credit: Gordon Wetzstein

Human Eye Muscles and Optical Controls

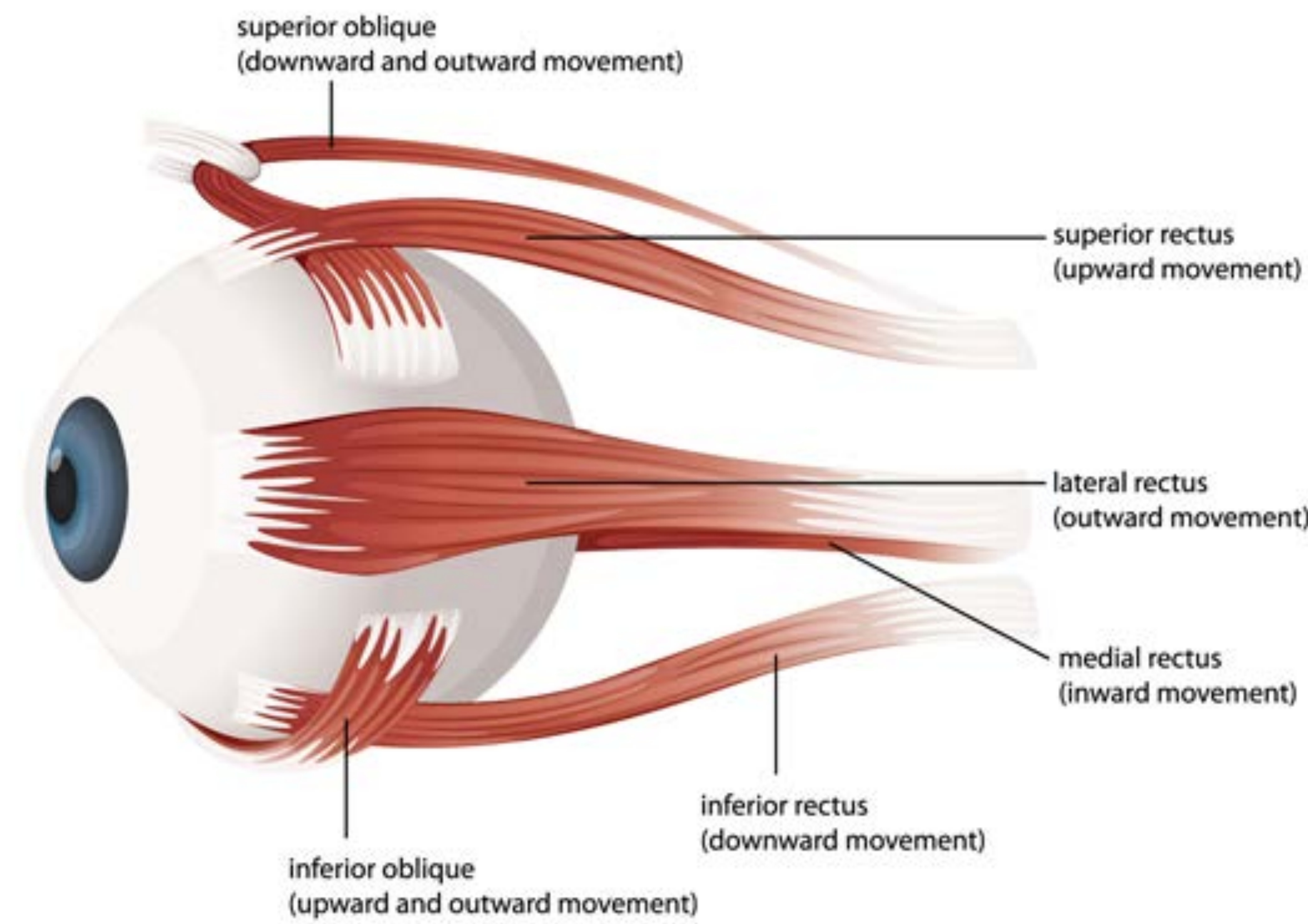
far focus

near focus

16 years: ~8cm to ∞
50 years: ~50cm to ∞ (mostly irrelevant)



adithyakiran.wordpress.com



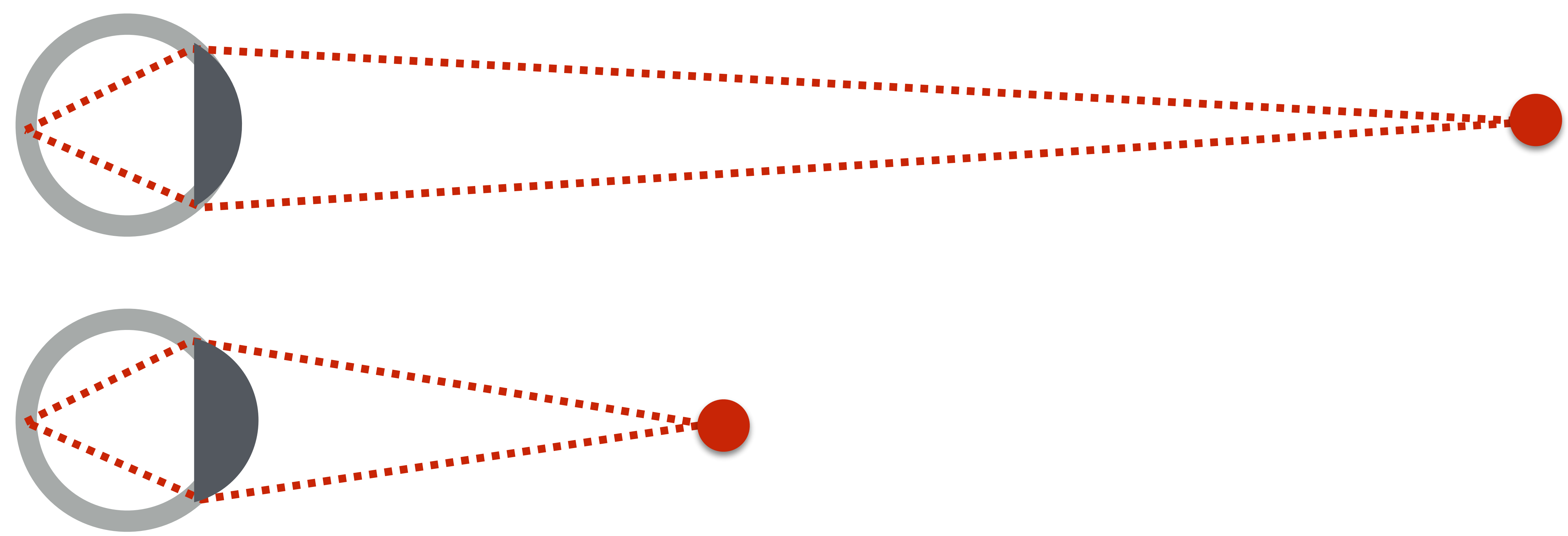
Slide credit: Gordon Wetzstein

Accommodation and Vergence

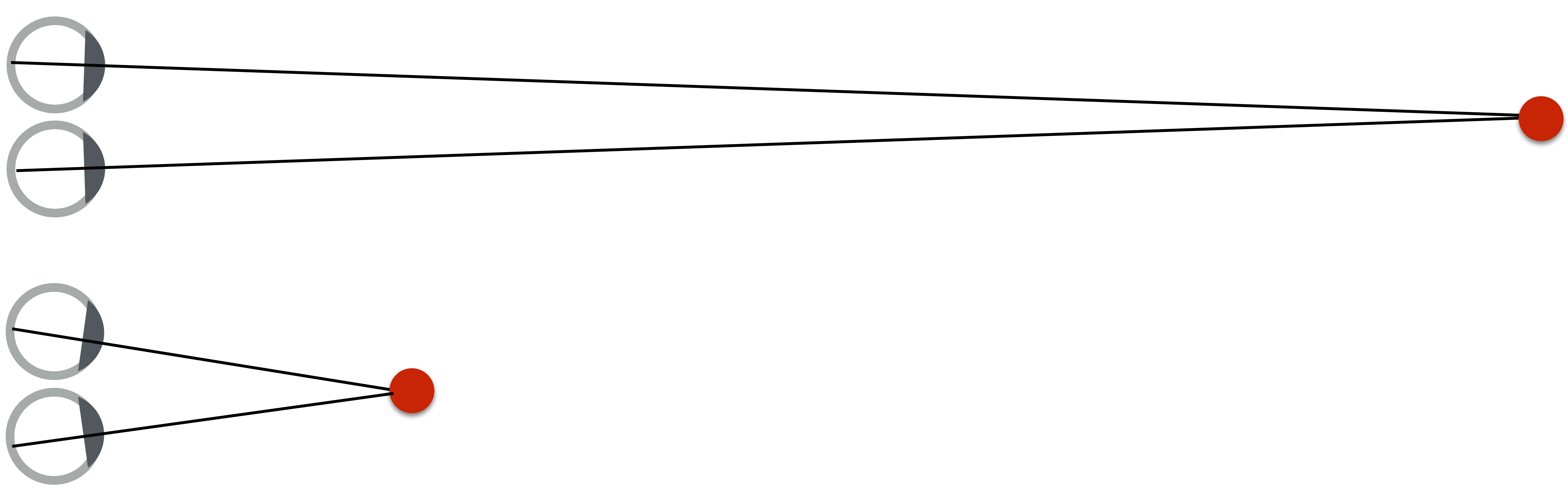
Accommodation: changing the optical power of the eye (lens) to focus at different distances

Eye accommodated to focus on a distant object

Eye accommodated to focus on a nearby object



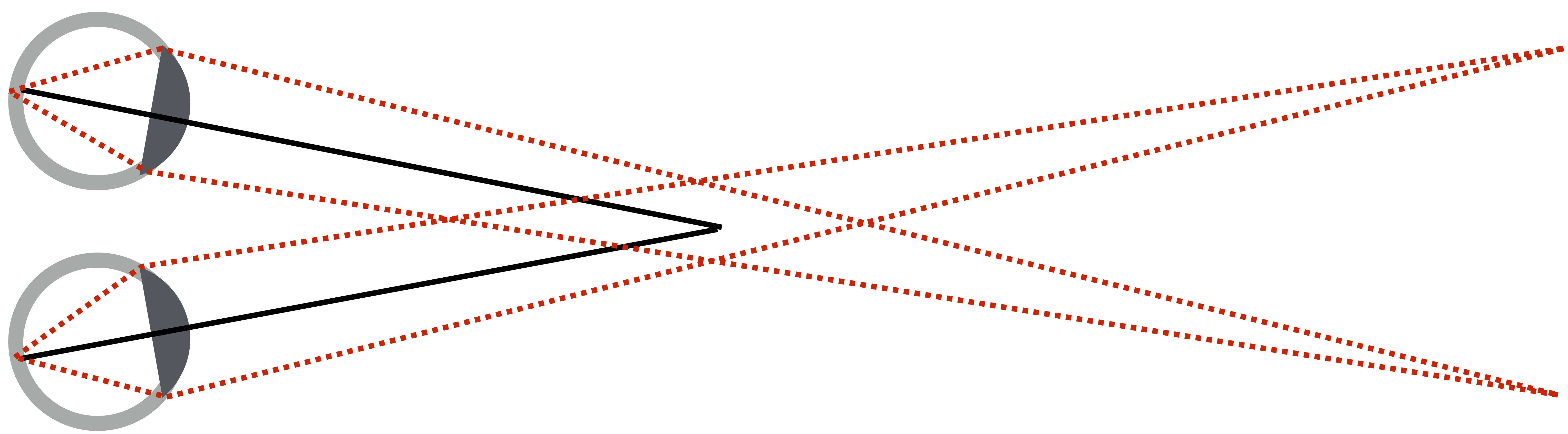
Vergence: rotation of the eye in its socket to ensure projection of object is centered on the retina



Accommodation – Vergence Conflict

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

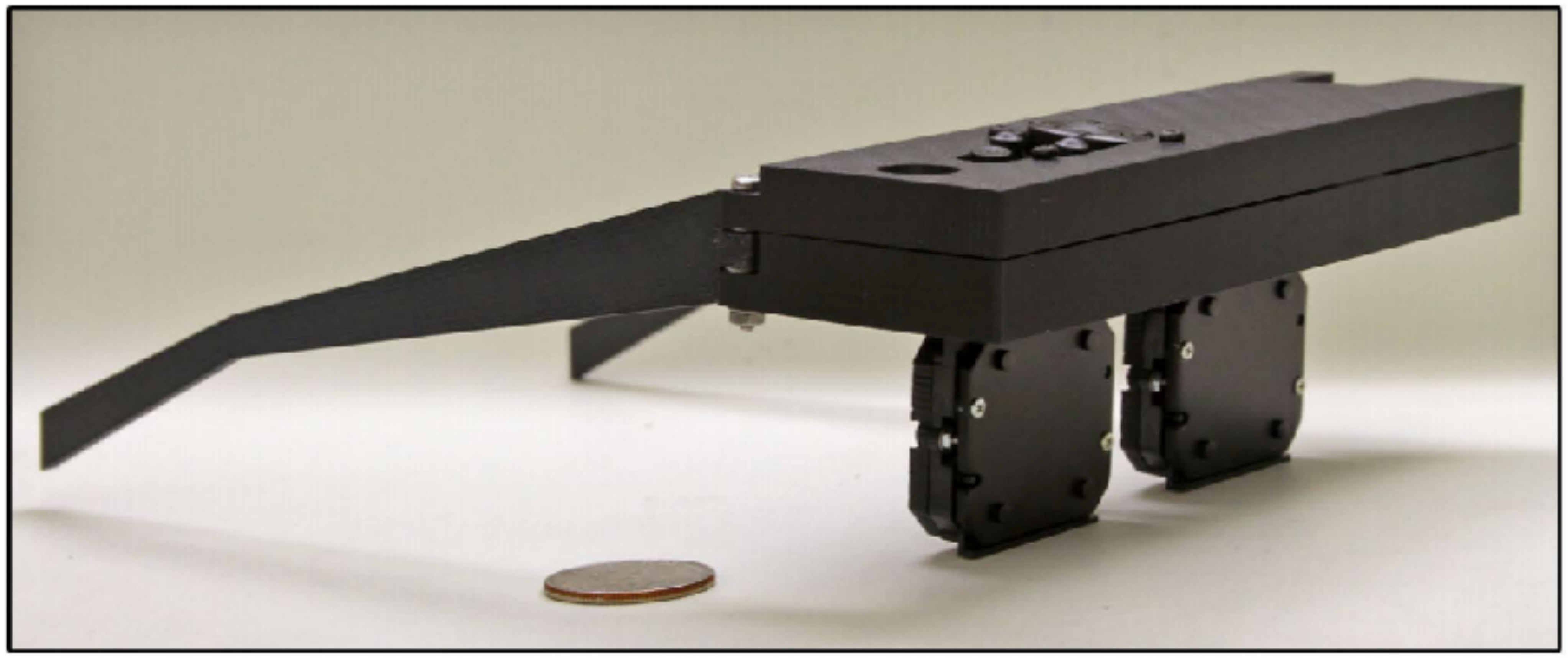
- Eyes must remain accommodated to far distance (otherwise image on screen won't be in focus)
- But eyes must converge in attempt to fuse stereoscopic images of object up close
- Brain receives conflicting depth clues... (discomfort, fatigue, nausea)

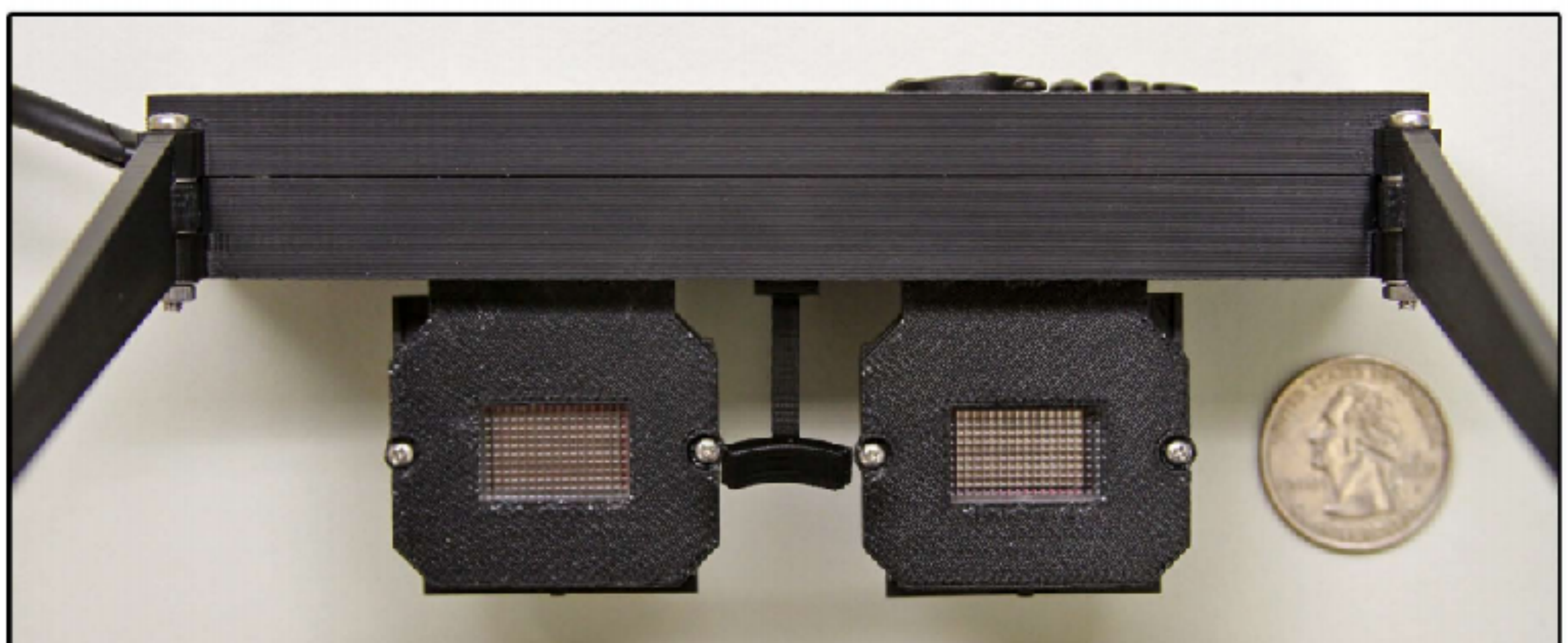


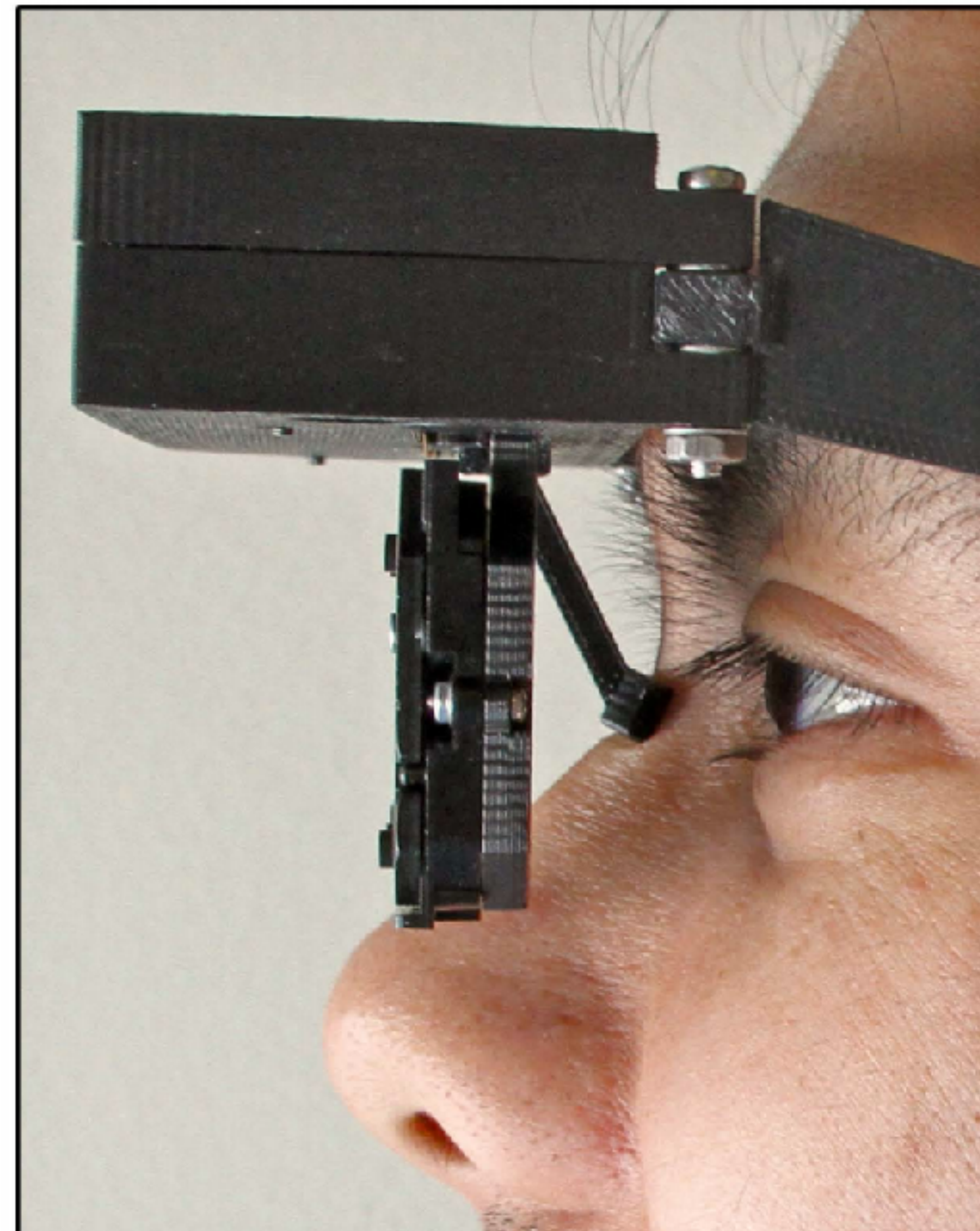
This problem stems from nature of display design. If you could just make a display that emits the light field that would be produced by a virtual scene, then you could avoid the accommodation - vergence conflict...

Aside: Research on Near-Eye Light Field Displays

Goal: recreate light field in front of eye





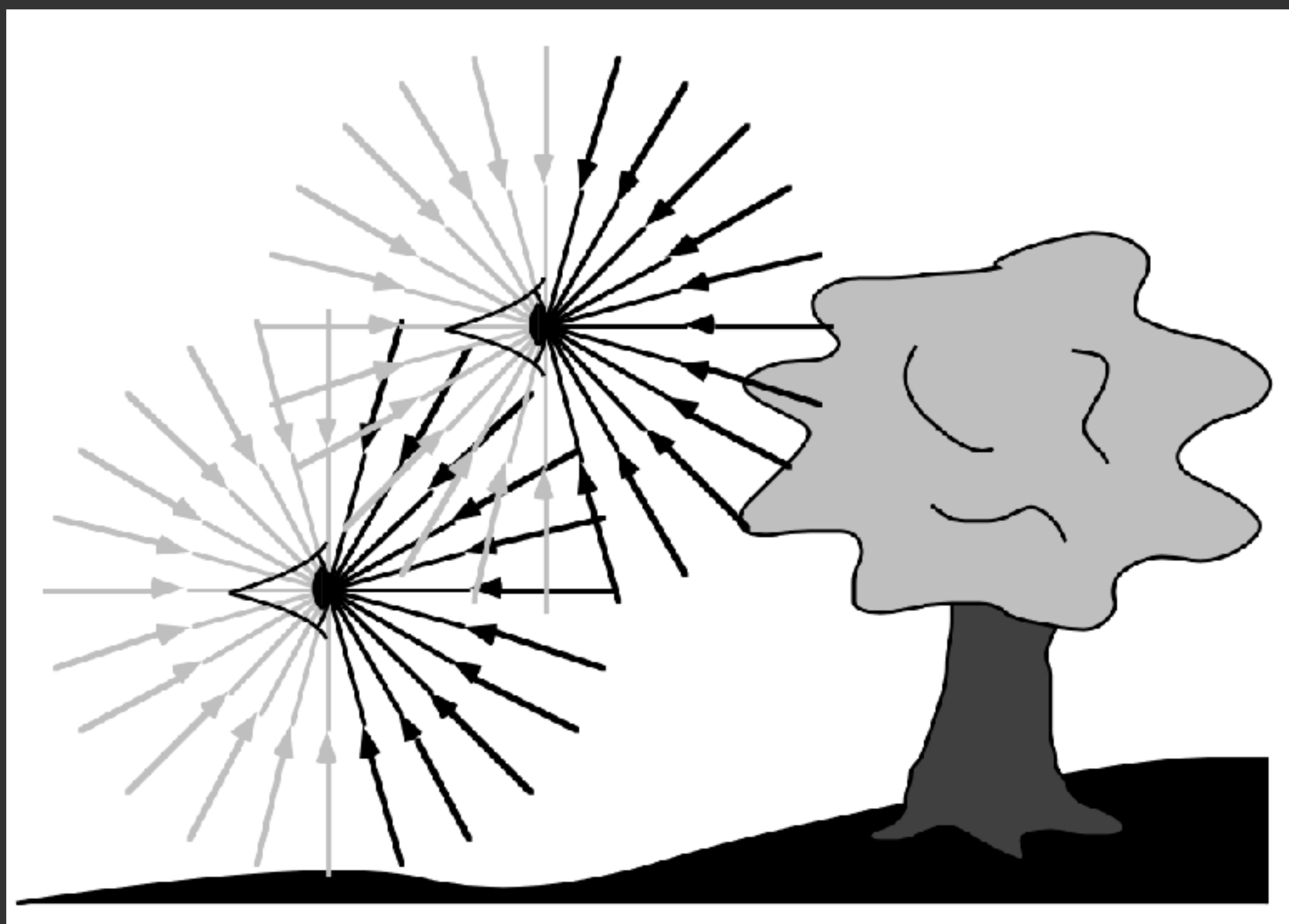


Lanman and Luebke, SIGGRAPH Asia 2013.

Display Requirements Derive From Human Perception

Example #4: Motion Parallax from Eye 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

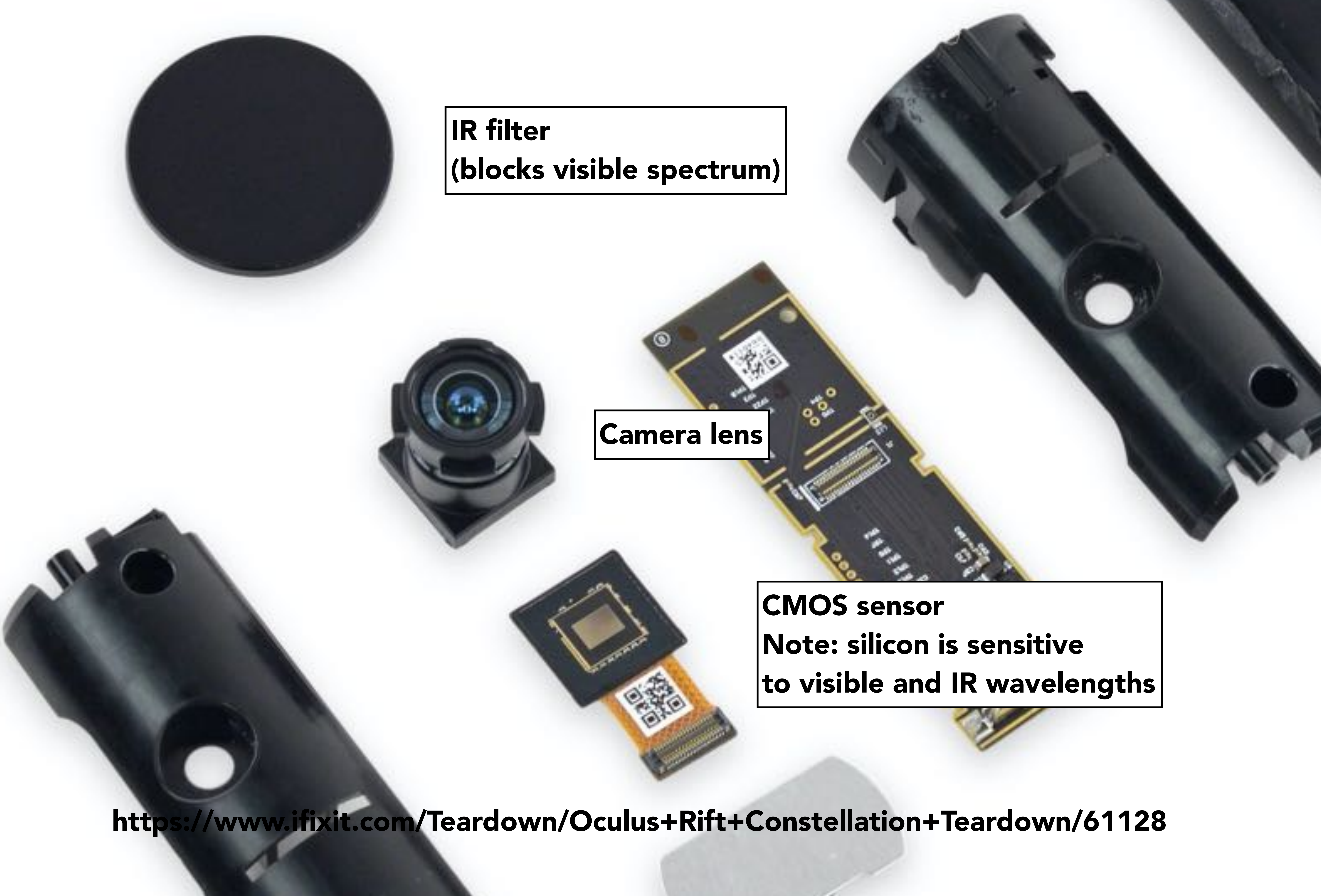
Oculus Rift + IR LED sensor

Oculus Rift IR LED Tracking Hardware

Photo taken with IR-sensitive camera

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

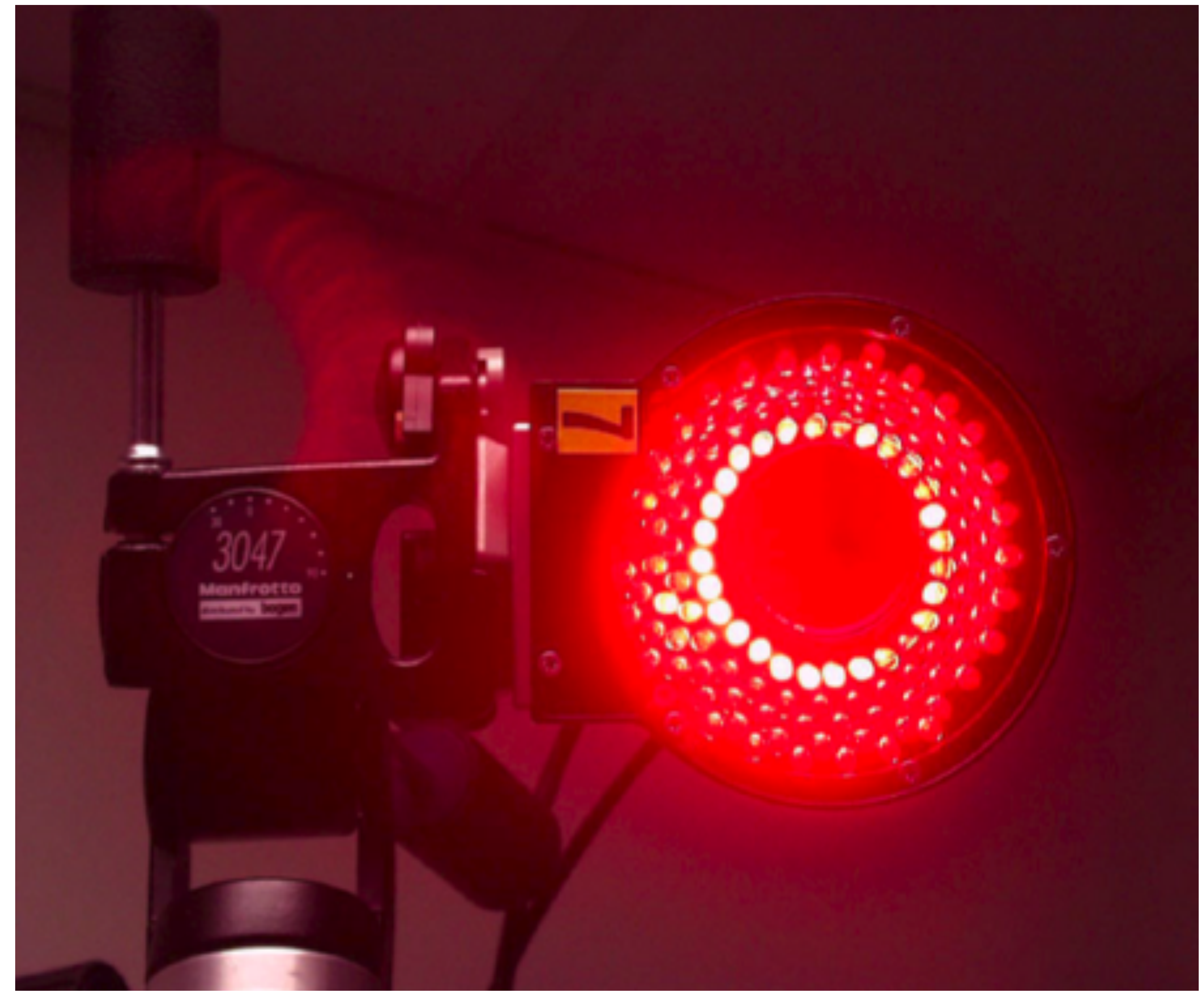
Oculus Rift IR Camera



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

Recall: Passive Optical Motion Capture

Retroflective markers attached to subject



IR illumination and cameras

- Markers on subject
- Positions by triangulation from multiple cameras
- 8+ cameras, 240 Hz, occlusions are difficult

Slide credit: Steve Marschner

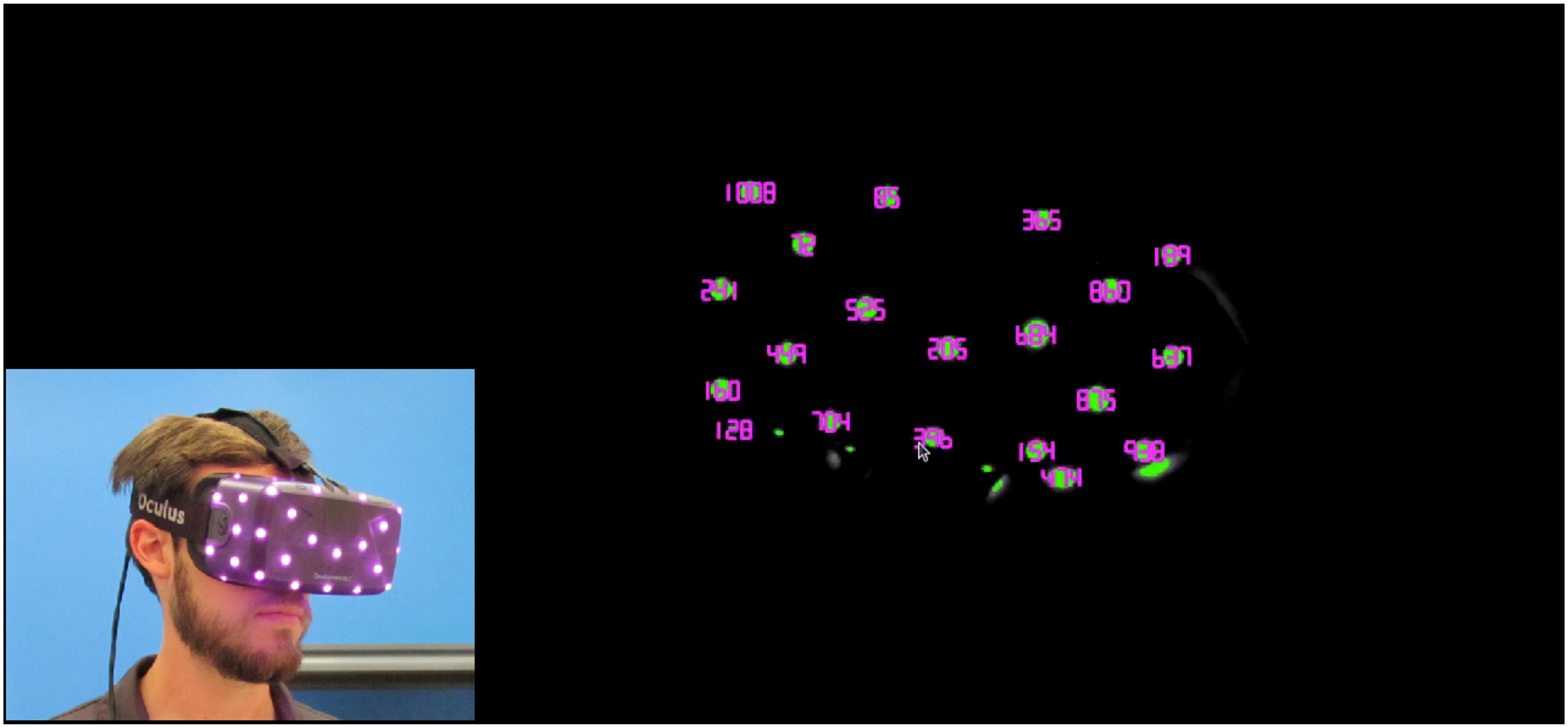
Active Optical Motion Capture

- Each LED marker emits unique blinking pattern (ID)
- Reduce marker ambiguities / unintended swapping
- Have some lag to acquire marker IDs

Phoenix Technology


Phase Space

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

6 DOF Head Pose Estimation

Head pose: 6 degrees of freedom (unknowns)

- 3D position and 3D rotation of headset (e.g. can represent as 4x4 matrix)

Inputs:

- Fixed: relative 3D position of markers on headset (e.g. can represent each marker offset as 4x4 matrix)
- Fixed: camera viewpoint (ignoring distortion, also a 4x4 projective mapping of 3D scene to 2D image)
- Each frame: 2D position of each headset marker in image

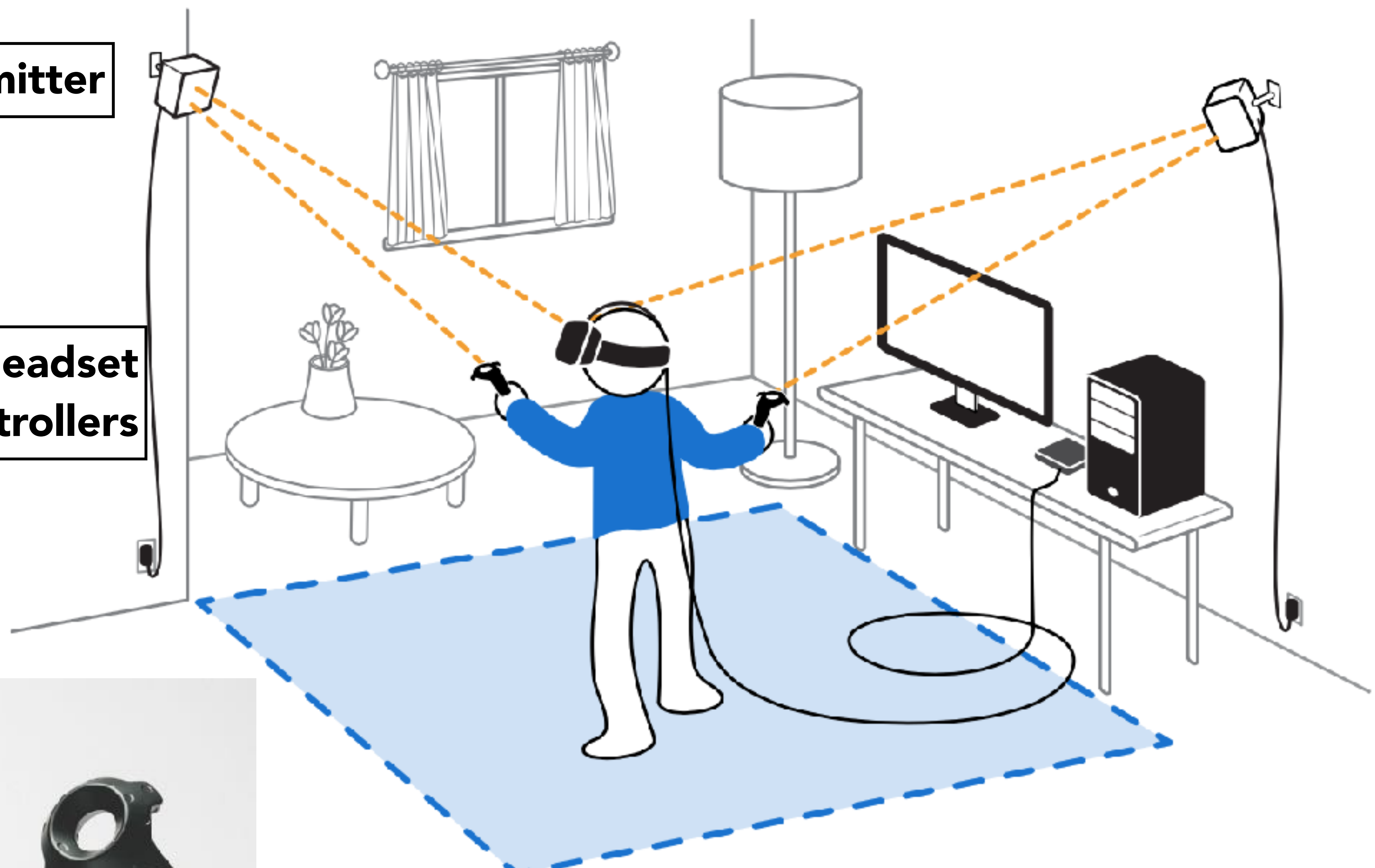
Pose calculation:

- Write down equations mapping each marker to image pixel location as a function of 6 degrees of freedom
- Solve for 6 degrees of freedom (e.g. least squares)

HTC Vive Tracking System ("Lighthouse")

Structured light transmitter

Photodiode arrays on headset
and hand-held controllers



To Be Continued

Acknowledgments

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