Università di Genova
DIBRIS DIPARTIMENTO
DI INFORMATICA, BIOINGEGNERIA,
ROBOTICA E INGEGNERIA DEI SISTEMI

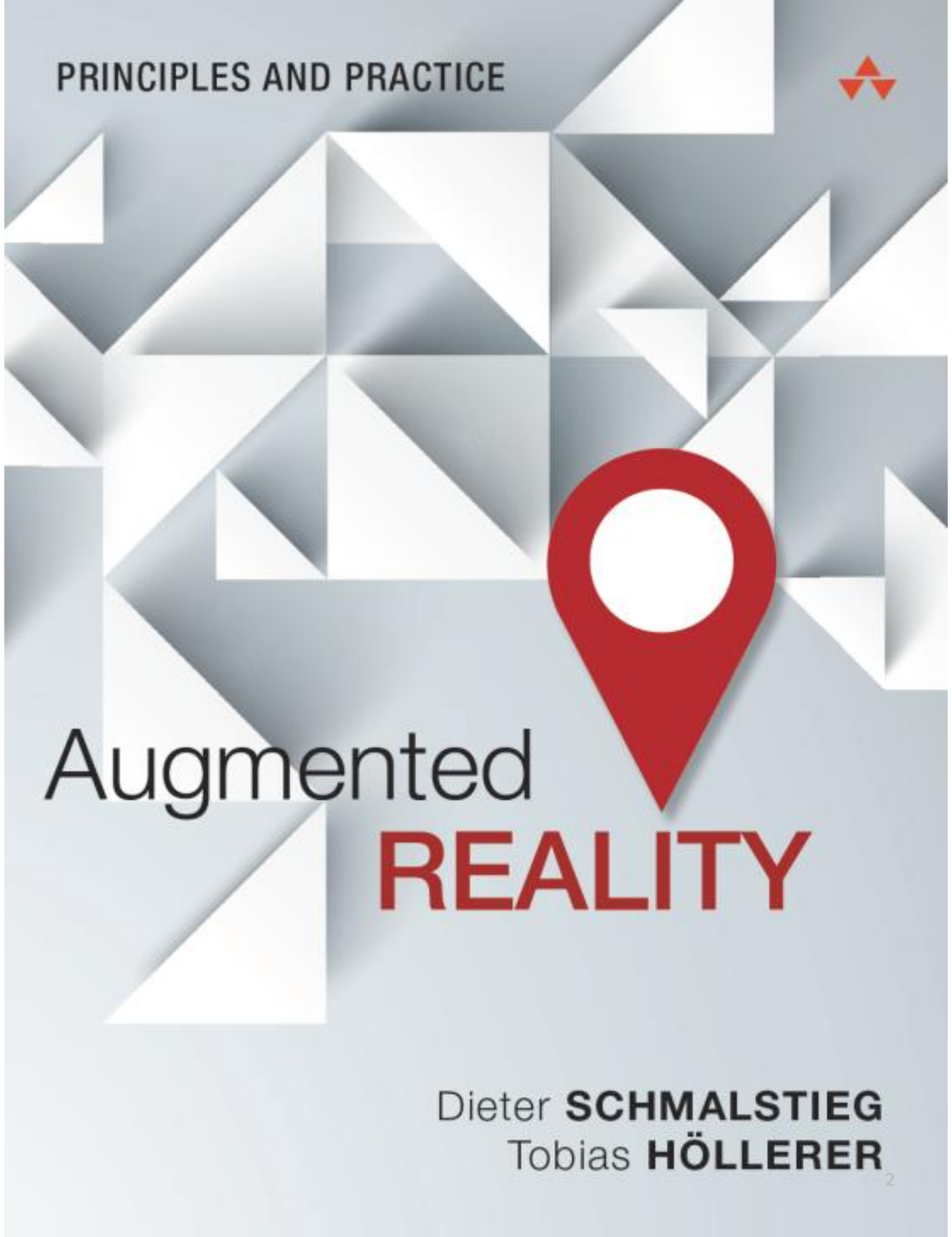
Augmented Reality

Lecture 2

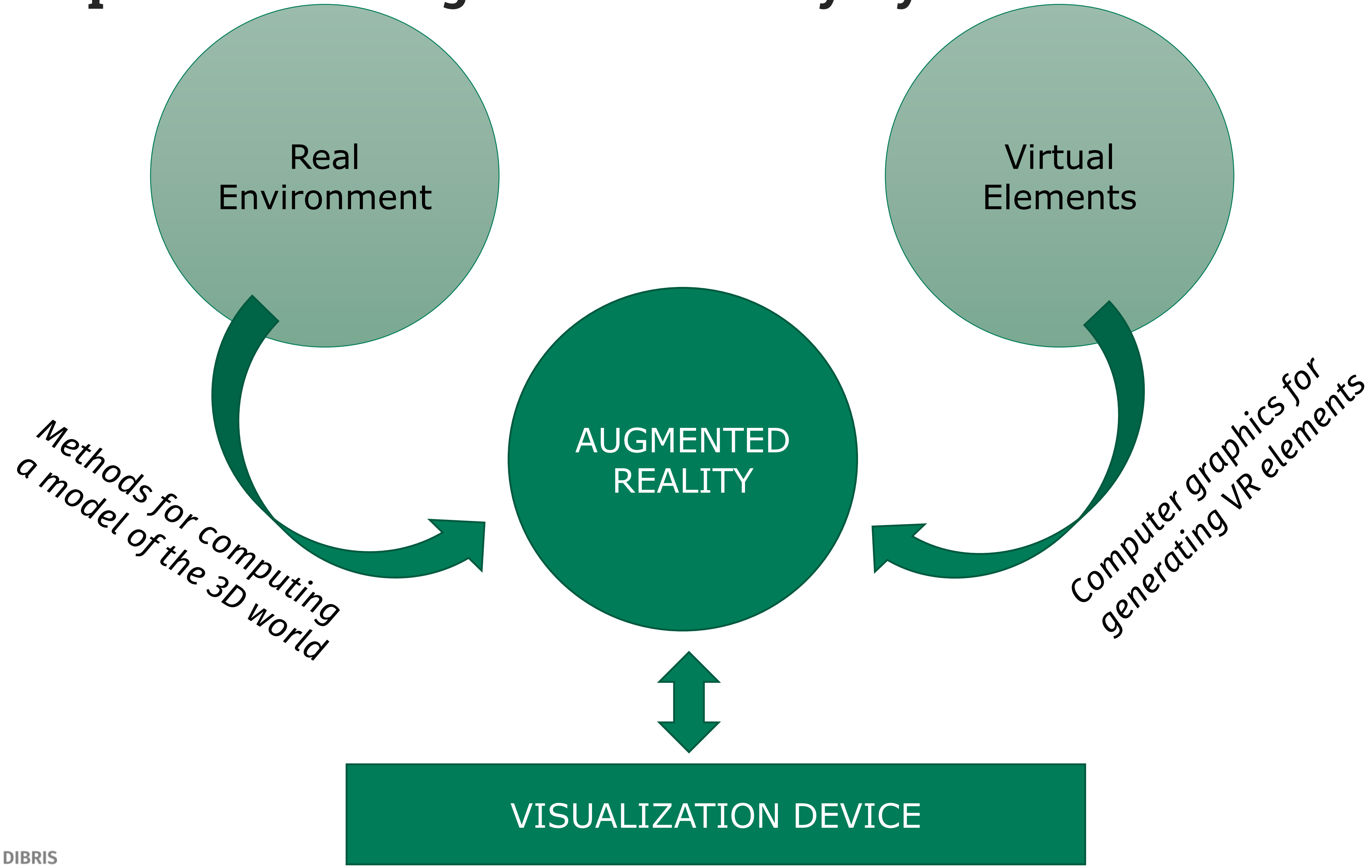
Manuela Chessa – manuela.chessa@unige.it
Fabio Solari – fabio.solari@unige.it

Credits

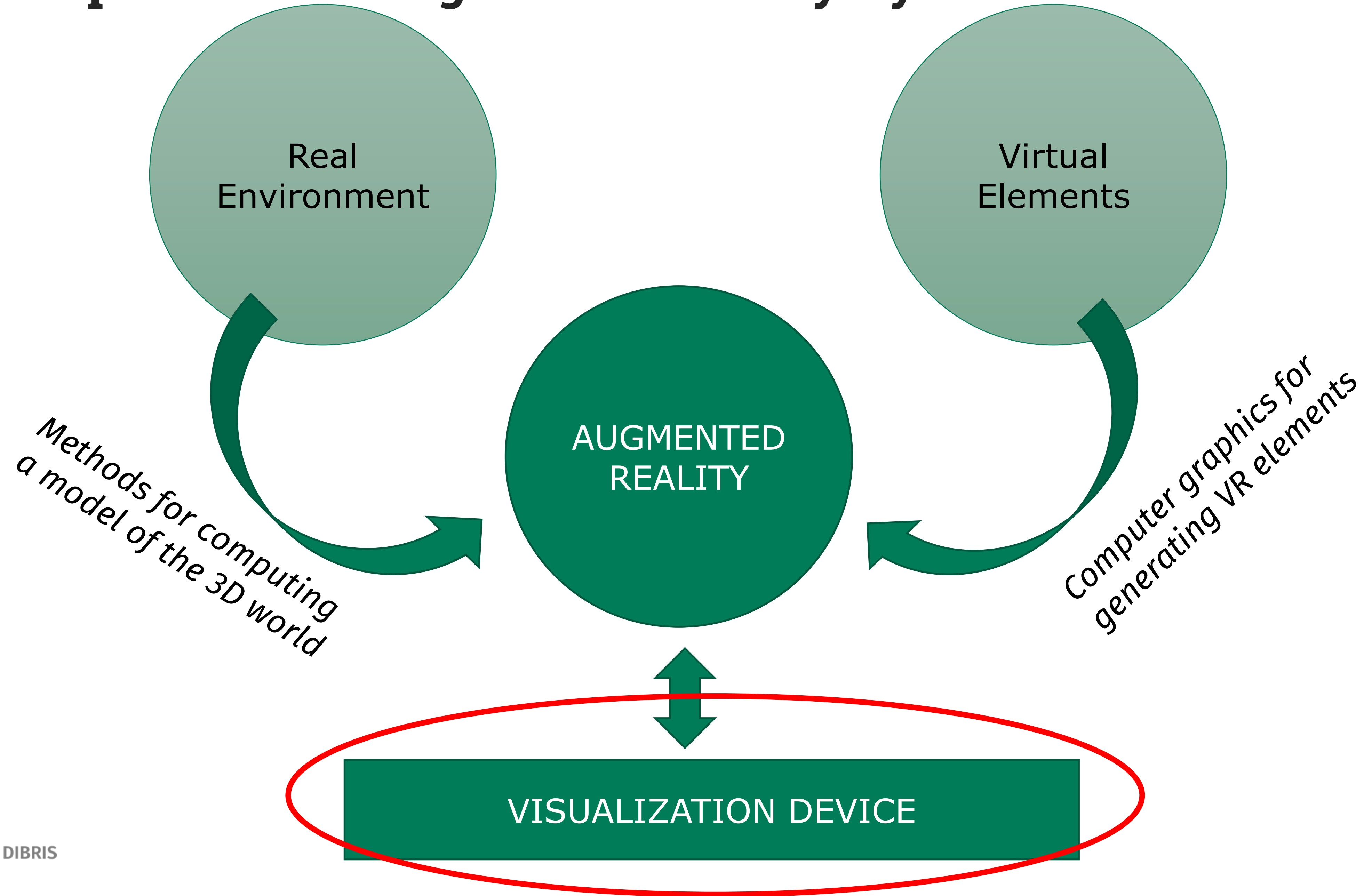
Augmented Reality: Principles and Practice (Usability) - Dieter Schmalstieg Tobias Hollerer



Description of an Augmented Reality System



Description of an Augmented Reality System



Visualization Devices

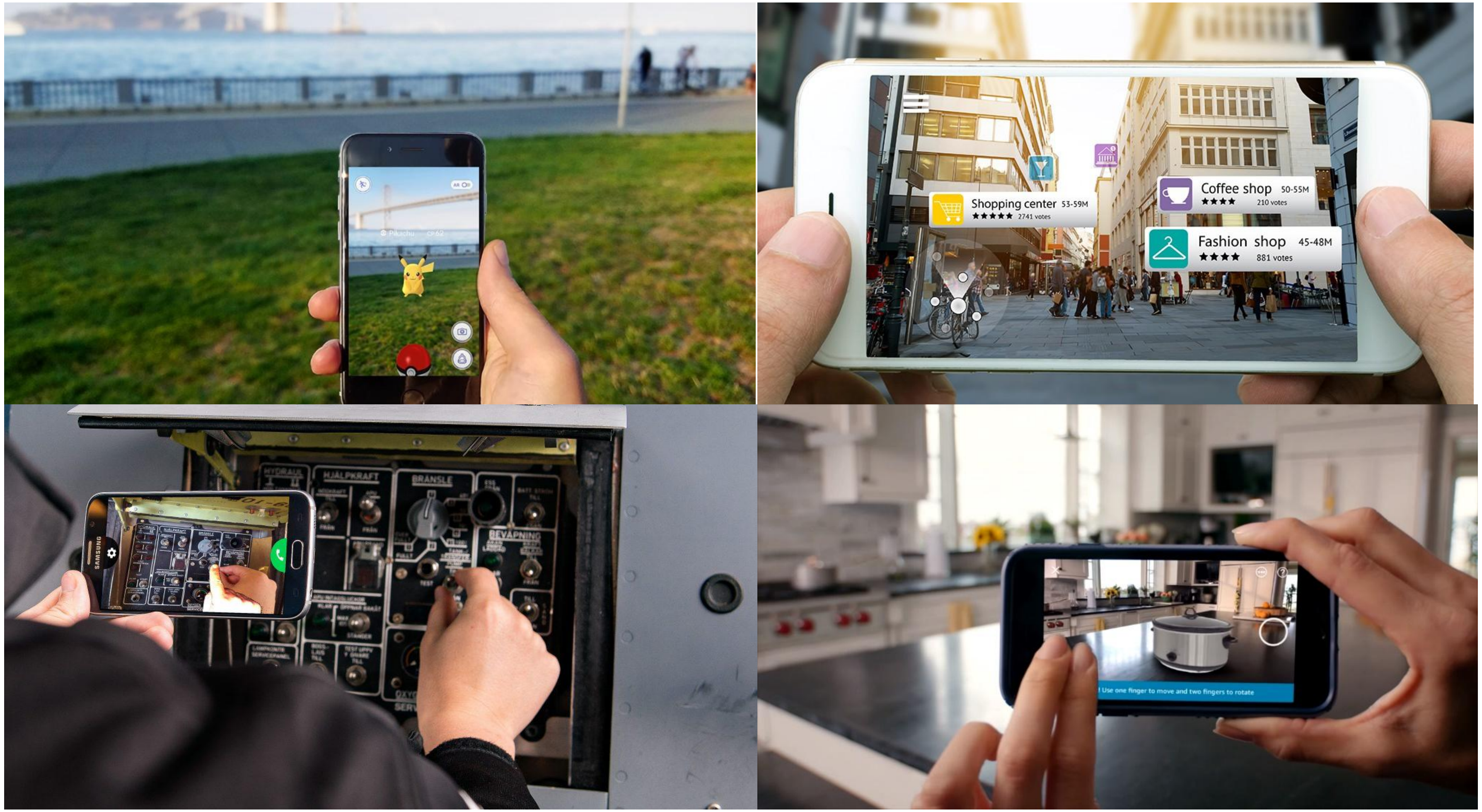
AR devices: head up displays

Devices that are worn by the user

AR devices: smart glasses

Similar to head-up, usually simpler and lighter

AR devices: handled/smartphone based



AR devices: holographic displays

No device required by the user

AR Displays

AR Displays

Requirement: the real environment and the virtual environment need to be combined.

1. When this combination of real and virtual content happens via a lens through which the user is viewing the environment, the result is described as a ***see-through*** display.
   - ***optical see-through*** *display*
   - ***video see-through*** *display*

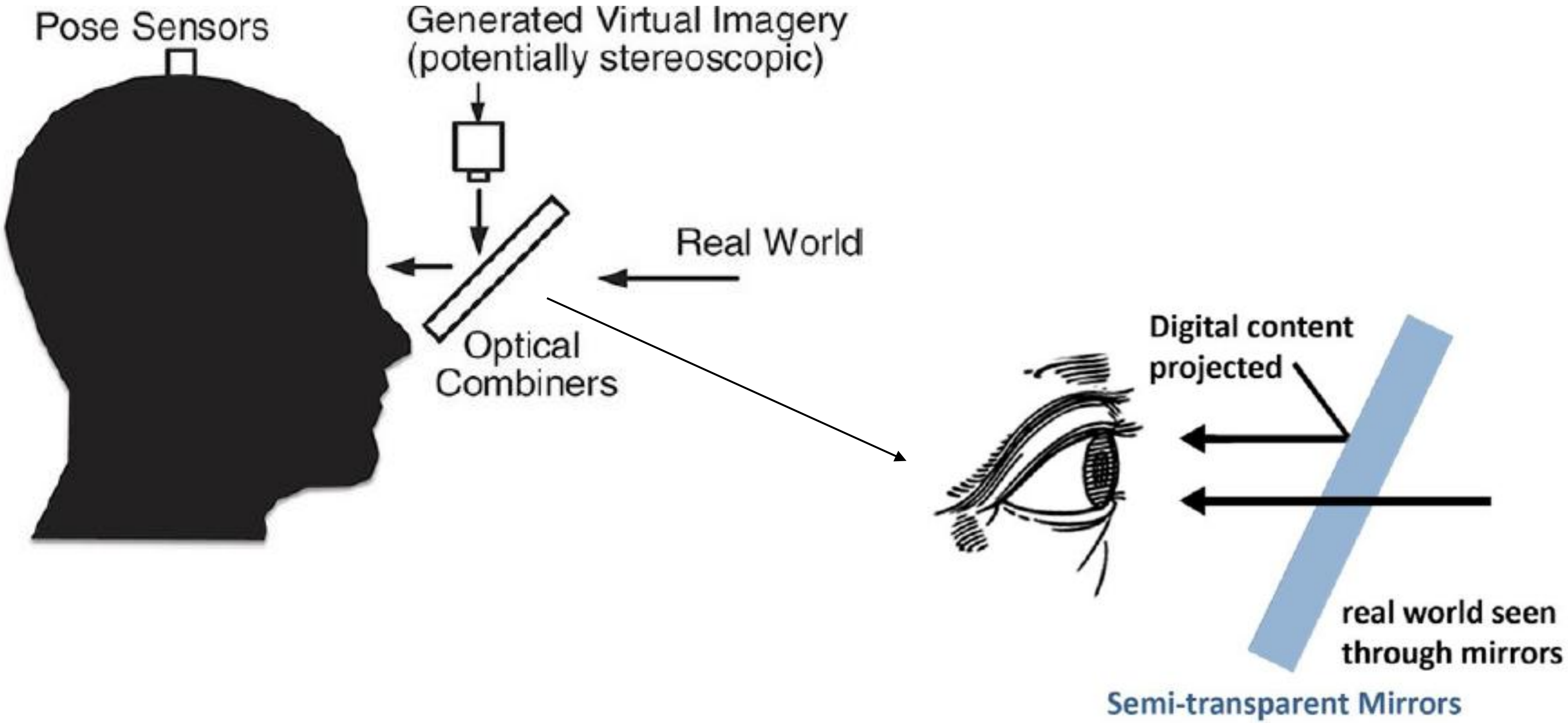
2. If the ***augmentations*** *are* ***projected*** *onto* ***actual physical geometry*** (be it dummy placeholder objects or natural parts of the real world), the technology is described as ***spatial AR***, *projection-based AR*, or *spatial projection*.

Optical see-through (OST) displays

- They rely on an **optical element** that is partially **transmissive** and partially **reflective** to achieve the combination of virtual and real.
- A half-silvered mirror is a simple example of it.

- The mirror lets a sufficient amount of light from the real-world pass through so that the **real world can be viewed directly**.
- A computer-generated display showing virtual images is placed overhead or to the side of the mirror, so that **the virtual images are reflected in the mirror** and overlaid on the real image.

Optical see-through (OST) displays



Optical see-through (OST) displays: examples

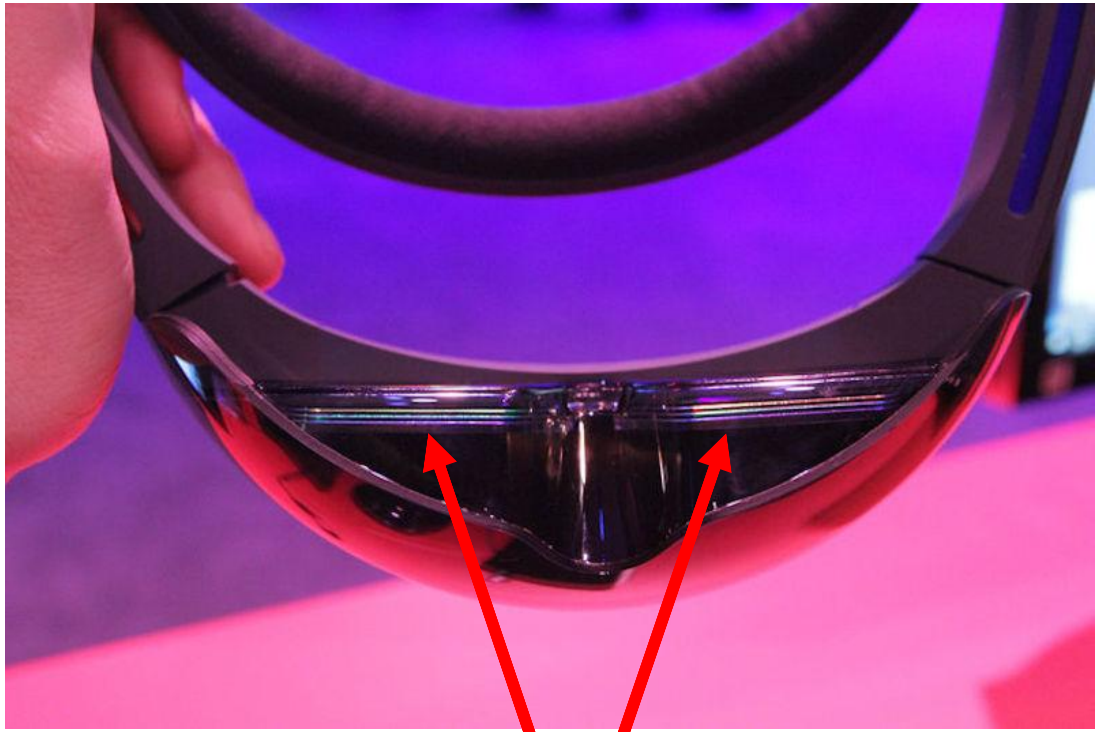
Microsoft HoloLens

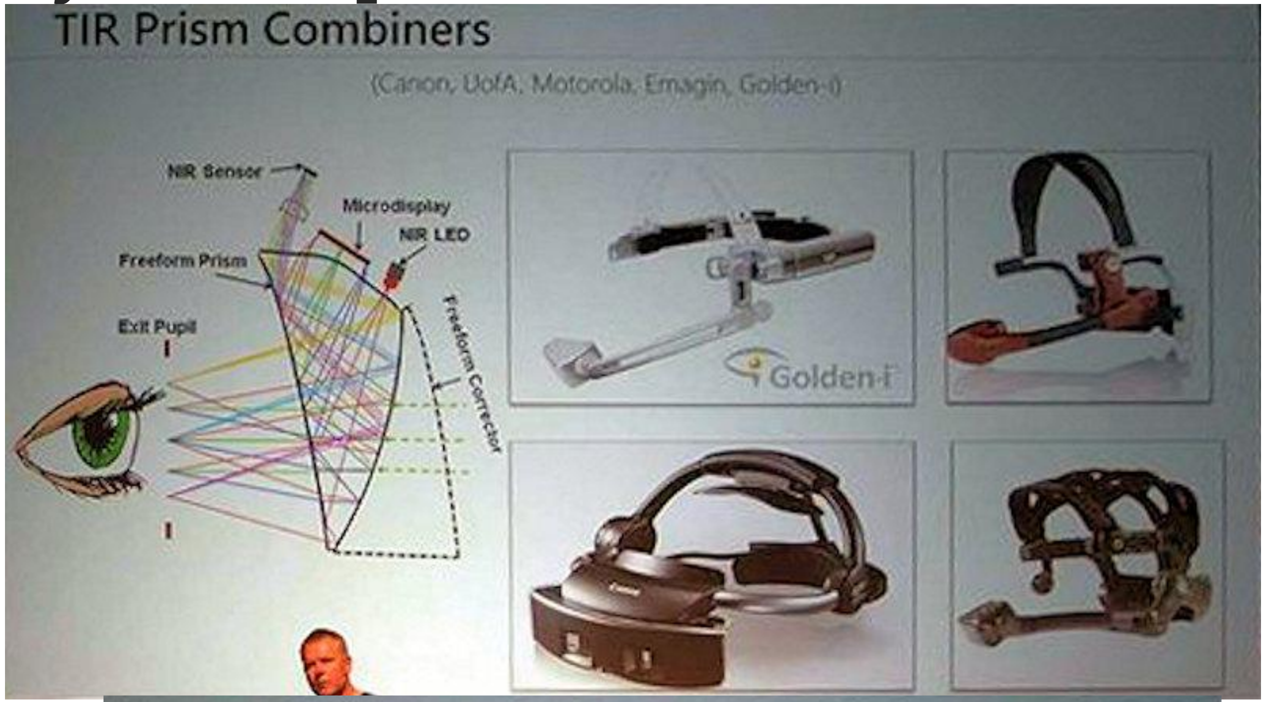
Epson Moverio

Meta Vision

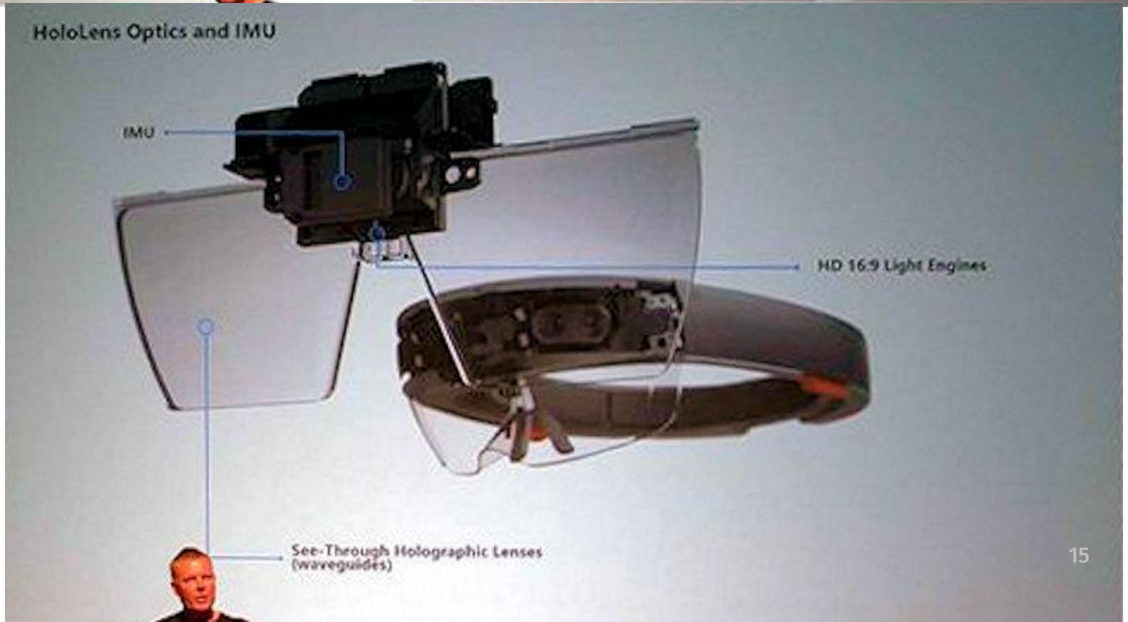
Vusix

Optical see-through (OST) displays: example the HoloLens





DISPLAYS and semitransparent MIRROR



Optical see-through (OST) displays: examples

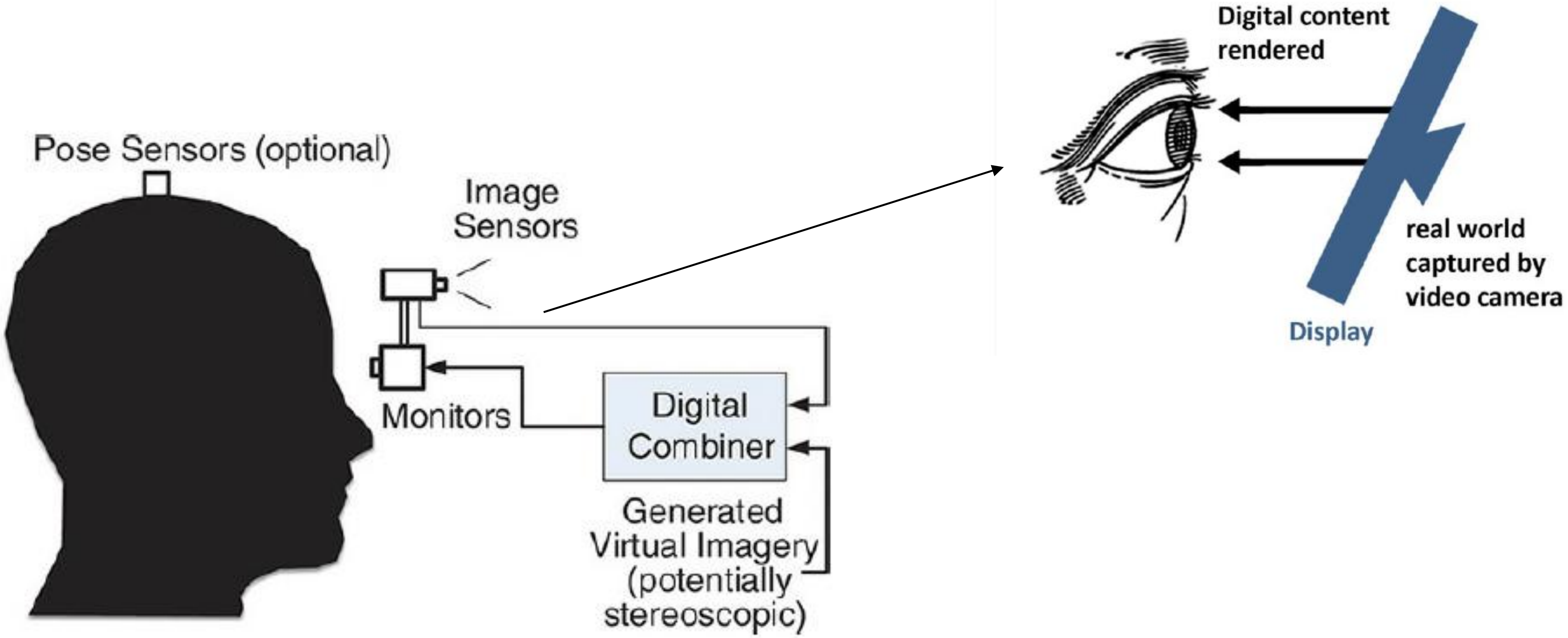
Video see-through (VST) displays

- They achieve the **combination of virtual and real <u>electronically</u>**.

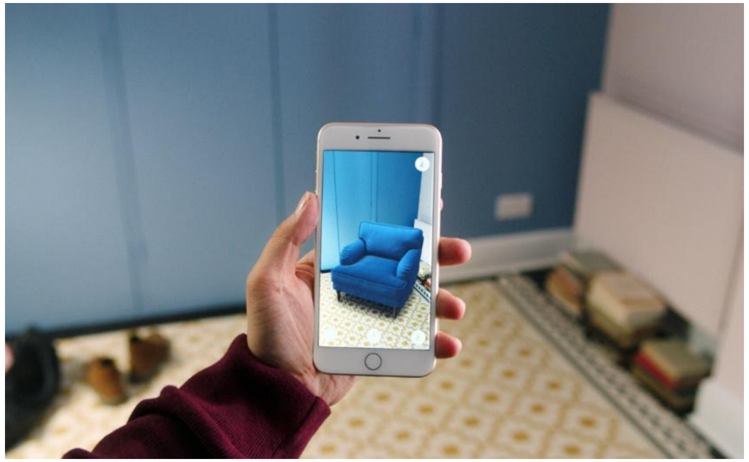
- A **digital video image of the real world** is captured through a video camera and transferred to the graphics processor.
- The graphics processor **combines the video image with the computer-generated images**, often by simply copying the video image into the frame buffer as a background image, with the computer-generated elements drawn on top.

- The **combined image** is then presented using a **conventional viewing display**.

Video see-through (VST) displays



Video see-through (VST) displays: examples



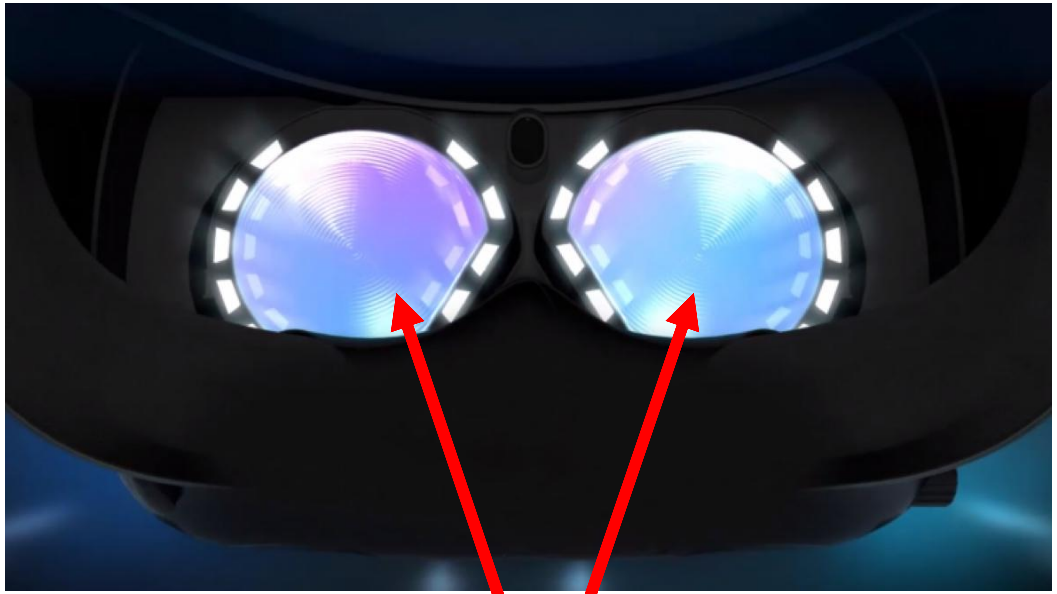
Video see-through (VST) displays: examples (HMD)

Meta Quest 3

Apple Vision Pro

Video see-through (VST) displays: examples (HMD)

HTC Vive



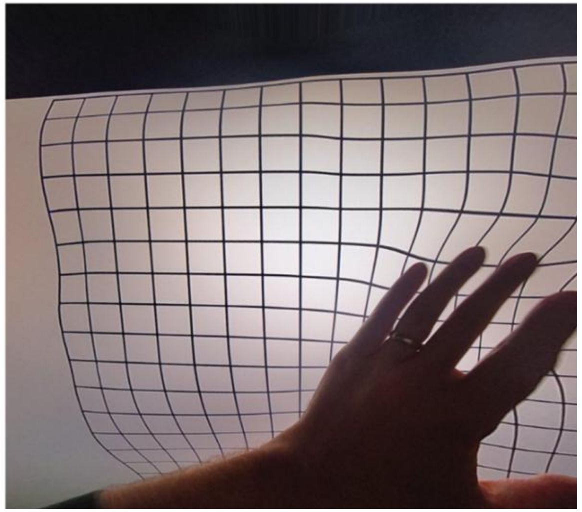
CAMERAS

DISPLAYS

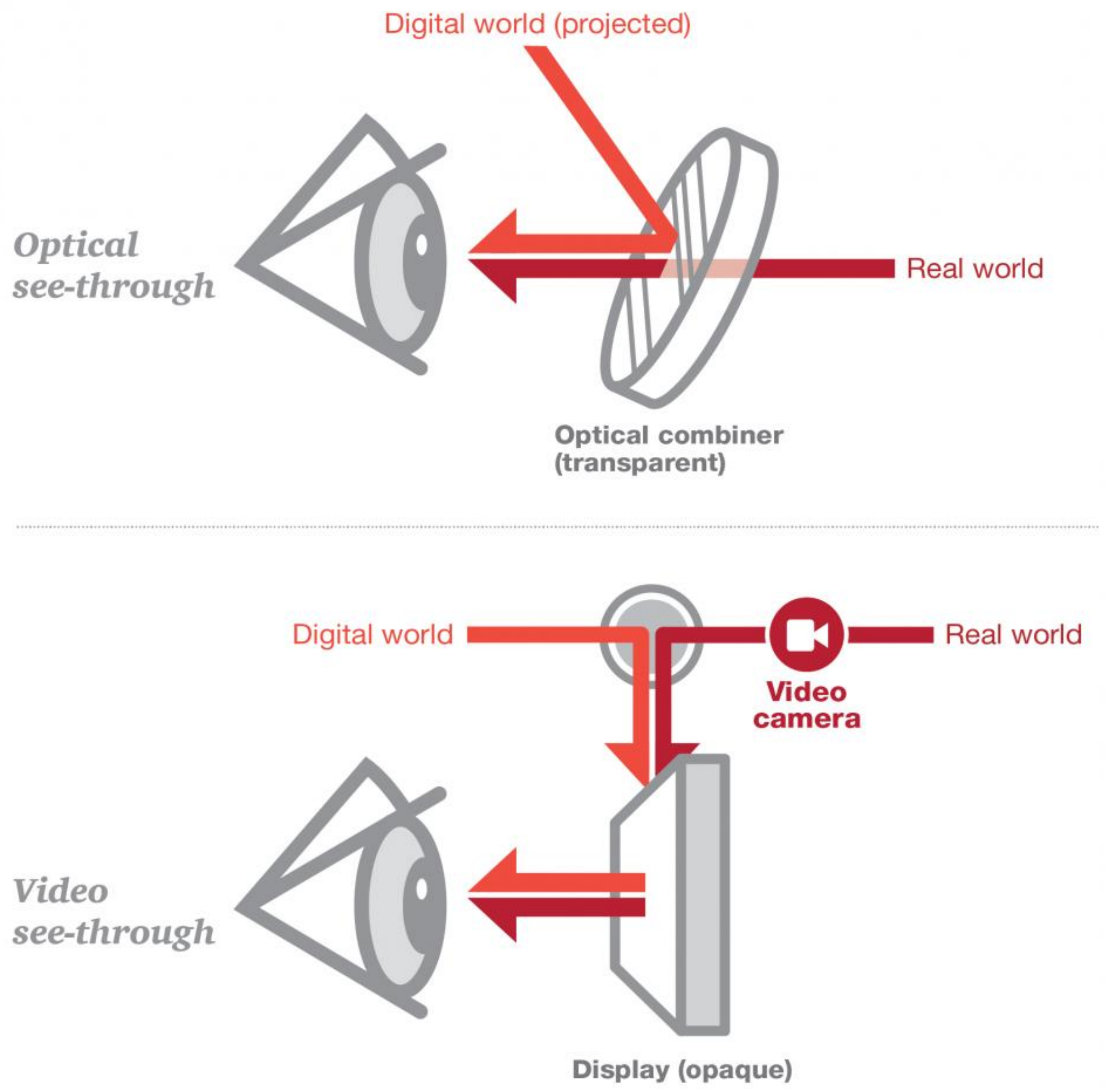
Video see-through (VST) displays: examples

Video see-through (VST) displays: examples (and issues)

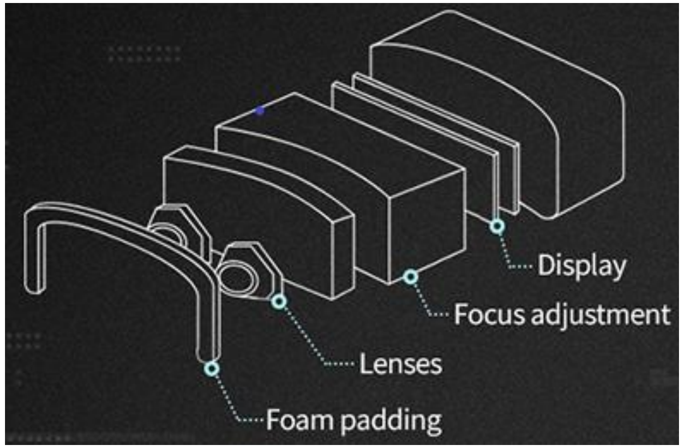
Meta Quest 3

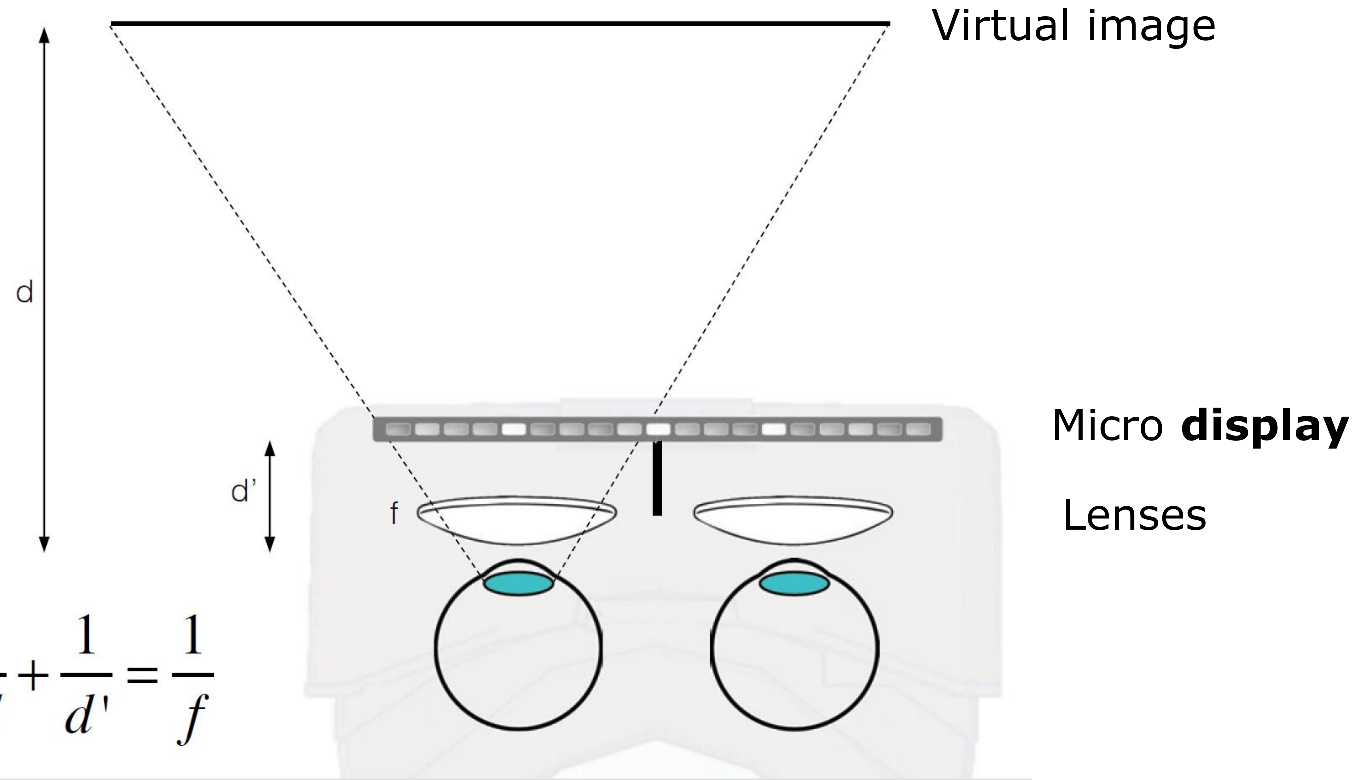


OST and VST displays



Virtual reality and VST displays





$$\frac{1}{d}+\frac{1}{d'}=\frac{1}{f}$$

Virtual reality and VST displays

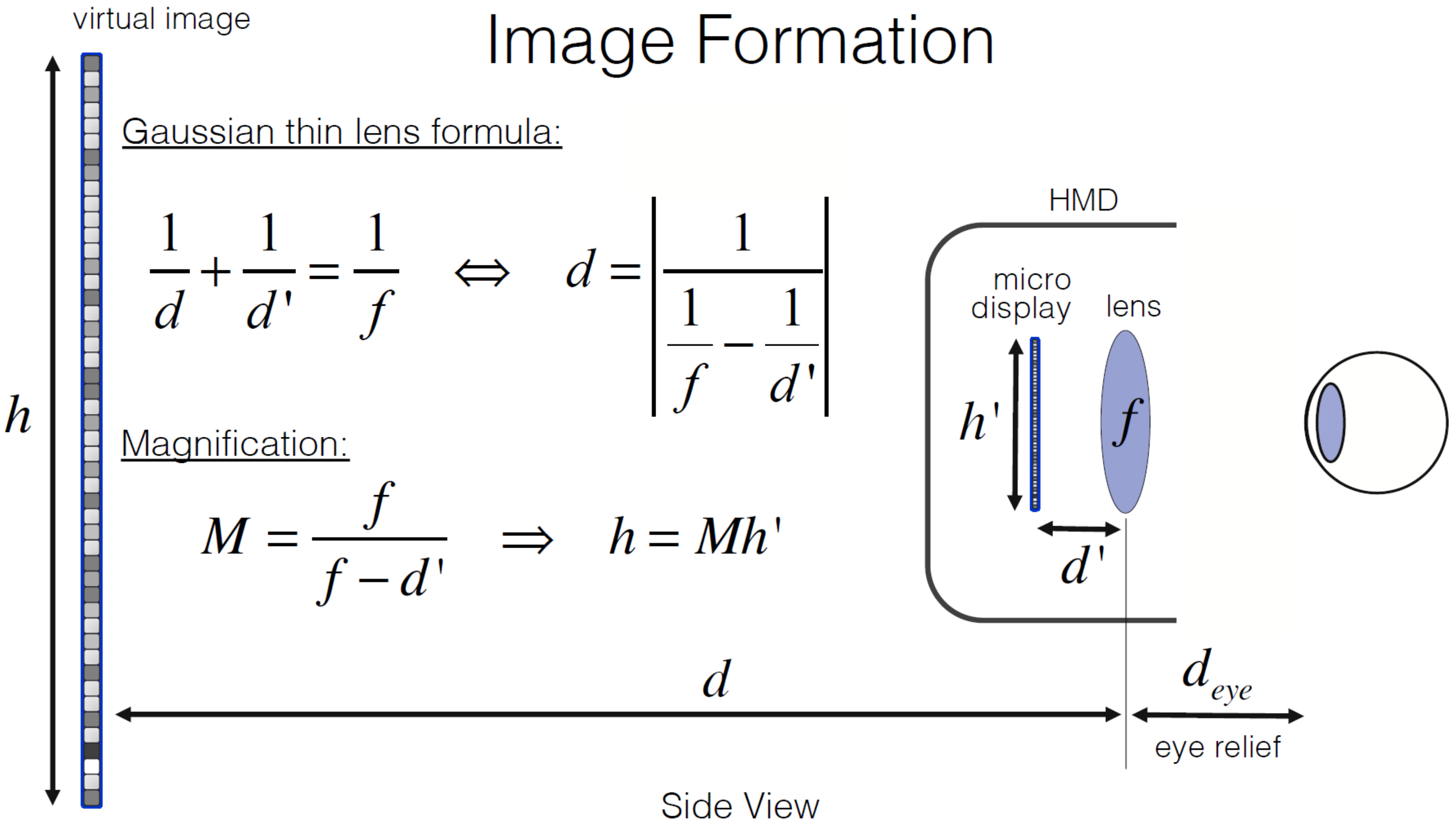
Image Formation

Gaussian thin lens formula:

$$\frac{1}{d}+\frac{1}{d'}=\frac{1}{f} \quad \Leftrightarrow \quad d=\left|\frac{1}{\frac{1}{f}-\frac{1}{d'}}\right|$$

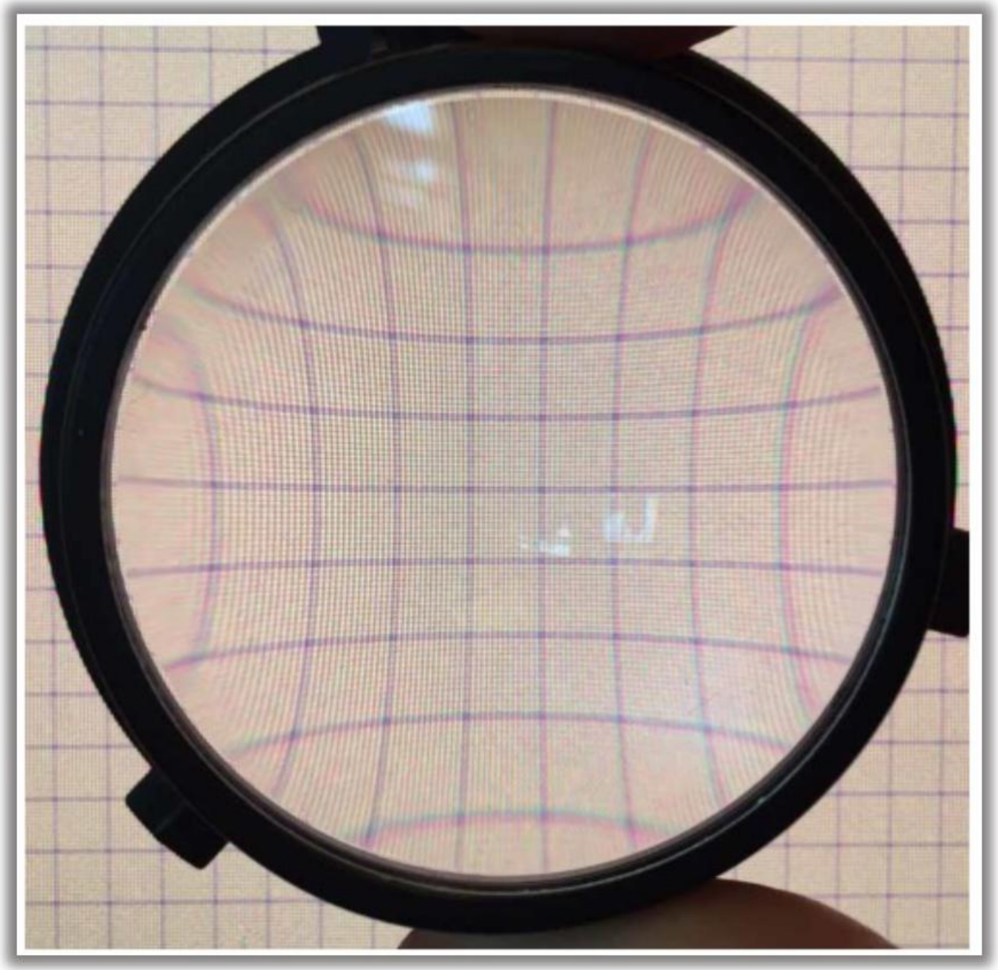
Magnification:

$$M=\frac{f}{f-d'} \quad \Rightarrow \quad h=Mh'$$



Side View

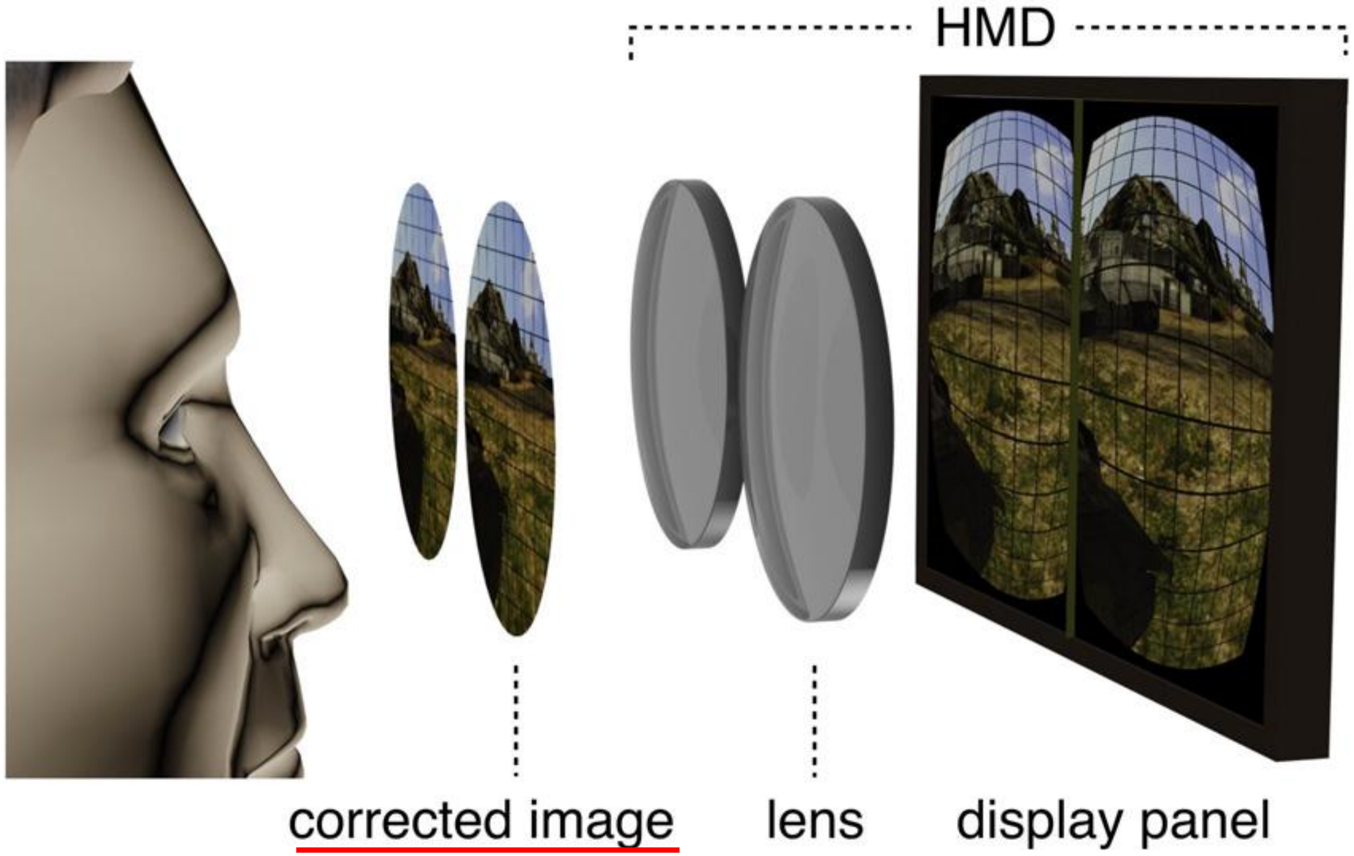
Virtual reality and VST displays



Lens distortion

- grid seen through HMD lens
- lateral (xy) distortion of the image
- chromatic aberrations: distortion is wavelength dependent!

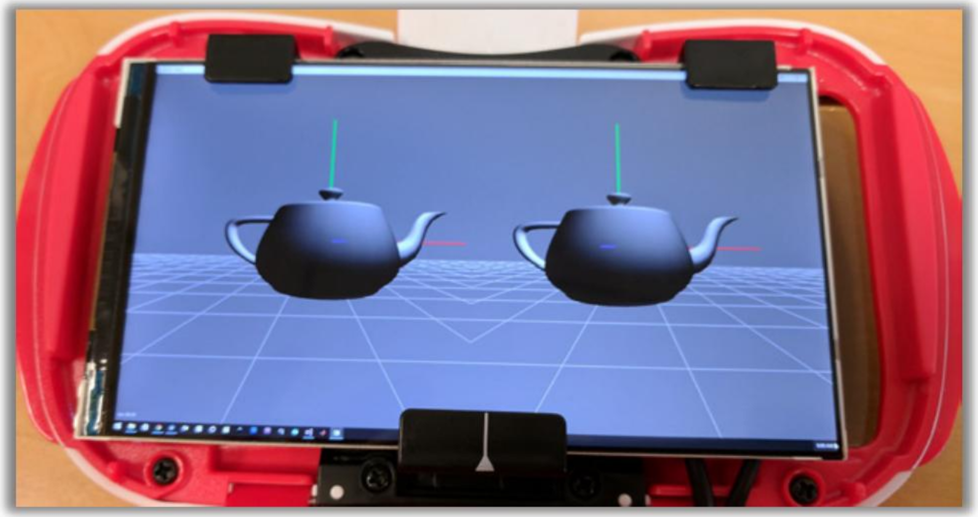
Virtual reality and VST displays



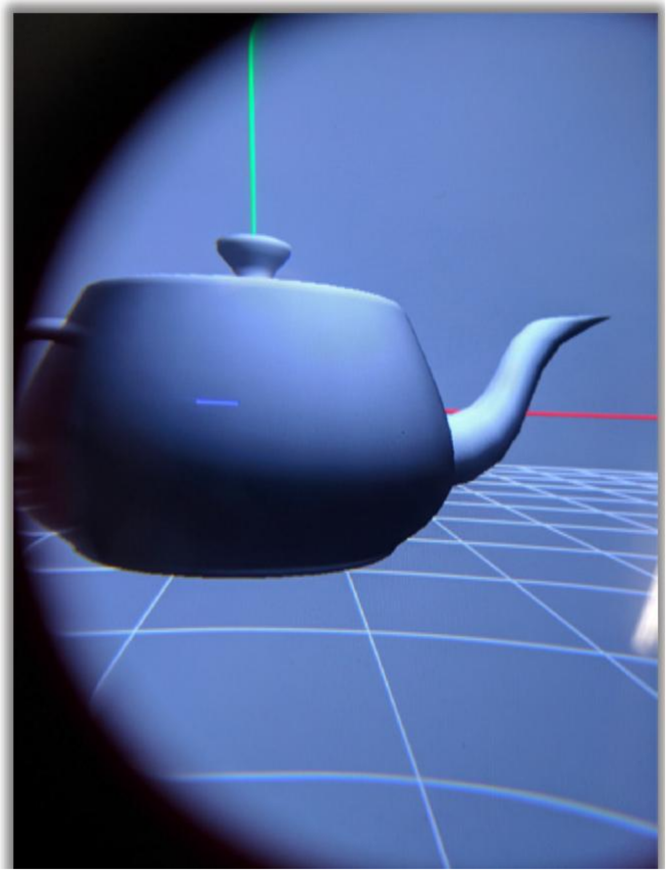
(perceived by the user)

(software pre-corrected image)

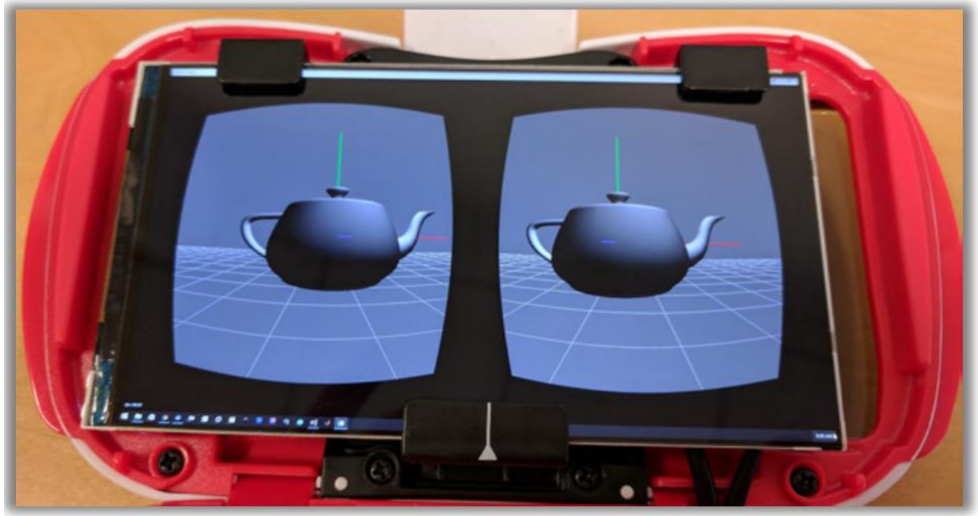
Virtual reality and VST displays



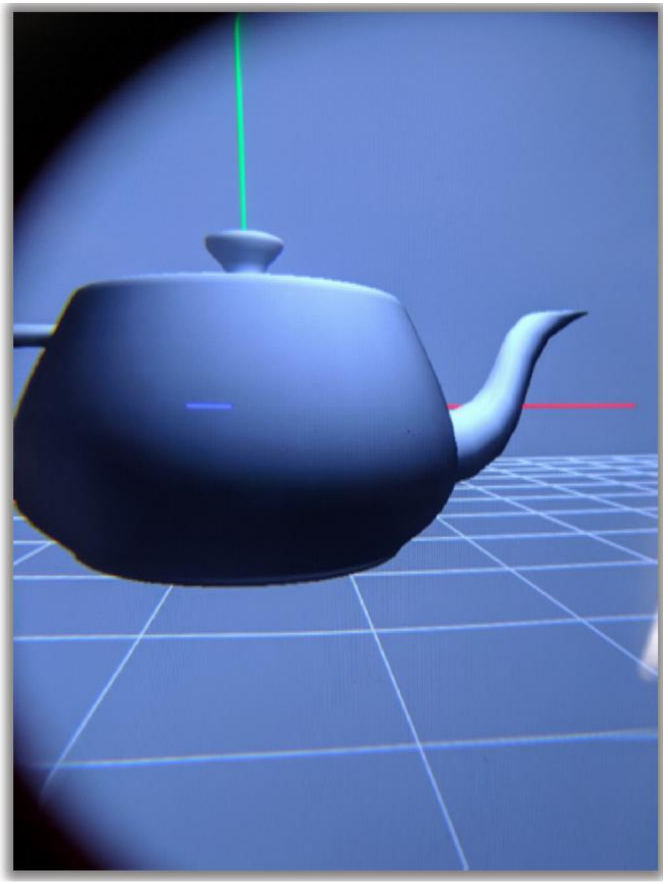
stereo rendering without lens distortion correction



Virtual reality and VST displays



stereo rendering with lens
distortion correction

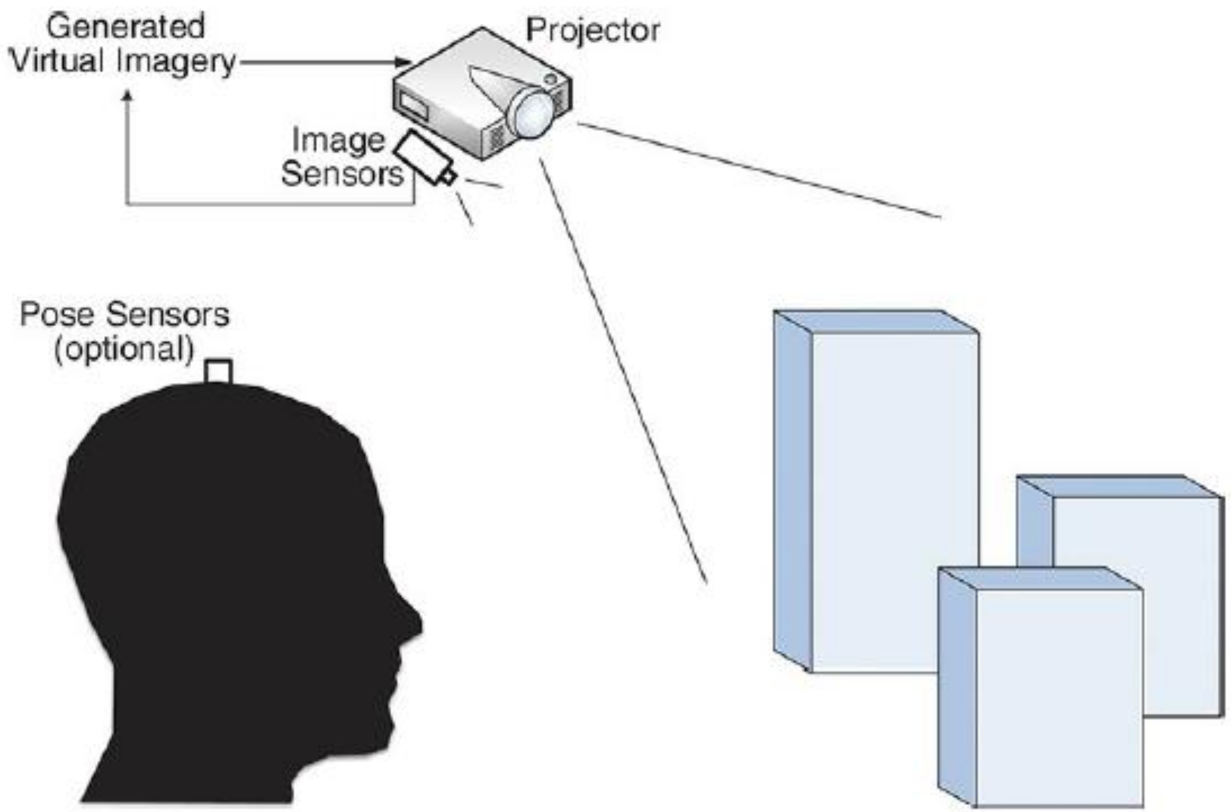


Spatial Projection (SAR)

- The **virtual part** of the AR display is generated by a **light projector**.
- Instead of using a special screen the **virtual image is projected directly on real-world objects**.

- This is also a form of **optical combination**, but here we do not need a separate optical combiner, and no electronic screen is involved.
- This **display paradigm** is an example of a ***volumetric 3D display*** *for which the points of light defining perceivable objects are physically distributed throughout a 3D volume.*

Spatial Projection (SAR)

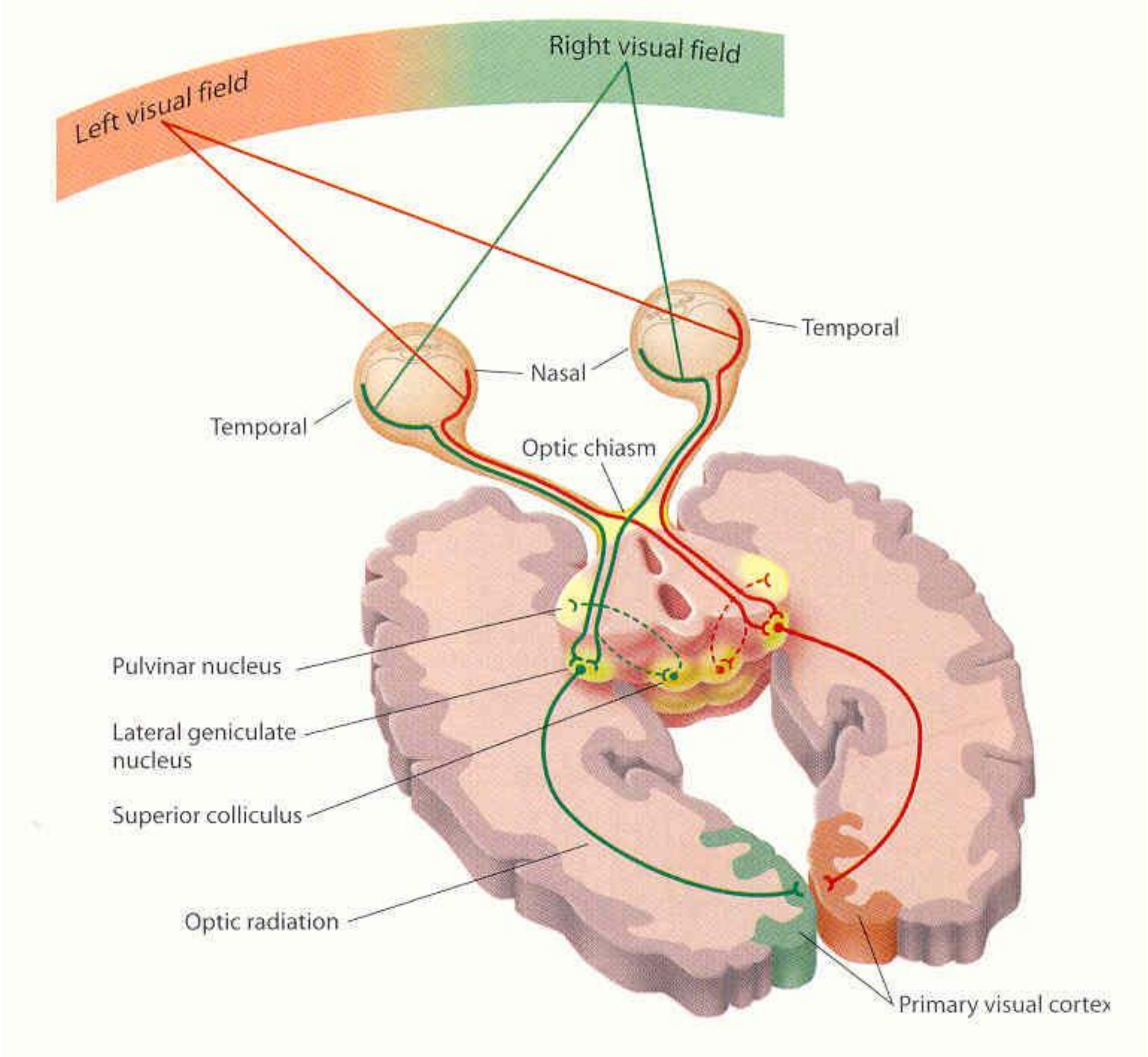


SAR: examples

plus sound effects, lights, etc…

AR devices and the human visual system

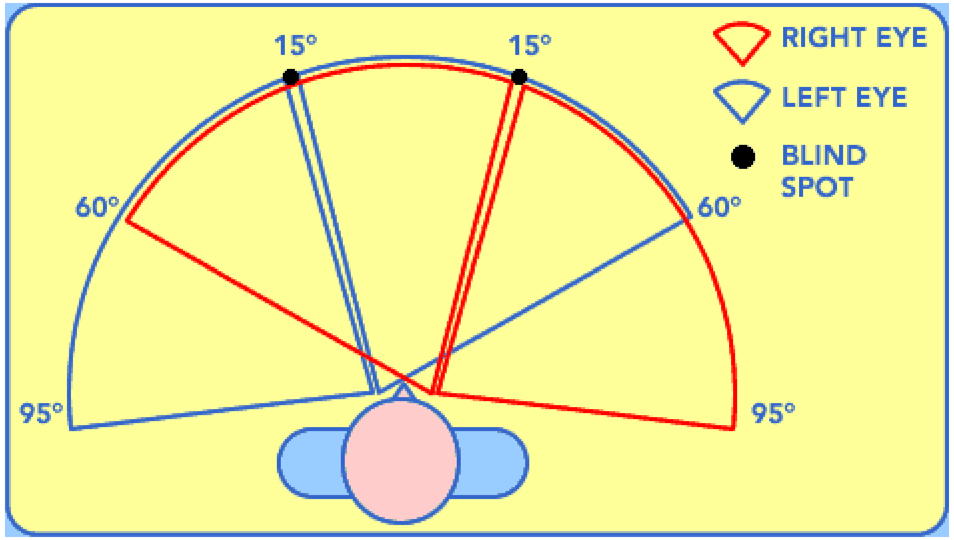
Human visual system

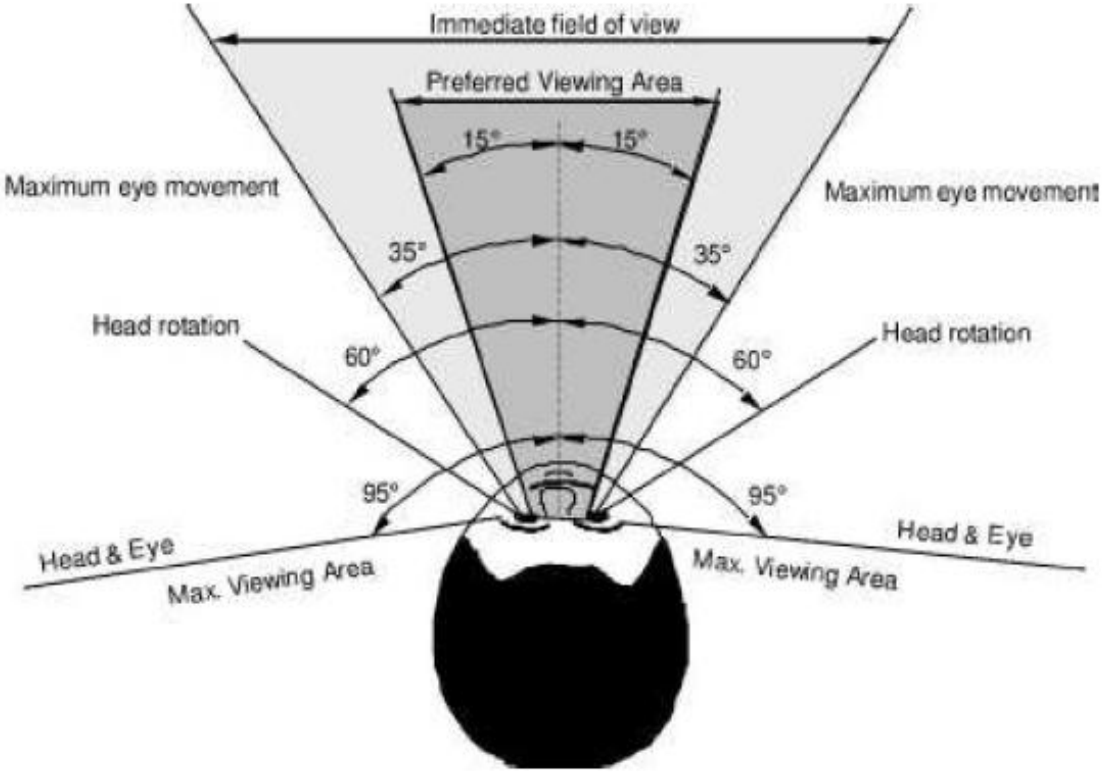


Human visual system

The **human field of view (FoV)**, from both eyes combined, spans usually 200–220° horizontally

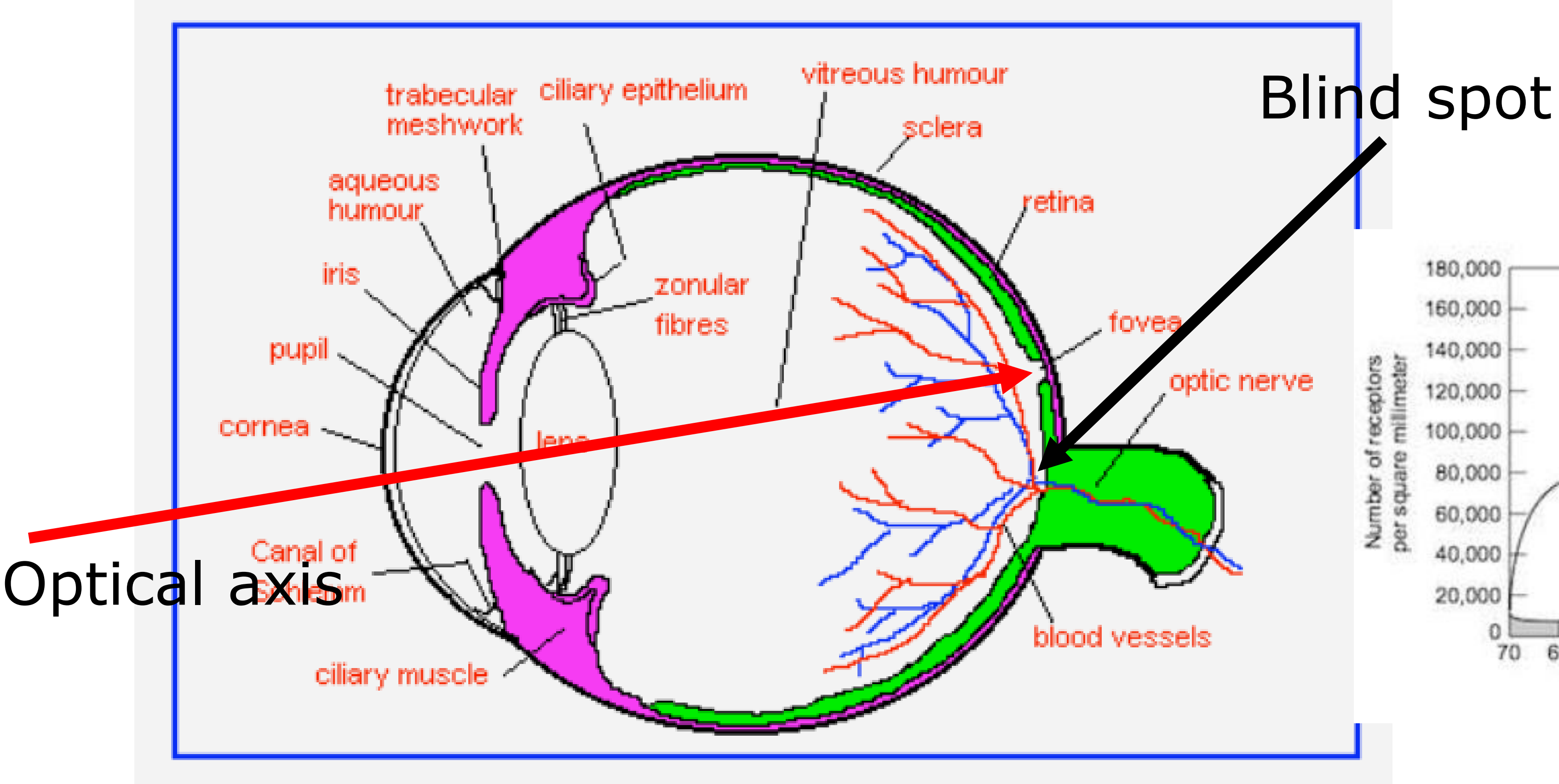
Human field of view

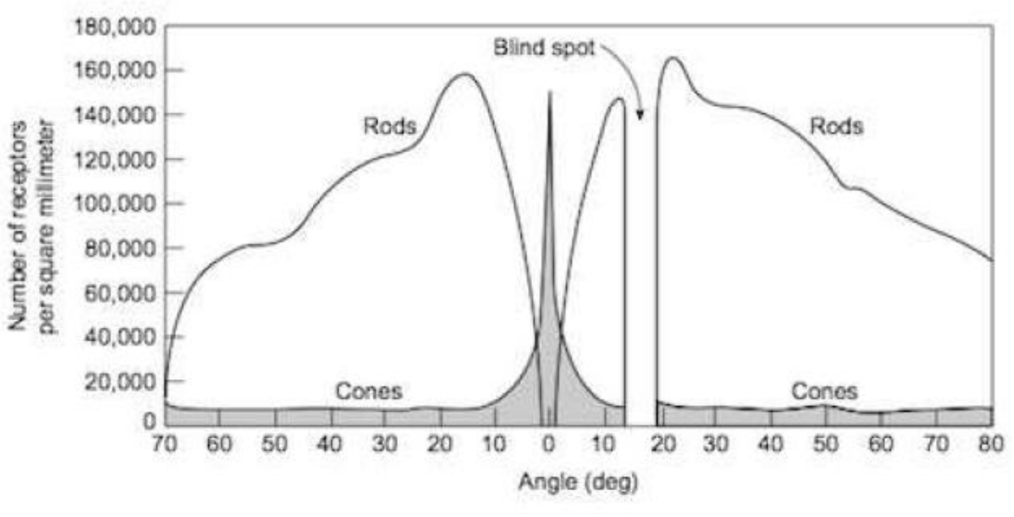




Human visual system

- The **fovea** (i.e., the area of **best visual acuity**) covers only 1–2°, with acuity peaking in the center 0.5–1°.
- Outside the fovea, the **visual acuity falls off quickly** with increasing viewing angle.
- Humans compensate for this effect **by moving their eyes**: active vision





AR devices and Human visual system

- The **fovea** (i.e., the area of **best visual acuity**) covers only 1–2°, with acuity peaking in the center 0.5–1°.
- Outside the fovea, the **visual acuity falls off quickly** with increasing viewing angle.
- Humans compensate for this effect **by moving their eyes**: active vision

The **perceived egocentric view** of the surrounding environment

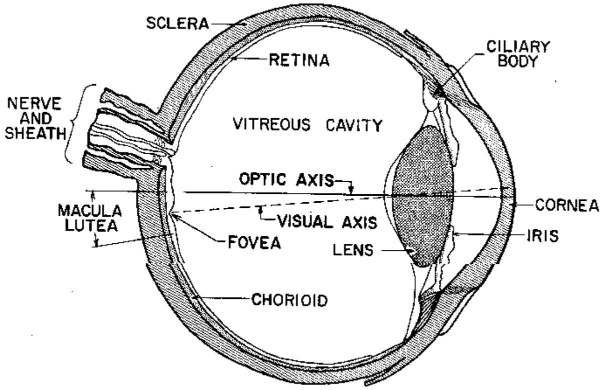
The **standard device view** of the surrounding environment

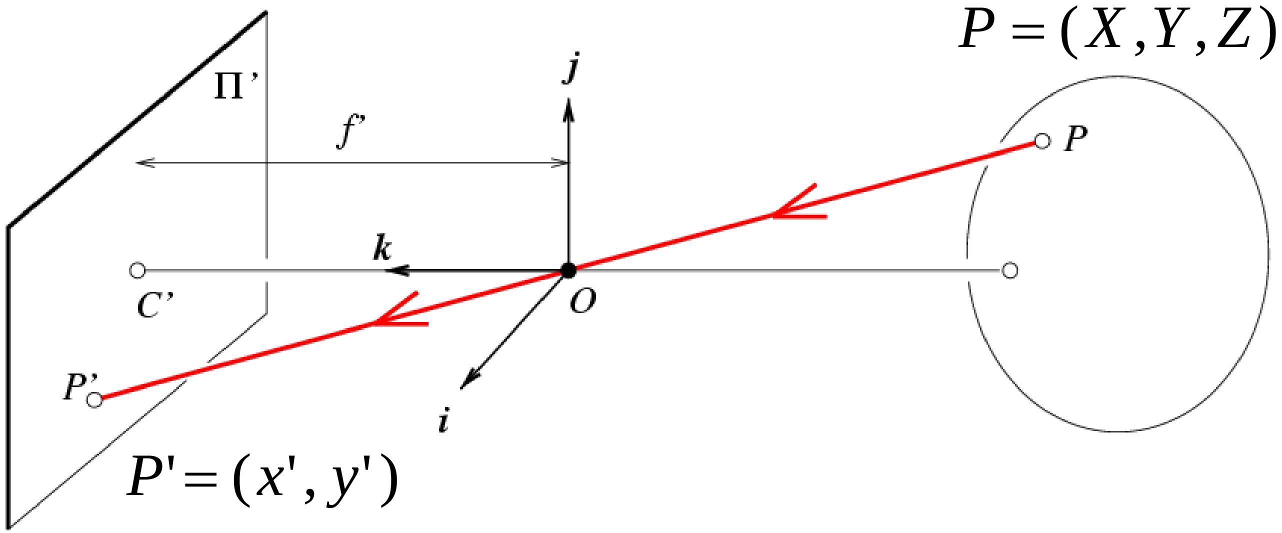
High-quality AR, therefore, requires a viewing **device** that can present **sufficient resolution** in the area of high acuity and a **sufficiently large field of view**.

Human visual system

Human perception of the environments: **perspective** and **distance**.

Pinhole Perspective Equation





$$\begin{cases} x' = f' \dfrac{X}{Z} \\ y' = f' \dfrac{Y}{Z} \end{cases}$$

Human visual system

– By adjusting the **pupil diameter**, humans can control the amount of light that enters the eye.

– This allows us **to accommodate a dynamic range** (the ratio of maximum and minimum perceivable light intensities) of up to 10^10, covering viewing conditions from dim starlight to extremely bright sunlight.

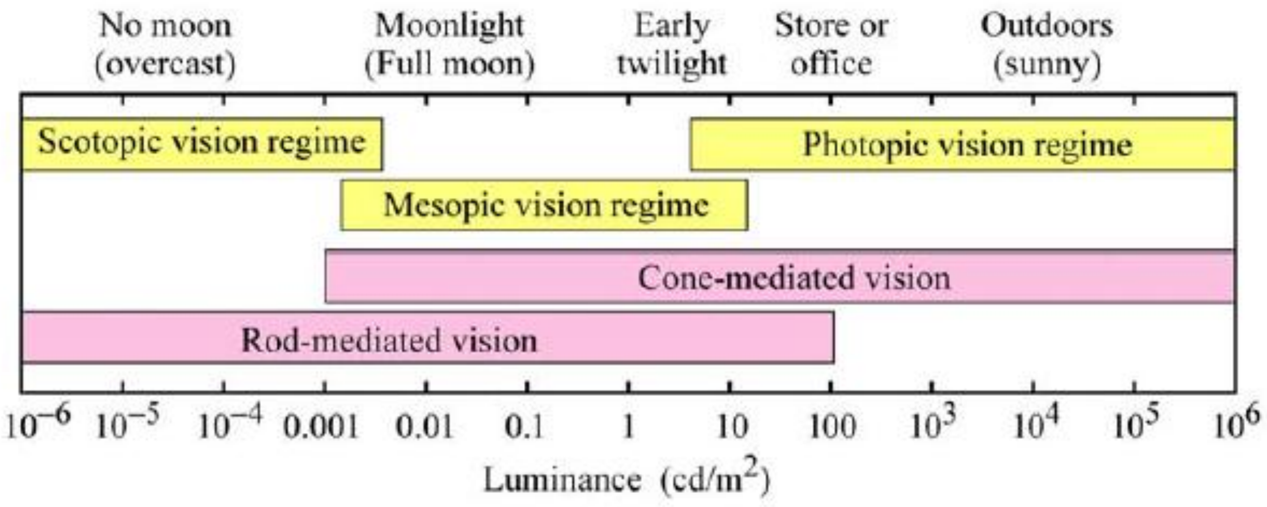
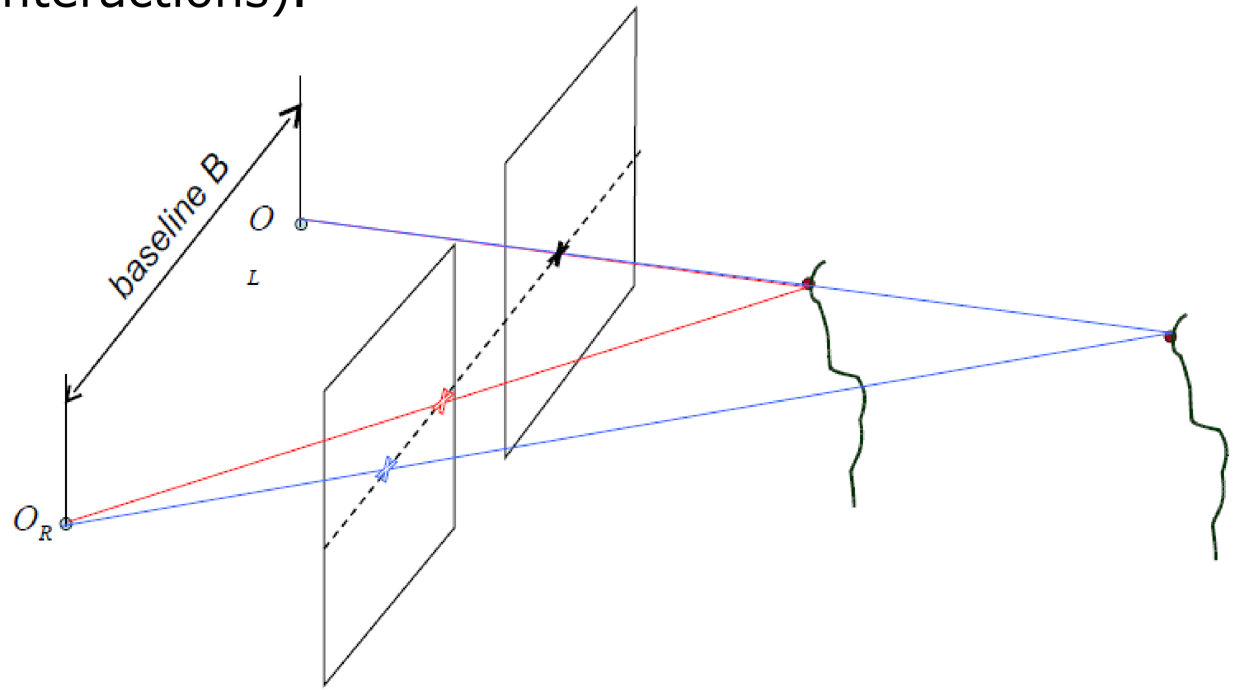


Fig. 16.2. Approximate ranges of vision regimes and receptor regimes (after Osram Sylvania, 2000).

Thus, a truly versatile AR display will need to be able to adapt to a wide range of viewing conditions.

Human visual system

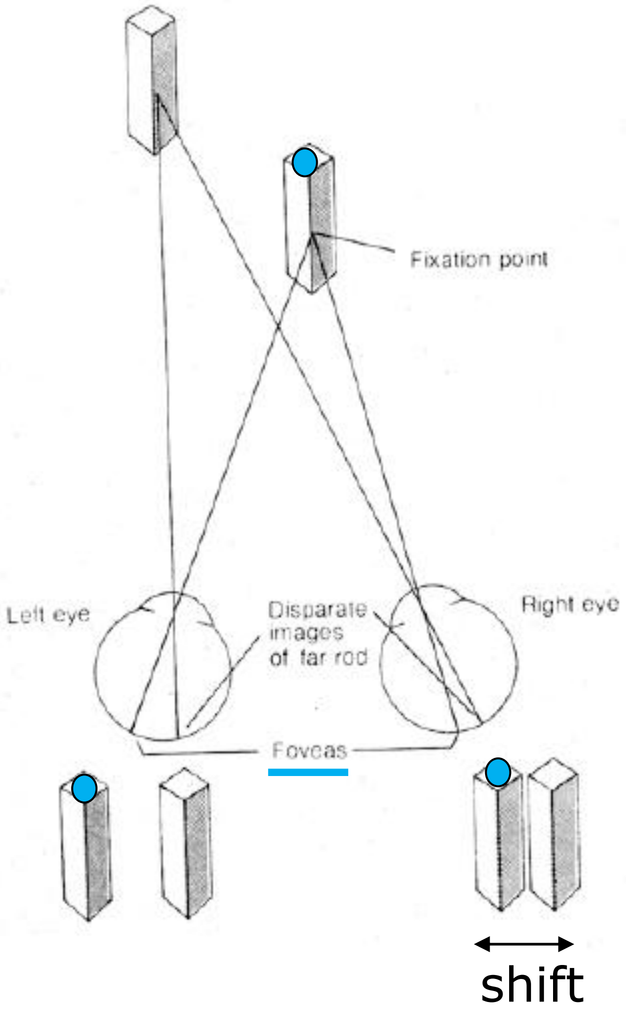
The use of **two eyes** means that humans are able to perceive **binocular depth** cues for 3D vision (e.g. useful for hand interactions).



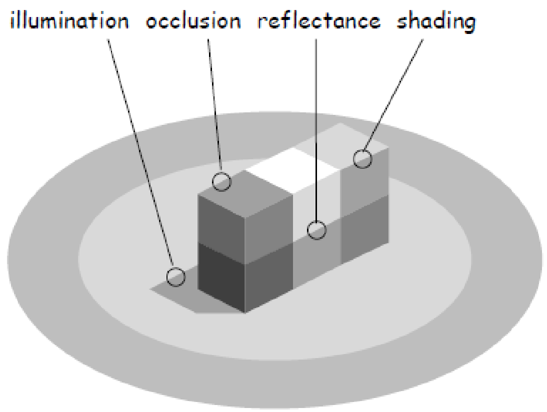
Pinhole Perspective Equation

$$\begin{cases} x' = f'\dfrac{X}{Z} \\ y' = f'\dfrac{Y}{Z} \end{cases}$$

Binocular disparity

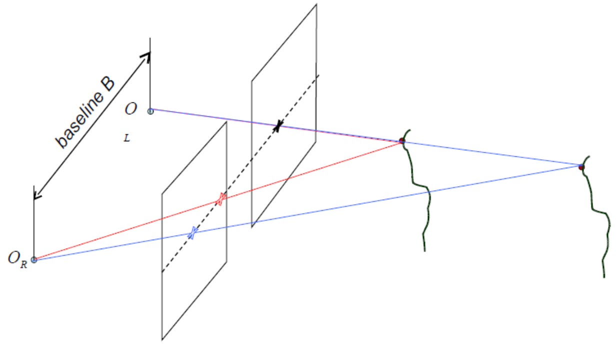


AR devices and Human visual system



- **Monocular depth cues** (size in the image, linear perspective, height in the field of view, occlusions, or shadows and shading) can be encoded in a single image by **conventional computer graphics**.

- **Binocular depth** cues require **display hardware** that can present **separate images to both eyes simultaneously** (stereopsis)



- **Stereopsis**: the closer an object is to the eyes, the bigger the angular offset, or parallax, of the object's projections in the two image planes (see Lecture on Stereoscopy and Disparity)

Università di Genova | DIBRIS DIPARTIMENTO DI INFORMATICA, BIOINGEGNERIA, ROBOTICA E INGEGNERIA DEI SISTEMI

AR devices and perception

Ocularity and Stereoscopy

If the physical world is seen through some kind of lens or mediated through cameras, two questions arise regarding **scene dimensionality**:

(1) Is the three-dimensionality of the real world maintained?

(2) Are the augmentations displayed with stereopsis, exploiting binocular vision?

Ocularity and Stereoscopy (VST)

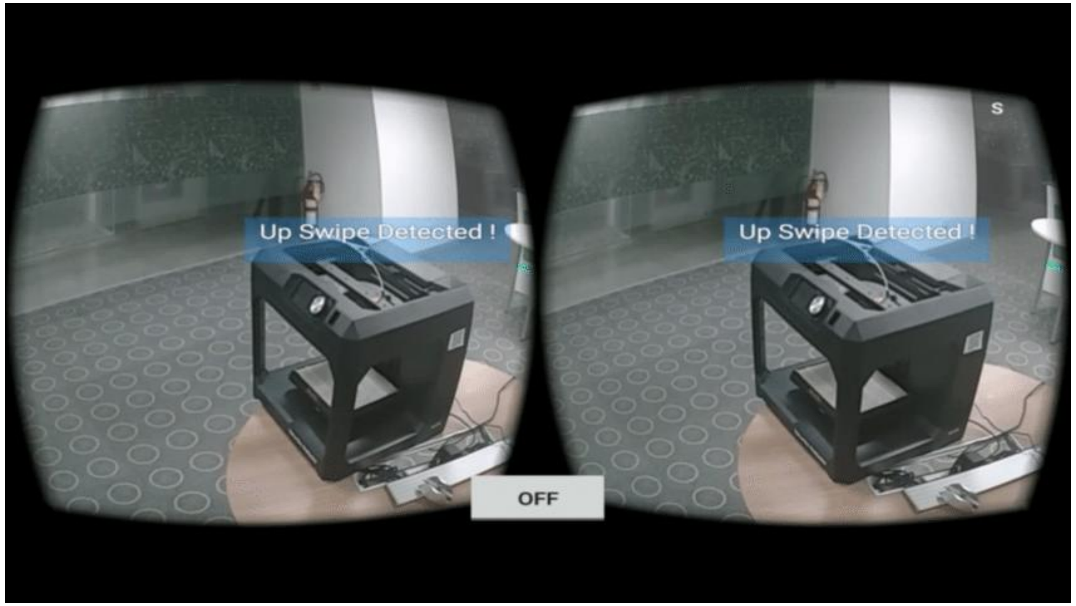
A **monocular** device presents images to only one eye.

A monocular display can be used for AR, but **it lacks immersion (no binocular cues)**.

Ocularity and Stereoscopy (VST)

A **bi-ocular** display presents the **same image to both eyes**, resulting in a **monoscopic impression**.

This approach is sometimes used for VST HMD, *because only a single camera stream is required*, and sensing and processing requirements are minimized.



Ocularity and Stereoscopy (VST)

Binocular HMD presents a **separate image to each eye**, resulting in a **stereoscopic effect**.

They have a significantly increased technical cost.

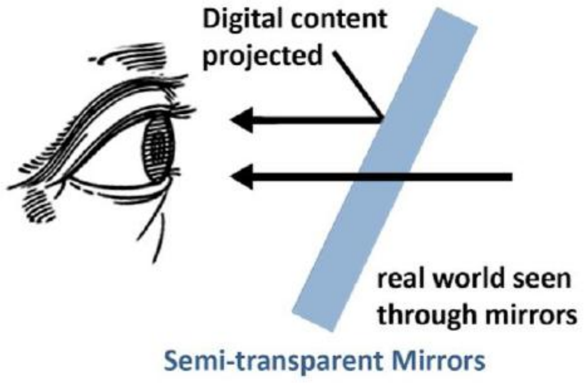
Pairs of video cameras and pairs of displays must be synchronized to deliver the images at the same time.

Ocularity and Stereoscopy (VST)

- A VST display with two cameras, aligned with the eyes' optical axes can produce imagery that **lets the user perceive** the physical world under **stereopsis**.
- But, it is difficult to give a **realistic impression** of a three-dimensional space using stereoscopic camera feeds.

- **Humans** are used to viewing the physical world **unmediated and unencumbered** and have **high expectations for realism**, especially regarding the interplay of varying depth cues.
- **Unavoidable deviations from the ideal values for other immersion factors, such as field of view, resolution, and focus ability, will make for a slightly or even considerably unnatural experience.**

Ocularity and Stereoscopy (OST)

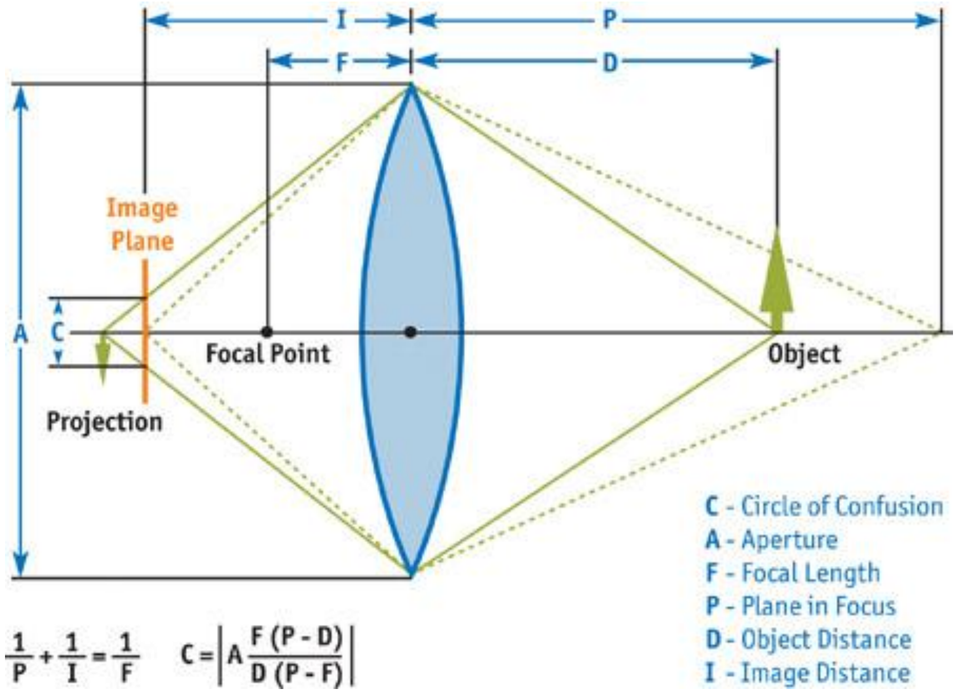


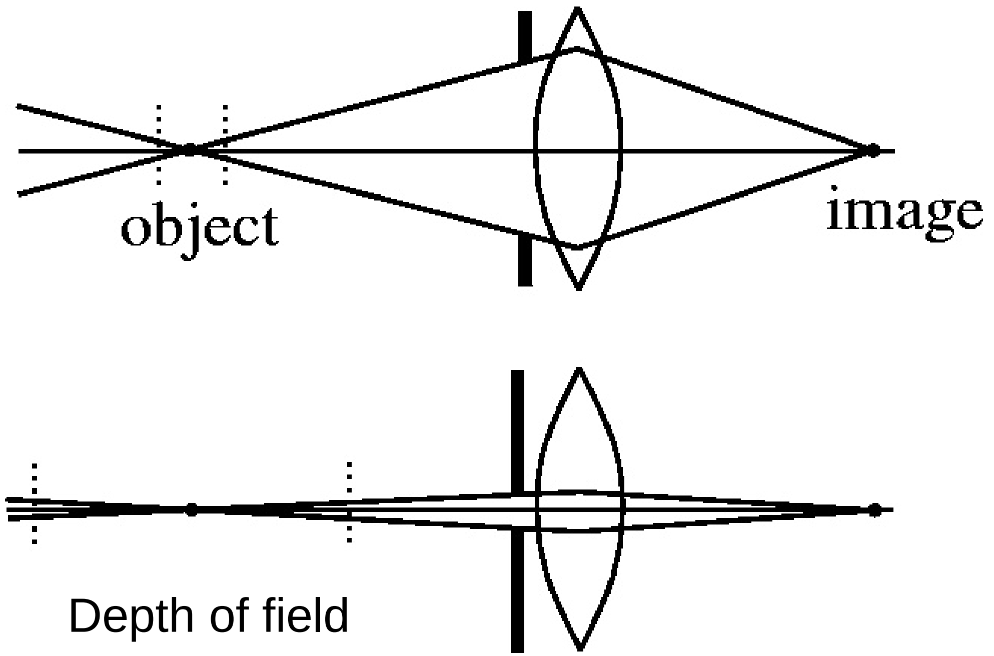
- With OST displays, the **real-world** backdrop is **viewed directly**, and thus naturally **exhibits binocular disparity**.

- The **augmentations** can be rendered either **with or without stereopsis**.

Focus

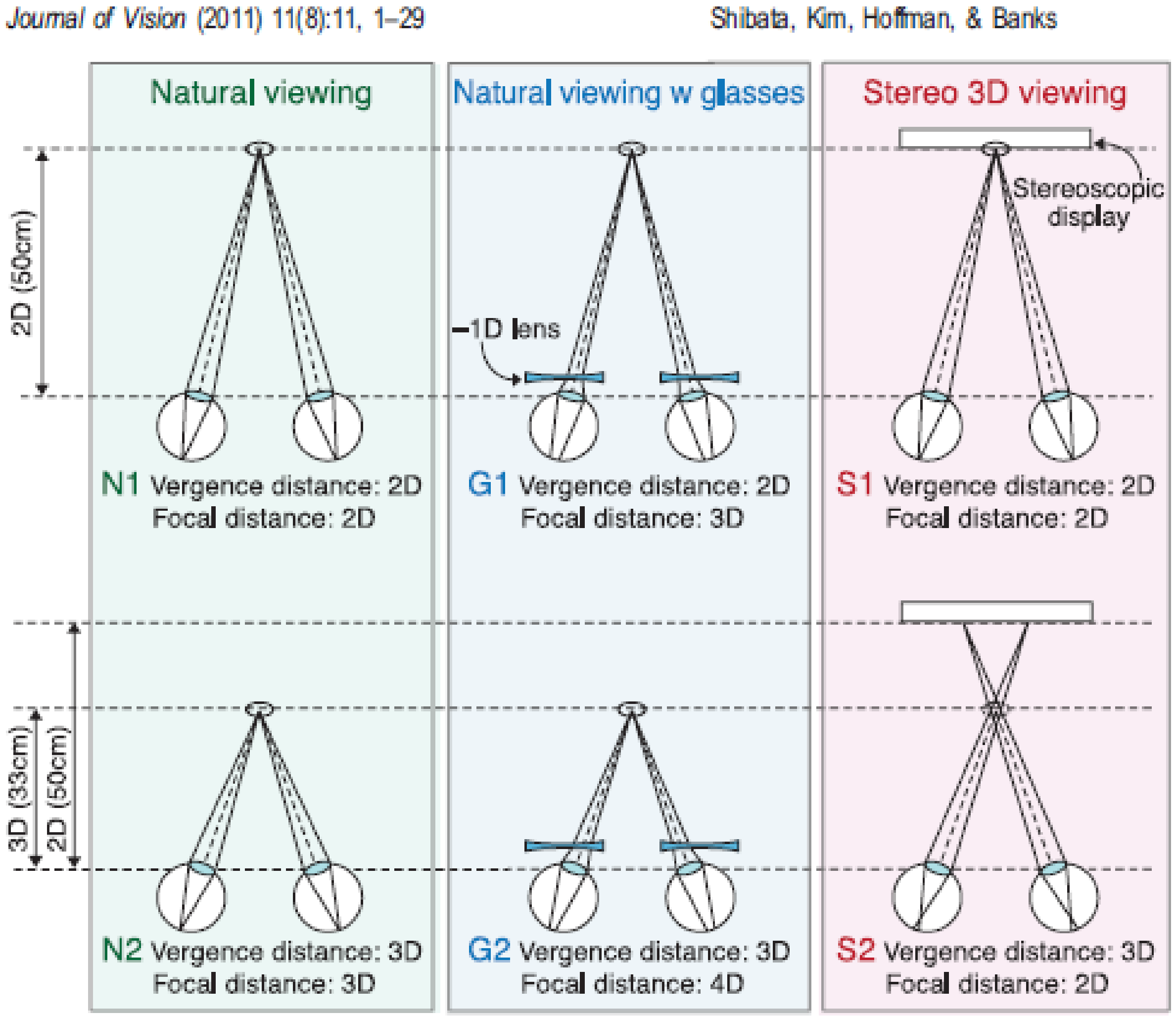
- **Rendering of virtual objects** with the pinhole camera model commonly used in computer graphics will result in **perfectly sharp images of all objects**, irrespective of the focal depth.
- **Our eyes** and actual cameras have a certain aperture size and, therefore, have to cope with the problem of limited ***depth of field***.

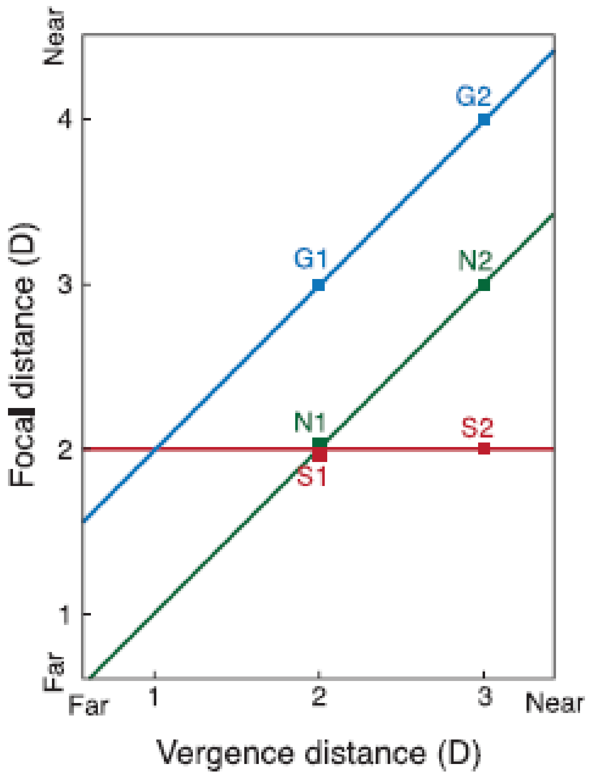




Vergence-accommodation conflict

- **Our eyes can *accommodate* (focus) at varying distances.**
- Accommodation can occur as a reflex to *vergence*, independent rotation of eyeballs to fixate a point in space: **vergence-accommodation reflex**.





D[diopters] = 1 / d [metres]

Vergence-accommodation conflict

- **vergence distance** is the distance to which the eyes must converge for both to foveate the same point in space

- **focal distance** is the distance to which the eyes must accommodate to bring the image of that point in space to sharp focus

- the eyes of the observer have to maintain accommodation on the **display screen**, thus **lacking the natural relationship between accommodation and vergence** eye movements, and the distance of the objects

Vergence-accommodation conflict

– NATURAL VIEWING: vergence and focal distance are equal to one another, so they lie on the green line of slope 1 in the plot on the right (In the optometric/ophthalmic literature, this line is called the demand line or Donders’ line (Donders, 1864). Because **vergence and focal distance are the same** in natural viewing, it is not surprising that the responses are neurally coupled.

– OPTICAL CORRECTION: The introduction of the correction (-1D) increases the focal demand on the crystalline lens within the eye without changing the vergence distance. This creates a **constant difference** in magnitude between the vergence and focal stimuli. Thus, stimuli like G1 and G2 are shifted vertically from the natural viewing line by the dioptric power of the correction.

Vergence-accommodation conflict

- VIEWING ON STEREO DISPLAYS: **Focal distance is fixed** at the distance from the eyes to the display screen. **Vergence distance varies** depending on the distance being simulated by the contents of the display. Thus, **vergence–accommodation conflict is created by viewing stereo displays**, but the magnitude of the conflict depends on the image contents relative to the viewer's distance from the display.

- Because of the neural coupling between vergence and accommodation, *the conflicts induced by optical correction and by stereo viewing need to be resolved*. The viewers must accommodate to a different distance than the distance to which they must converge.

Vergence-accommodation conflict

Possible solutions:

- A display that can ***shift the focus plane* in real time** [Liu et al., 2008]. One would need eye tracking to identify the objects the user is focusing on and then adjust the focus plane shift according to the user's attention.
- ***Multifocus displays*** [Schowengerdt and Seibel, 2012] present another possibility to avoid the accommodation–vergence conflict.
- ***Inverse blurring*** [Hussain, Chessa and Solari, 2023] a ***software solution*** for *off-the-shelf HMD* to mitigate the accommodation–vergence conflict.

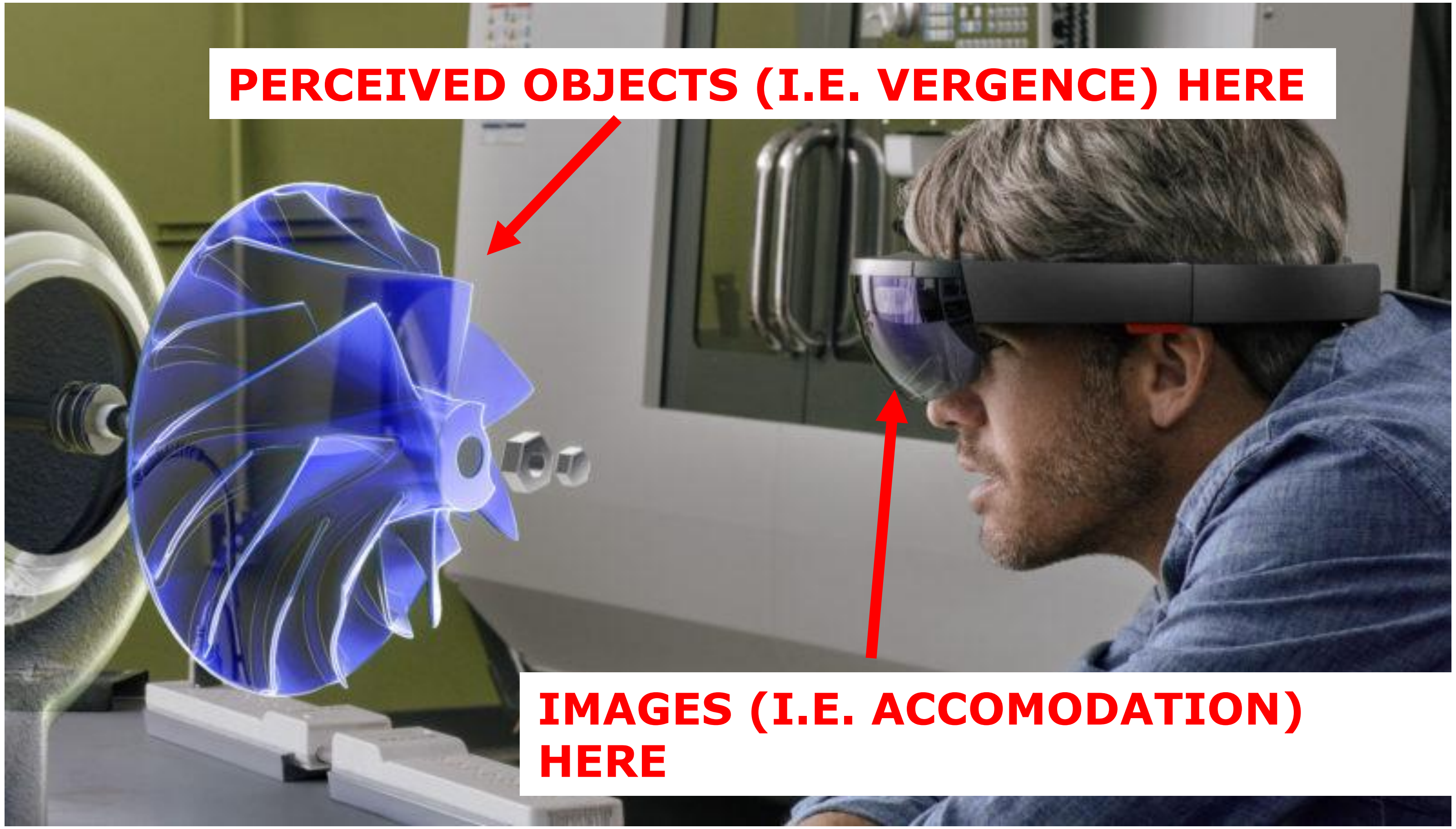
- No such a problem in the case of ***volumetric displays*** [Blundell and Schwartz 1999], where light emanates or reflects from actual points in a 3D volume (a specific example of this is *spatial projection*)

Vergence-accommodation conflict in OST AR

- For OST AR with such stereoscopic rendering there is an additional, related problem:
  - *The user will view the real world with correct accommodation cues, whereas to see the virtual annotations in focus, the user will need to accommodate to the display image plane.*

- **People manage to adjust accommodation at will quite effortlessly, but increased visual fatigue and discomfort might ensue with medium- to long-term use.**

Vergence-accommodation conflict in OST AR



Occlusions

- Occlusion between virtual and real objects is an important cue to convey the scene structure.

- Correct occlusion among real objects is naturally given.

- Correct **occlusion among virtual objects** is easily achieved by means of a **z-buffer**.

- **Correct occlusion of virtual in front of real, or vice versa, requires special consideration (later in the course)**

- In **OST systems**, where augmentations often appear as semi-transparent overlays, it is ***more difficult*** *to make virtual objects appear as if they are truly in front of real objects*.
  - Virtual objects can be rendered very brightly relative to the intensity at which real objects are visible, so the virtual objects will be dominant.

Occlusions in OST

The relevant part of the real scene can be illuminated with a computer-controlled projector, while the rest of the scene (in particular, real objects that are occluded by virtual objects) remains in the dark and, therefore, imperceptible

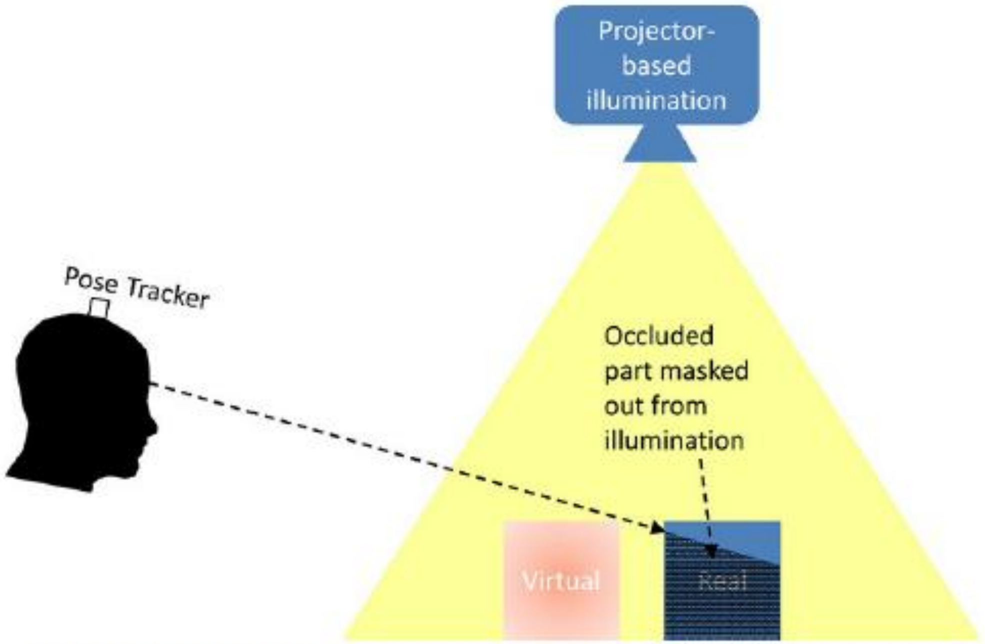


Figure 2.8 The occlusion shadows technique uses controlled illumination to blank out those portions of the real world where opaque computer graphics should be visible.

Resolution and Refresh Rate

- **Spatial resolution** is restricted by the type of display and by the optical system.

- In **VST** systems**, the resolution of the real world** is additionally **restricted** by the resolution of the camera.
- **Sufficient resolution is desirable** to suppress disturbing artifacts of computer-generated images (such as *pixelated lines or text*) that stand out in comparison to the user's perception of the real world.

- *Take into account that humans have a space variant resolution.*

Resolution and Refresh Rate

- The **temporal resolution**—that is, the native refresh rate of the display—is important to minimize perceived *flicker* and to avoid *image lag* and *ghosting*.

- The **flicker fusion threshold** is the frequency at which an intermittent light stimulus appears to be *completely steady* to the average human observer.
- The human flicker fusion threshold is usually taken as 16 Hz. TV cameras operate at 25 or 30 frames per second.

- For rendering **fast motions without blur** 120Hz.
- For **AR** and **VR** frame rates **above 60Hz** are commonly desired (take into consideration multiplexing).

Field of View

- FOV and resolution are interrelated, as more **pixels** at the same **density** are needed to fill a wider FOV

- In **AR**, we distinguish between an **overlay FOV** (where there is augmentation) and a **peripheral FOV** (non-augmented portion of the observed environment)

- **FOV limitations** are **common** in VR and, especially, AR displays

- The resulting reduction in immersion **limits the user's presence** in the displayed scenes and content.

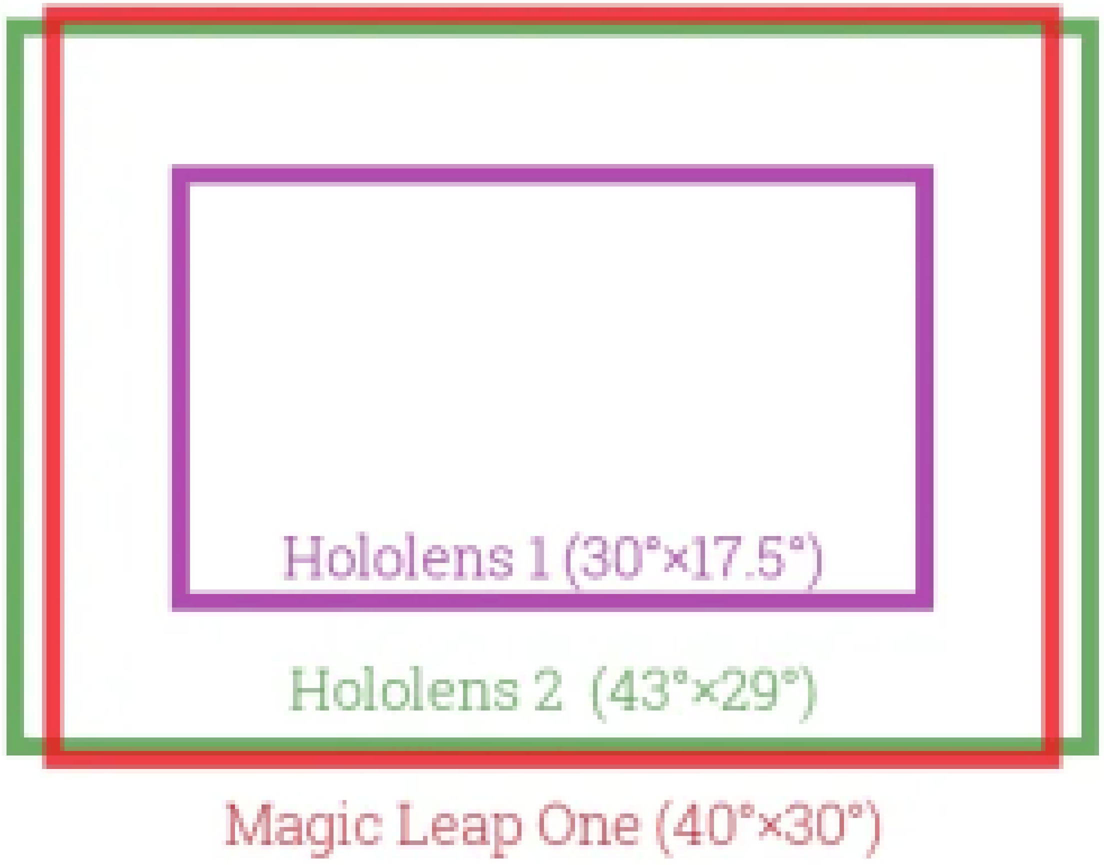
Field of View

- In **VST systems**, the camera often has a larger FOV than the display, so that the camera image actually appears compressed. For HMDs, the goal is to have as wide an overlay FOV as possible.

- The field of view together with the spatial resolution of the display determines the **angular resolution**.
- A display with a very high pixel density can use rather simple magnification optics to present a wide field of view.

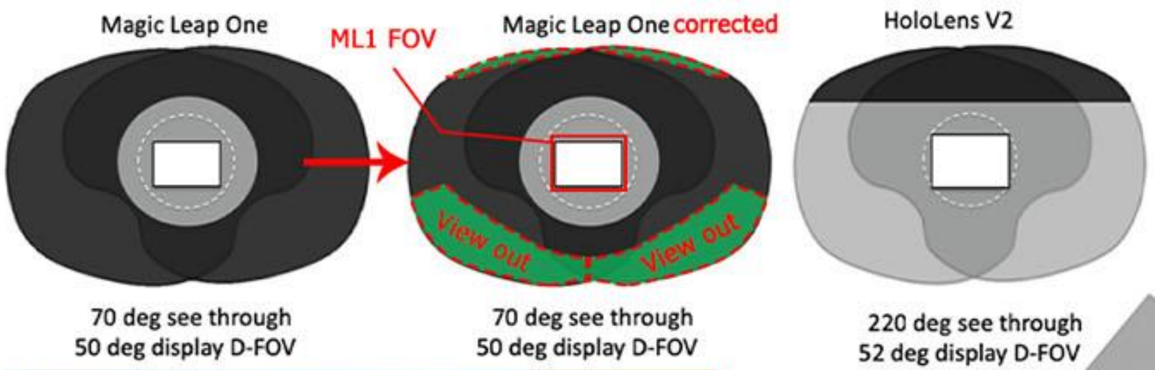
- *Pilots may prefer a wider field of view, while surgeons may require higher resolution.*

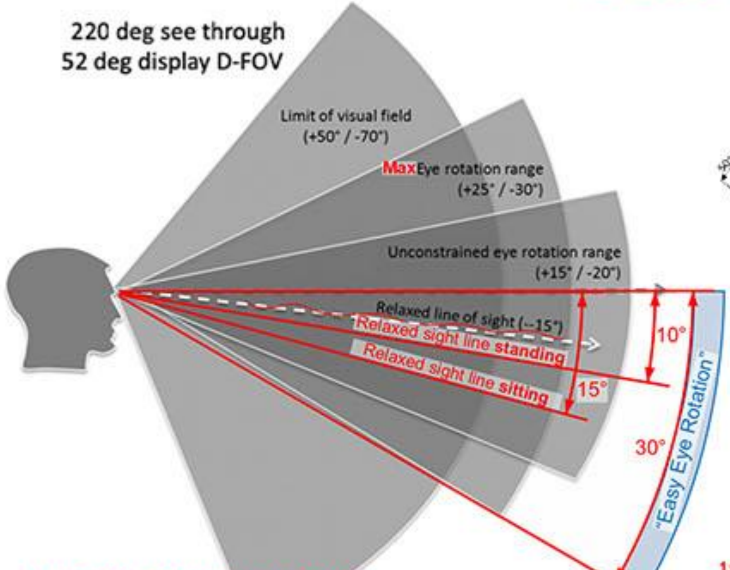
Field of View



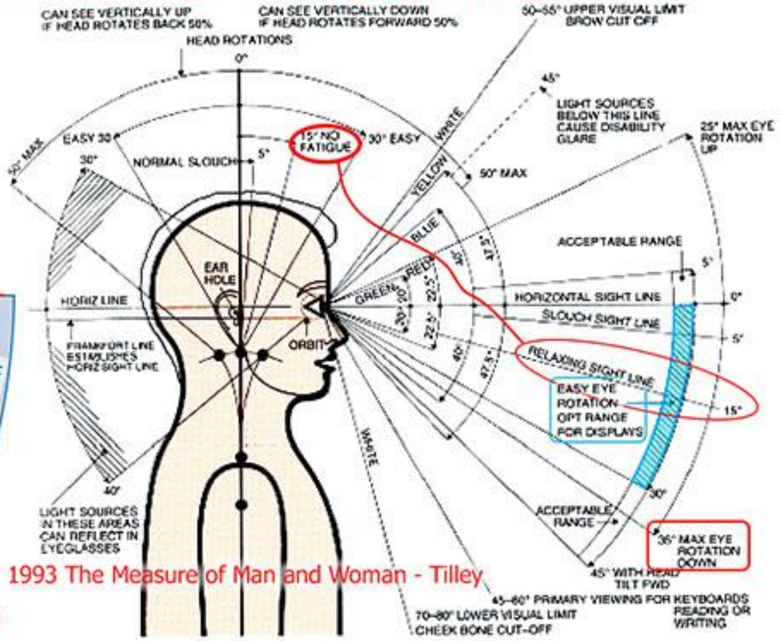
https://uploadvr.com/hololens-2-field-of-view/

FOV: AR and The View of the Real World



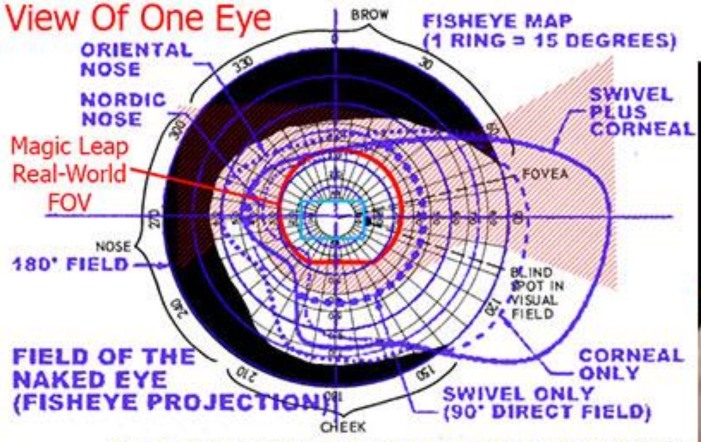


Bernard Kress (Microsoft Hololens) -
PROCEEDINGS OF SPIE - Digital optical elements and technologies



1993 The Measure of Man and Woman - Tilley

View Of One Eye



With eye swivel: Visual Field of the Naked Eye, section 3.1 of Howlett, SID 1992
Retina vision map: www.are.na/david-boyd/human-vision
Magic Leap Real-World FOV: KGOnTech (www.kguttag.com)

Field of View

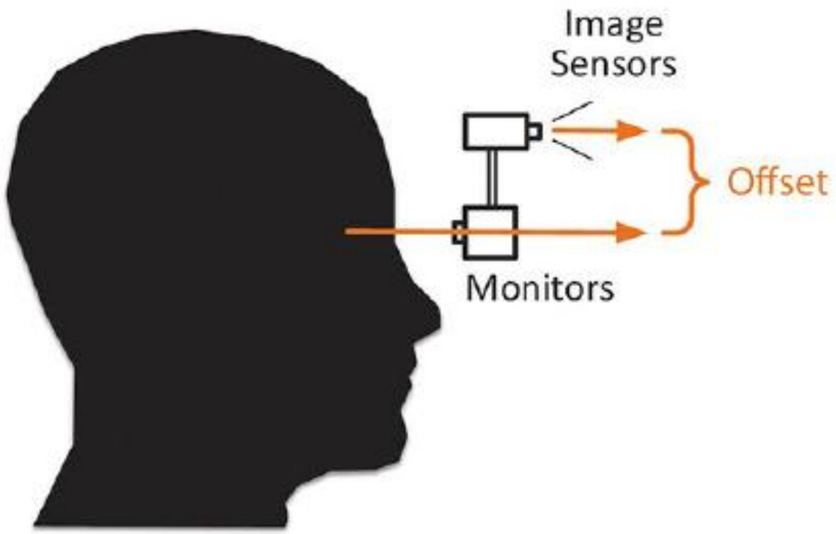
Viewpoint Offset

- **OST displays** fuse the optical paths of the virtual and the real into one, so the resulting images are *aligned by design*.
- However, it requires **calibration** of the **virtual camera**, which is used to generate the virtual part of the AR display, **to the eye** of the user.

- In **VST displays**, the camera frame can be used for **computer-vision–based registration** (see later in the course) leading to pixel-accurate annotations

Viewpoint Offset

VST configurations will often introduce a significant **offset** between the viewing direction of the camera and the viewing direction of the screen on which the camera image is presented.



Brightness and Contrast

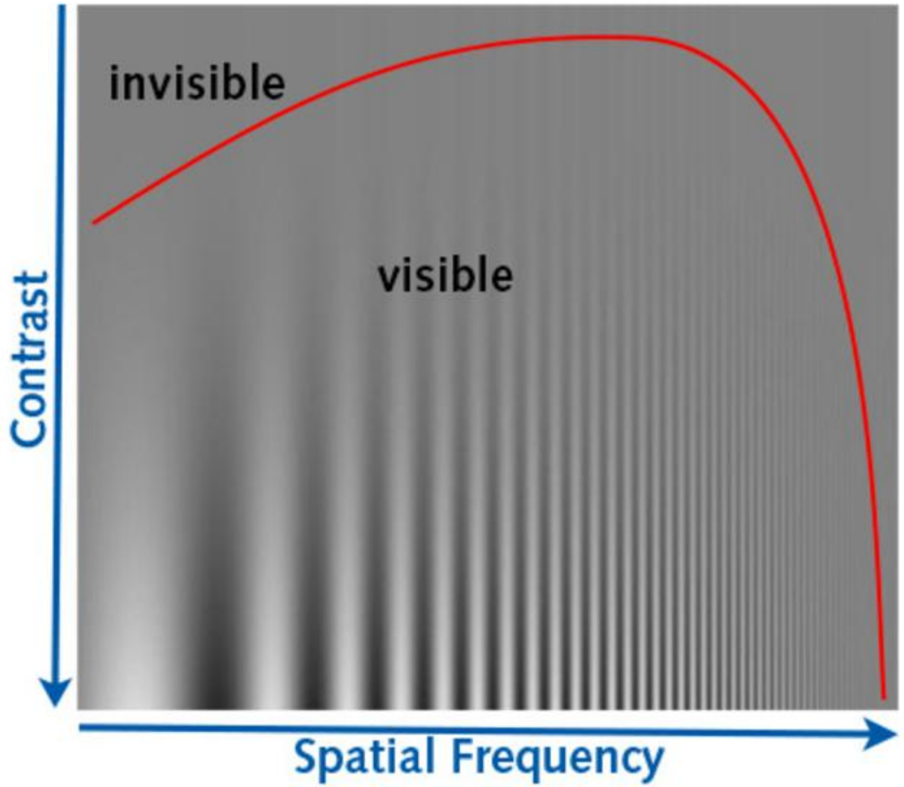
- In outdoor situations or in situations with an abundance of natural light, **most computer displays are not bright enough** to achieve sufficient contrast.

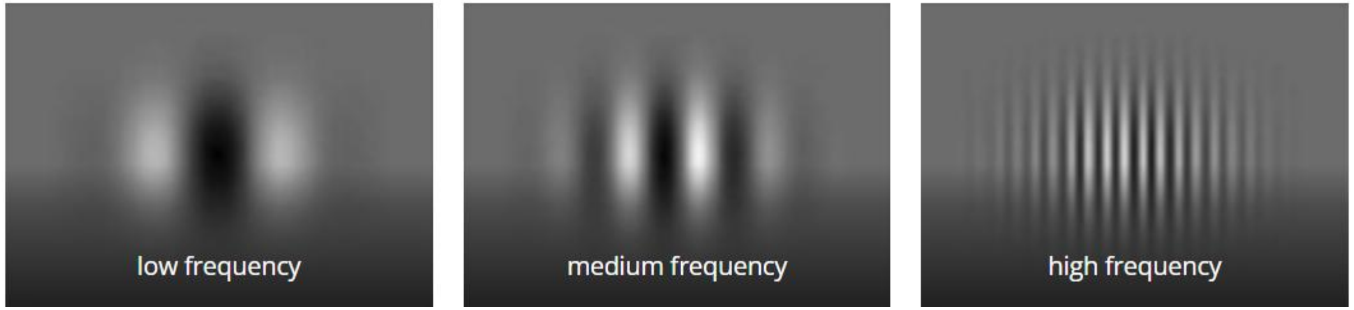
- **OST displays** depend on the transparency of the optical combiner

- **VST displays** can change brightness and contrast arbitrarily, as long as the display itself can deliver sufficient contrast.
- Note that VST is critically dependent on working electronic components. *If the camera or the display fails, no meaningful image is shown at all (!)*

Brightness and Contrast

Human contrast sensitivity functions





Brightness and Contrast

Microsoft HoloLens

Meta Quest 3