Computer Vision and Mixed Reality

Displays

Chapter 2

Pedro Mendes Jorge

Multimodal Displays

- The human experience of the physical world is intrinsically multimodal
- AR devices tend to be multimodal
- However, the human vision system is responsible for delivering roughly 70% of the overall sensory information to the brain

Multimodal Displays

Types of modalities displays:

- Visual
- Audio
- Haptic, Tactile and Tangible
- Olfactory and Gustatory

Multimodal Displays - Audio

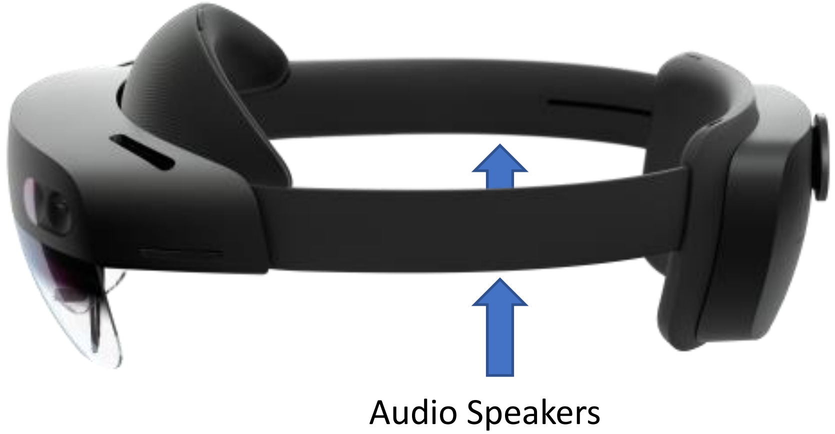
- Audio Displays

Audio guides and electronic multimedia guides

Assistive audio guidance systems for people with visual impairment [S. Real, 2019]

Multimodal Displays - Audio

- Modern AR/VR displays use 3D spatial location enabling 3D sound modelling and effects

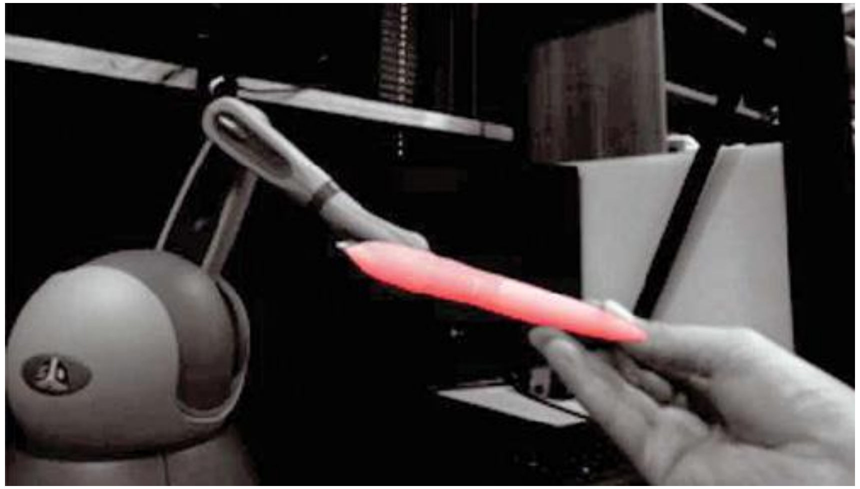


Microsoft HoloLens 2

HTC Vive Pro

Multimodal Displays - Haptic, Tactile and Tangible

- Extrinsic and intrinsic **haptic displays**
- Haptic gloves, shoes, vests, jackets, and exoskeletons



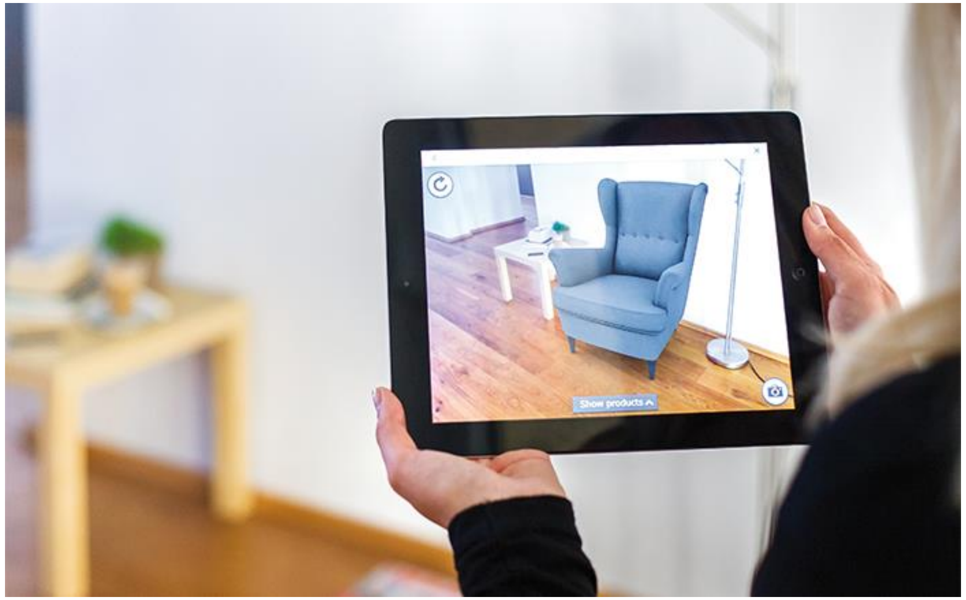
Example of visuo-haptic registration. The stylus of a Phantom Omni (now Geomagic Touch) haptic device is highlighted by visual augmented reality. Courtesy of Ulrich Eck and Christian Sandor.

Haptic gloves: SensorialXR

Kinesthetic force feedback: Dexmo exoskeleton glove

Multimodal Displays - Haptic, Tactile and Tangible

- Tactile



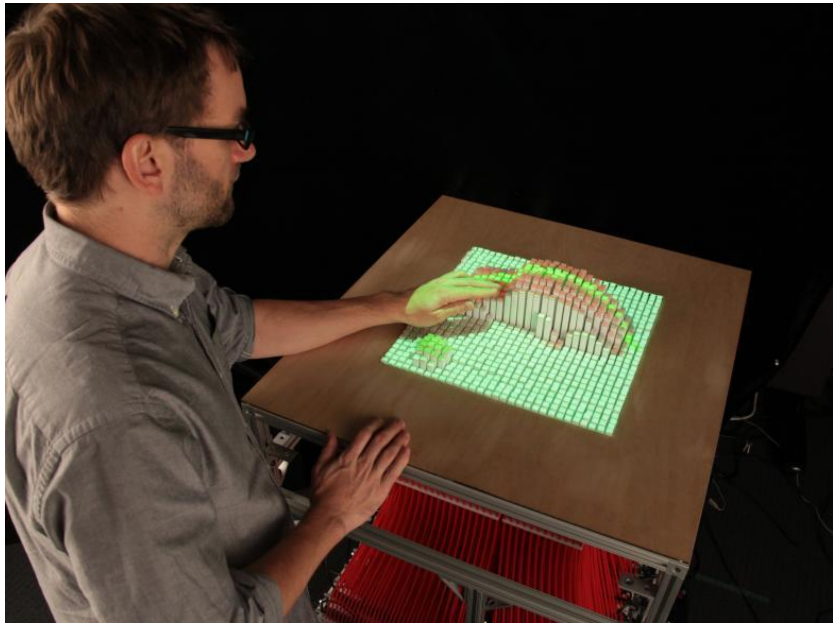
Handheld display with touchscreen

Turning an ordinary surface into a touchscreen with a projector–camera system. Courtesy of Claudio Pinhanez (copyright IBM 2001)

Multimodal Displays - Haptic, Tactile and Tangible

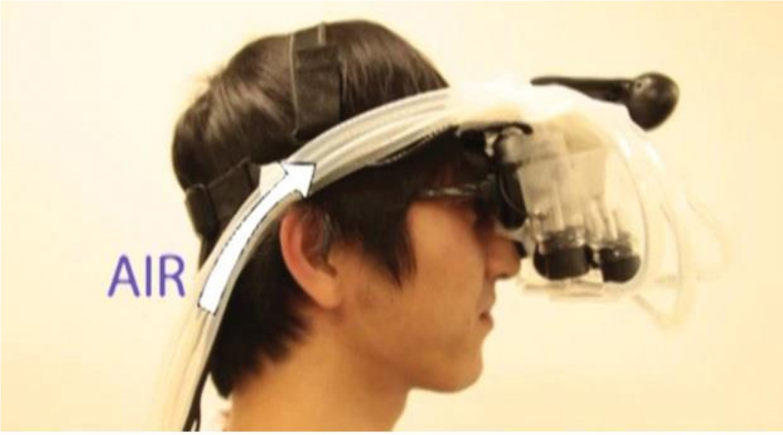
- Tangible: Incorporate the physical reality surrounding a user into the interaction with virtual objects.

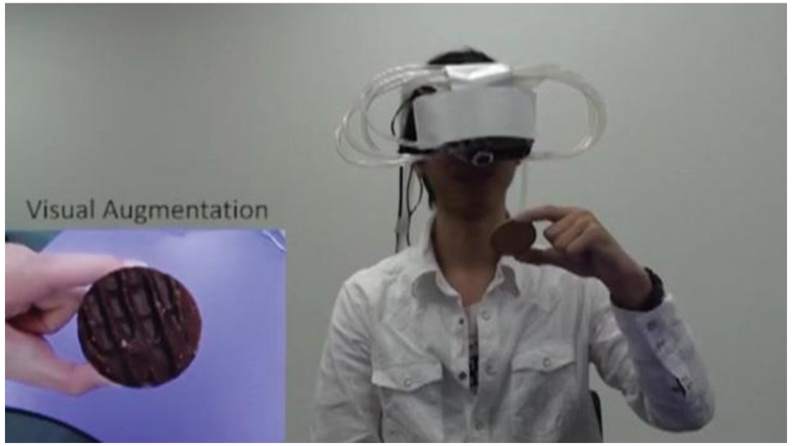
Reactable, an electronic musical instrument example of tangible user interface.



inFORM: MIT Media Lab, Tangible Media Group

Multimodal Displays - Olfactory and Gustatory



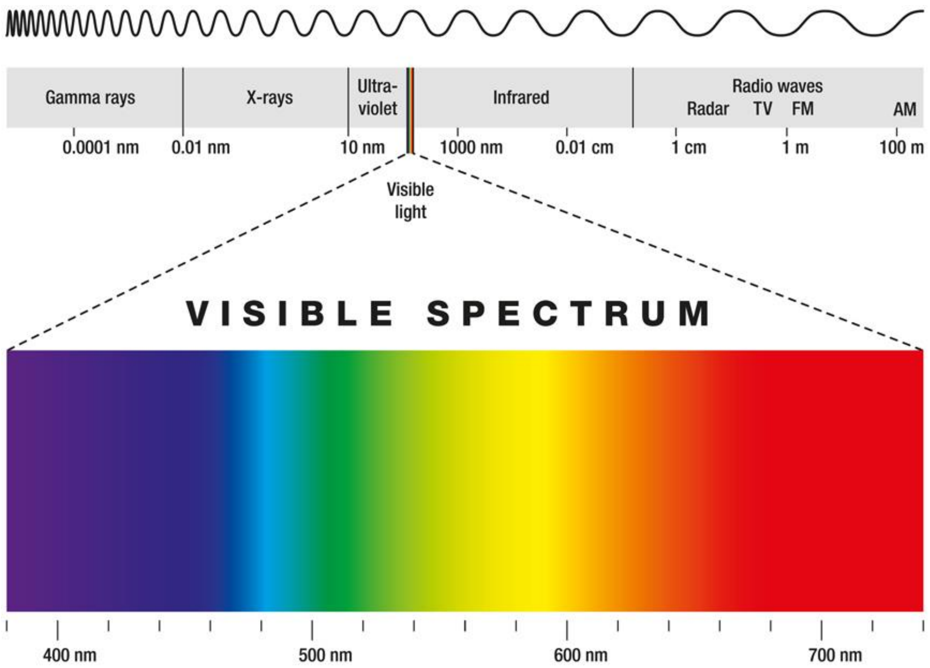


MetaCookie: An olfactory display is combined with visual augmentation of a plain cookie to provide the illusion of a flavored cookie (chocolate, in the case of the bottom image). Courtesy of Takuji Narumi.

Olfactory Virtual Reality - OVR Technology
9 scent actuators with interchangeable cartridges

Visual Perception - Properties of the human visual system

Light Sensitivity

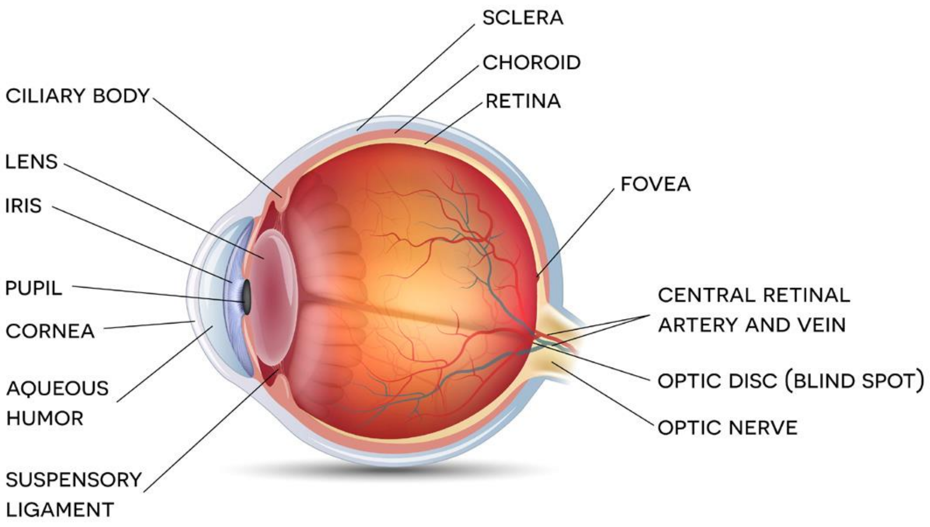


- The human eye is only sensitive to a narrow band within the electromagnetic spectrum falling between wavelengths roughly measuring 380 nm (violet) – 740 nm (red).

(Based on [2])

Visual Perception - Properties of the human visual system

Entering the Eye

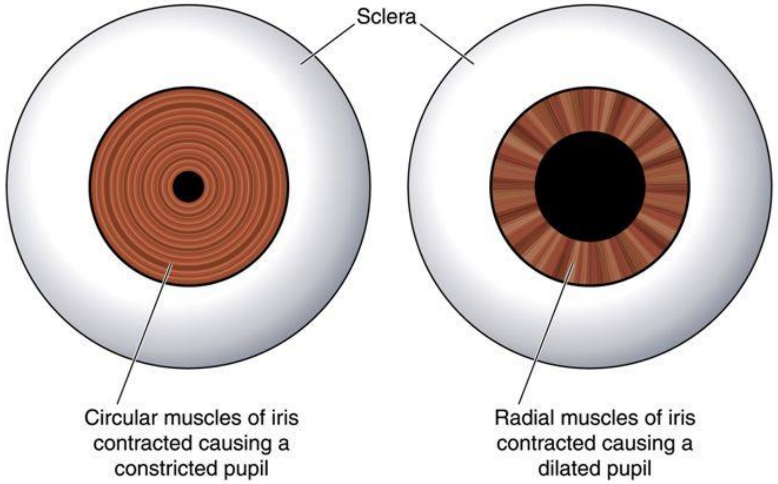


- Light enters the eye through the transparent, dome-shaped outer surface known as the **cornea**.
- Most of the refraction of light by the eye (~80%) takes place at the air-cornea interface because of its curvature and the large difference in the indexes of refraction.

(Based on [2])

Visual Perception - Properties of the human visual system

Entering the Eye

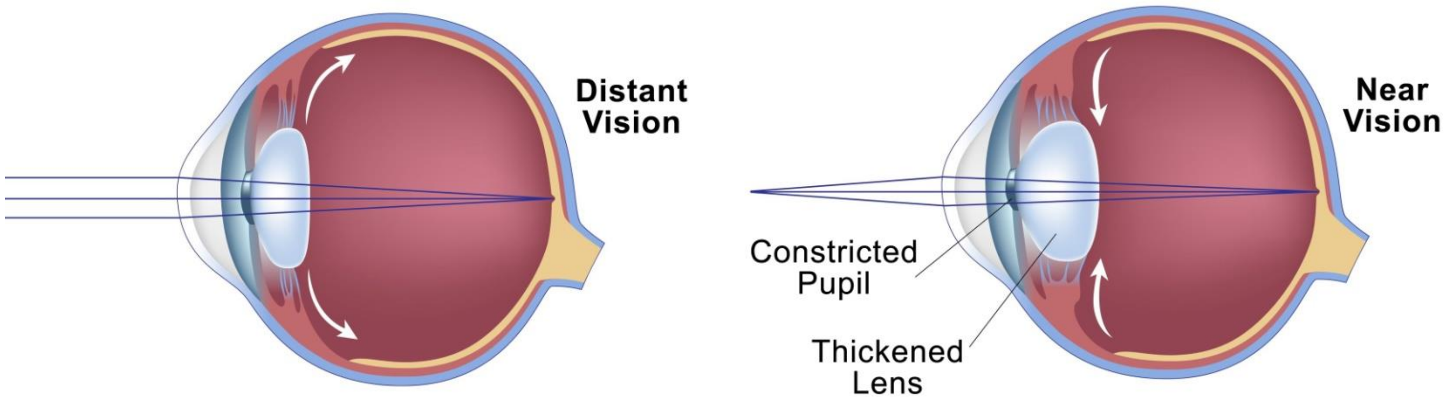


- The **pupil** functions very similar to the aperture of a camera diaphragm, expanding and contracting to control the amount of light entering the eye.
- In brightly lit conditions, the pupil contracts in size. This involuntary reaction is called the pupillary reflex.

(Based on [2])

Visual Perception - Properties of the human visual system

Entering the Eye

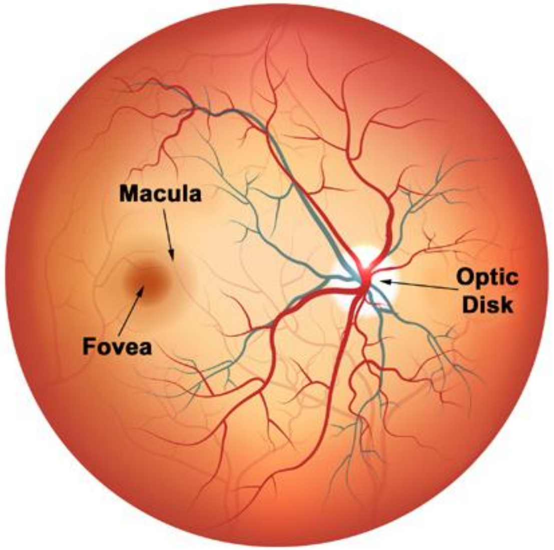


- The **crystalline lens** is the fine-focus mechanism as its shape can be changed, enabling the focal length of the optical system to be changed.
- This ability to rapidly switch focus between objects at different depths of field is referred to as **accommodation**.

(Based on [2])

Visual Perception - Properties of the human visual system

Image Formation and Detection



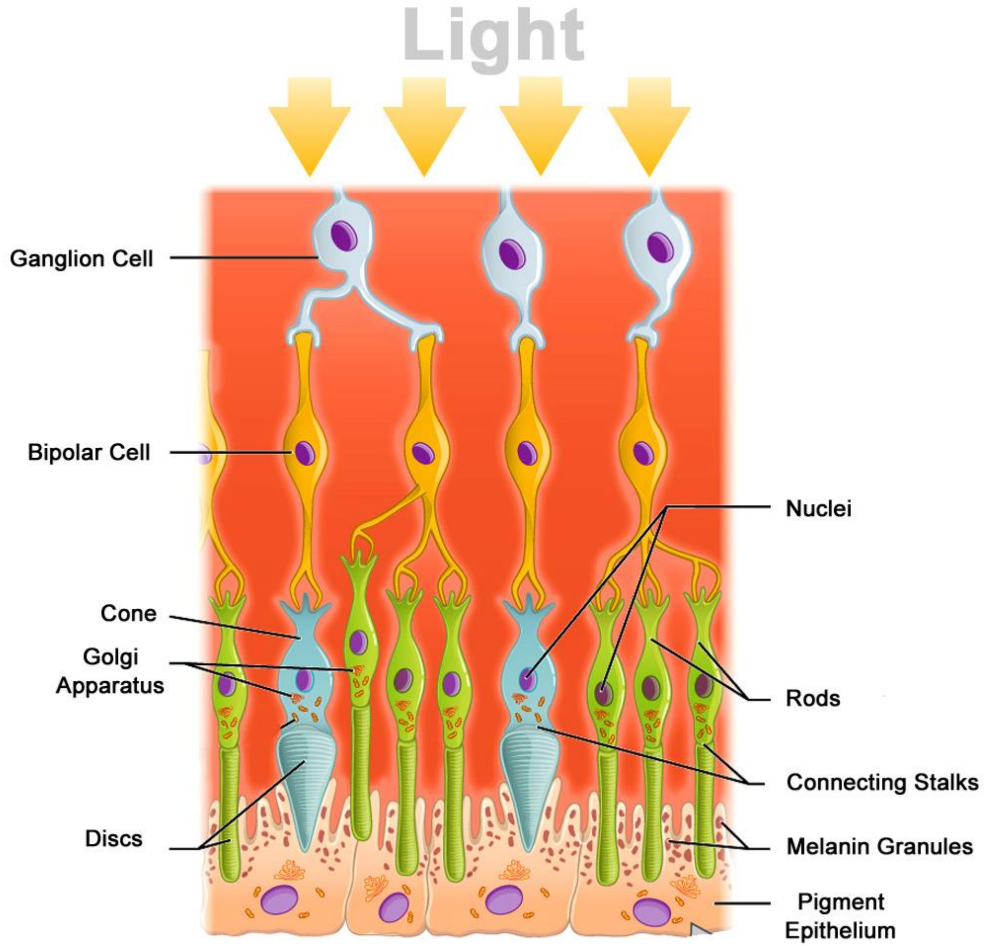
- Visual perception begins as the optical components of the eye focus light onto the **retina**.
- The **fovea** is naturally centered on objects when we fixate on them and is the point on the retina with the greatest acuity.

(Based on [2])

Visual Perception - Properties of the human visual system

Facts About the Retina

- Covers ~65% of the interior surface of the eye.
- The thickness of the retina ranges from 0.15 mm to 0.320 mm.
- The retina is almost completely transparent.
- As light hits the retina, it passes straight through and is reflected back into the photoreceptor neurons.



(Based on [2])

Visual Perception - Properties of the human visual system

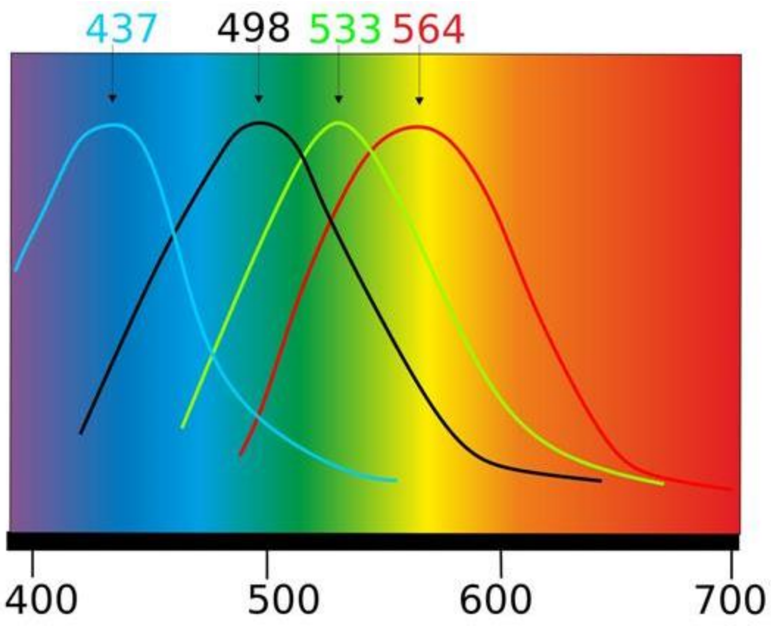
Rods and Cones

- Rods are more numerous, responsible for vision at low light levels, and are highly sensitive motion detectors.
- Rods are found mostly in the peripheral regions of the retina and are responsible for our peripheral views.
- Cones are active at higher light levels, have a high spatial acuity, and are responsible for our color sensitivity.
- There are approximately 100-120 million rod and 7-8 million cone photoreceptors in each retina.

(Based on [2])

Visual Perception - Properties of the human visual system

Rods and Cones



- Individual cones are sensitive to one of three light conditions: red (most numerous) shows peak sensitivity at 564 nm, green at 533 nm, and blue at 437 nm.
- Rods show a peak sensitivity to wavelengths around 498 nm (green-blue).

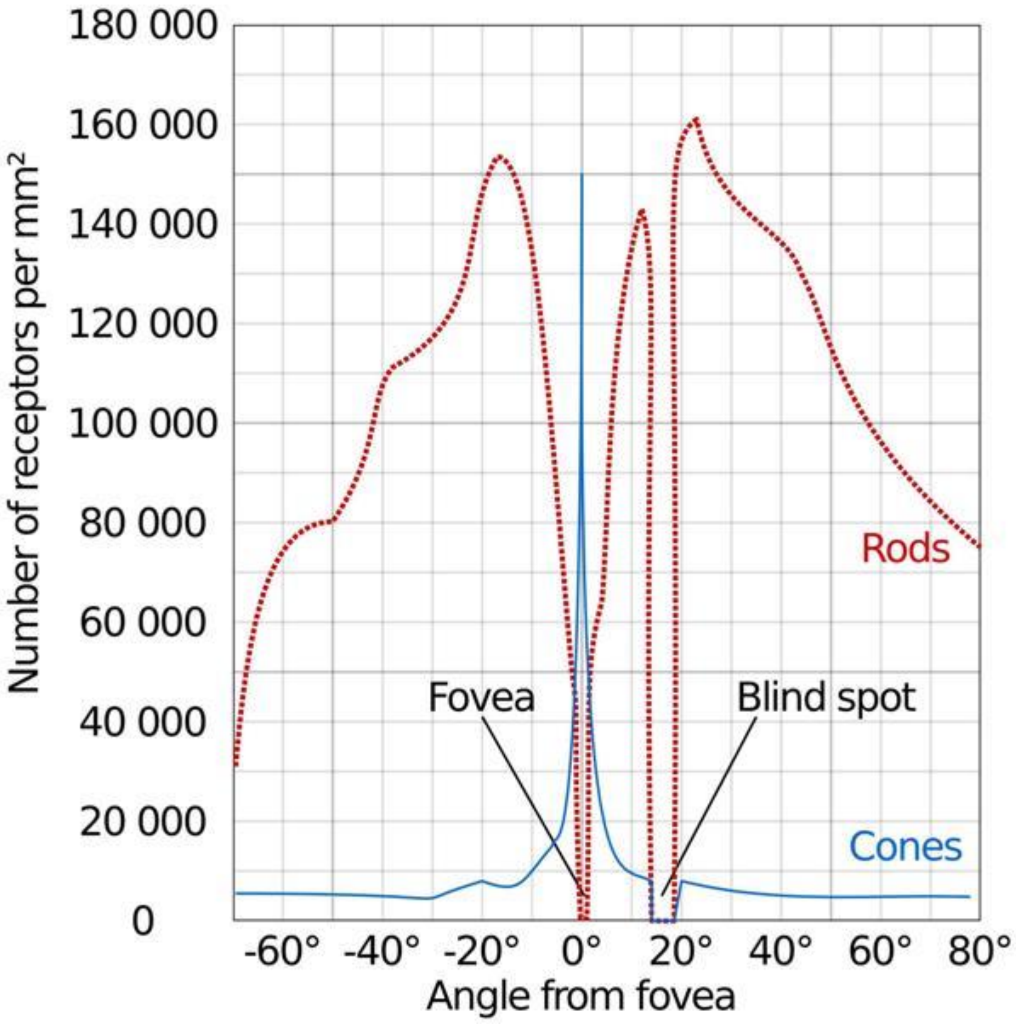
(Based on [2])

Visual Perception - Properties of the human visual system

Rod and Cone Distribution

- Most cones are concentrated in the fovea, whereas rods are absent there but dense elsewhere.
- While perception in daytime light levels is dominated by cone-mediated vision, the total number of rods in the human retina far exceeds the number of cones.
- There are no photoreceptors in the optic disk.
- This lack of photoreceptors means there is no light detected in this area, resulting in a blind spot for each eye.

(Based on [2])



Visual Perception - Properties of the human visual system

Blind Spot

To find the blind spot for each eye, start by placing this book flat on a table. Cover your left eye and look at the dot on the left side of this image. Remain aware of the cross on the right without looking directly at it. Slowly move your face closer to the image. At some point, you will see the cross disappear. Reverse the process to find the blind spot for the right eye.
Credit: Illustration by S. Aukstakalnis

(Based on [2])

Visual Perception - Properties of the human visual system

Depth Cues – Extraretinal

Extraretinal depth cues are those triggers or pieces of information that are not derived from light patterns entering the eye and bathing the retina, but from other physiological processes.

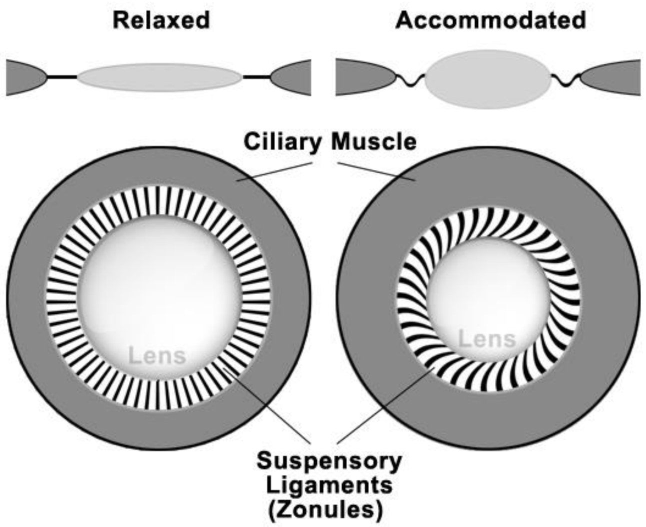
There are two extraretinal depth cues:

- Accommodation
- Vergence

(Based on [2])

Visual Perception - Properties of the human visual system

Accommodation



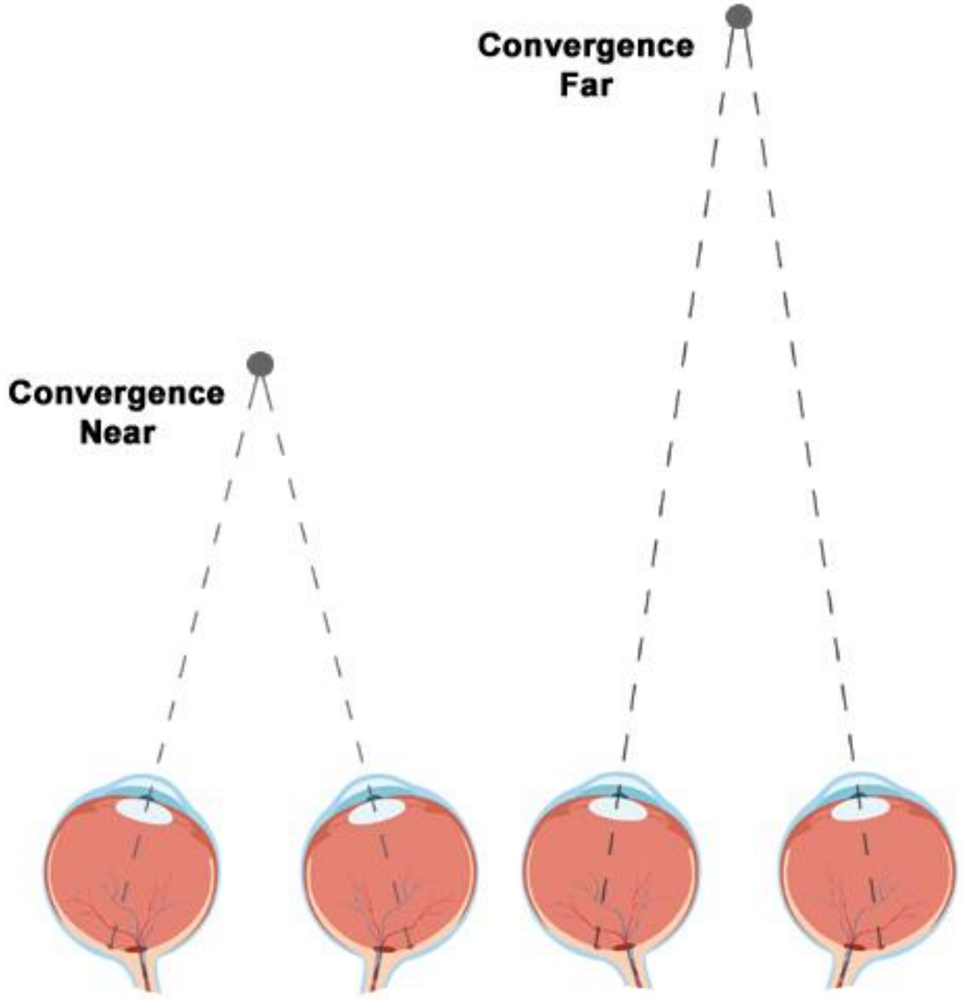
- **Accommodation** is an involuntary physiological process by which the optical power of the eye lens changes to focus light entering the eye and falling on the retina.
- It is widely believed that blurring on the retina is the stimulus for accommodation.

(Based on [2])

Visual Perception - Properties of the human visual system

Vergence

- Vergence is the pointing of the fovea of both eyes at an object in the near field.
- Vergence entails the eyes simultaneously rotating about their vertical axis in opposite directions such that the projected image of that object is aligned with the center of the retina of both eyes.
- When looking at an object in the near field, the eyes rotate toward each other, or converge.
- When looking at an object in a far field, the eyes rotate away from each other, or diverge.



(Based on [2])

Visual Perception - Properties of the human visual system

Coupling of Cues

- Accommodation and vergence are normally tightly coupled physiological processes, with vergence triggering accommodation.
- The design of flat panel-based stereoscopic head-mounted displays causes a decoupling of these processes, sometimes resulting in headaches and eye strain – accommodation-vergence conflict.

(Based on [2])

Visual Perception - Properties of the human visual system

Depth Cues – Binocular

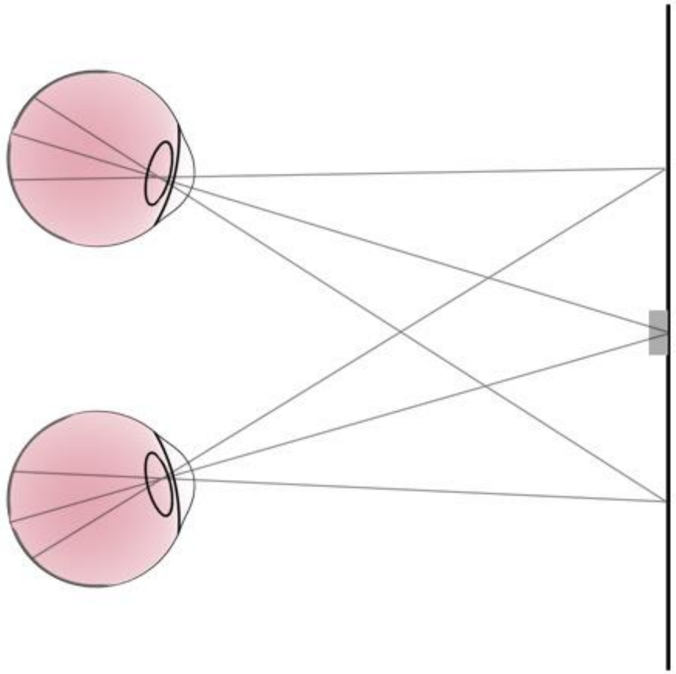
- Binocular depth cues are those triggers or pieces of information that are detected as a result of viewing a scene with two eyes, each from a slightly different vantage point.

- These two scenes are integrated by our brain to construct a 3D interpretation of our real or virtual surroundings.

(Based on [2])

Visual Perception - Properties of the human visual system

Stereopsis



Within a binocular view, each point in one retina is said to have a corresponding point in the other retina.

(Based on [2])

Visual Perception - Properties of the human visual system

Depth Cues – Monocular

- Monocular depth cues are those triggers or pieces of information that are derived from light patterns on the retinas but are not dependent on both eyes.
- Monocular cues are divided between those requiring movement of light patterns across the retina (that is, viewer or scene motion) and those that can be discerned from a fixed viewing position.

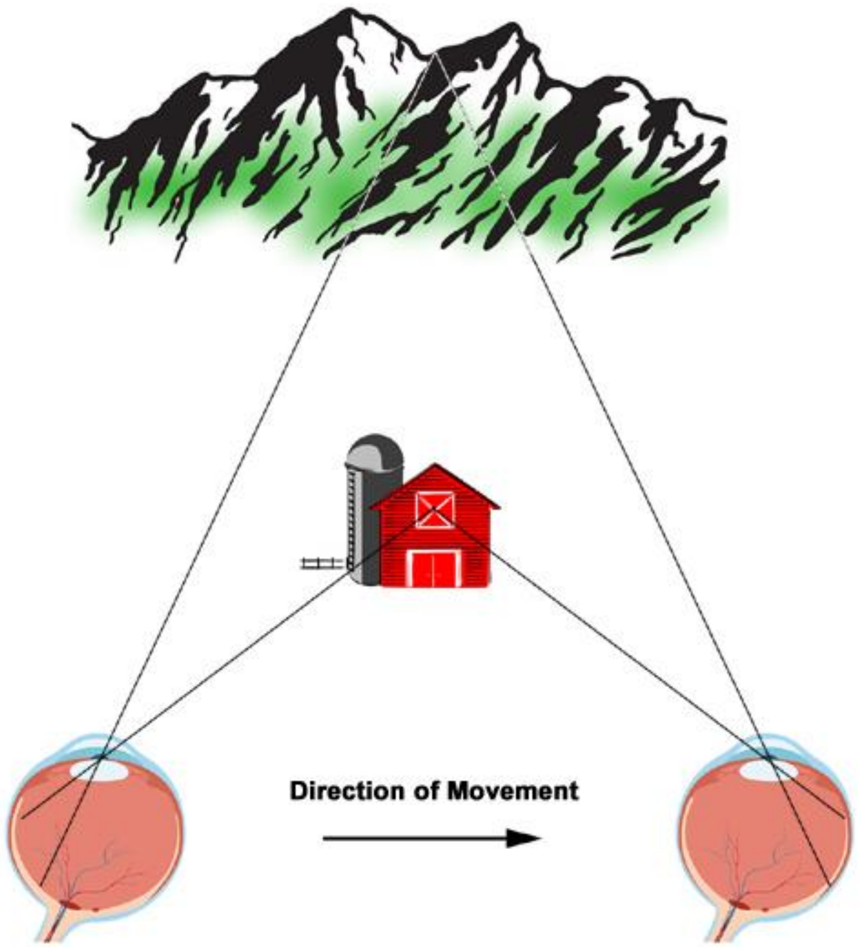
(Based on [2])

Visual Perception - Properties of the human visual system

Motion Parallax

- Motion parallax is a strong, relative motion cue within which objects that are closer to a moving observer appear themselves to move faster than objects that are farther away.

- Motion parallax is the result of the speed at which an image moves across the retinas of the eyes.



(Based on [2])

Visual Perception - Properties of the human visual system

Occlusion

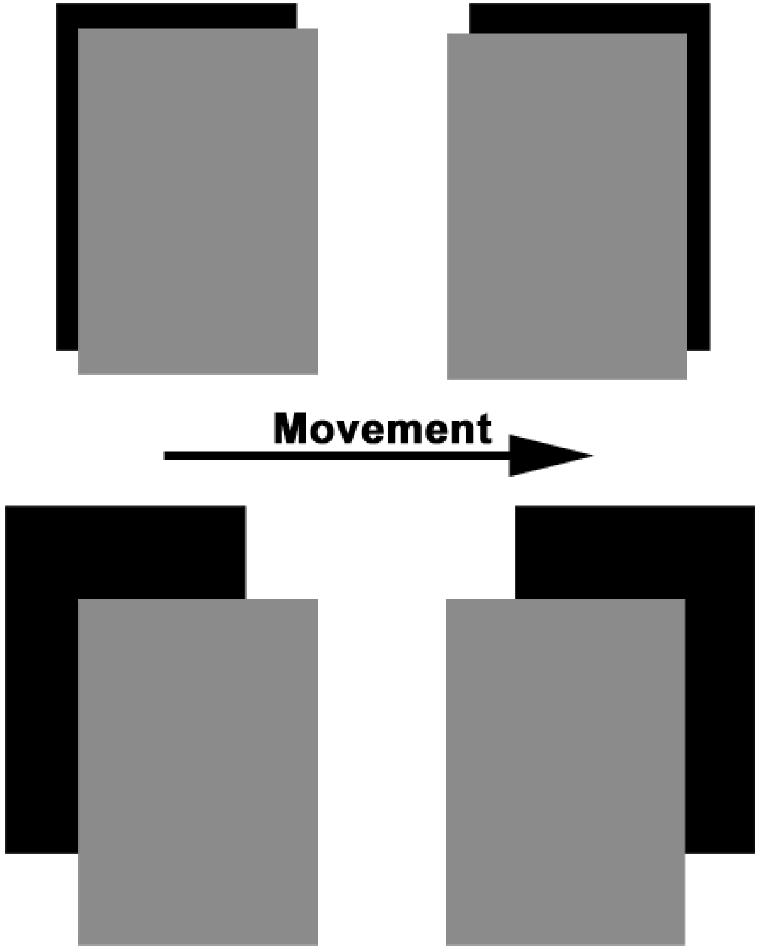
- Also known as interposition, occlusion cues are generated when one object blocks an observer's view of another object.
- In such a situation, the blocking object is perceived as being closer to the observer.

(Based on [2])

Visual Perception - Properties of the human visual system

Deletion and Accretion

This refers to the degree to which an object or a surface in the near field reveals or covers objects or surfaces in the far field as your viewpoint translates past their position.

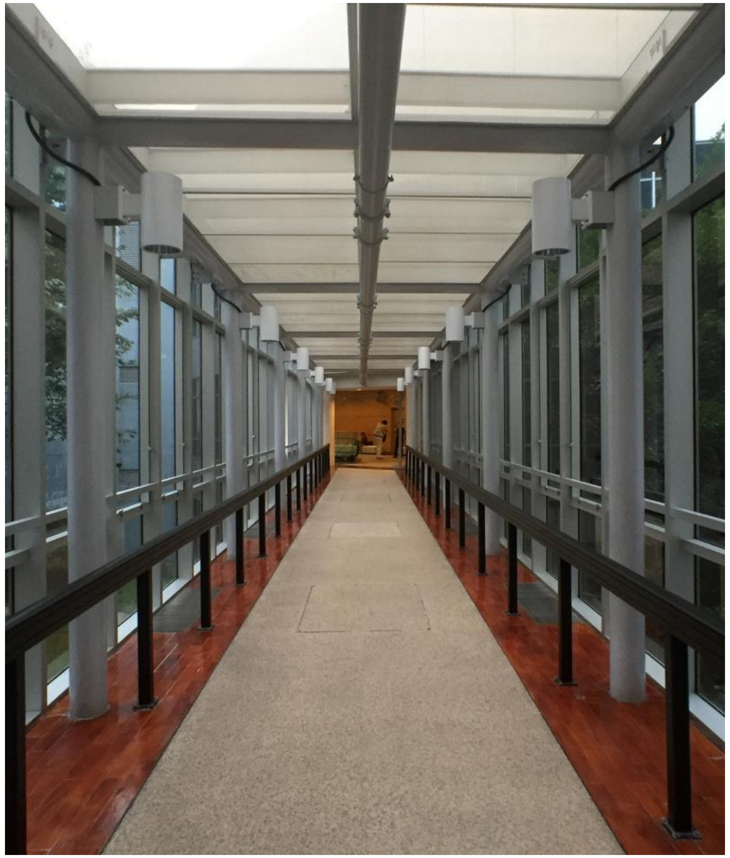


Visual Perception - Properties of the human visual system

Linear Perspective

- Linear perspective is a depth cue within which parallel lines recede into the distance, giving the appearance of drawing closer together.

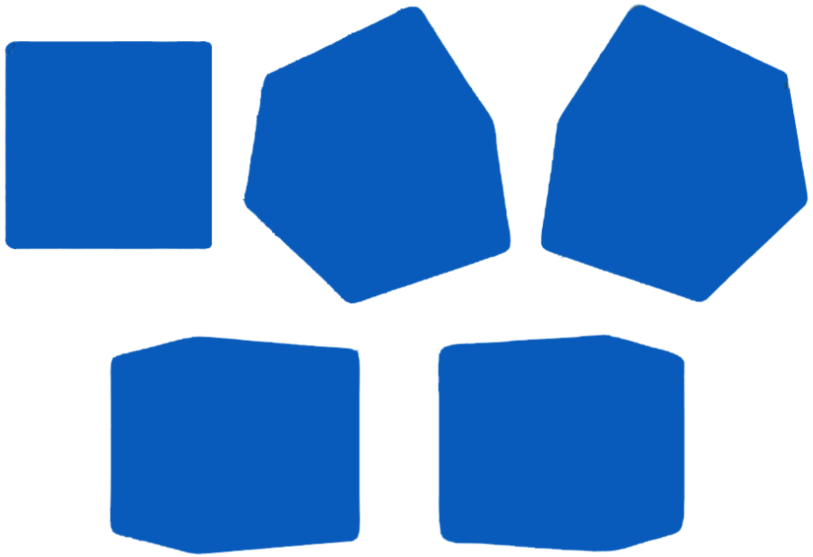
- The more the lines converge, the farther away they appear.



(Based on [2])

Visual Perception - Properties of the human visual system

Kinetic Depth Effect (Structure from Motion)



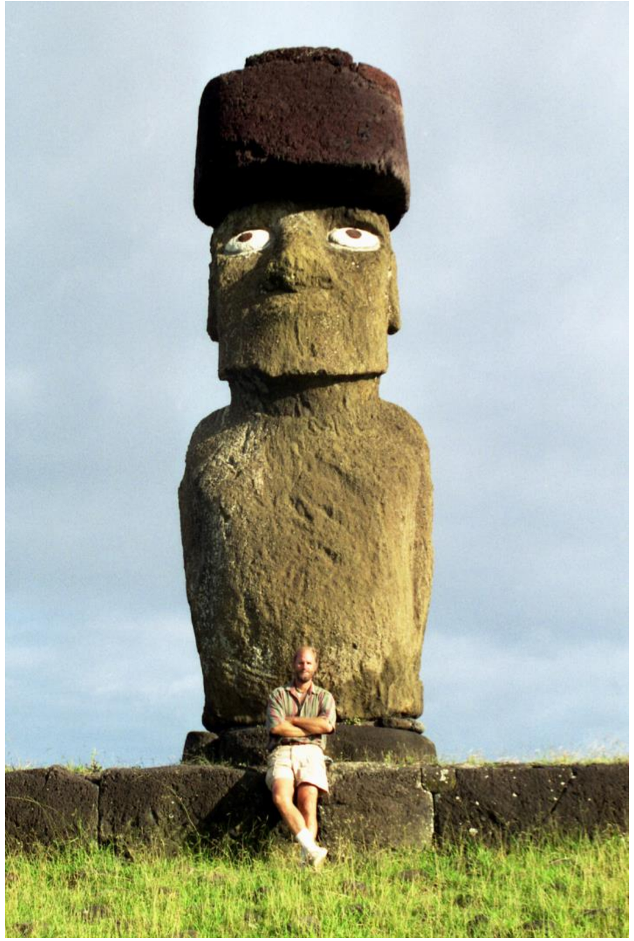
Kinetic depth effect is the perception of an object's complex, 3D structure from that object's motion.

(Based on [2])

Visual Perception - Properties of the human visual system

Familiar Size

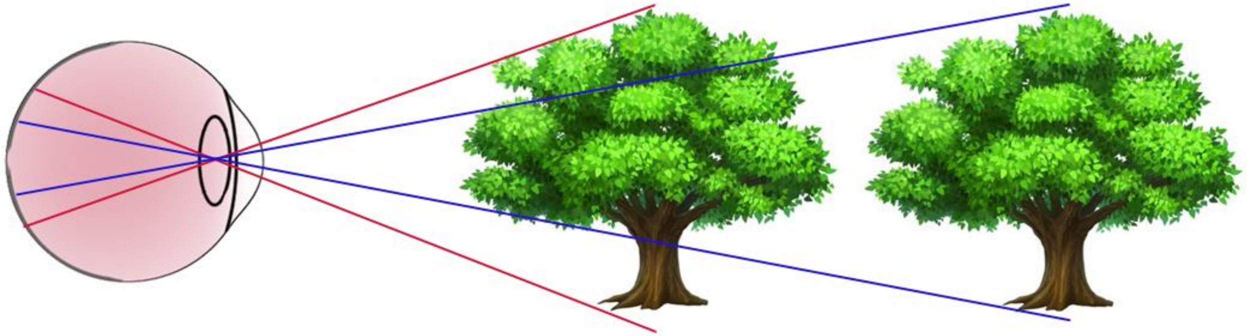
- If we know the size of an object in the far field, our brain uses that understanding to estimate absolute distances.
- This cue draws upon existing observer knowledge of an object in view.



(Based on [2])

Visual Perception - Properties of the human visual system

Relative Size



- If two objects are similar in size but offset in terms of their distances from the position of the observer, we perceive the one that casts a smaller image on the retina as being farther away, and the one with the larger image as being closer.
- This depth cue is heavily weighted based upon personal experience.

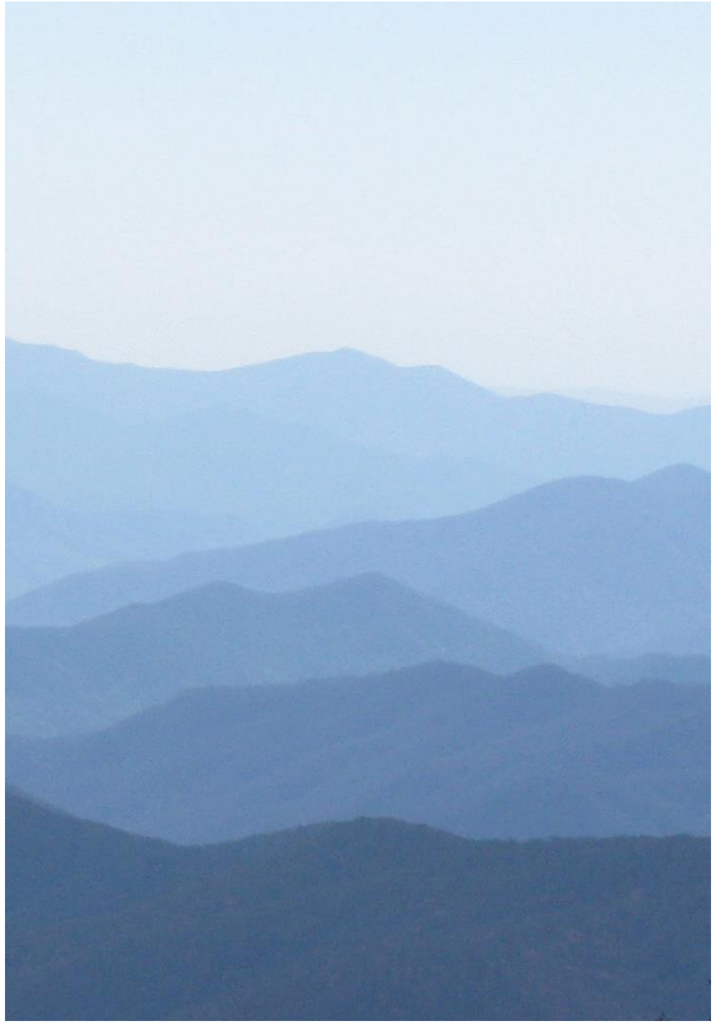
(Based on [2])

Visual Perception - Properties of the human visual system

Aerial Perspective

- This cue refers to the increased scattering of light by particles in the atmosphere as distance grows.
- Features become progressively less saturated and shift toward the background color with distance.
- Leonardo da Vinci referred to this cue as "the perspective of disappearance."

(Based on [2])



Visual Perception - Properties of the human visual system

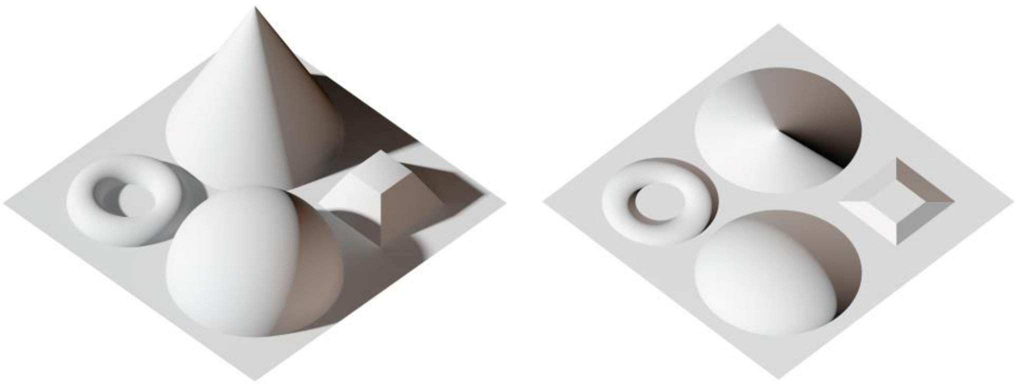
Texture Gradient

A gradual change in the appearance of textures and patterns of objects from coarse to fine (or less distinct) with distance.

(Based on [2])

Visual Perception - Properties of the human visual system

Lighting/Shading/Shadows



- Shadows and reflections cast by one object on another provide information about distances and positioning.
- Smaller, more clearly defined shadows typically indicate a close proximity of an object to the surface or object upon which the shadow is cast.
- Similarly, the perception of greater depth can be influenced by the enlarging of a shadow and the blurring of edges.

(Based on [2])

Visual Perception - Properties of the human visual system

Optical Expansion

As an object moves closer to an observer, the image projected on your retina also grows in size isotropically and occludes an increasing amount of the background. This cue allows an observer to perceive not only that an object is moving, but the distance of the object as well.

(Based on [2])

Visual Perception - Properties of the human visual system

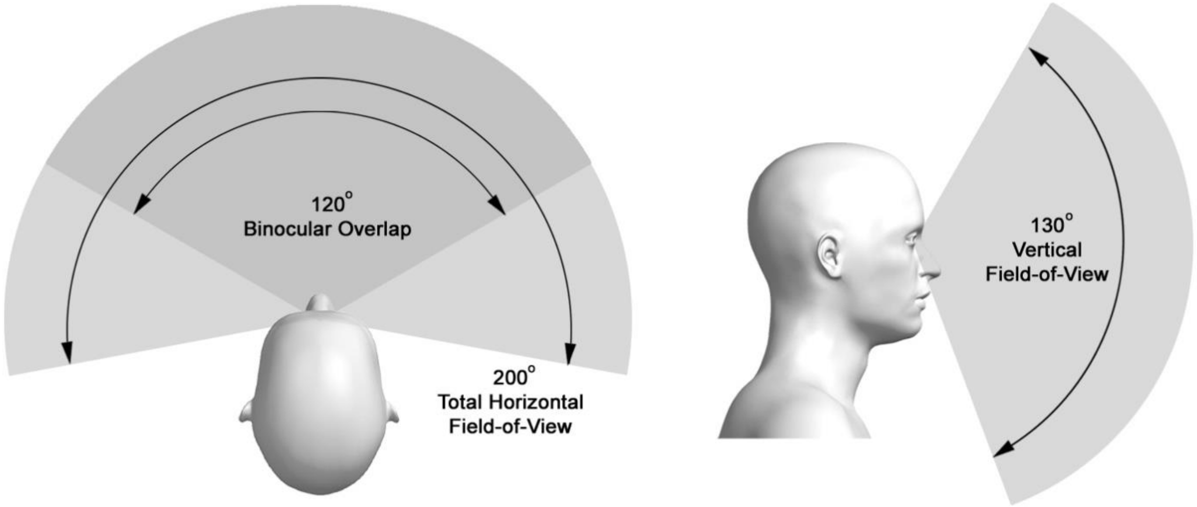
Relative Height

- Objects on a common plane in the near field of your vision are projected onto a lower portion of the retinal field than objects that are farther away. The key is your height relative to the objects in your field of view.
- Conversely, if objects being viewed are on a common plane above your viewpoint, such as a line of ceiling lanterns, objects closest to you will appear in a higher portion of the retinal field than objects that are farther away.

(Based on [2])

Visual Perception - Properties of the human visual system

- **Field of View (FOV)** - Defined as the total angular size of the virtual image visible to both eyes, expressed in degrees.
- For a healthy adult, the average horizontal binocular FOV is 120 degrees, with a total FOV measuring approximately 200 degrees. Our vertical FOV is approximately 130 degrees.

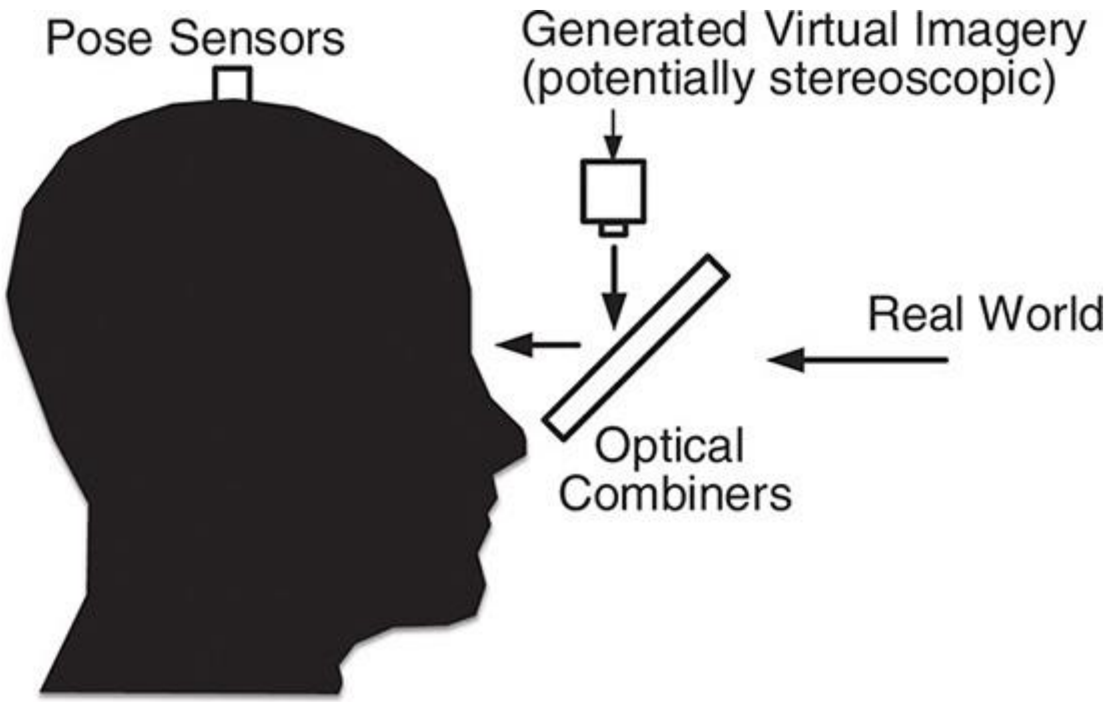


(Based on [2])

Requirements and Characteristics

Method of Augmentation

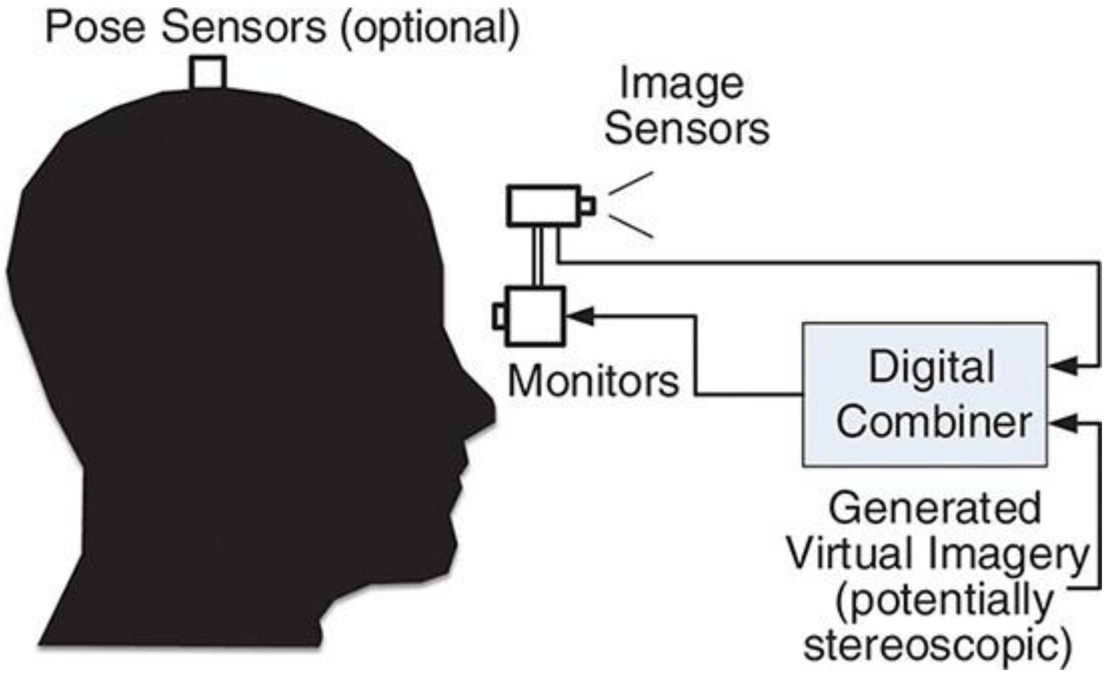
- Optical see-through (OST)



Requirements and Characteristics

Method of Augmentation

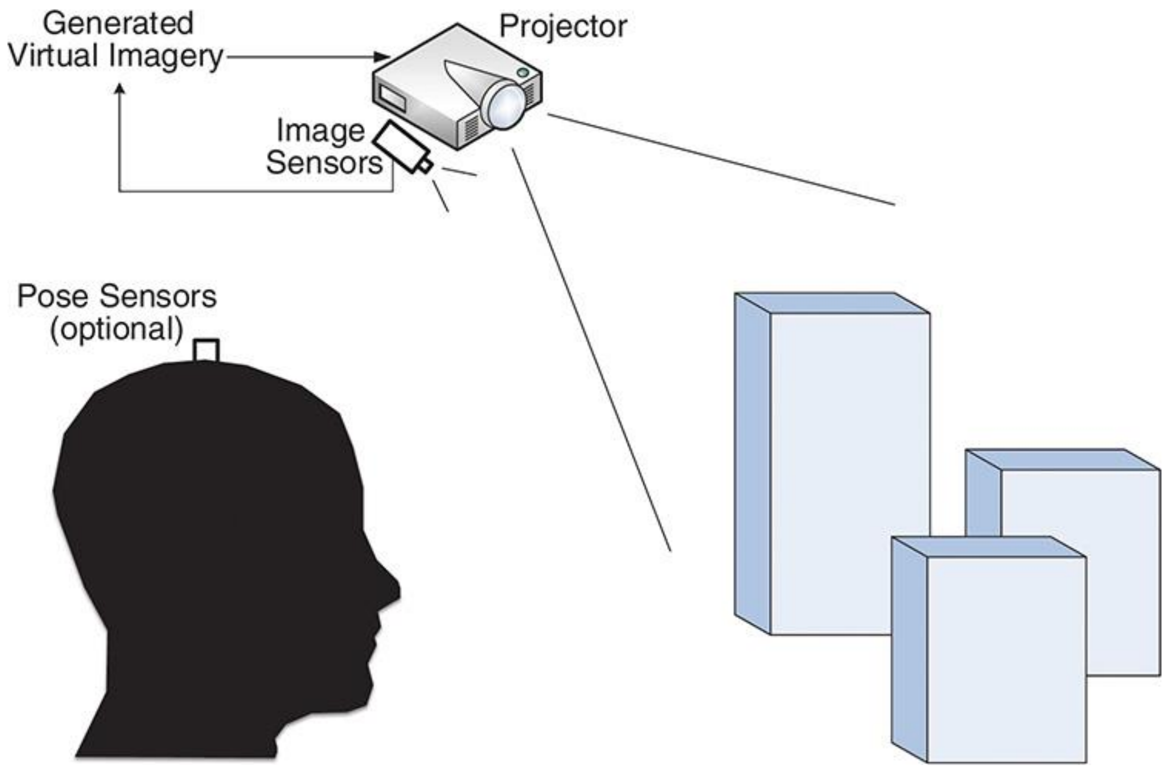
- Video see-through (VST)



Requirements and Characteristics

Method of Augmentation

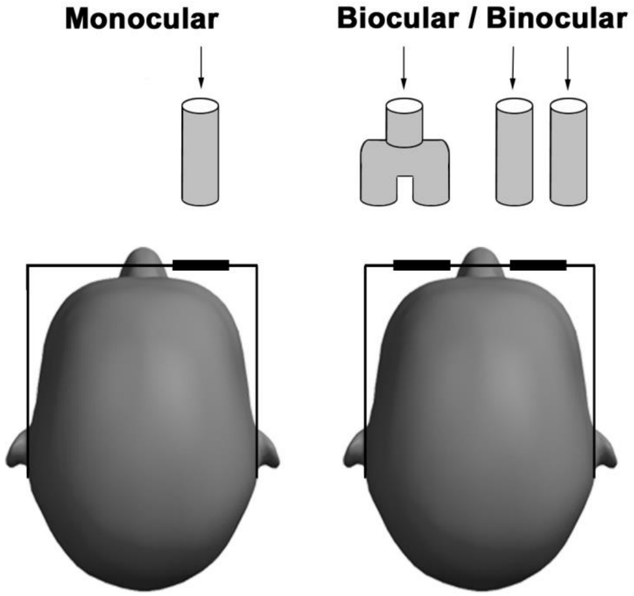
- Spatial Projection (SP)



Requirements and Characteristics

Ocularity and Stereoscopy

- Ocularity



- Head-mounted displays can be categorized by ocularity, or the specification of their design in serving one or two eyes.

Requirements and Characteristics

Ocularity

- Monocular – Presents images to only one eye

New Google Glass Enterprise Edition 2 with safety frames by Smith Optics.

Requirements and Characteristics

Ocularity

- Biocular – Presents the same image to both eyes

Samsung Gear VR

Requirements and Characteristics

Ocularity

- Binocular – Presents a separate image to each eye, resulting in a stereoscopic effect.

Oculus Quest 2

Magic Leap 1

Requirements and Characteristics

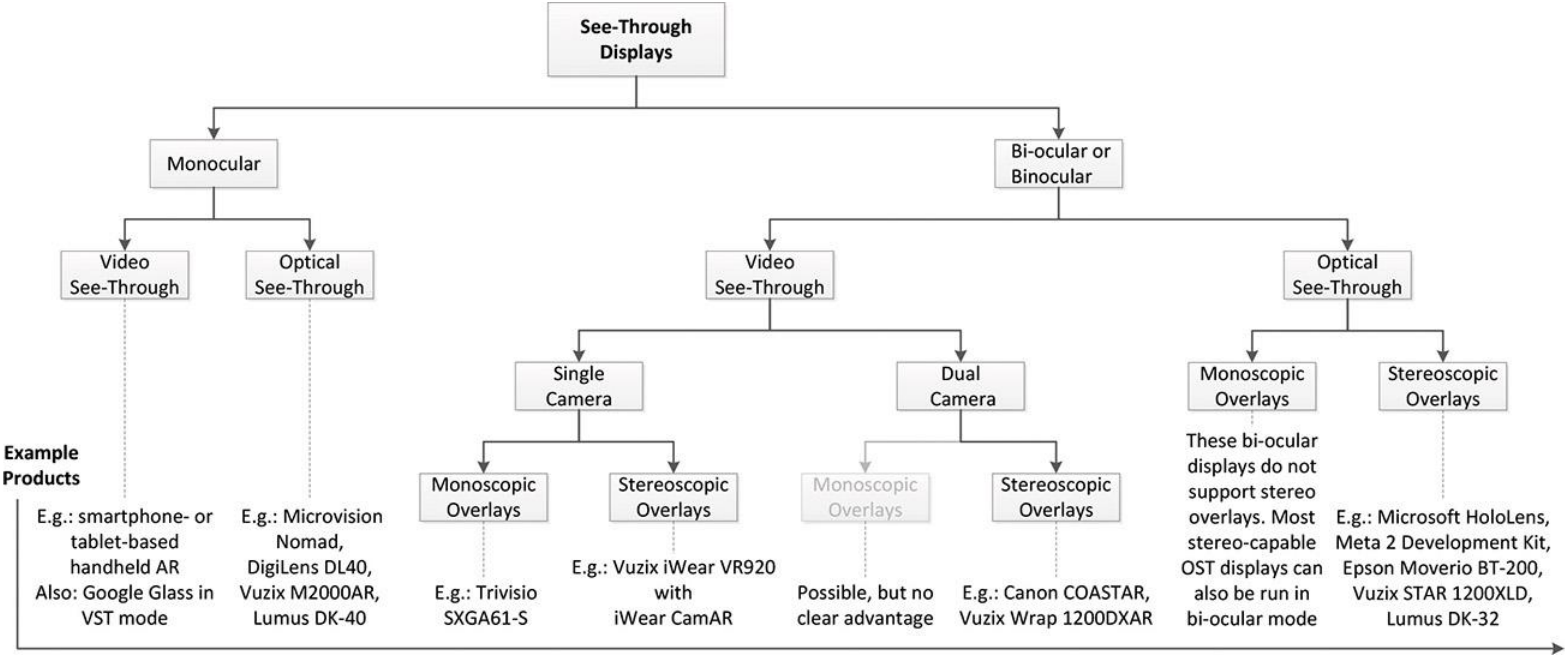
Ocularity

Table 4.1 Ocular Design Advantages and Disadvantages

Ocularity	Advantages	Disadvantages
Monocular	Low weight, small form factor, least distracting, easiest integration, least computational overhead, easiest alignment.	Possibility of binocular rivalry and eye-dominance issues, small field of view (FOV), no stereo depth cues, asymmetric mass loading, reduced perception of low-contrast objects, no immersion.
Biocular	Less weight than binocular, no visual rivalry, useful for close-proximity training tasks requiring immersion, symmetric mass loading.	Increased weight, limited FOV and peripheral cues, no stereo depth cues, often lens is larger to accommodate larger eye box.
Binocular	Stereo images, binocular overlap, larger FOV, most depth cues, sense of immersion.	Heaviest, most complex, most expensive, sensitive to alignment, computationally intensive operation.

Requirements and Characteristics

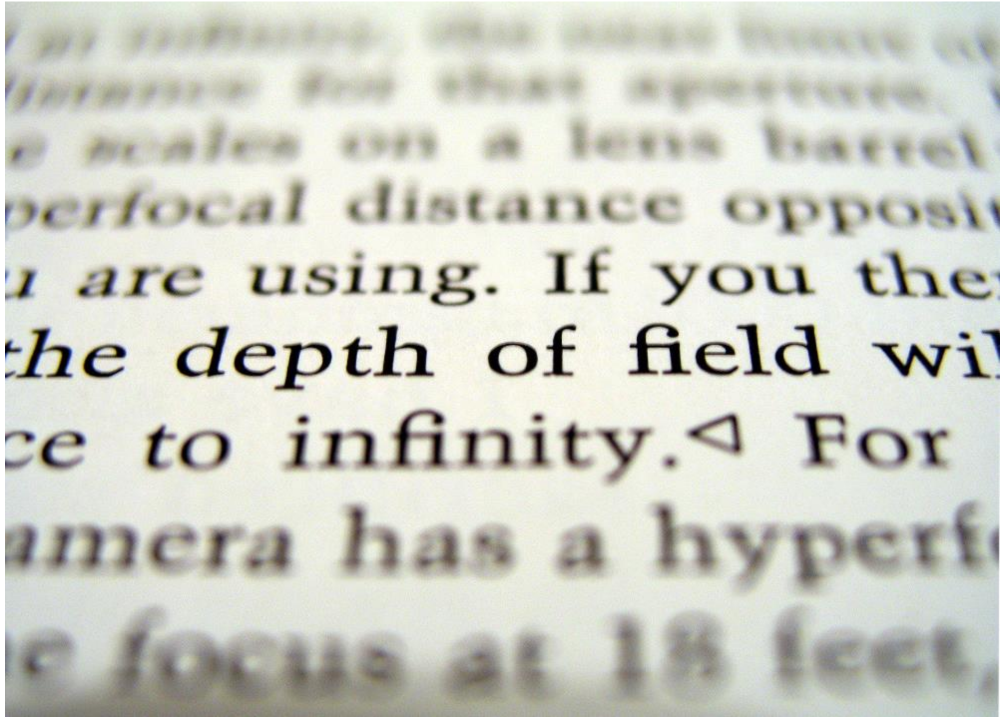
Categorization of see-through displays based on stereo capabilities.



Requirements and Characteristics

Focus

- Human visual system has a limited ***depth of field (DOF)***: only a certain range of objects will be in focus, and everything outside this range will be blurred.
- ***Accommodation-vergence*** processes allow the human eye to vary the depth of field to focus on a wide variety of distances (7cm to infinite).

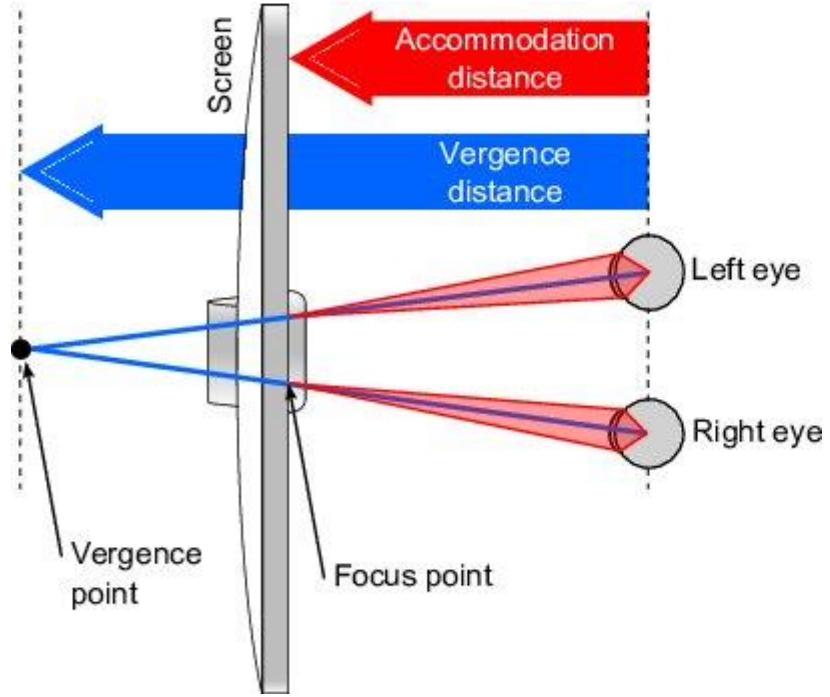


[Wikipedia]

Requirements and Characteristics

Focus

- Any display screen viewed with the naked eye or through a conventional optical system has a fixed focal depth.
- In a stereoscopic display, the object's actual distance is signaled to the human visual system via stereo parallax, resulting in a certain vergence response.
- The problem of accommodation to a fixed focal depth and the vergence to deal with stereo parallax cause the accommodation-vergence conflict (AVC)!



[Matthieu Urvoy *et al.* 2013]

Requirements and Characteristics

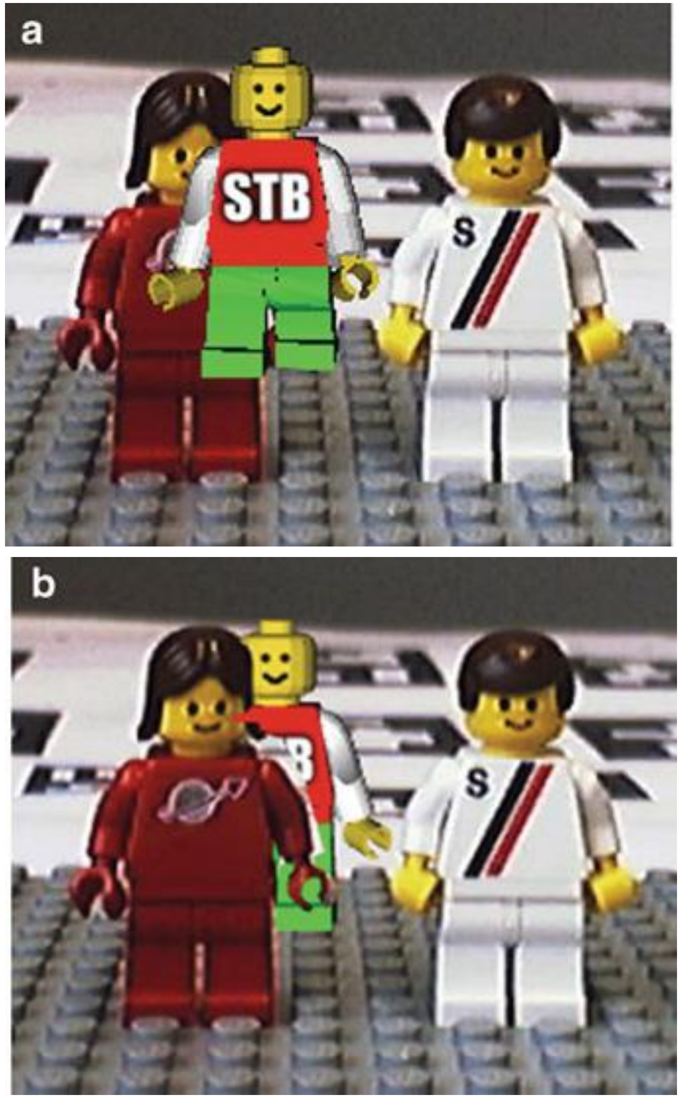
Focus

- One solution to this problem might be a display that can shift the focus plane in real time.
- With such technology, 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.
- In conjunction with eye tracking, the real-time rendering of realistic ***blurring effects*** outside the depth of field may alleviate the AVC.
- In video see-through systems, the camera optics are responsible for delivering images that have the right focus.

Requirements and Characteristics

Occlusion

- Occlusion among real objects is naturally given
- Occlusion among virtual objects is easily achieved by means of a z-buffer
- Occlusion of virtual and real objects is more challenging!



Requirements and Characteristics - Occlusion

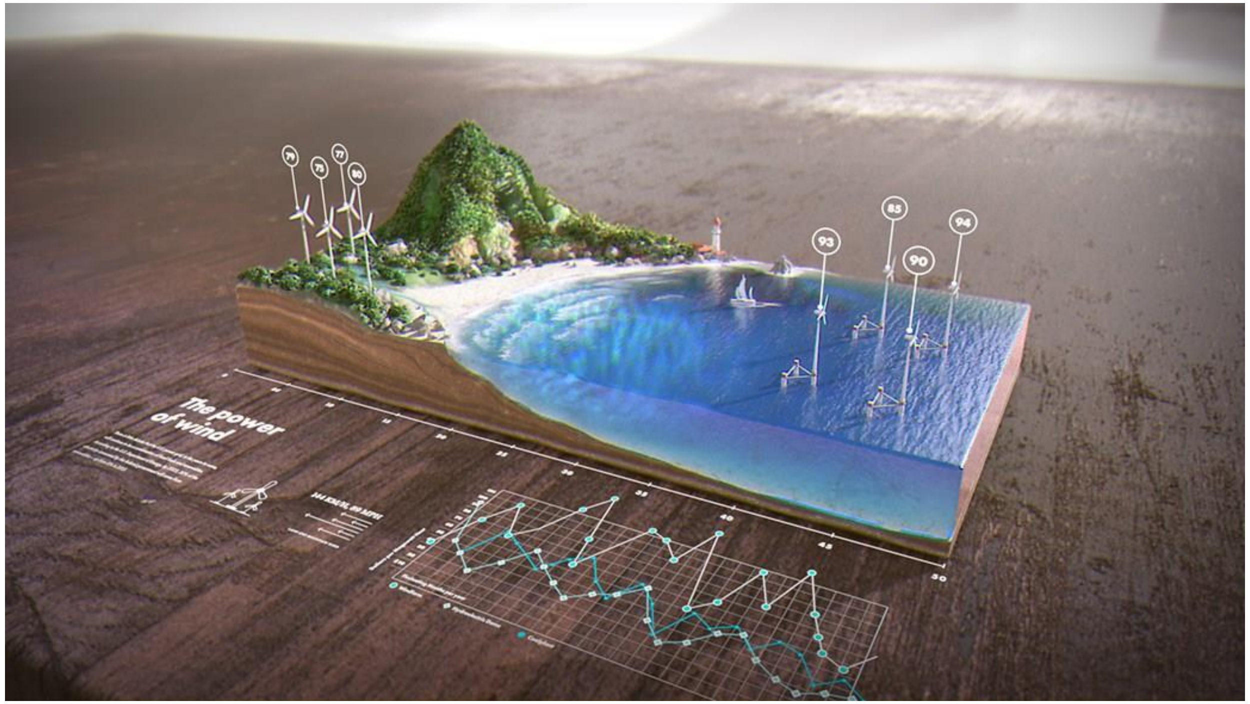
Occlusion of virtual and real objects

- A geometric representation of the real scene is necessary
- In a video see-through system is also possible with a z-buffer
- In optical see-through (OST) systems is more difficult to make virtual objects appear as if they are truly in front of real objects

Requirements and Characteristics - Occlusion

OST systems

- Virtual objects are rendered brighter relative to the intensity at which real objects are visible

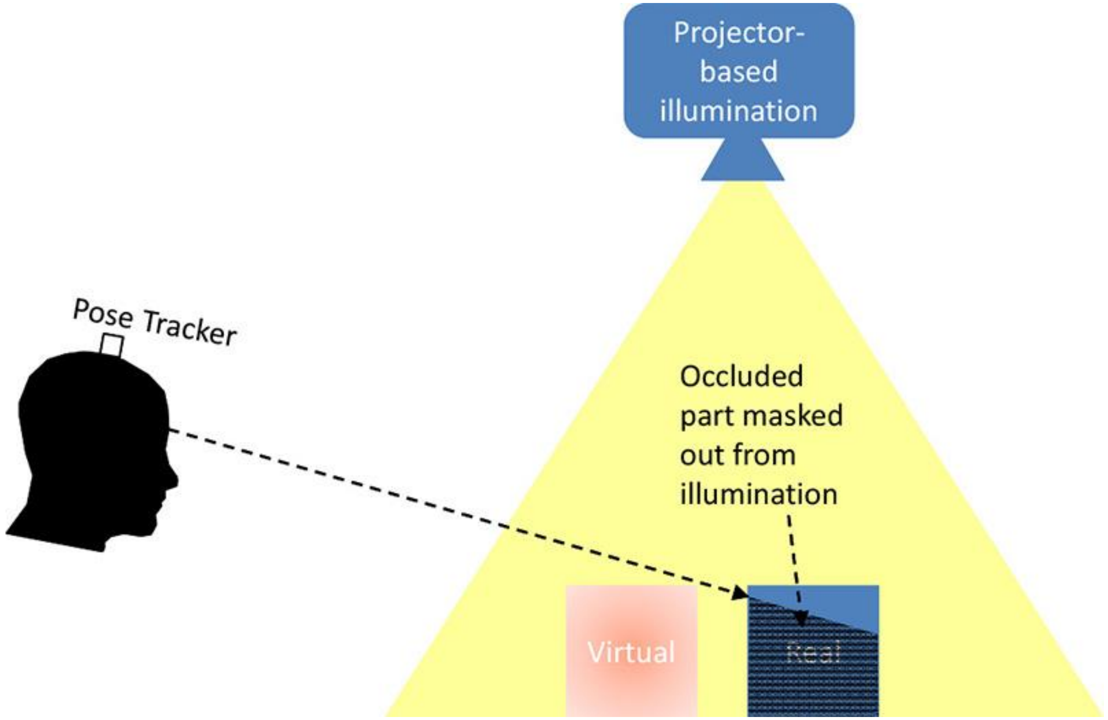


https://www.microsoft.com/en-us/hololens/hardware

Requirements and Characteristics - Occlusion

OST systems

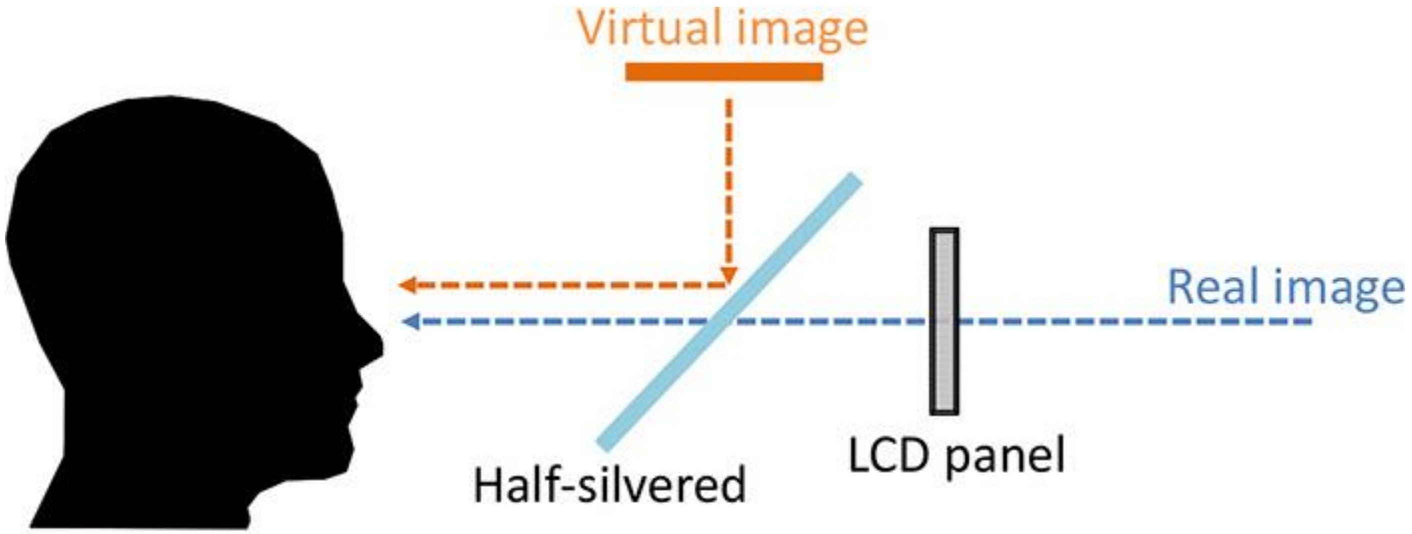
- In a controlled environment, the relevant part of the real scene can be illuminated with a computer-controlled projector



Requirements and Characteristics - Occlusion

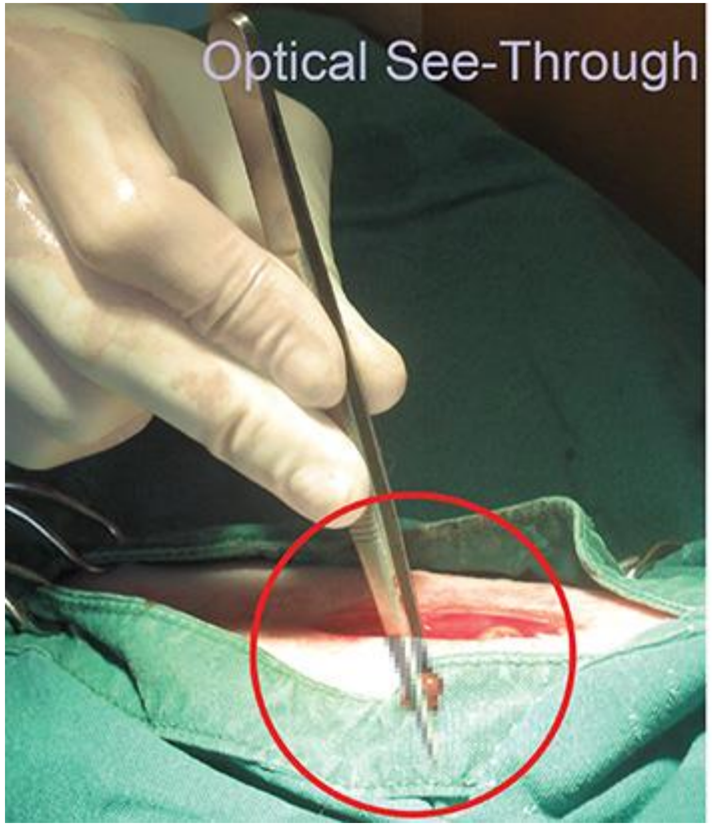
OST systems

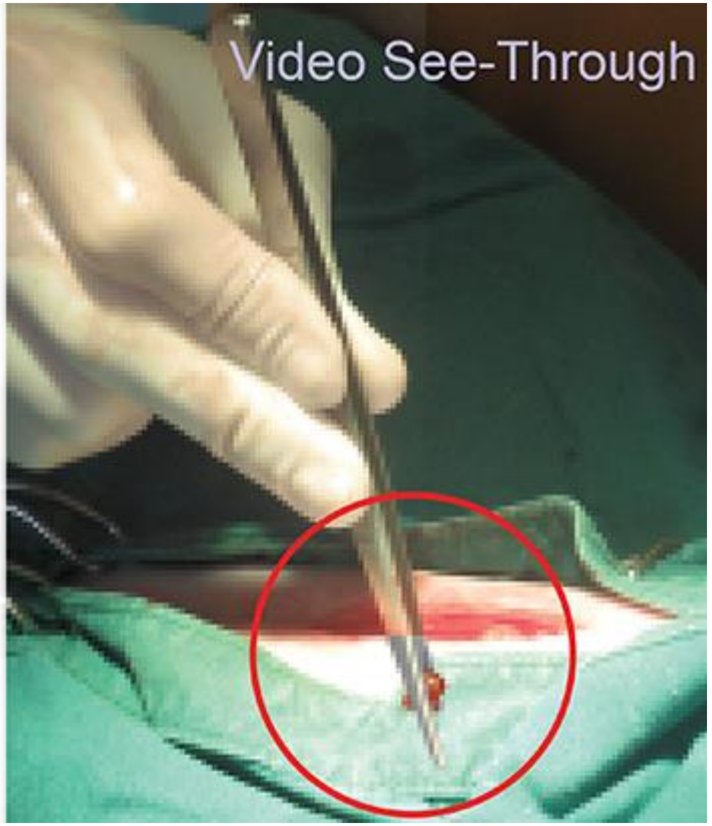
- An optically transparent display can be enhanced with a liquid crystal screen, which allows for selectively making individual pixels transparent or opaque.



Requirements and Characteristics

Resolution and Refresh Rate



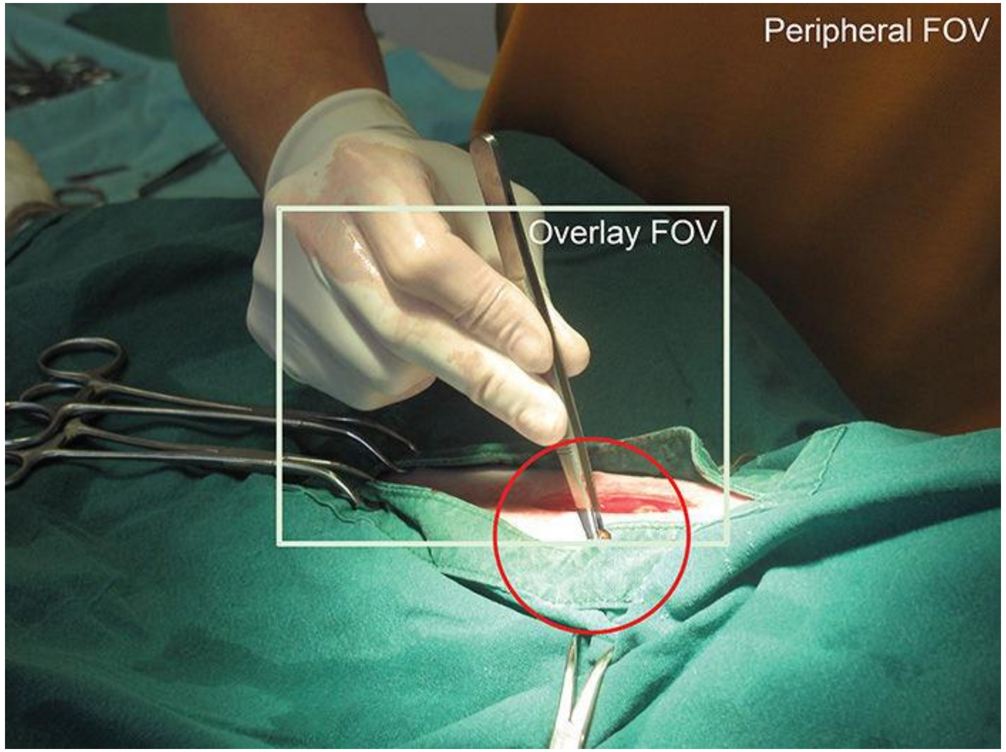


Requirements and Characteristics

Field of View

- HoloLens 2 has a diagonal FOV of 52° (34° first edition of HoloLens) and a resolution of 47 pixels per degree (2048 x 1080 per eye)

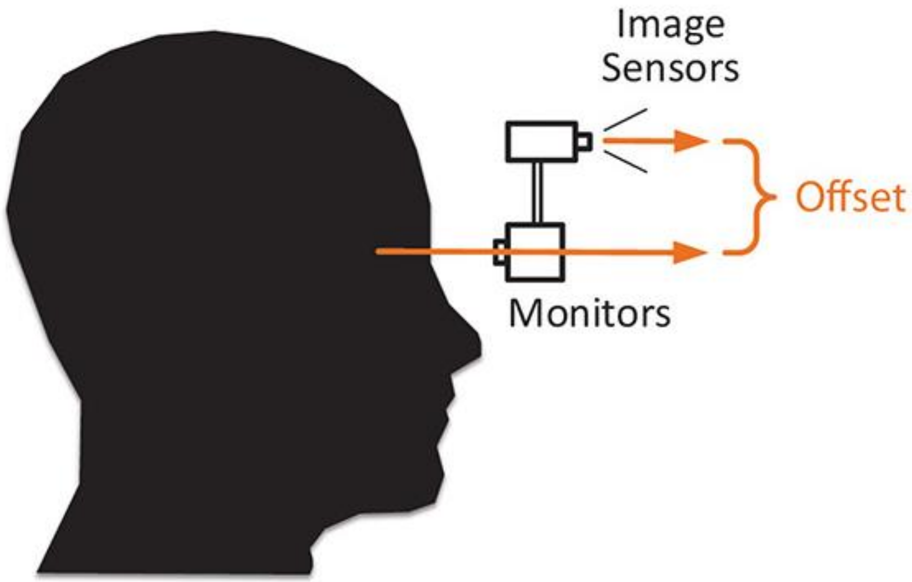
- Magic leap 1 has a diagonal FOV of 50° and a resolution of 1280 x 960 per eye.

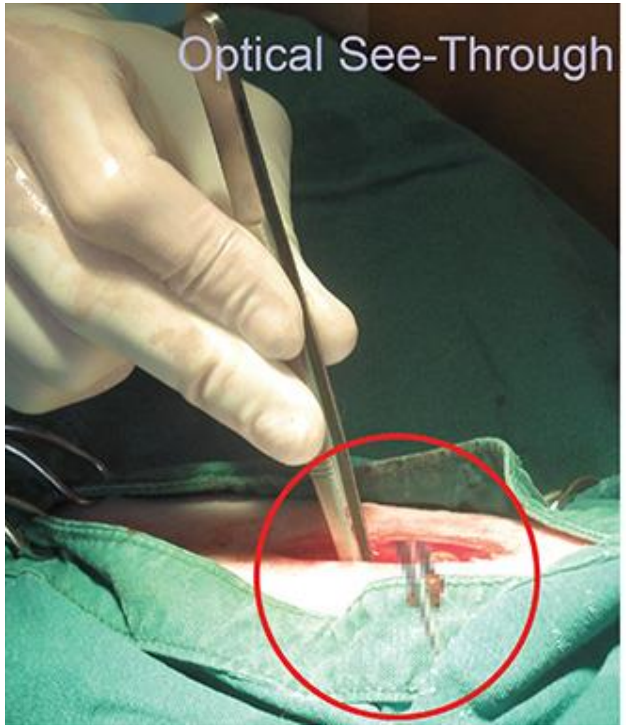


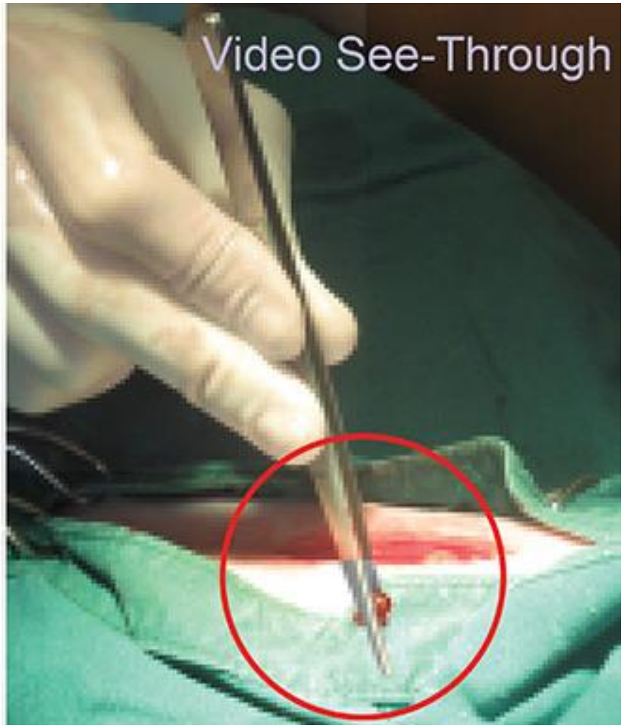
Peripherical FOV of 62° (diagonally) and overlay FOV is approximately 30°

Requirements and Characteristics

Viewpoint Offset



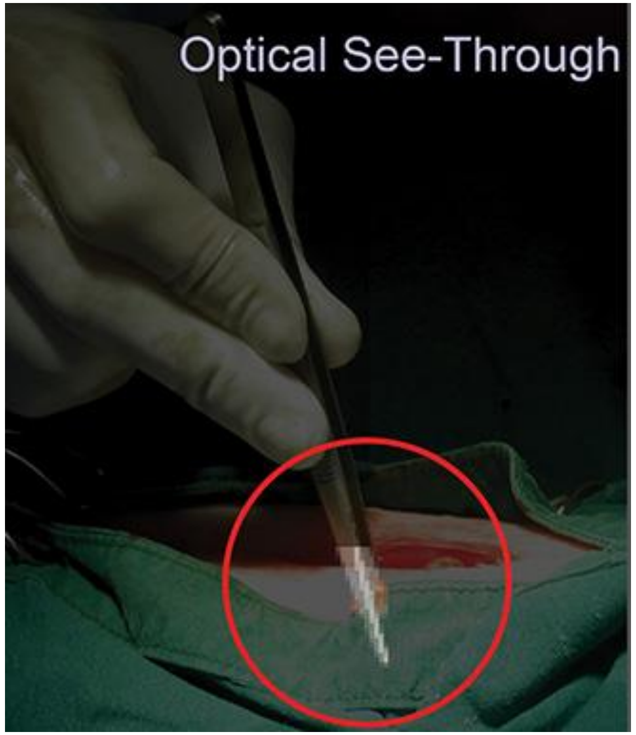


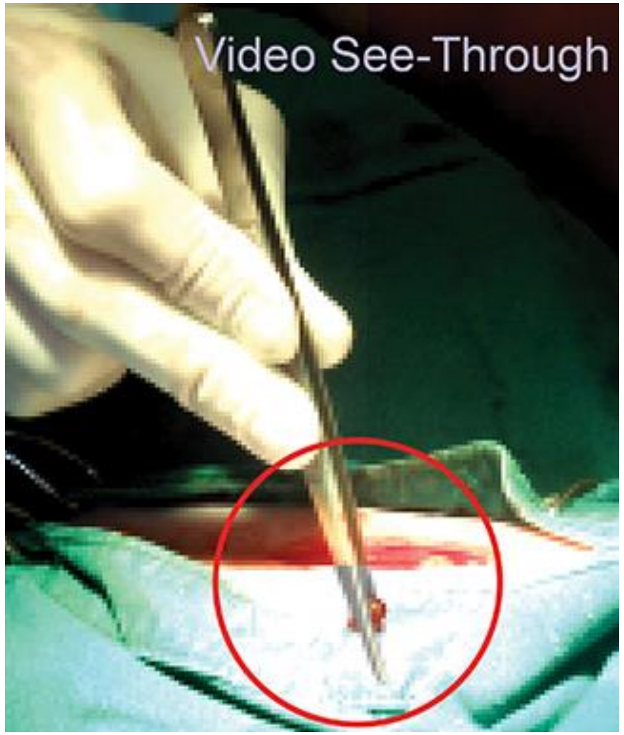


Requirements and Characteristics

Brightness and Contrast

- Optical see-through displays depend on the transparency of the optical combiner, while video see-through displays can change brightness and contrast arbitrarily

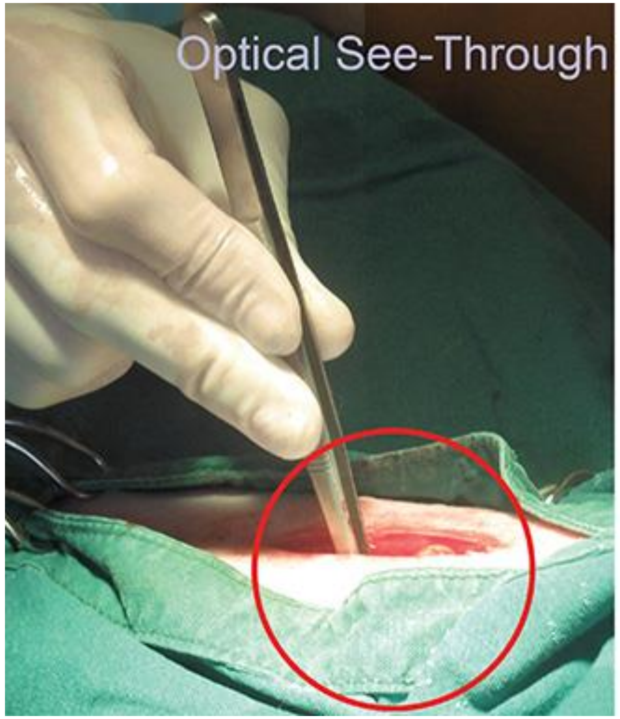


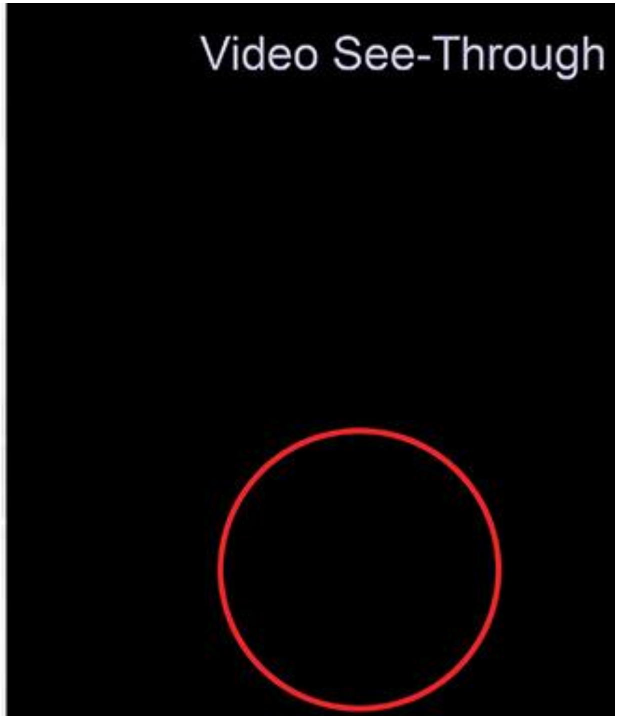


Requirements and Characteristics

Brightness and Contrast

- If the display fails, video see-through will not allow the user to see anything.





Requirements and Characteristics

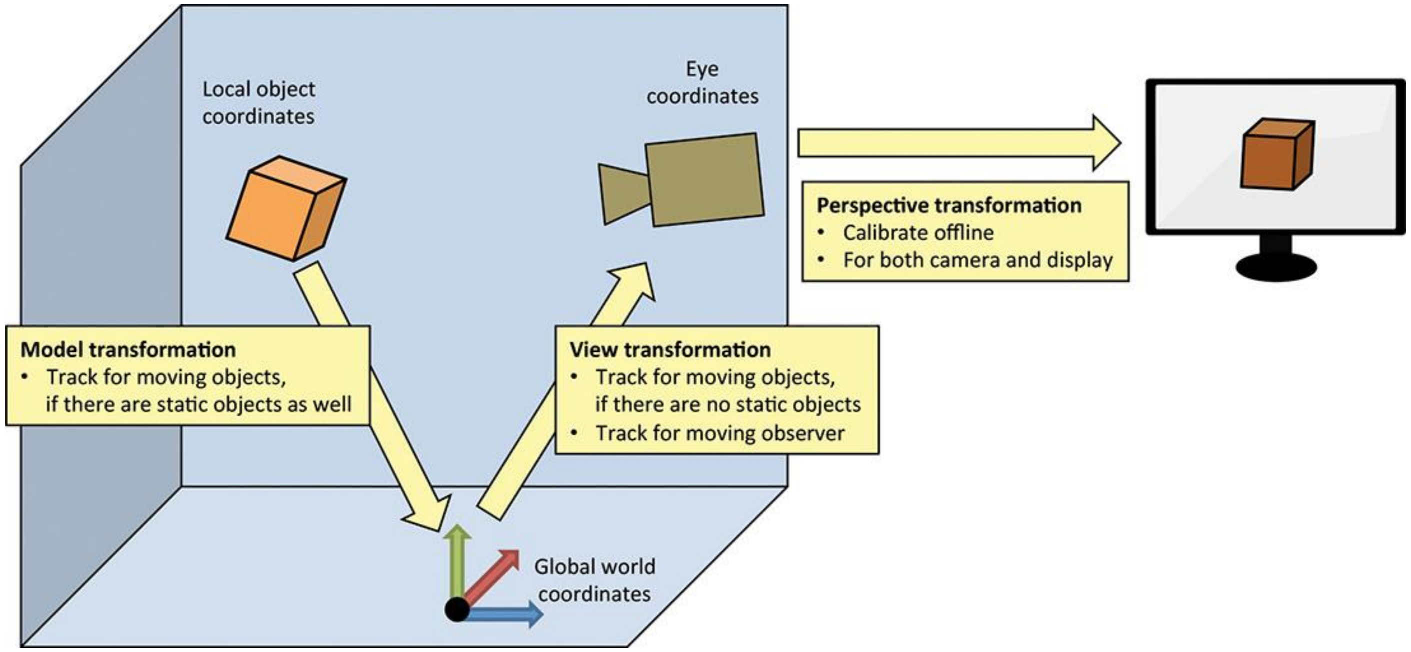
Other Characteristics

- Distortion and Aberrations
- Latency
- Ergonomics
- Social Acceptance

Spatial Display Model

- In AR, we rely on a standard computer graphics pipeline to produce overlays on the real world.

- This pipeline consists of:
  - Model transformation
  - View transformation
  - Projective transformation



Spatial Display Model

- **Model transformation**: describes the relationship of 3D local object coordinates and 3D global world coordinates. The model transformation describes how objects are positioned in the real world.
- **View transformation**: describes the relationship of 3D global world coordinates and 3D view (observer or camera) coordinates.
- **Projective transformation:** describes the relationship of 3D view coordinates and 2D device (screen) coordinates.

Spatial Display Model

Model transformation

- **Object tracking** is required if we are interested in moving real objects in the AR scene, while static object positions can be measured once and need not be tracked.

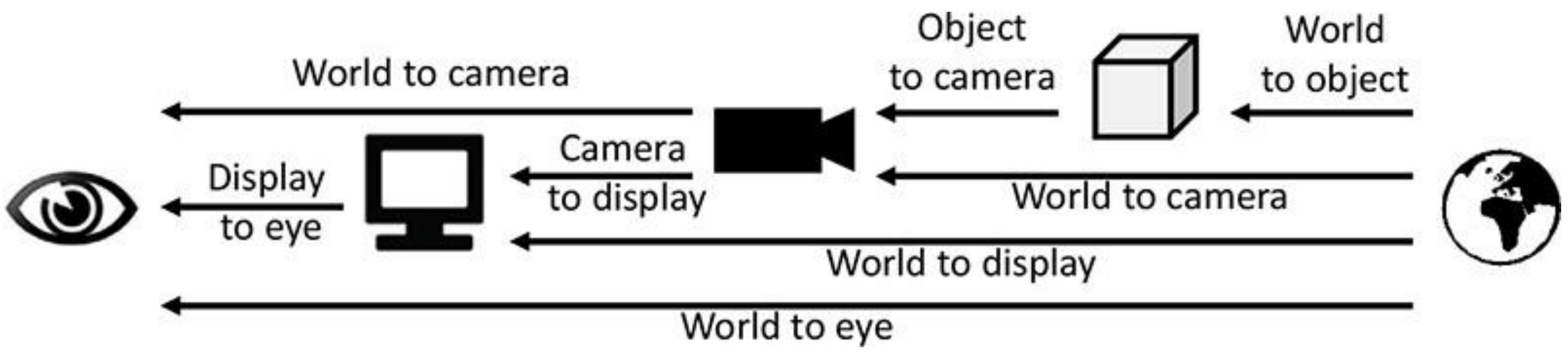
- Object tracking is used to set the model transformation.

Spatial Display Model

View transformation

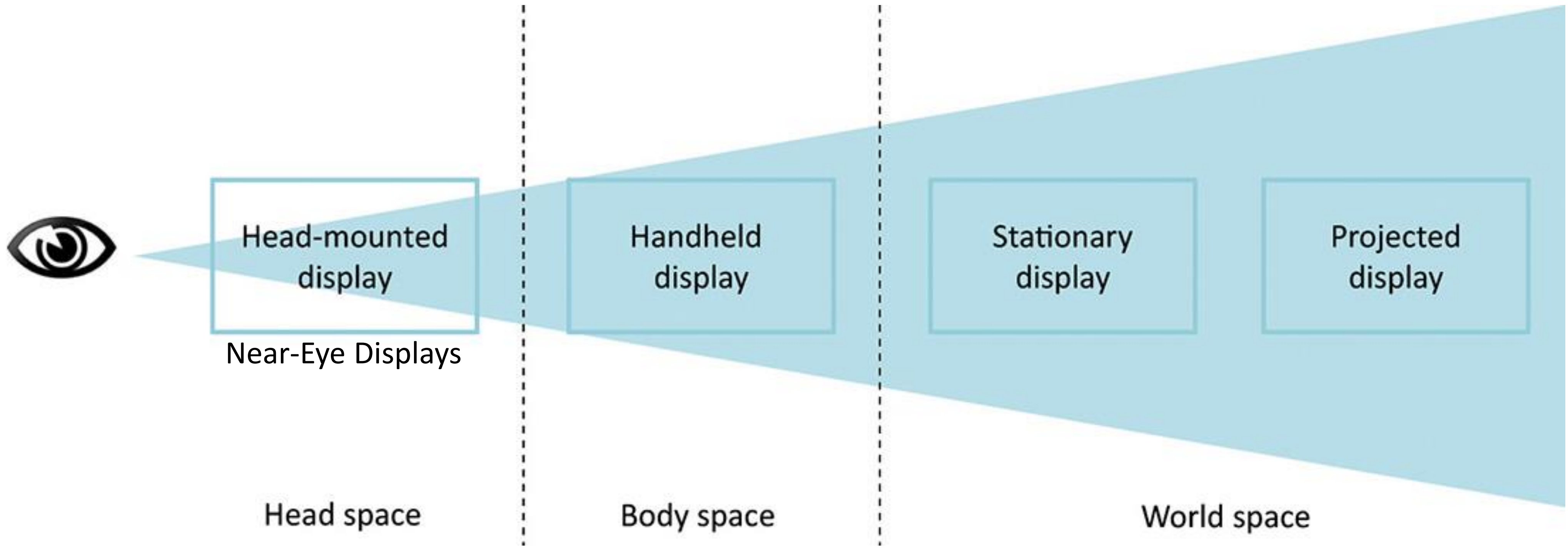
Can involve:

- head tracking and eye tracking (user is moving relative to the display)
- display tracking (display is moving relative to the world)
- camera tracking (a VST display is used)



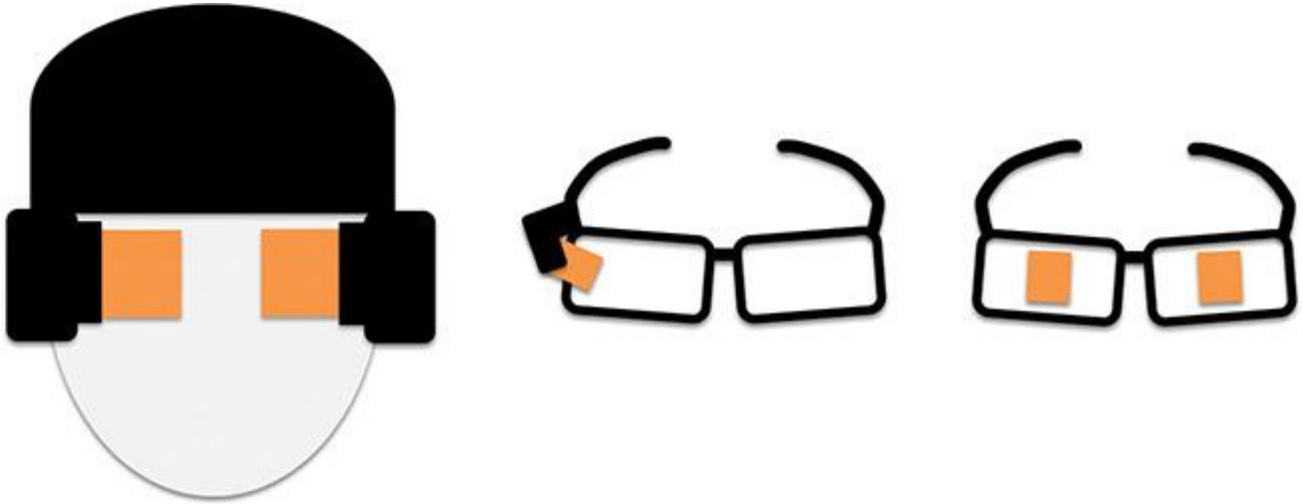
The spatial model of most AR displays can be defined as the spatial relationship of up to five components: the user’s eye, the display, the camera, an object to be augmented, and the world.

Visual Displays



Visual Displays – Near-Eye Displays

The most prominent class of displays for AR is probably the HMD



Different display mounting options. (left) Helmet-mounted display, like that used by [Rockwell Collins SimEye]. (middle) Clip-on display, like that used by [Google Glass]. (right) Visor display, like that used by [Epson Moverio].

Visual Displays – Near-Eye Displays

Optical See-Through Head-Mounted Display

Lumus DK-50 developer kit glasses

AtheerAiR Glasses

SmartEyeglass SED-E1

Visual Displays – Near-Eye Displays

Optical See-Through Head-Mounted Display

DAQRI Smart Helmet

nVisor ST50

Meta 2

Visual Displays – Near-Eye Displays

Optical See-Through Head-Mounted Display

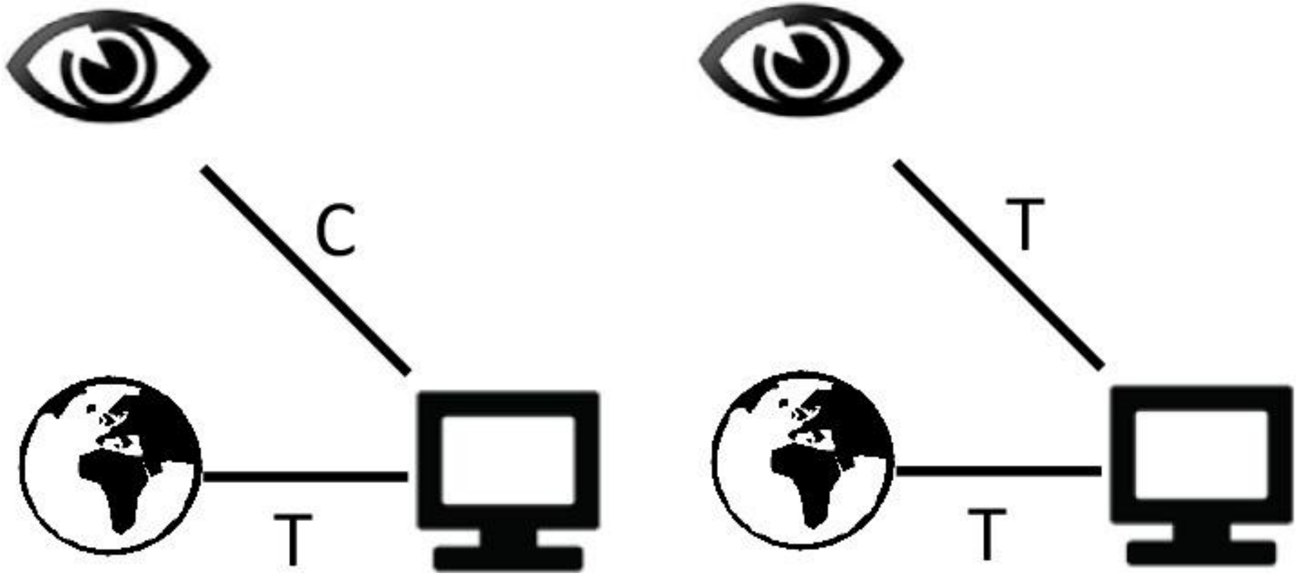
Microsoft HoloLens 2

Magic Leap 1

Visual Displays – Near-Eye Displays

Optical See-Through Head-Mounted Display

C – constant and calibrated transformations
T – transformation requires tracking



(left) Optical see-through head-mounted display. (right) Optical see-through head-mounted display with eye tracking.

Visual Displays – Near-Eye Displays

Video See-Through Head-Mounted Display

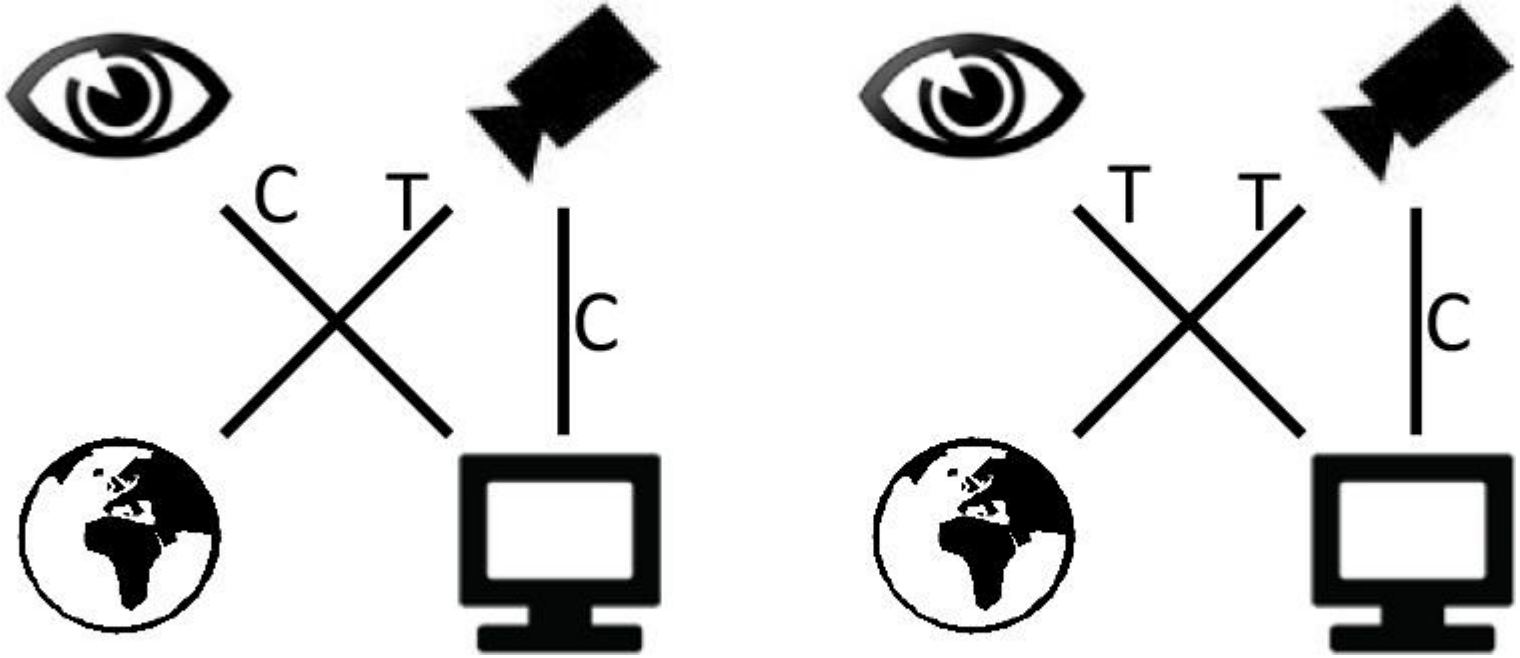
HTC Vive Pro

Samsung Gear VR

Google Cardboard

Visual Displays – Near-Eye Displays

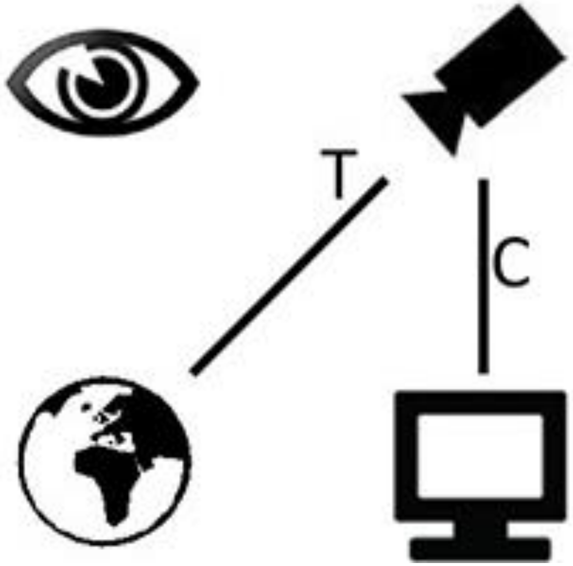
Video See-Through Head-Mounted Display



(left) Video see-through head-mounted display. (right) Video see-through head-mounted display with eye tracking.

Visual Displays – Handheld Displays

A handheld AR display can be built from an unmodified smartphone or tablet computer.

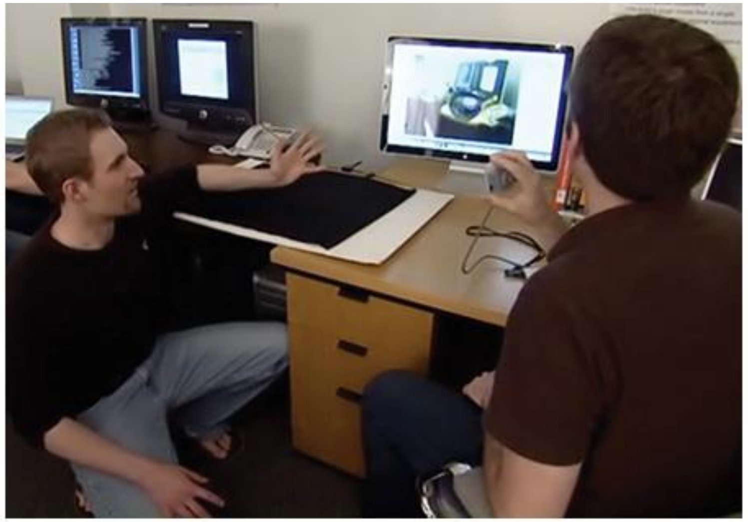


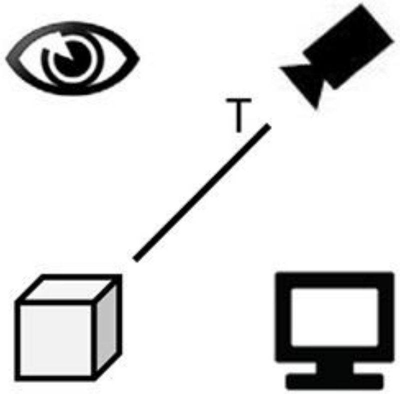
Visual Displays – Handheld Displays

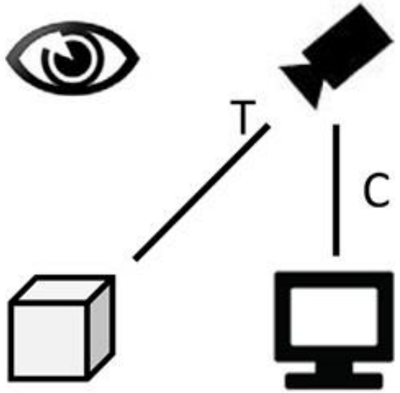
(left) Handheld display with device perspective. (middle) Handheld display with user perspective. (right) A user-perspective handheld display requires tracking both the camera and the user's viewpoint.

Visual Displays – Stationary Displays

Desktop Displays

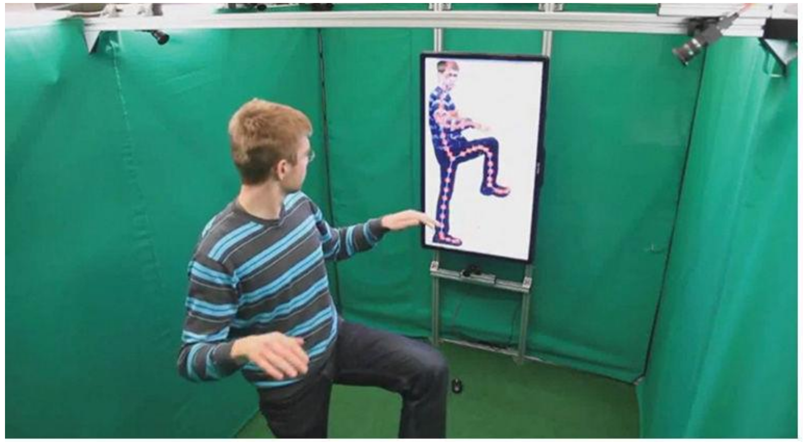


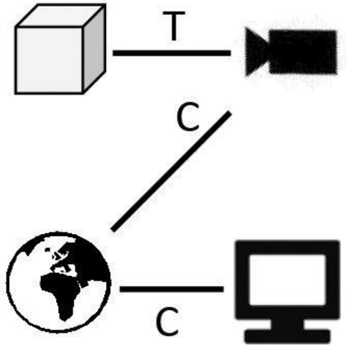


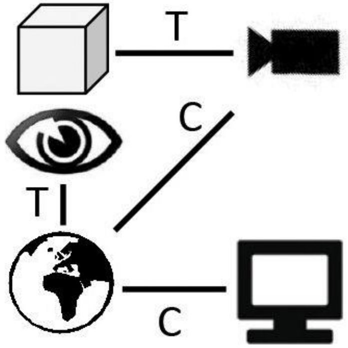


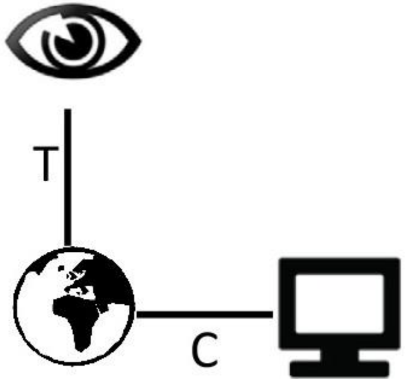
Visual Displays – Stationary Displays

Virtual Mirror



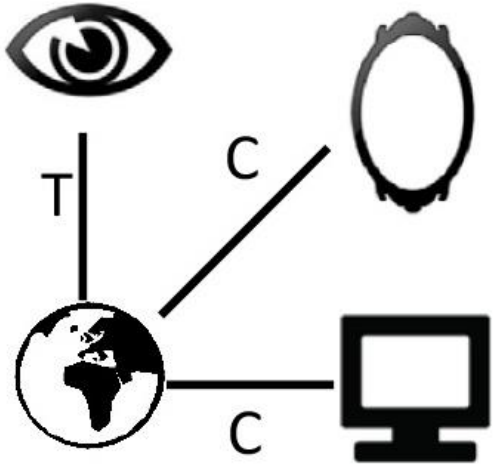






Visual Displays – Stationary Displays

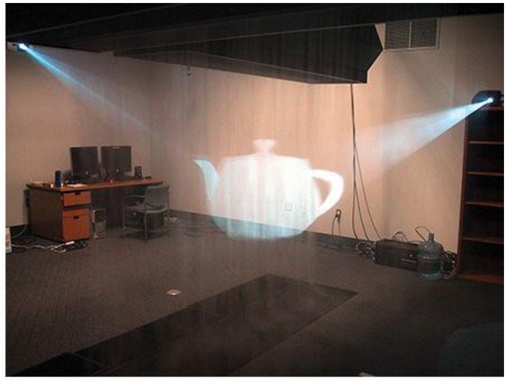
Virtual Showcase

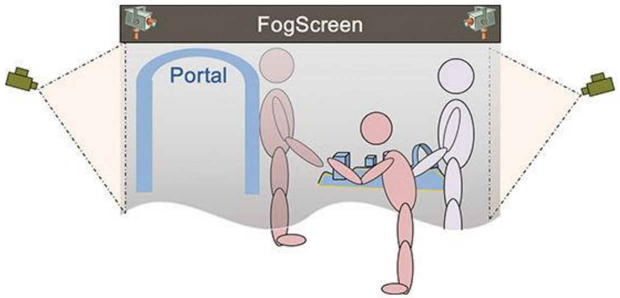


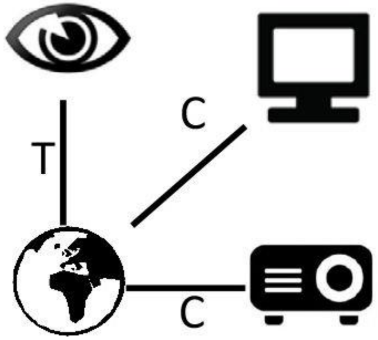
Visual Displays – Stationary Displays

Windows and Portal Displays

Samsung Transparent Smart Window display, showcased at CES 2012.



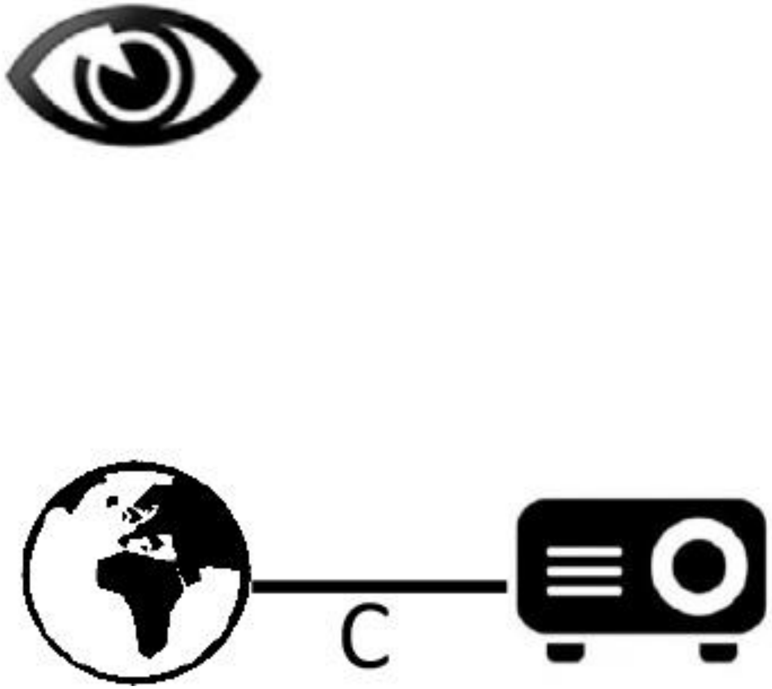




Visual Displays – Projected Displays

Spatial Augmented Reality

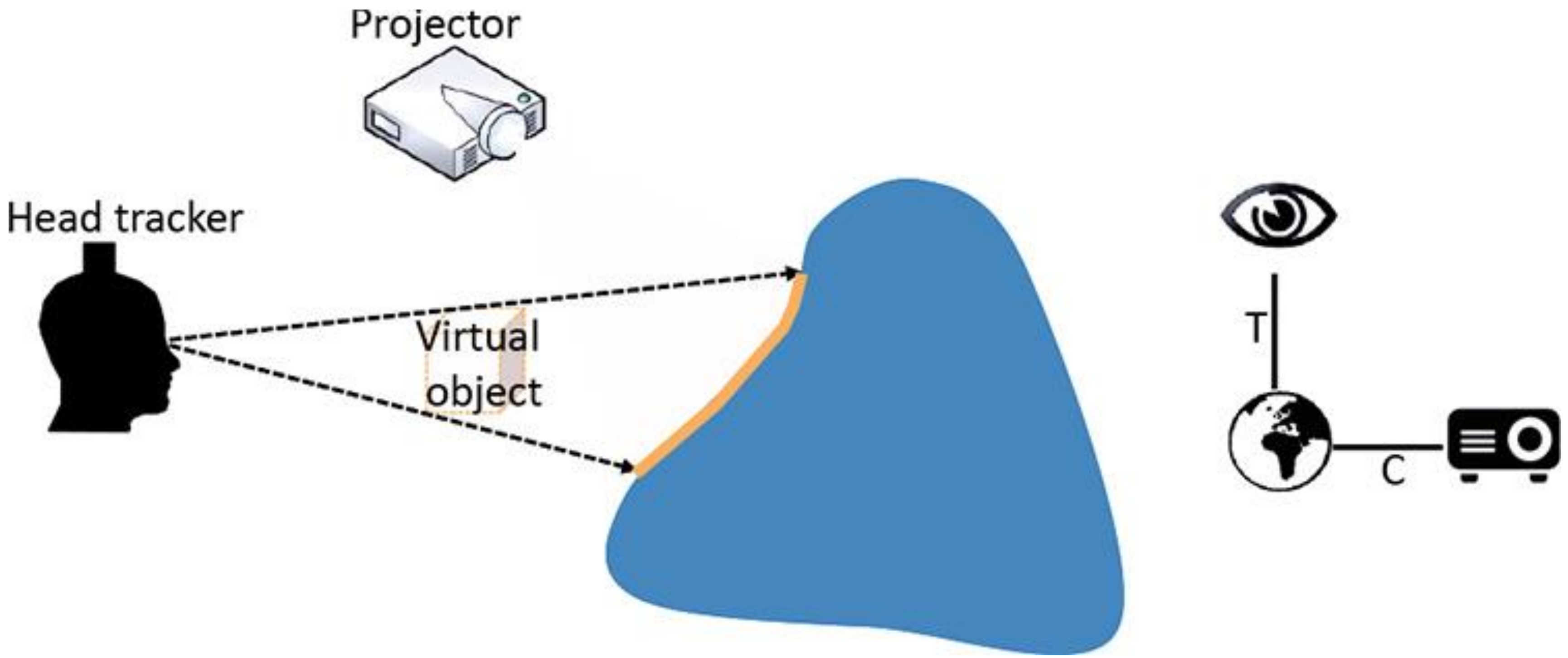
Spatial AR can be used to turn generic objects into textured models.



Simple spatial AR does not require any tracking, as long as the augmented scene is static.

Visual Displays – Projected Displays

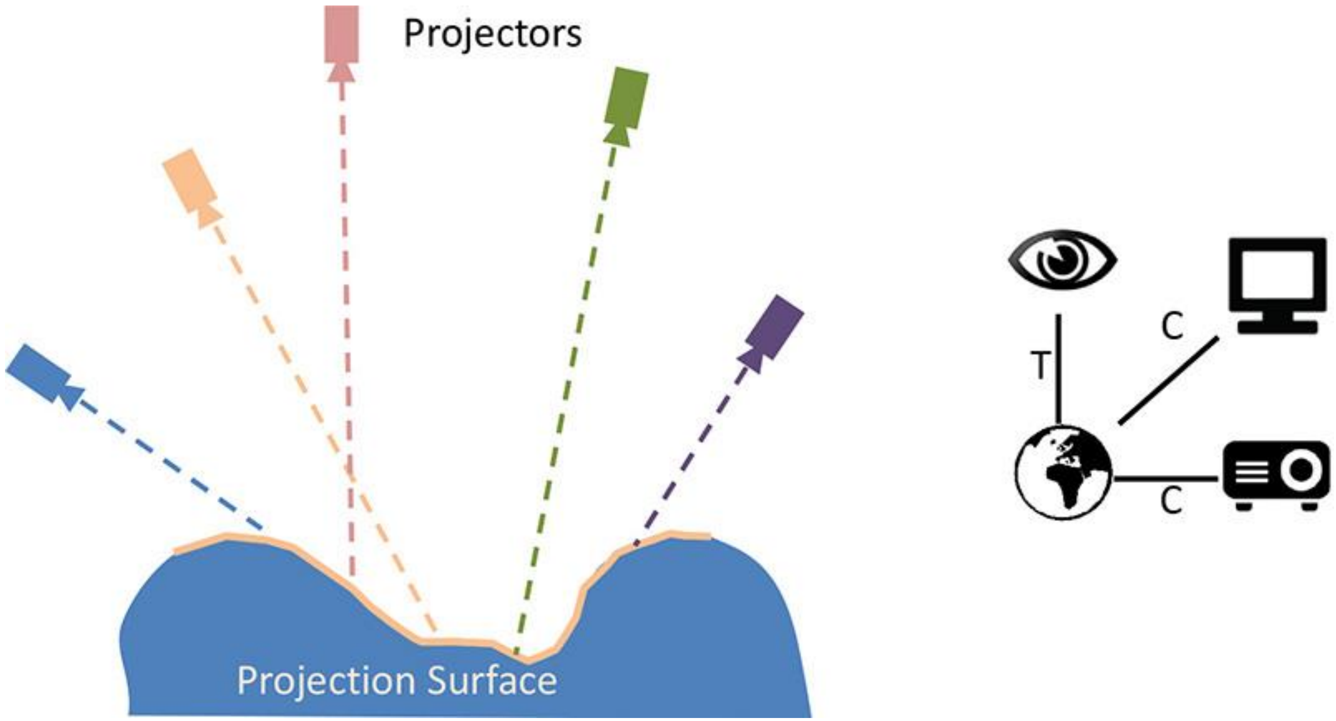
View-Dependent Spatial Augmented Reality



View-dependent spatial AR requires tracking the user, but can present free-space 3D objects.

Visual Displays – Projected Displays

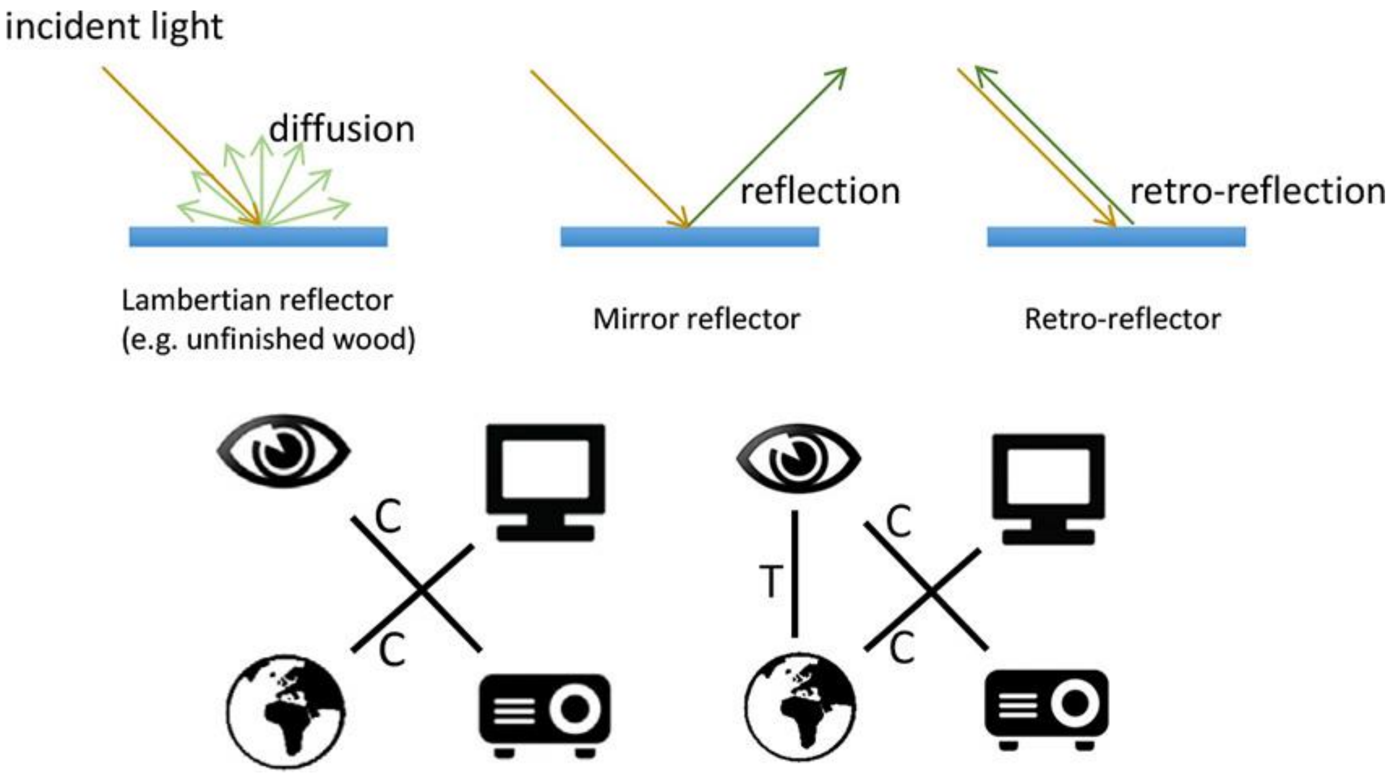
Multiple projectors



(left) Multiple projectors can be combined to minimize pixels projected out of focus. (right) The geometry of the projection surface needs to be known; it is represented here as a display calibrated to the world.

Visual Displays – Projected Displays

Head-Mounted Projector Displays

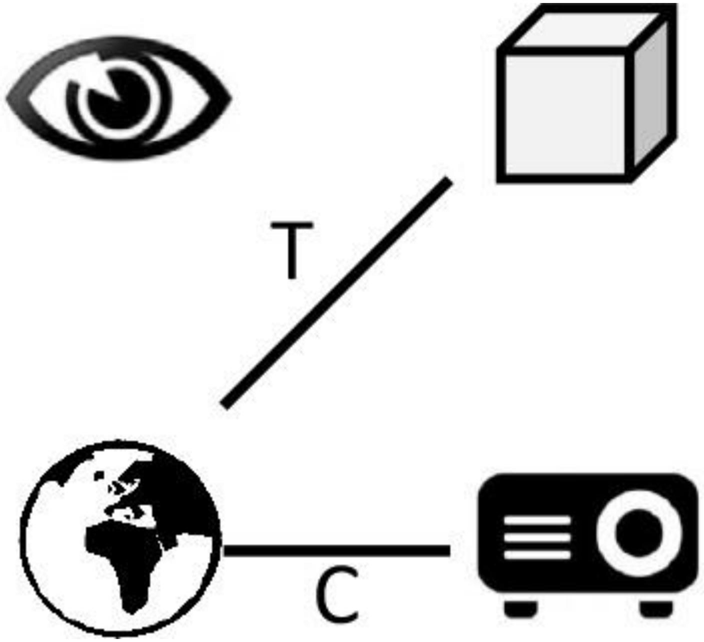


(top) Retro-reflective materials send incident rays back to the illuminating source, so they work well with head-mounted projector displays. (bottom) Spatial relationship schematics for HMPDs without and with head tracking. In the latter case (shown at right), virtual objects can be kept stable in space, while the viewer is moving.

Visual Displays – Projected Displays

Dynamic Shader Lamps

Two applications of dynamic shader lamps. (left) Painting with light on real surfaces. Courtesy of Michael Marner. (right) Animatronic character with animated facial projection. Courtesy of Greg Welch, University of North Carolina–Chapel Hill.

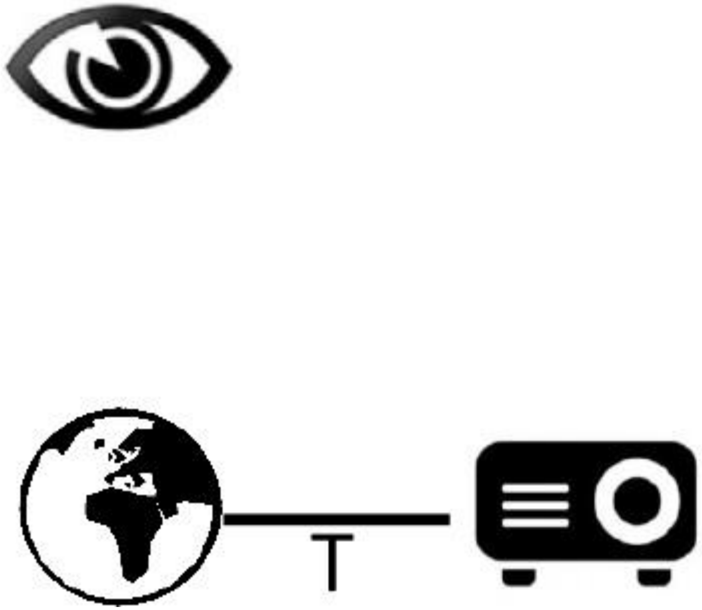


Dynamic shader lamps deliver spatial AR on tracked objects.

Visual Displays – Projected Displays

Everywhere Displays Projector

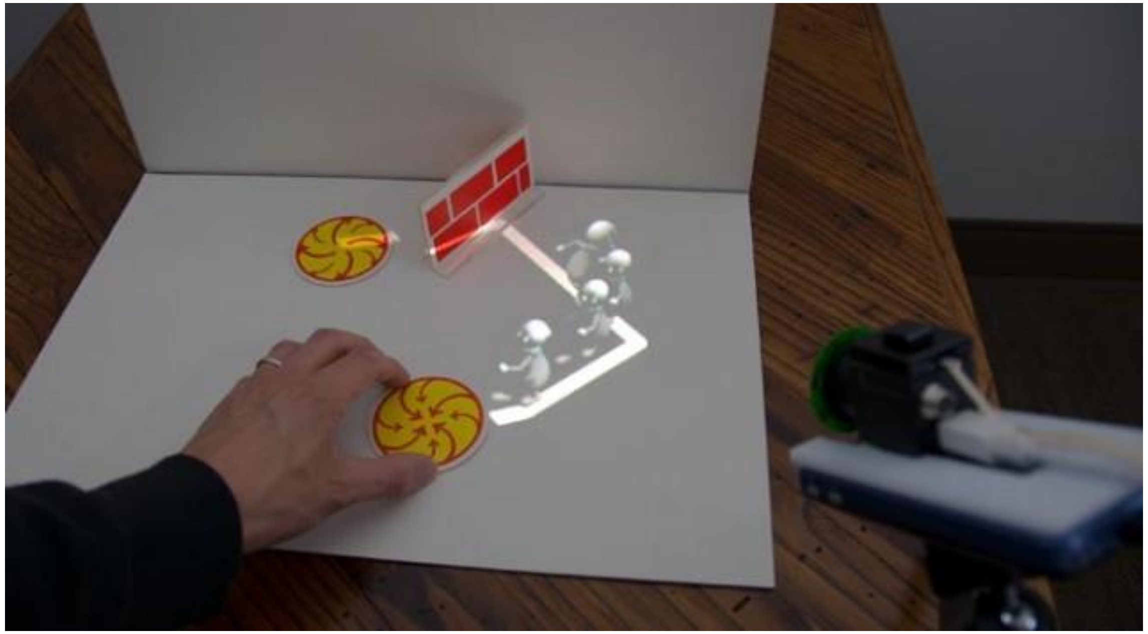
The Everywhere Displays Projector is based on tracked, steerable projection and can deliver content on any surface. For example, navigational hints can be displayed on product shelves.



A steerable, tracked projector can display images anywhere.

Visual Displays – Projected Displays

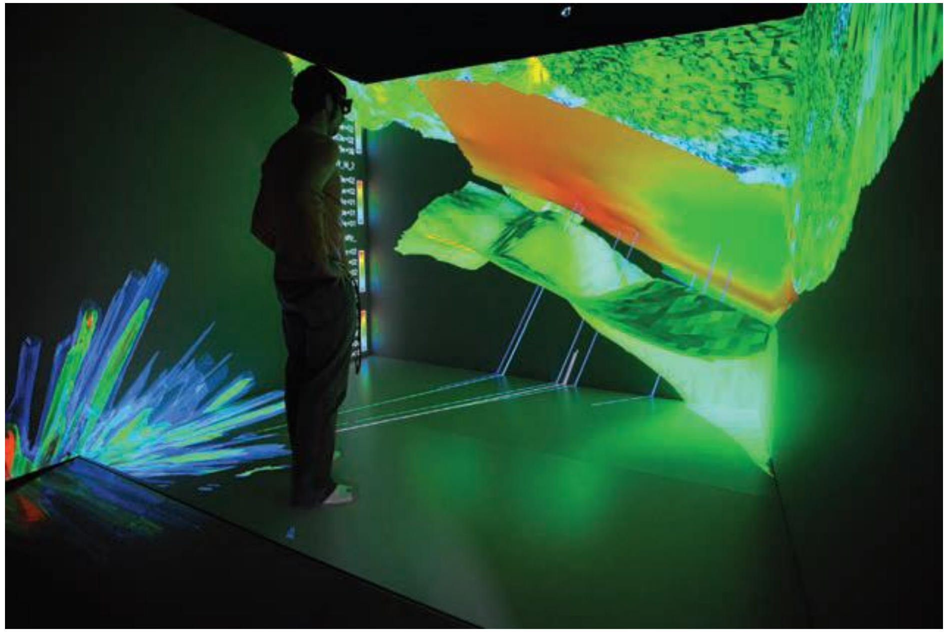
Portable Projector Displays



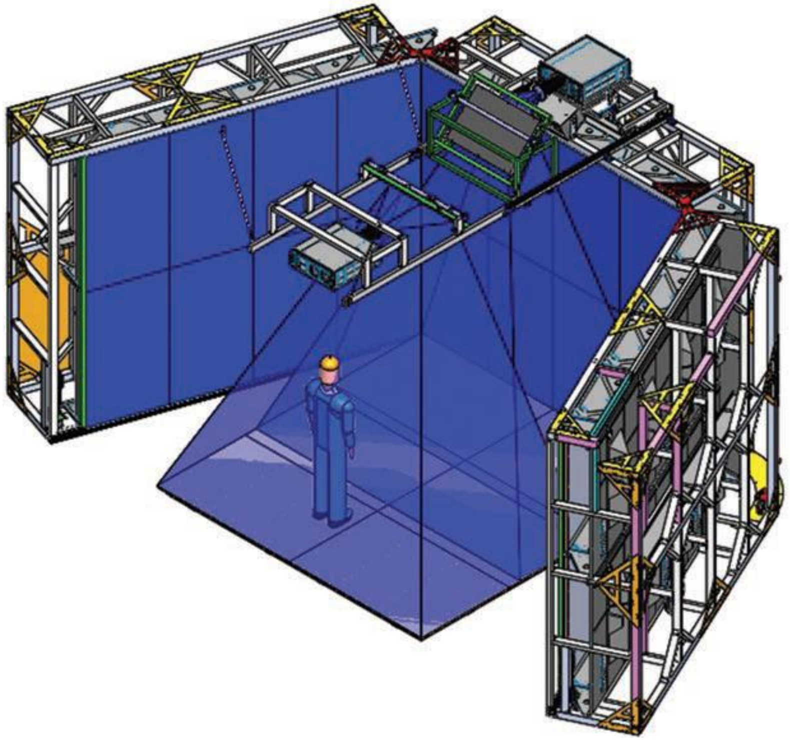
Disney uses pico projector to bring its magic to [augmented reality storybooks]

Visual Displays – Projected Displays

CAVES and Walls



A computer-assisted virtual environment (CAVE) display at Idaho National Laboratory.



The EmergiFLEX CAVE display is composed of an articulating wall of 24 high-definition Barco LED DLP projection cubes as well as overhead projectors to create floor imagery.

Bibliography

Based on:

[1] - Augmented Reality: Principles and Practice, 1st Edition by Dieter Schmalstieg and Tobias Hollerer, June 2016, Pearson Education

[2] - Practical Augmented Reality: A Guide to the Technologies, Applications, and Human Factors for AR and VR, Steve Aukstakalnis, 2017, Addison-Wesley