Overview of some Patents on AR/VR Systems

Madan Y N

June 2022

Contents

1 Introduction **3**

2 Qualcomm **4**
2.1 Head Mounted Device for Virtual & Augmented Reality Combining Reliable Gesture Recognition with Motion Tracking Algorithm 4

3 Microsoft **6**
3.1 Augmented Reality 6

4 Magic Leap **9**
4.1 Multi-Resolution Display Assembly for Head-Mounted Display(HMD) Systems 9

5 Snap Inc **14**
5.1 Wearable Device with In-Eye Display Capacity 14

6 Apple **17**
6.1 Augmented Reality Display 17
6.2 Audio System and Method of Augmenting Spatial Audio Rendition . . . 19

1 Introduction

The terms "augmented reality"(AR) and "virtual reality"(VR) is making quite a round on the internet. Social media giant Facebook changing its name to Meta added to the buzz around these terms. Many tech pioneers say AR & VR(or Extended reality, XR together) are(is) the future of computing. An ever-increasing capital and human ingenuity directed to develop XR technology by companies like Meta, Apple, MagicLeap, and others are clear as to where the technology is heading. That said, AR & VR are not very new. They have been around since 1968 when American computer scientist Ivan Sutherland invented the head-mounted display. That was a very rudimentary form of XR. Now technology is mature enough to realize the true powers of extended reality. The advancements in computer vision, 3D rendering, and near-eye display technology are making it possible to use XR technology in education, entertainment, healthcare, business, and whatnot. In this short document, I summarize some patents about XR systems with an emphasis on AR. I begin with an overview of AR systems by summarising two patents, one each of Qualcomm and Microsoft; describe head-mounted display systems by summarising two patents, one each of MagicLeap and Snap; a real-world application and spatialized audio by summarising two patents of Apple.

2 Qualcomm

2.1 Head Mounted Device for Virtual & Augmented Reality Combining Reliable Gesture Recognition with Motion Tracking Algorithm

[Jea21]This patent is about a general head-mounted display(HMD) for virtual or augmented reality applications.

A HMD system should consist of,

- Suitable display element
- Cameras, gyroscopes, accelerometers and other sensors to sense the environment
- Illuminators to illuminate the environment with infrared rays
- Lens apparatus

It is important that cameras, illuminators and lens apparatus have common pass-bands

- 6 Degree of Freedom(DoF) algorithm to detect translational and rotational movements.
- Gesture tracking algorithm for gesture(hand gesture or any other) recognition

note:

- *The Field of Vision(FoV) of illuminator should be more than that of the cameras*
- *The illuminator is activated at a frame-rate lower(half or one-third depending on the application) than that of camera module.*
  - *When illuninator is OFF, camera captured images are used to implement 6 DoF algorithm*
  - *When illuminator is ON, camera captured images are used to implement the gesture tracking algorithm.*

Block diagram:

The block diagram(taken from the patent document) of the HMD system is as below

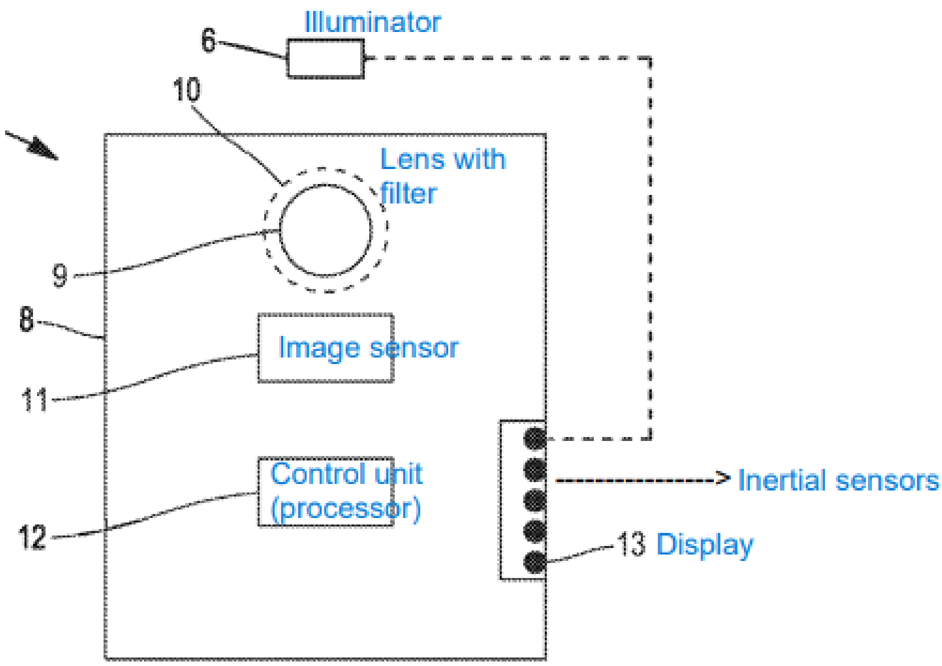

Figure 1:

A monochromatic, fish-eye lens is to be used to capture the world. The lens filter which is covered above the lens filters light of specified range of wavelength(eg. 95nm). Then the image sensor converts the input of lens into image pixels. The control unit uses the output of image sensor and inertial sensors to run algorithms and sends the contents to be displayed to the display apparatus accordingly.

note: *The 6 DoF and gesture recognition algorithms may be run in parallel in separate processors or in sequence in the same processor*

3 Microsoft

3.1 Augmented Reality

[S19]This patent is about an augmented reality system that receives a plurality of local 3D models of the real world environment and process them generate rendering data so as to augment virtual objects on real world environment

An augmented reality(AR) system comprises of a computer interface, a global model generation module, and an augmentation sharing module. An interplay of these three modules is what makes an AR system. These three modules maybe implemented as software or hardware or a blend of the two. The computer interface receives a number of local 3D models generated by AR devices in a particular physical location. The global model generation module uses these local models to generate one global model of the said real world environment. This global model is used to explore and augment virtual objects on the real world. Then the augmentation sharing module uses the rendering data to render virtual objects in the real world and display it to the user through a display device.

Major components in an AR head-mounted device:

- Light engine: This comprises of a micro-display and imaging optics in the form of collimating lenses. The display can be of LCD, LCOS, LED etc. A circuit drives the diaplay to light up specific pixels. Light from each pixel is passed through collimated lenses and excited to the optical components

- Optical components: The beams coming from the light engines undergo multiple diffraction & TIR and are finally directed into user's eyes.
  note: *By using transparent pixels we can bypass the diffraction and TIR stages of light beams*

Light engine & optical components together constitute display apparatus.
note: *projecting 3D images requires displaying slightly different images for both the eyes called "Stereoscopic imaging". This can be done by either using two different light engines or by using suitable optics to split light into two different beams*

- Computational block: This comprises of processors(CPU and/or GPU) to process input data, memory to store algorithms to run, and input-output pathways to receive data from various sensors and send data to light engine and other output devices.

One to one communication in AR world:

Consider two people with one of them with an AR headset and the other with or without one but with a computer device like a mobile phone. Their devices are either connected directly or are both connected to a common cloud. They can communicate as below,

i AR user moves around the environment and creates a local mesh of the real world environment.

ii The local mesh is either uploaded to the cloud or transmitted directly to the companion user along with video.

iii The companion user receives the mesh data of world of AR user on his/her display and adds annotations based on local mesh and uploads to cloud or transmits directly to AR user.

iv Then the AR device receives the annotations and renders virtual objects it using the locally generated mesh

Concept of Global mesh:

Now consider a situation in which there are many people in the same real environment each with an AR device. These users communicate with themselves and with a far away companion user by the following method,

i Each user in the environment creates a local mesh and either uploads to cloud, to which all the AR devices and companion device is connected or shares with other users in the environment and companion user directly.

ii These local meshes are stitched together to create a global mesh, which is always bigger than any of the local meshes of the real world environment.

iii The companion device receives this global mesh from cloud or directly and its user makes annotations to the world and uploads back to cloud or send directly.

iv Then an AR user in the real world environment receives annotations and the global mesh that the global mesh companion device used to annotate. The user uses that global mesh and his/her local mesh to generate an aggregated mesh(in which local mesh would always override global mesh) and renders the annotations in the generated aggregated mesh

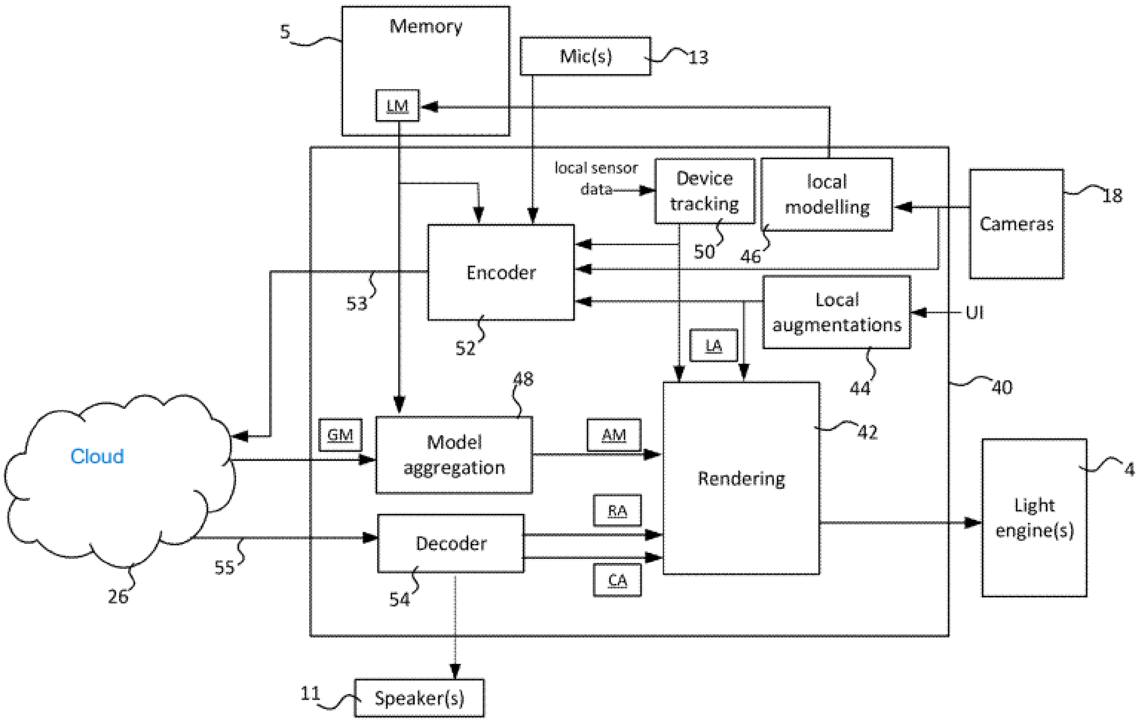

Figure 2: Block diagram of an AR system taken from the patent document

4 Magic Leap

4.1 Multi-Resolution Display Assembly for Head-Mounted Display(HMD) Systems

[Chu21]Displays which cannot project beyond a small central portion of the user's field of view can reduce the level of immersion. This patent describes a HMD display with a display assembly configured to display content to most or all of the user's field of view. While the user's ability to focus is limited to about 30 and 50 degrees off-axis, most user's eyes are capable of detecting content and particularly fast movement past 100 degree off-axis in some direction. For this reason, a display need to cover the outer periphery to create truly immersive experience.

- The HMD described here includes a first display; a second display that at least partially surrounds the first and an attachment to couple these displays to the user's head
- The visual acuity of a user is substantially reduced in the peripheral region. For this reason, the peripheral display(or second display) can run in lower spatial & angular resolution, lower color resolution, a different intensity and lower refresh rate than the main display(or first display)

note: *We can either have a display whose resolution decreases towards its periphery or we can employ a eye-tracking mechanism and have customizable resolution according. Displays like OLEDs consume less power and provide low resolution and can thus be used for peripheral display*

Combination of main & peripheral displays

- When a high resolution content is provided to the user, main displays and peripheral display **130** is used and peripheral display **132** and that part of the peripheral display right in front of the main displays are not used as in figure 5
- When interactive or high resolution content is not being displayed, we can use only peripheral display and not main displays. The peripheral display made of OLED consumes much less power than main displays that are driven by high energy consuming light projectors
- In some other occasions, whole of peripheral and main displays can be used in tandem to enhance user experience

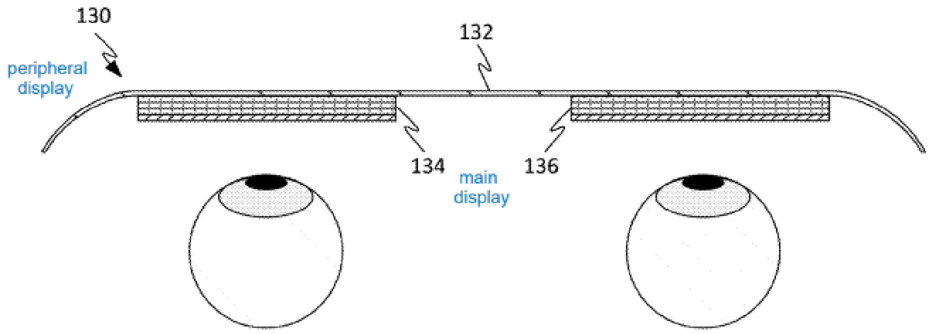

Figure 3: Schematic of main & peripheral displays taken from the patent document

Average human FoV of left eye:

Fovea is a place in retina where the virtual acuity is the highest. As we rotate our eyes, the fovea rotates. This angular field in which fovea moves is called the foveal field or field of regard. Outside of this is the peripheral field of vision. A user can only focus in this field of regard. So main display can be designed to cater only to this field of regard and peripheral display can be used for the remaining FoV.

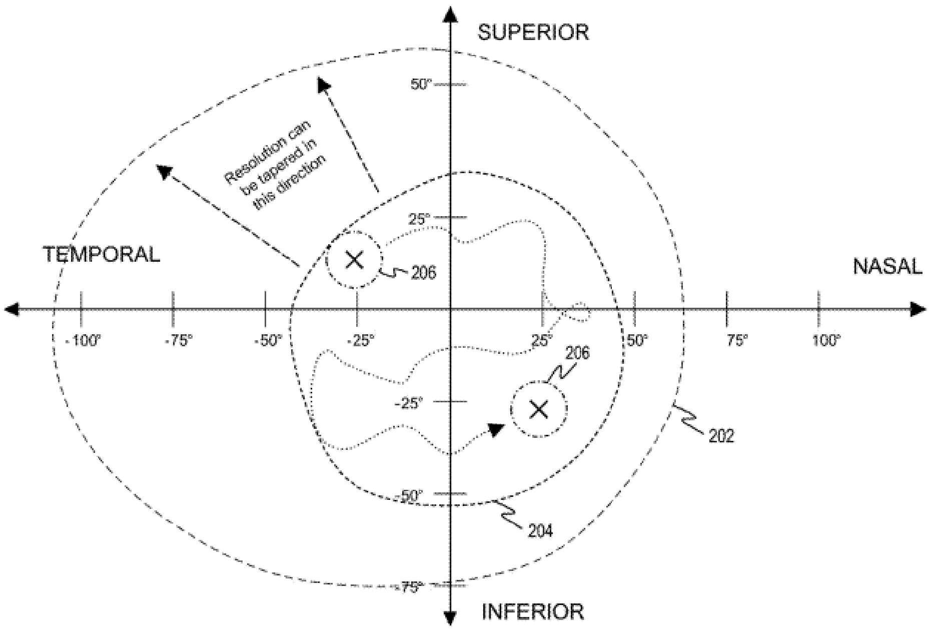

Figure 4: Schematic diagram of FoV of human left eye taken form the patent document

In the diagram, **206** - Fovea, **204** - Foveal field, **202** - Peripheral field

AR system of which the HMD is a part of:

An AR system contains the following important elements in it,

- Image Generating Processor: This consists of a memory block, a CPU, a GPU & other circuitry. This element takes in data from all sensors, processes it and generates virtual contents that is to be displayed to the user. It should generate all the data to make stereoscopic imaging to be displayed. This can be either on-board the HMD or connected through wire or connected wirelessly.

- Light Source: This consists of a plurality of sub-light sources. These sub-light sources may include specific color LEDs & lasers in various geometric configurations. The light source may also include condenser to collimate the light coming out of light source.
- Spatial Light Modulator(SLM): This may consist of reflective and/or transmissive and/or emissive optical elements. SLM improves the creation of 3-D perception. The digital light processing(DLP) done on light coming from the light source encodes the desired virtual content.
- Injection Optical System: This consists of one or more lenses that are configured to direct light from light source & SLM into light-guiding optical elements(LOEs).
- Light-guiding Optical Element: These are optical elements like waveguides that guide the light into user's eyes

Thus by sequentially projecting images using plurality of LOEs and sub-light sources at sufficiently high frame rate the system in figure 7 can generate a 3-D image of virtual objects at various depths that appear to exist simultaneously in the 3-D image.

- Controller: It is operationally coupled to image processor, the light source and SLM to coordinate synchronous display of images to the user's eyes.
- Eye-Tracking System: This is configured to track the user's eyes and determine the user's focus point. The input from this subsystem can be used to activate only a subset of sub-light sources corresponding to a particular LOE that is configured to deliver collimated light to the user's eyes such that high resolution image appears to be coming from the point the user is focusing. This way the other sub-light sources can be deactivated to save power.

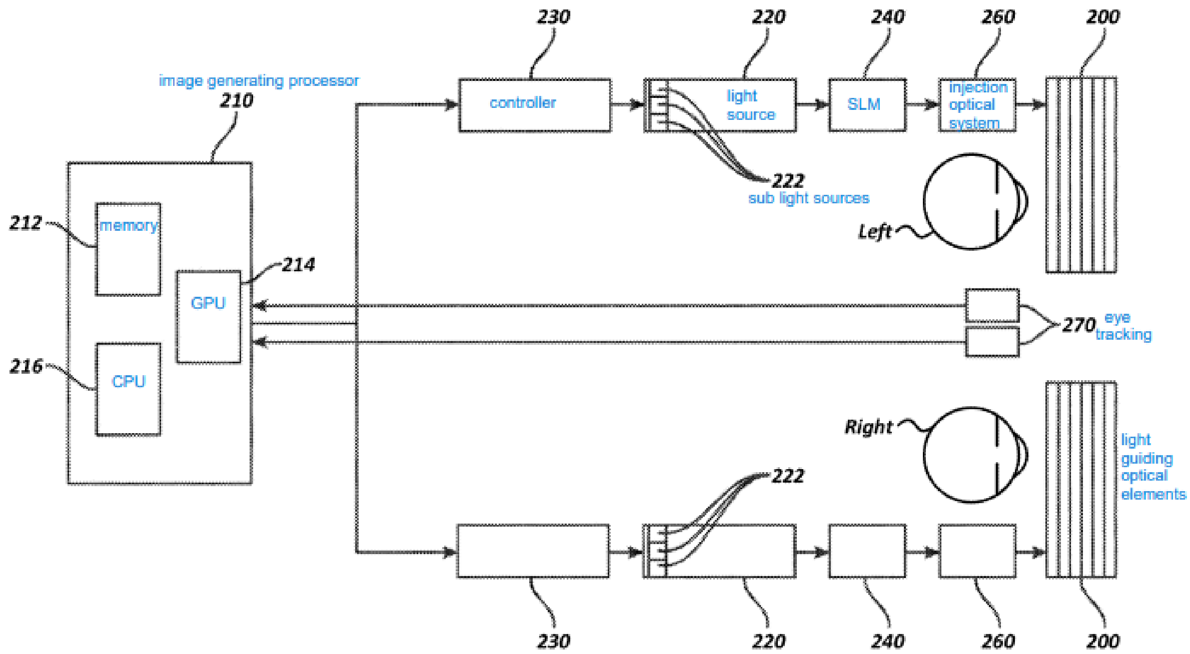

Figure 5: Block diagram of AR system described taken from the patent document

Different parts of display:

The display of a HMD can have multiple regions of varying properties as shown in the below figure

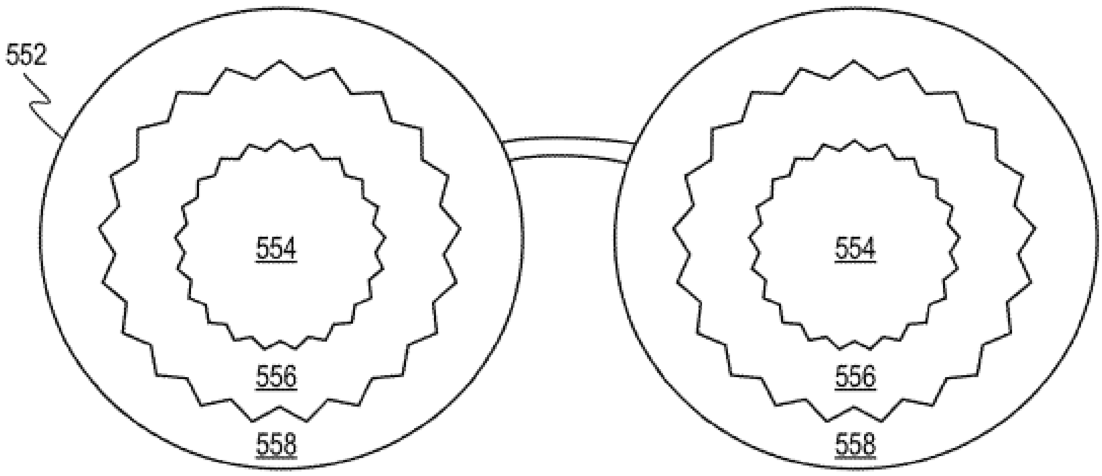

Figure 6:

- The region **554** can be designed to distinguish color & spatial resolution. Since the center of the eye has the highest concentration of cones, region **554** can have highest resolution and truest color reproduction.
- The region **556** can have lower spatial & color resolution and region **558** can have

further lower resolution or can even be displayed on grey scale. The zig-zag lines represent the transition between two regions so that the change in resolution is not disturbing

5 Snap Inc

5.1 Wearable Device with In-Eye Display Capacity

[Sam22]The present patent describes examples of various apparatus and systems that can be used for augmented reality. Many headsets have been developed for augmented and/or virtual reality using the projectors and waveguide technology for projection and display to the viewer. Unfortunately projectors consume a relatively large amount of power during use and due to this the use time is limited or headset must be tethered with a power cable. The system described herein reduce power consumption by using one or more arrays of light emitting diodes(LEDs) which consume much smaller amount of power relative to projectors.

LED display

Every near-eye display system like a smart glass will have optical elements that would facilitate the user to see the world through it while digital objects are overlaid on the display. Over this optical element, a layer of LED display is attached which has the following properties,

- Electrically conductive elements made of copper or silver or gold make up an array in a substrate. This array consists of a plurality of rows and columns
- These plurality of rows and columns can be configured to form girds of various shapes like square or triangle or circle, etc and of various sizes.
- The LED display has a plurality of LEDs or groups of LEDs at the intersections of the rows and columns of the electrically conductive elements. Each group of LEDs consists of red, green and blue LEDs
- These groups of LEDs are controlled to light up specific color LEDs through the grid of electrically conductive elements to generate desired virtual objects to be displayed to the user

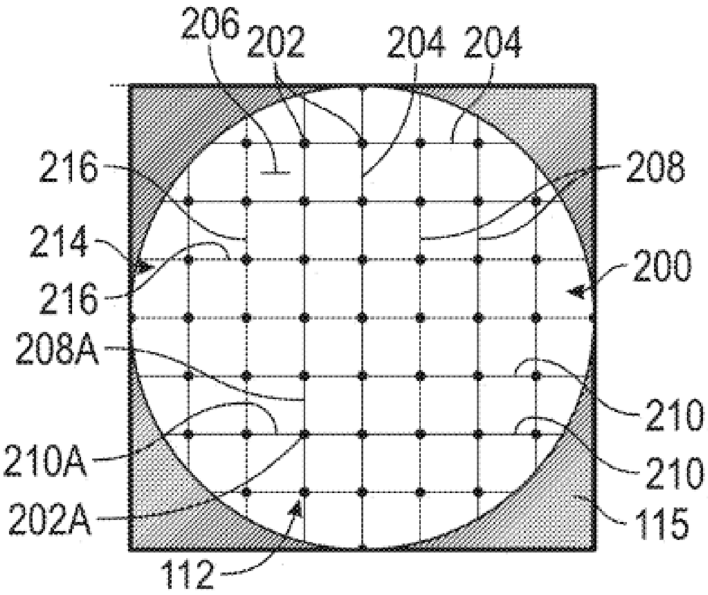

Figure 7: An example of array of LEDs and conducting elements(not to scale)

In the above image,
206 - Substrate, **202** - LED, **208** - Row of conducting elements, **210** - Column of conducting elements

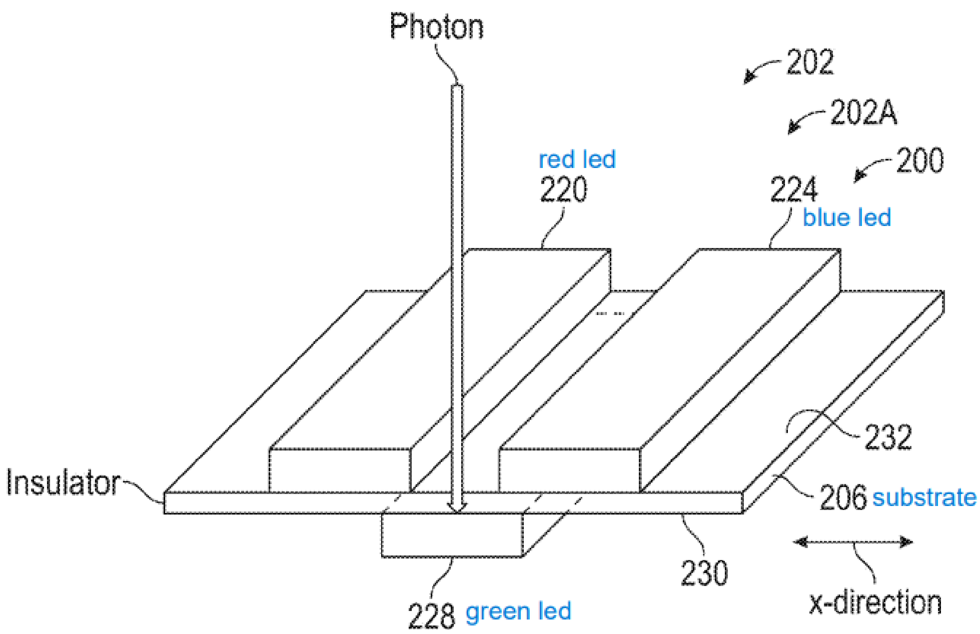

Figure 8: A schematic of a group of LEDs

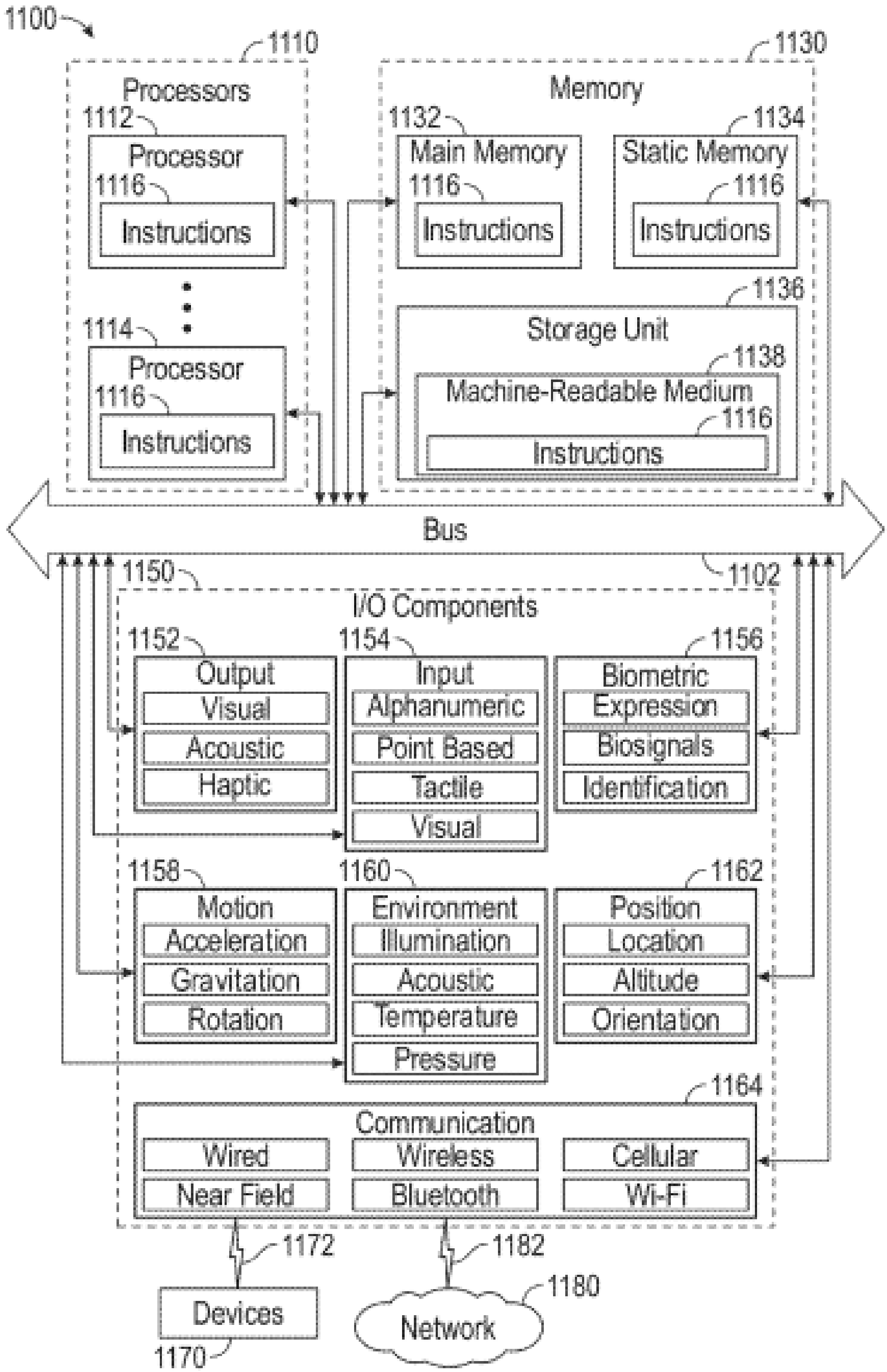

Figure 9: A schematic diagram of a general AR/VR system

6 Apple

6.1 Augmented Reality Display

[Pet21]The patent is regarding AR display in automobiles. An AR display in automobile, say a car can be used to guide the drivers by showing directions to a particular location, by showing upcoming curves or intersections especially in hilly regions and many more things. In order to achieve this we will be needing the 3D mesh map data of the real world environment in which the driver is in.

We can use the pre-generated 3D mesh data of that particular location and the locally developed mesh data by the sensors on the vehicle to augment virtual objects. Pre-generated data will be particularly useful when local sensors cannot fully map the environment due to various reasons like poor visibility, short range, occlusion, vehicle motion, etc.

note: *The 3D mesh data of a particular real world location can be generated by using images collected from large number of sources over time. The data should comprise if 3D meshes, textures & other geometric information. This 3D data may be uploaded on cloud with location and time tags, so that AR users in that particular location can download from cloud by providing location tag*

3D Tiles:

In the cloud, the 3D data is stored in the form of 3D tiles. A 3D tile is a portion of real world tagged with appropriate location and time information. Each 3D tile contains 3D meshes(planes and other geometric shapes in the environment), texture(properties like reflectivity, refractive index, etc), etc. Multiple 3D tiles makeup a whole 3D data of a location.

Benefits of 3D data from cloud along with local data

- Having 3D mesh map from cloud reduces the required number of point clouds(local 3D data) to generate full map and thus saves a lot of computational power
- Data from cloud can be used to fill-up the regions occluded to the on-board sensors.
- Data from cloud can also be used to augment virtual objects in distant regions which are out of range of the on-board sensors

Coloring and placing of virtual objects:

- The 3D data received from the cloud contains data about color of the surface. This can be used in deciding the color of the virtual content to be augmented on that surface so that it shouldn't be hard to recognize it.
- Knowing about where the sun is, is very important so that we do not place any virtual object on or near the sun. This piece of information can be determined by using GPS location of the vehicle, time and day of the year, direction in which the vehicle is moving.

- We should also not place any virtual content upon surfaces which are highly reflective and their surface normal is such that it will reflect sunlight towards the user. These data about reflectivity and surface normal will be present in the 3D data received from cloud.

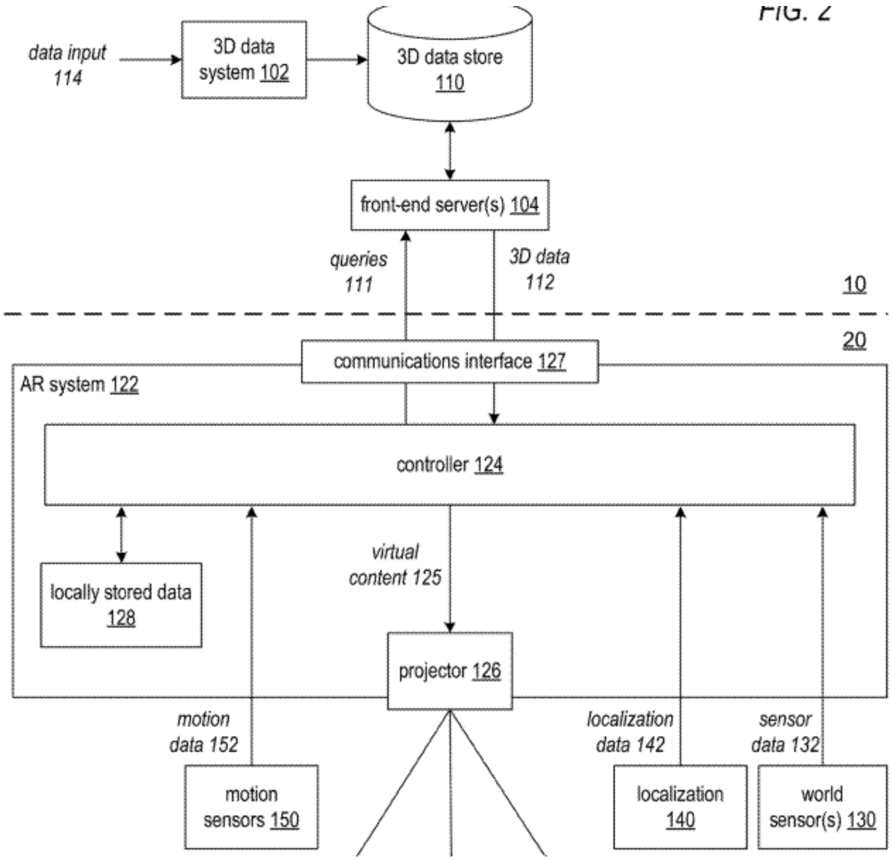

Figure 10: Block diagram of an AR display system taken from the patent document

6.2 Audio System and Method of Augmenting Spatial Audio Rendition

[M21]This patent is about an audio system and method of using audio system to augment spatial audio rendition. When the AR user augments a sound source at a location in the mesh, a binaural audio signal is generated using some transfer functions to render spatialized audio i.e different audio signals to both ears, so that the sound appears to be coming from that particular location in real world in which the virtual sound source is placed.

note: *In normal headphones, audio will be perceived as coming from within the listener's head but spatialized audio will be perceived as coming from a location outside the listener's head*

3 key things needed for spatial audio

i Head-Related Transfer Functions(HRTF): It is a model of a filter that contains all the acoustic information required to describe how sound from different points in space reflects or diffracts around a listener's head and outer ear before entering the auditory system. It is important to note here that HRTF is unique to an individual based on distinctive physical shape of the ear.

ii Environment-Related Transfer Function(ERTF): It is a model containing acoustic information required to describe how sound reflects and reverberates within a listening environment. It is to noted here that when we augment virtual objects on real world the ERTF of that location should be changed accordingly in order to maintain immersion.

iii Head Tracking: Tracking position and orientation of the listener's head and using the information to shift the virtual audio source accordingly

Process flow of rendering audio:

i Receive a user input designating a location in a listening environment

ii Present visual representation of a sound source at the location in the listening environment

iii Apply a head-related transfer function corresponding to the location in the listening environment to an audio input signal to generate a binaural audio.

iv Render spacialized audio using binaural audio signal to localize audio to the location

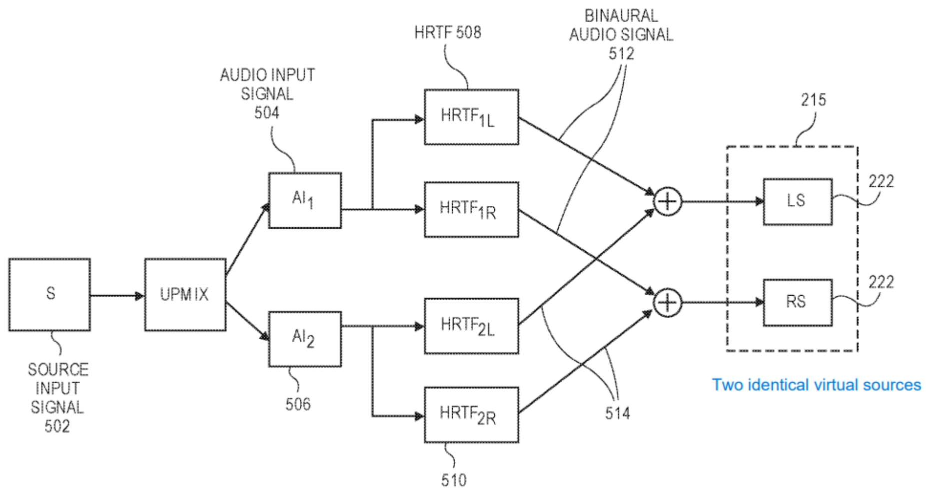

Figure 11: Block diagram rendering two similar virtual audio sources from the patent document

References

[S19] Chen Henry Yao-Tsu Taylor Brandon V Swift Mark Robert Lee Austin S Menezes Ryan S. US Patent 10,169,917 B2. Jan. 2019. URL: https://www.lens.org/lens/patent/096-615-149-631-024/frontpage?l=en.

[Chu21] Miller Samuel A Welch William Hudson Edwin Lionel Ernest Yeoh Ivan Li Chuen. US Patent 2021/0239991 A1. Aug. 2021. URL: https://www.lens.org/lens/patent/037-635-980-998-410/frontpage?l=en.

[Jea21] Guignard Jean-Baptiste. US Patent 10,902,627 B2. Jan. 2021. URL: https://www.lens.org/lens/patent/134-775-570-434-47X/frontpage?l=en.

[M21] Lindahl Aram M. US Patent 11,102,578 B1. Aug. 2021. URL: https://www.lens.org/lens/patent/170-055-842-623-136/frontpage?l=en.

[Pet21] Piemonte Patrick S De Rocha Rosario Daniel Gosnell Jason D Meier Peter. US Patent 10,922,886 B2. Feb. 2021. URL: https://www.lens.org/lens/patent/100-701-708-715-739/frontpage?l=en.

[Sam22] Bates Robert Matthew Canberk Ilteris Carrillo Brandon Fliszar David G Greengard Adam Douglas Kubala Kenneth Meisenholder David Rodriguez Ii Jonathan M Singh Amit Thompson Samuel. US Patent 11,327,310 B1. May 2022. URL: https://www.lens.org/lens/patent/183-474-801-546-494/frontpage.