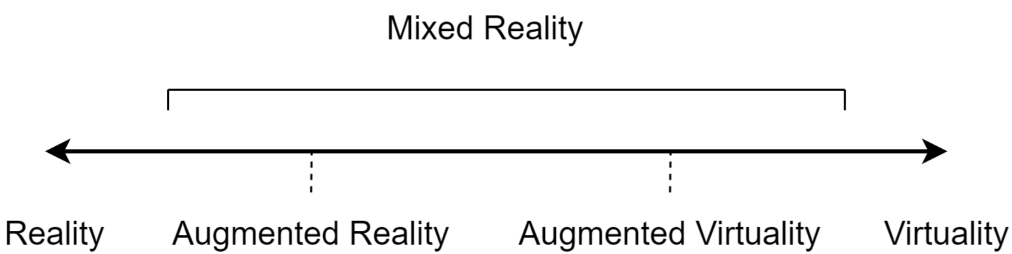
Chapter 20

Augmented Reality and Virtual Reality

What is Augmented/Virtual/Mixed Reality?

- Reality is one end of a spectrum
- Virtuality (total simulation) is the other
- AR, VR, and MR are all in-between

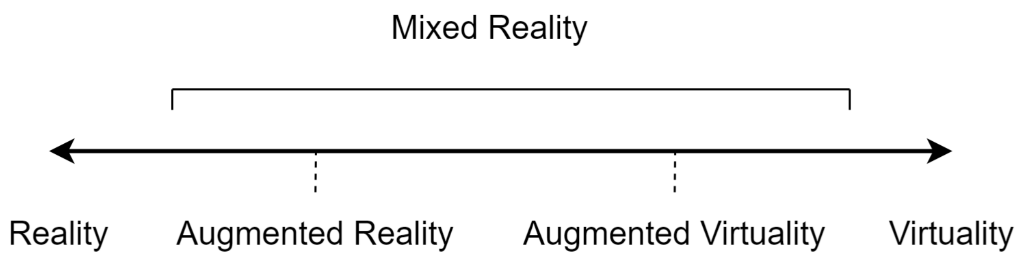


What is Augmented/Virtual/Mixed Reality?

- AR closer to reality → real elements >> virtual elements
- Example: standing in field of grass
  - virtual guidelines
  - virtual ball
  - predominantly real environment

What is Augmented/Virtual/Mixed Reality?

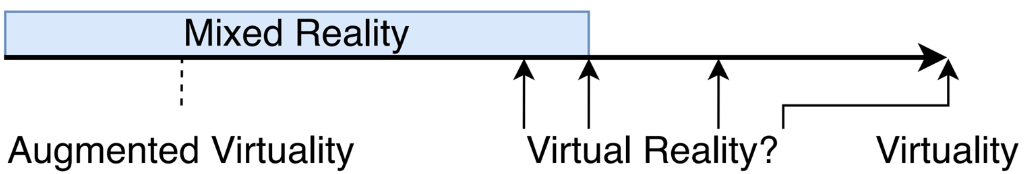
- Conversely, augmented virtuality closer to virtuality - predominantly virtual
- Example: standing in cubic white room, but
  - real surroundings (room) replaced with rectangular virtual soccer arena
  - virtual players
  - real soccer ball
- Both AR and AV are examples of mixed reality



What is Augmented/Virtual/Mixed Reality?

- What about virtual reality?
- VR *attempts* to replace reality with simulated reality.

- Problematic: not quite virtuality, but not quite mixed reality?



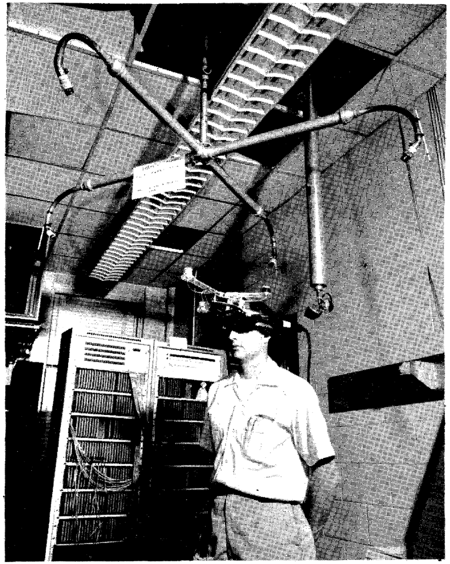
What is Augmented/Virtual/Mixed Reality?

- True virtuality is entirely simulated, à la Matrix
  - Simulated bleeding === real bleeding
- However, some aspects are undesirable (dying, injuries)
- VR eliminates these

- Mixed reality is a blend of reality and virtuality
- VR does not necessarily achieve this, but often lumped in for convenience

20.1 Defining AR and VR

- AR first defined by Ivan Sutherland in 1968:
  - Illusion of a 3D object presented in 2D that changes in exactly the way a real object would change when the user's head moves.
- Sutherland also made first optical head-mounted display & AR system
- Definition was too broad though, VR fits under this definition



Defining AR and VR

- Revised definition in 1997 of AR to be a system that
  - Combines real and virtual objects in a real environment
  - Aligns both real and virtual objects with each other
  - Runs interactively in real-time
  - Is registered in three dimensions
- Opposite problems: this definition was too narrow
  - Smartphone driven AR completely excluded (lack of 3D registration)
- But also too broad
  - Spatial audio systems emitting sound of a virtual ball rolling fits the definition.

Defining AR and VR

- Take the best of both worlds and define AR as
  - a predominantly visual system
  - that combines real and virtual objects in a real environment
  - while aligning the real and virtual together
  - running in real-time
  - allows 6 degrees of freedom of interaction (position, orientation) with virtual objects
- What about VR?
  - Challenging to define outside of a simulated experience of reality
  - Future implementations of VR may be vastly different than today

Defining AR and VR

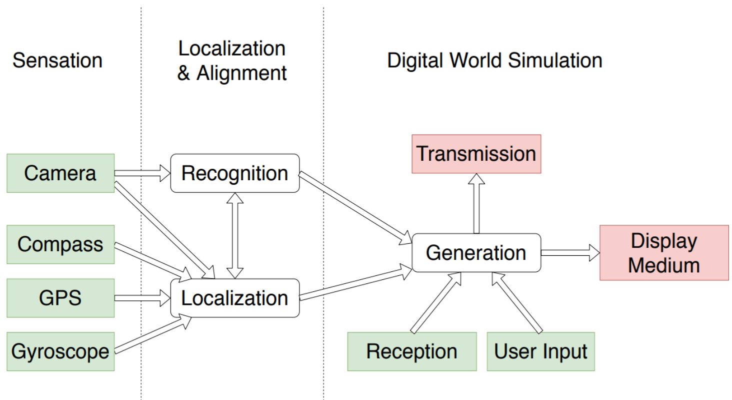
- Try defining VR from first principles
  - Compared to AR systems, VR replaces everything seen
  - Existence of telepresence
- Telepresence necessitates head-mounted displays
- Milgram provides a taxonomy of displays

Defining AR and VR

Class	Characteristics	Real/Virtual world	Overlays
1	Monitor-based, showing video of reality	R	CGI
2	Head-mounted, showing video of reality	R	CGI
3	Head-mounted, optical view of reality	R	CGI
4	Extension of 3 with stereoscopic video	R	CGI
5	Monitor+head displays showing video of CGI with video cutouts of reality	V	Video of reality
6	Extension of 5 with real object interactions	V	None

20.2 Anatomy of AR and VR Systems

- AR and VR systems have 3 principle tasks
  - Sensation
  - Localization & Alignment
  - Digital World Simulation
- Tasks are pipelined, source-sink data flow



Anatomy of AR and VR Systems

- **Sensation**
  - Analog-to-digital conversion of various aspects of the real environment
  - Continuous stream of raw information
  - Individual sensors provide windows to the world
  - Combined, they provide meaning

Anatomy of AR and VR Systems

- **Localization & Alignment**
  - Hardest task
  - Many aspects to consider
  - User-centric? World-centric? Single-user? Multiuser?
  - Sensor data contains noise and error
  - Will look at different systems over time to see different solutions

Anatomy of AR and VR Systems

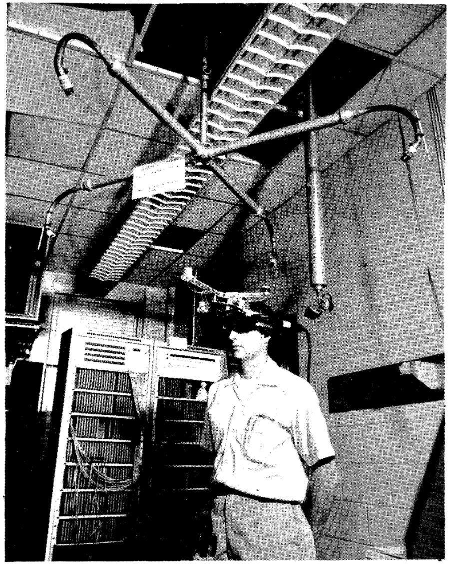
- **Virtual World Simulation**
  - Similar to video games
  - User input processed at this stage
  - Virtual objects are placed according to localization
  - In multiuser systems, external user actions are applied
  - World advances one simulation step forward

Anatomy of AR and VR Systems

- Application or Platform?
  - Security and privacy issues
  - Different stakeholders: enduser, system maintainer, AR world developer
  - Data stewardship - who protects what?

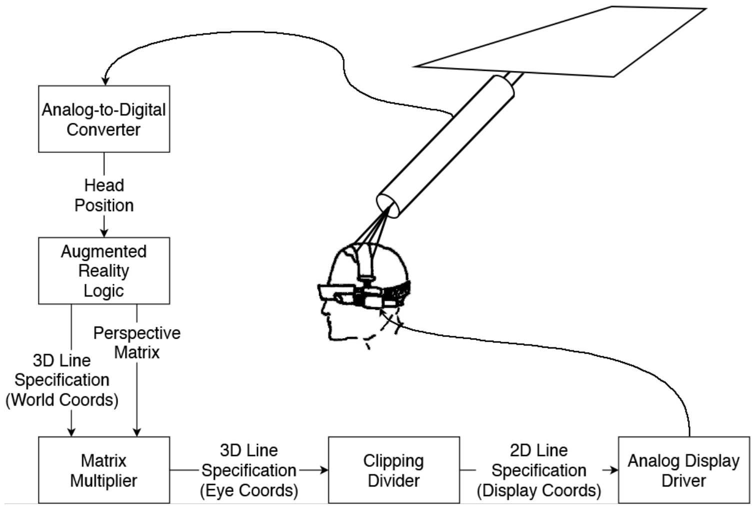
20.3 Early Foundational Systems

- Swords of Damocles – Sutherland 1968
  - Pioneering work, ~20 years ahead of time
  - Computers still used punch cards
  - Possibly first pseudo-GPU
  - Used mechanical linkage for localization



Early Foundational Systems

- Swords of Damocles - Sutherland 1968
  - A2D conversion of mechanical tension
  - Computer calculates perspective matrix, generates vertex list for virtual objects
  - Proto-GPU takes matrix, list and generates 2D image
  - Display driver sends image to headset



Early Foundational Systems

- Swords of Damocles - Sutherland 1968
  - Limited to simple lines
  - No interaction outside of moving head
  - Restricted movement to range of linkage
- Main innovations
  - Proto-GPU pipeline and clipping algorithms
  - Utilization of "kinetic depth effect" to imitate real objects

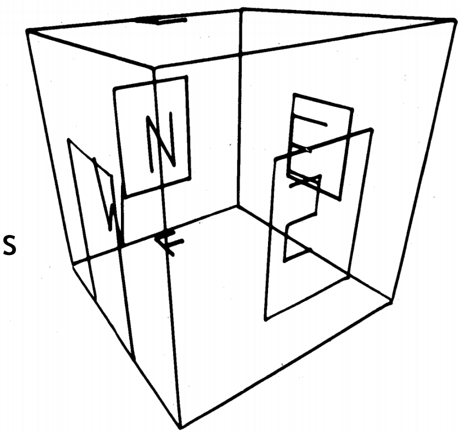


FIGURE 9—A computer-displayed perspective view of the "room" as seen from outside

Early Foundational Systems

- Visually-coupled airborne systems simulator - US Air Force 1977
  - Early flight simulator for air force pilots
  - Part of a larger “Super Cockpit” program for targeting HUDs
  - No position localization - pilots stationary in seats
- Many AR applications of this time were military
  - Computers still in infancy stage, just transitioned from vac tubes to transistors
  - Required expensive purpose-built hardware

20.4 Enabling Hardware and Infrastructure

- Three key fundamental advancements
  - Global Positioning System
  - Graphical Processing Unit
  - Internet
- Each lowered the barrier to entry for AR

Technological Revolutions of the 2000s

- Global Positioning System
  - Solved general localization problem
- Localization is always relative to a reference point
  - How would you tell someone where you are now?
  - Room numbers are relative to a building, buildings to a street, streets to a city, etc.
- Simultaneous Localization And Mapping is difficult for this reason
  - To localize, need a reference point
  - To map/obtain a reference point, need localization

Technological Revolutions of the 2000s

- GPS provides “global” reference point: Earth’s spherical surface
  - Bisect sphere twice to get meridian and equatorial axes (not *quite* true, but close enough)
  - Positioning now relative to these two axes
  - Report arc length up/down and left/right of principal axes
- Localize via triangulation, then plot on map (axes)
  - GPS uses constellation of satellites
  - Minimum 3 satellites visible at any location on Earth

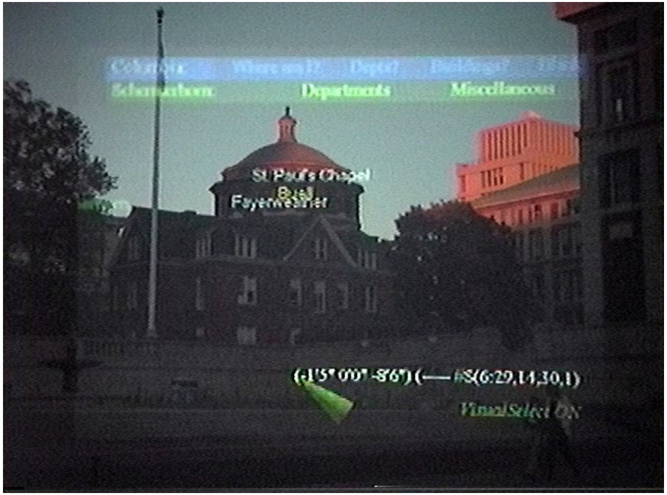
Technological Revolutions of the 2000s

- Initial accuracy of ~100 metres
  - US Gov feared misuse for weapons guidance
  - Added uniform noise to distort signals
  - The limit was later turned off, the accuracy can now be within 30 centimeters
- GPS enabled portability to AR systems
  - Differential GPS (DGPS) can be used for higher accuracy

* Images of "The Touring Machine" with permission of Dr. Stephen K. Feiner, Columbia University.

Technological Revolutions of the 2000s

- The Touring Machine, Feiner 1997
  - First notable overhaul of AR systems
  - Designed as a digital assistant for marking areas/buildings of interest
  - Used over-the-counter parts to build
  - Localized using GPS + gyroscope
  - Accurate to ~1 metre due to DGPS



Technological Revolutions of the 2000s

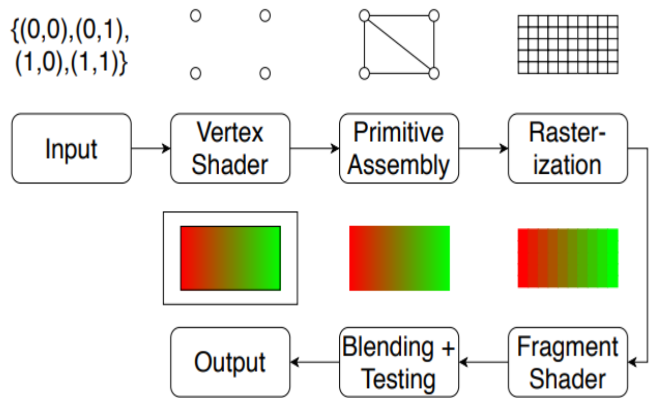
- Differential GPS (DGPS)
  - Requires known fixed-location receiving station
  - Station requests position from GPS
  - Computes difference against known location
  - Correction factor is broadcast over shortwave radio

Technological Revolutions of the 2000s

- The Touring Machine, Feiner 1997
  - Still had challenges that needed resolving
- Energy Consumption
  - Increase in visual fidelity = increase power draw
  - Battery power is limited
- Display Quality
  - Direct sunlight is very bright, difficult to overpower
  - Neutral density filters are a necessity
- Tracking Quality
  - Analog sensors still have noise
  - Compounded noise over time severely degrades tracking

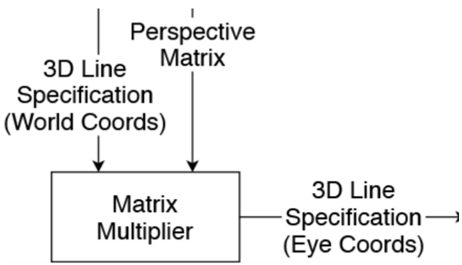
Technological Revolutions of the 2000s

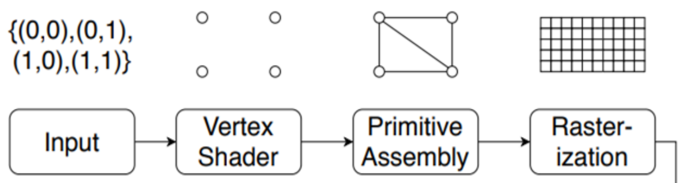
- Graphics Processing Unit
  - Provided acceleration for and standardized graphical processing pipeline
  - Alleviates some energy concerns
  - All AR systems use dedicated GPUs now
  - 2-stage pipeline: geometry + pixelization
  - Pipeline has drawbacks for AR work



Technological Revolutions of the 2000s

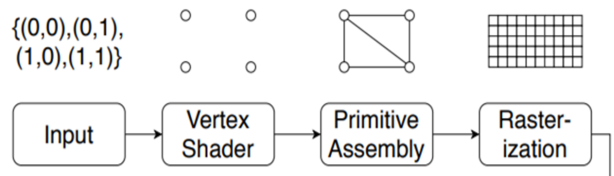
- Geometry Stage
  - User supplies vertex list, view + projection matrices, primitive shape
    - Valid primitives: Triangle, line, point
  - Vertices are created in 2D/3D space based on list by Vertex Shader
  - Shader applies view matrix to supplied vertex list
    - Programmable version of Sutherland's matrix multiplier



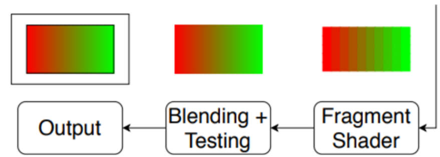


Technological Revolutions of the 2000s

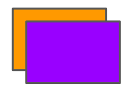
- Geometry Stage
  - Processed vertices are checked if still in view, vertices not in view are removed
  - Remaining vertices are connected in the order supplied, using the supplied primitive
  - Primitives are then rasterized
    - 1 pixel = 1 fragment (not really, but good enough conceptually)
    - *N* fragments per 1 primitive
  - Similar function performed by Sutherland's Clipping Divider



Technological Revolutions of the 2000s



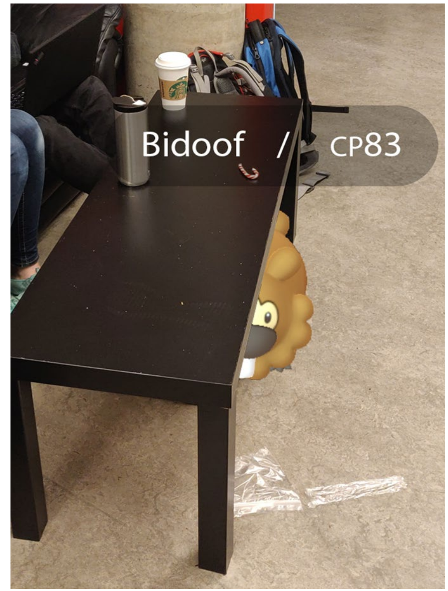
- Pixelization Stage
  - Fragments are coloured in by Fragment Shader
    - Lighting, texturing, reflectivity, etc. all handled by programmable fragment shader
  - At blending + testing, case of multiple fragments at a single pixel are dealt with
  - Example: overlapping shapes - 2 fragments, 1 pixel
  - Solve by either blending colours, or depth



- Pipeline Drawbacks in AR
  - Pipeline was made for *generating* imagery from nothing, not for merging imagery
  - Depth + image output buffers are located on GPU

Technological Revolutions of the 2000s

- Pipeline Drawbacks in AR
  - Consider case where virtual object is hidden behind real object
  - Occlusion must be pixel-perfect, can be achieved in 2 ways
    - Render image → remove bad pixels on CPU
    - Create depth mask on CPU → use mask on GPU
  - 1st method stalls rendering due to data transfer to CPU
  - 2nd method requires depth "shading" (custom GPU/pipeline)
  - HoloLens took latter approach



Technological Revolutions of the 2000s

- Internet
  - Simplified multiuser systems
    - Virtual objects no longer had to be static
    - Client-server architecture allowed for multiuser coordination
    - Enabled higher mobility systems
  - Enabled offloading of difficult/complex computations

Technological Revolutions of the 2000s

- Studierstube - Szalavári 1998
  - One of the first AR systems to support multiple simultaneous users
  - Achieved by limiting scope of problem
    - Similar to Sutherland's system, only worked in a specific room
    - Utilized various magnetically tracked devices: 3D cursor, HMD, 2D control plane
  - Centralized servers maintained spatial coordinates of each object
  - HMD client responsible for local POV rendering
    - Desync was possible, synchronization was not pixel-perfect
  - System architecture was very similar to how MMORPGs were.

Technological Revolutions of the 2000s

- AR Quake - Thomas 2001
  - One of the first AR systems to work outdoors
  - PC users could connect with AR users using the campus as the "level"
  - Hybrid tracking approach
    - Used GPS for coarse location data
    - Fiducial markers for context-specific location data
  - Documented difficulties of AR in uncontrolled environments

Technological Revolutions of the 2000s

- Outdoor AR Drawbacks
  - Inaccurate colour blending models due to additive/subtractive light projection methods
    - Notice how yellow face doesn't appear yellow
  - Direct sunlight is very, very bright and over-powers most light projection methods
    - Image is very dim due to having to block most of the sunlight
  - Power-scarce environment due to being battery powered
    - System could not run for particularly long without needing to recharge

* Image with permission of Dr. Bruce H. Thomas, University of South Australia.

Technological Revolutions of the 2000s

- Applications
  - Computing power became exponentially cheaper
  - Enabled plethora of applications for AR
- Navigational
  - Touring Machine, NaviCam, etc.
- Militaristic
  - Fight jet cockpits
- Industrial
  - Assembly plants at BMW, Boeing, Airbus, etc.
- Educational
  - Virtual tour guides
- Entertainment
  - 1&10 yellow line marker in American Football

Systems from 2010 onward

- Hardware initially focused on miniaturization, but began shifting to ASICs
- Current transistors are fabricated <7 nm apart
- Computation began to turn into a commodity
- New form-factors for AR systems emerged as a result

Systems from 2010 onward

- iPhones and Mobile Computing
  - Smartphones became ubiquitous during this time
  - Smartphone sensors became more pervasive
  - Smartphones became incrementally more affordable
    - Low end phones are generally equipped with magnetometers and gyroscopes now

- Pokémon GO
  - One of the first AR applications to take advantage of the above factors
  - At its peak, installed on over 10% of all US Android devices
  - Demonstrated that smartphone-based AR is sufficient for the general public
  - Influenced the creation of ARCore and ARKit from Google/Apple

Systems from 2010 onward

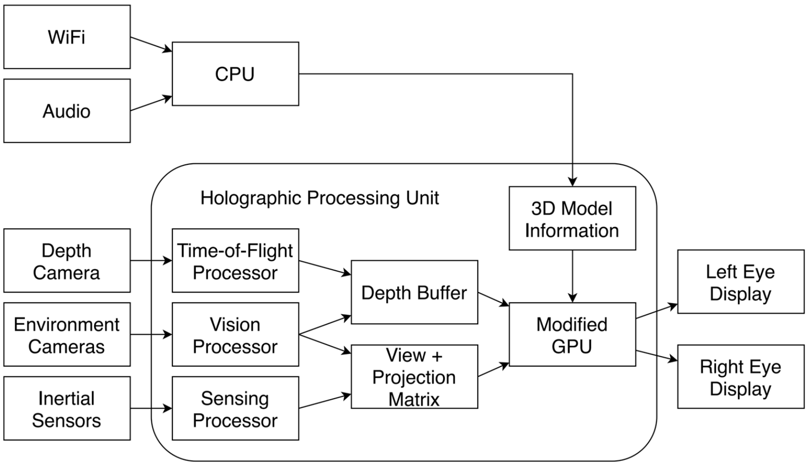
- Pokémon GO
  - Used magnetometer to determine compass orientation (yaw)
  - Used gyroscope to determine angle relative to ground (pitch) and head tilt (roll)
  - GPS provided relative location to nearest Pokémon
  - Rear smartphone camera acted as “real world” backdrop
  - Pokémon would then be rendered on top of this image

Systems from 2010 onward

- Commodification of Computing
  - Incremental advancements in manufacturing processes and CPU architecture design
  - Opened market for fully custom CPU fabrication
    - Customizable SoCs and instruction set architectures
    - Significant power savings through 1 chip = 1 function
  - iPhones have a Neural Processing Unit
  - HoloLens have a Holographic Processing Unit
  - Dedicated processing chips have allowed HoloLens to achieve 720p at 60 FPS per eye

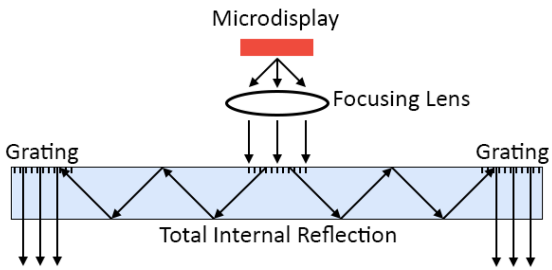
Systems from 2010 onward

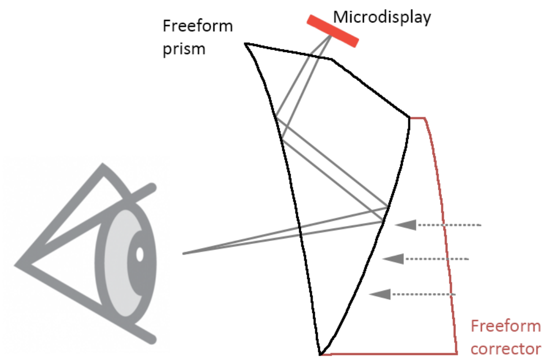
- HoloLens: HPU
  - Custom SoC that bridges gap between conventional graphics pipeline and AR pipeline
  - Adds dedicated vision processing for environmental semantics
  - Supports depth shading via custom support for capturing depth



Systems from 2010 onward

- HoloLens: Display Mechanism
  - Principle mechanism of action is a light additive microdisplay, transferred to both eyes
  - Bulky prism acts as optical carrier with total internal reflection (like a fibre optic cable)
  - Real world light is emitted into the prism via an additional correctional prism
  - Major drawback of this system is the inability to render pure black objects (cannot remove light)
  - Secondary drawback is that all objects will be transparent to a degree since light is multiplexed immediately prior to entering retina



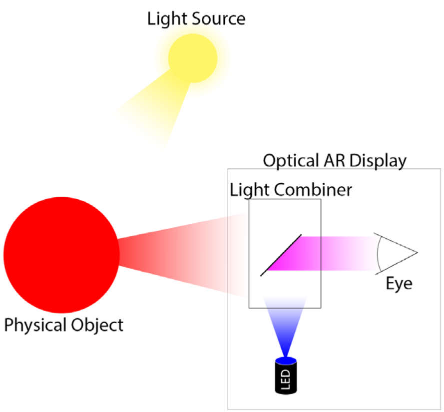


20.6 Limitations and Challenges

- Color Perception
- Depth
- Localization
- Information Presentation
- Social Acceptance

Limitations and Challenges

- Color Perception
  - Noted by Feiner and Thomas (and demonstrated by HoloLens), optical AR displays have many challenges in color
  - Three intrinsic, core problems:
    - Sunlight is extremely bright
    - Light can be either additive or subtractive w/ displays, but not modified vis-a-vis software graphics style
    - AR "backgrounds" are dynamically colored and uncontrollable
  - Optical AR display must be able to account for these issues, otherwise unexpected results may occur

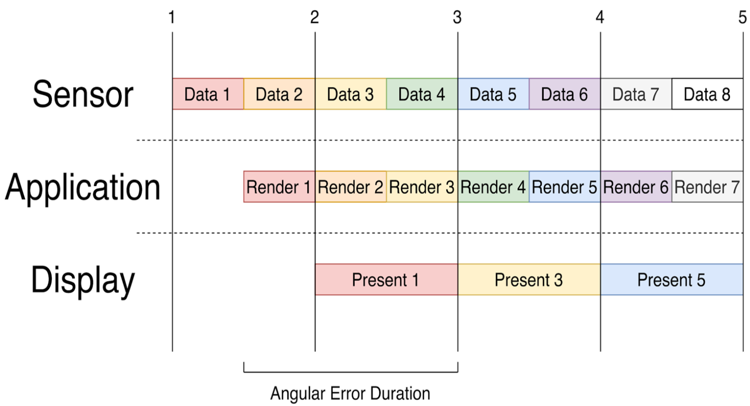


Limitations and Challenges

- Depth
  - Extremely important aspect of high fidelity AR that is still not yet fully solved
  - Currently, real objects cannot be in front of virtual objects; real-time segmentation is too hard
  - Possible in specific specialized cases, but not in general
  - HoloLens uses infrared sensors to pick up depth, IR absorbing surfaces, like water, cannot be picked up
  - Software methods, like stereo triangulation, may prove to be more suitable, but finding corresponding points between cameras is extremely hard

Limitations and Challenges

- Localization
  - One of the largest contributing factors to localization error is latency
  - Stale sensor data leads to stale image data
  - By the time an image is displayed, it could be 2 frames stale in terms of sensor data
  - Simple example is the smoothness between 30 Hz and 144 Hz monitors



Limitations and Challenges

- Localization
  - Following latency is sensor accuracy and precision related errors
  - Accuracy - how close estimate of X is to X
  - Precision - how close estimate of X is to prior estimate of X
  - One can be accurate, but not precise
    - Think arrows hitting close to bulls-eye, but bad clustering
  - And also precise, but not accurate
    - Arrows hitting nowhere near bulls-eye, but tight clustering
  - Very difficult to be both precise AND accurate
  - Even harder to be precise, accurate, AND fast

Limitations and Challenges

- Information Presentation
  - AR conveys visual data, which can be information-dense
  - Balancing act in providing sufficient information without overloading the user
  - Similarly, new UI paradigms will be necessary
    - Transitioning between different AR contexts?
    - AR interface control inputs?
    - Control mechanisms?

Limitations and Challenges

- Social Acceptance
  - Google Glass predated HoloLens by almost 5 years, but was a universal flop
  - Received widespread criticism and social rejection, despite being almost identical to HoloLens
  - Reasons were twofold
    - Innocuous camera that could record everything (Privacy)
    - Non-invasive design meant the wearer could zone-out on the device (Safety)
  - Some users may also experience virtual reality sickness
    - Consequences of completely replacing/augmenting a person's view with a digital one is not entirely well understood yet
  - In today's techno-centric world, AR treads on a fine line between utility and privacy/safety
  - We bear the burden to ensure that public good outweighs the negatives