

# Chapter 5

## VR/AR Output Devices



Wolfgang Broll, Paul Grimm, Rigo Herold, Dirk Reiners,  
and Carolina Cruz-Neira

**Abstract** This chapter discusses output devices and technologies for Virtual Reality (VR) and Augmented Reality (AR). The goal of using output devices is to enable the user to dive into the virtual world or to perceive the augmented world. Devices for visual output play a crucial role here, they are of central importance for the use of VR and AR. First and foremost, Head-Mounted Displays (HMD) must be mentioned, the different types of which are discussed in detail here. However, VR also uses different forms of stationary displays, which are another major topic of this chapter. Finally, output devices for other senses are reviewed, namely acoustic and haptic outputs.

### 5.1 Introduction

How can virtual content be transformed into sensory experiences? What possibilities and alternatives exist to address individual senses? Output devices serve to present the virtual world to users or to expand the real world by generating appropriate stimuli. In this chapter, output devices for VR and AR are presented. A VR or AR system must react to user actions, which are recognized by the use of suitable input devices (see Chap. 4), and generate a corresponding representation, which in turn appeals to the senses of the users (see Sect. 2.1). Commercially available output devices address in particular the visual, acoustic and haptic senses. Here, we will focus on the visual output of the output devices, because it is of outstanding importance for VR and AR. Figures 5.1 and 5.2 show two typical representatives of the most important visual device categories: an HMD and a CAVE-like large

---

Dedicated website for additional material: [vr-ar-book.org](http://vr-ar-book.org)

---

W. Broll (✉)

Department of Computer Science and Automation / Department of Economic Science  
and Media, Ilmenau University of Technology, Ilmenau, Germany  
e-mail: [wolfgang.broll@tu-ilmenau.de](mailto:wolfgang.broll@tu-ilmenau.de)



**Fig. 5.1** Typical consumer HMD with integrated head-tracking and accompanying controllers (© TU Ilmenau 2019, all rights reserved)



**Fig. 5.2** CAVE-like multi-sided projection to visualize a design study (© RWTH Aachen 2013, all rights reserved)

projection. We will then provide an overview of acoustic output devices and some haptic output devices used in VR and AR. In addition, there is a multitude of other, sometimes very special, output devices in the form of prototypes and demonstrators,

which address further senses. For example, there are olfactory displays, acceleration simulators based on galvanic-vestibular stimulation and specific solutions such as the event-controlled generation of wind or the splashing of water, which will not be considered further here due to their limited popularity so far. Pure motion platforms, such as those used for driving and flight simulators, or in amusement parks, are also not discussed here, although one might consider them a large-scale VR output device.

Sect. 5.2 introduces basic aspects of visual output for better understanding of the following sections. Section 5.3 deals with Head-Mounted Displays (HMDs). This includes those for VR as well as for AR. Furthermore, part of it deals with the technical properties of HMDs. In Section 5.4 stationary VR systems are considered. This also includes multi-sided projections such as CAVEs and tiled displays, and their technical challenges as well as technologies for stereo presentation. Sections 5.5 and 5.6 deal with audio output and haptic output devices for VR and AR, respectively. The chapter concludes with a short summary, including a list of questions, a list of recommended literature and the list of references.

## 5.2 Basics of Visual Output

The basic goal of the visual output is to present the *virtual world* (in the case of a VR system) or the *augmented world* (in the case of an AR system) to users in such a way that they can perceive it in a similar way to the real world. The term *display* is used in the following as an umbrella term for monitors, *projection systems* (i.e., projector with projection surface or canvas) and *head-mounted displays (HMDs)*. Monitors and projection systems are used in stationary systems. HMDs refer to displays mounted on the head, which can usually be viewed by both eyes, or sometimes only by one eye. In the following, *HMD* is used as a generic term for both VR and AR glasses. Smart glasses, which are also HMDs, but are primarily used to display information in a small area of the field of view, should be distinguished from HMDs as they are not suitable for VR or AR.

The classification of visual output devices can be based on different criteria. Possible criteria include quality, brightness, field of vision or perception, size of the area of use, uniformity, freedom of movement, usability or location. The following aspects can thus be used to describe visual VR/AR output devices. Starting with viewpoint-related aspects, we will look into the technical parameters of such systems before discussing more user- or usage-oriented aspects.

### Visual Field

The visual field is the area that can be perceived by the eyes of a user without moving the eyes or head. The human visual field is about  $214^\circ$  horizontally (see Fig. 5.3). Each eye covers an angle of approximately  $60^\circ$  towards the nose and  $107^\circ$  towards the outside. The area that can be perceived with both eyes, the so-called binocular cover field, is thus approximately  $120^\circ$  (horizontal). Vertically, the visual field is generally much smaller (approximately  $130^\circ$ – $150^\circ$ ).

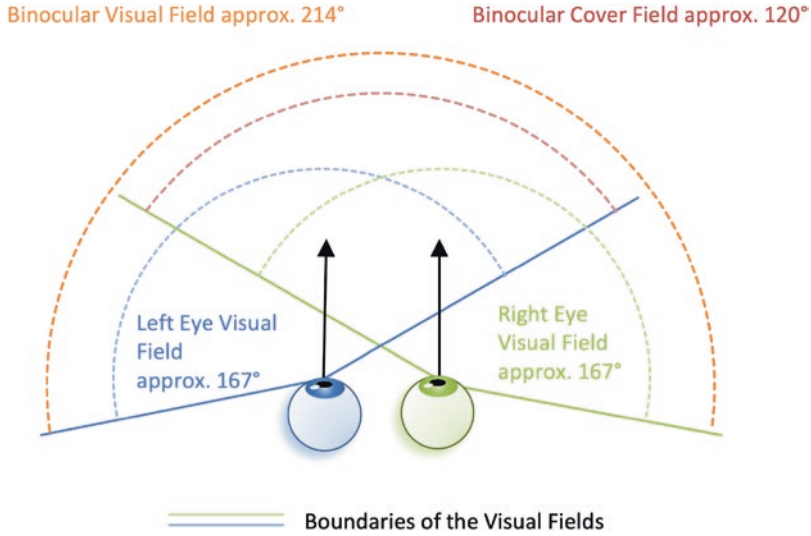


Fig. 5.3 Human visual field

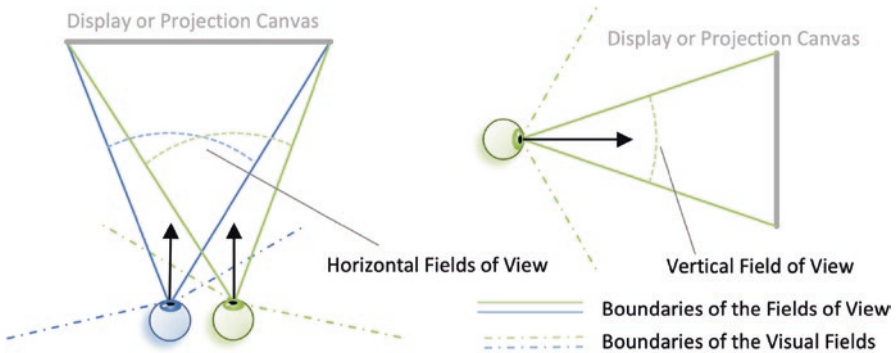


Fig. 5.4 Horizontal and vertical field of view (FoV)

**Field of View**

The field of view (FOV) is the angle of view that can be perceived using a technical device (e.g., an HMD). It is usually specified separately for horizontal and vertical angles (see Fig. 5.4); sometimes the diagonal angle of view is also used. When using the device, the visual field is either reduced to the field of view (e.g., in the case of HMDs) or the field of view covers only a part of the visual field. Thus, one criterion for evaluating visual output devices is the size of the field of view. Here, the absolute size of the display is irrelevant. For example, a smartphone has a very small display in relation to a large screen as a projection surface. However, if the smartphone is used in an HMD (see Sect. 5.3.1), the field of view can be much larger than when

standing several meters in front of a screen. The size of the field of view has a major influence on the degree of immersion and thus the sense of presence.

### **Frame Update Rate**

The *update rate* describes the resolution of an output device in time. The output is done in discrete time steps and is specified either in *Hertz* [Hz] or in *frames per second* (fps). The repetition rate can be different depending on the sensory channel.

### **Latency**

Each output device needs a certain amount of time to output the transferred data (e.g., time until the output is refreshed, due to signal propagation delays in cables or due to the processing of data by algorithms), causing a delay. This is called *latency* (see Chap. 7).

### **Brightness, Luminance and Dynamic Range**

*Brightness* is a subjective measure of the amount of light a user perceives. It is therefore only of limited use for the evaluation of displays. For projectors, the *luminous flux* (in lumen) is usually specified. However, the resulting impression of the user is significantly influenced by the size and nature of the canvas. *Luminous intensity* describes the luminous flux per solid angle (measured in candela). A better way to describe the brightness of planar light sources is therefore the luminous intensity in relation to the area. This describes the *luminance* (measured in candela per square meter). *Contrast* is a measure to differentiate the luminance. The *dynamic range* (DR) or *contrast ratio* describes the ratio between the minimum luminance to the maximum luminance of a display. For visual output devices, luminance and dynamic range are crucial for the capabilities of VR and AR applications: if they are too low, they can only be used in darkened areas (e.g., in a laboratory without direct sunlight). If they are large enough, they can even be used in daylight.

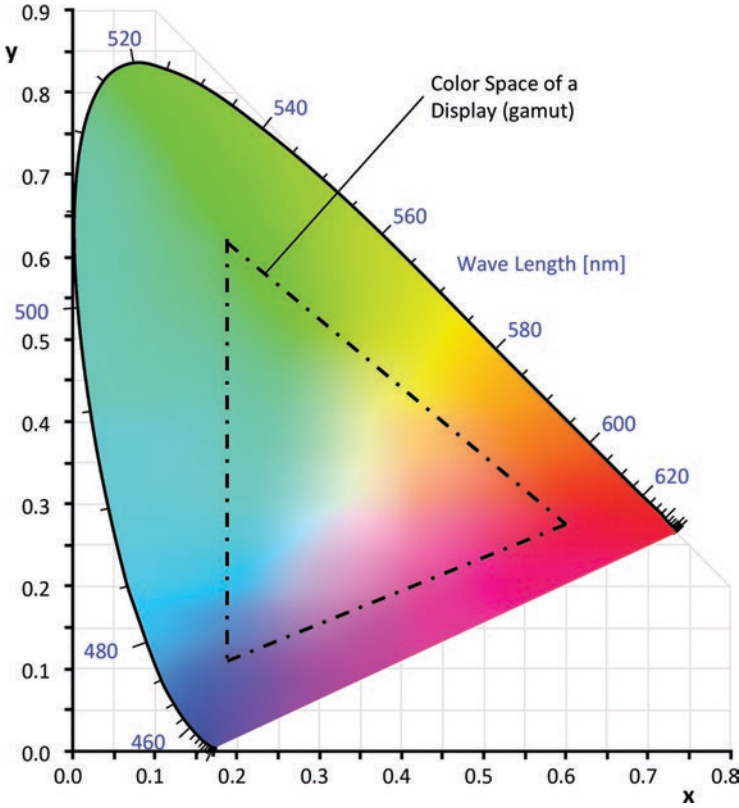
### **Ambient Light**

The ambient light represents the light in the environment of a user or a display. Here, this includes all light in the scene except that emanating from the display itself (i.e., the screen or projector), whether it is sunlight or lamps, directional or non-directional. A bright ambient light usually leads to reduced brightness of the display. Even though the luminance of the light source does not change as a result of this, of course, the perceived amount of light becomes less due to the lower contrast ratio. This more traditional view of ambient light should not be confused with ambient lighting in virtual worlds.

### **Color Reproduction**

To evaluate the quality of a display in terms of colors that can be displayed, the CIE Yxy color system can be used. Figure 5.5 shows this *color space*. In order to describe the display colors, a triangle is drawn in the color system, where the vertices correspond to the three basic colors of the display. The triangle includes all colors that can be displayed (called the *gamut* of the display).

The gamut always covers only a part of the colors perceived by the human eye, which are represented by the area enclosed by the curve. Here, the points on the



**Fig. 5.5** The CIE chromaticity diagram of the colors perceived by humans with the indicated color space of a display (gamut)

curve correspond to the wavelengths of light visible to the human eye, e.g., the wavelength 555 nm corresponds to a bright green color. So you can describe a color in the color system by  $xy$  coordinates. Displays with more than three primary colors exist, covering a larger part of the visible color spectrum. These are not in common use though, and therefore will not be discussed further here.

### Resolution

A visual display presents content using *pixels*. The resolution of visual displays is specified either by specifying the total number of pixels in (mega) pixels (similar to photos) or by specifying the horizontal and vertical number of pixels separately. The *resolution* of output devices is crucial for the details that can be displayed.

### Homogeneity

Output devices should reproduce the virtual world or the virtual parts of an augmented world in homogeneous quality independent of position and direction. With regard to visual output devices, this means that brightness uniformity is maintained as well as that the image sharpness and color reproduction are of constant quality.



### Location (in-)Dependence

Depending on their structure, VR/AR systems can be described as stationary, i.e., location-bound, or *mobile systems*. *Stationary systems* are usually permanently installed and cannot be used at another location (or only with substantial effort). An example of a stationary system would be a large multi-sided projection as shown in Fig. 5.2. Reasons for a stationary use can be, size, weight, dependence on connections (e.g., power supply), or the overall effort required for installation (e.g., a complex calibration process – see also Sect. 5.4.3). Mobile systems can be used independent of location. Systems that are used in stationary location, but can in principle be set up at another location with very little effort are called *nomadic systems*. An example would be a VR system consisting of an HMD, two controllers and a tracking system on tripods.

### Personal output devices vs. multi-user output devices

Generally, a distinction can be made between personal output devices that can only be used by one person (e.g., HMDs or headphones) and multi-user output devices that can be used by several people at the same time (e.g., projections). However, real multi-user output devices further require consideration of each user's individual viewpoint, which typically is not the case for most projection-based systems, where at most the viewpoint of a single (tracked) user is considered. Additionally, software can be used to give multiple personal output devices access to a shared virtual or augmented world (Collabative Virtual Environment – CVE – or Collaborative Augmented Environment).

### Usability

For the application it can be important to what extent users are restricted by the output devices. For example, it may be necessary to put on glasses or attach an HMD to the head. It also makes a difference for the application whether the respective devices are wired or connected via RF technology. The supported room size also has an influence on whether the user can immerse herself in the application or whether she must constantly take care not to exceed predetermined interaction areas. It can also be necessary for the user to always be oriented towards the output device in order to be able to see something. A detailed examination of *usability* is carried out in the context of the consideration of basics from the area of human–computer interaction in Sect. 6.1. The *obtrusiveness* of an output device can be seen as a measure of the extent to which it is considered disturbing. It makes a big difference whether a head-mounted display can be worn like a pair of sunglasses, or whether it can be used like a bicycle helmet due to its weight and dimensions. Ergonomics such as grip or weight distribution can also be critical.

## 5.3 Head-Mounted Displays (HMDs)

*Head-mounted displays* (HMDs) are generally understood to be personal displays that are worn on the head directly in front of the user's eyes. Depending on their design and weight, they are worn like glasses or more like a bicycle helmet. HMDs

often have an integrated tracking system or are combined with a tracking system (see Chap. 4) to continuously adjust the viewing direction and viewing position of the virtual camera according to the current position and orientation of the HMD. HMDs usually use binocular optics, so that the user can perceive the contents stereoscopically. A distinction is made between VR glasses, which isolate the user from the outside world and thus facilitate immersion in a virtual world, and AR glasses, which enrich the user's real environment by adding virtual content.

### 5.3.1 VR Glasses

This section deals with HMDs for VR applications. Figure 5.6 shows typical consumer VR glasses. VR glasses usually use a closed design so that the user is visually completely isolated from his environment, only allowing him to see the virtual world. The field of view here sometimes almost matches the natural visual field. This may result in complete immersion.

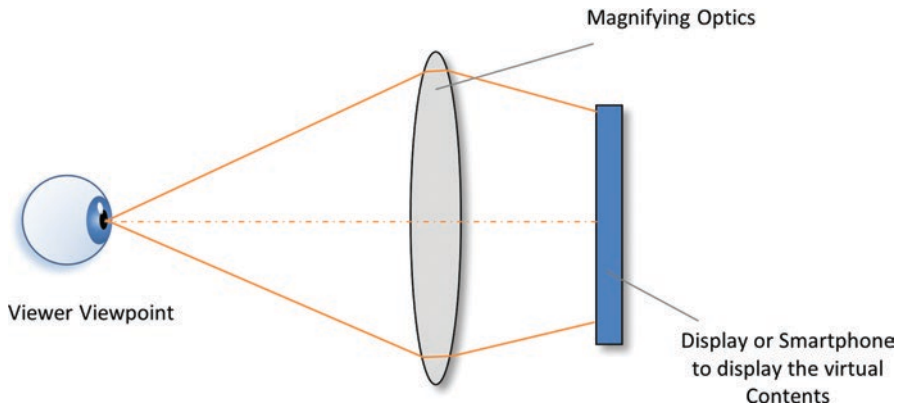
Consumer HMDs are often based on a simple magnifier design. Here, a simple magnifying optic is used for each eye, allowing the user to focus on the actual display (see Fig. 5.7). Depending on the individual design, a single display or two separate displays are used.

In a single display, the eyes see different areas of the same display, allowing stereo vision. In two separate displays, they are often slightly tilted towards each other to cover a larger field of view. Through the use of LCD or OLED displays, which are also used in smartphones, such displays now achieve high resolutions combined with high luminosity and reasonable prices.

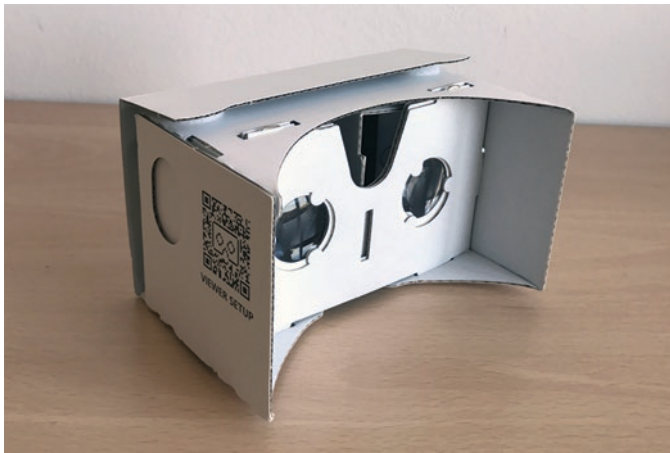


**Fig. 5.6** Typical representative of consumer VR glasses with integrated sensors for tracking (© TU Ilmenau 2019, all rights reserved)





**Fig. 5.7** Simple magnifier principle as typical design of current VR glasses



**Fig. 5.8** Simple Cardboard HMD using the magnifying glass principle

Low-cost versions do not have their own display. Instead, a smartphone is used as the display, which is inserted into an HMD rack. Cardboard displays (see Fig. 5.8) use only a holder made of cardboard in which two lenses are inserted for the optics.

In contrast, high-end systems may even use multiple displays per eye. See also Sect. 5.3.4. for such an approach. Besides the simple magnifying glass design, alternative designs exist. For example, prism-based VR glasses allow a very compact design because the prism optics significantly reduce the overall depth (see also Fig. 5.11). However, commercially available smartphone displays cannot be used for this purpose. Therefore, displays of this type are not very common for more recent HMDs.

The classical design of HMDs often used optics with mirrors. Here, a mirror and possibly an additional semi-transparent mirror were used to make the virtual content of the display visible to the eye.

### 5.3.2 AR Glasses

There are two basic approaches to glasses for Augmented Reality, or *AR glasses* for short: *optical see-through displays (OST displays)* and *video see-through displays (VST displays)*. While the first type optically superimposes real and virtual images in the user's view, the latter type uses video cameras to capture the environment and then superimpose it with virtual content during rendering (see Chap. 8). In the following sections, we will first take a closer look at OST displays and their construction methods, before we go on to discuss VST displays in more detail.

#### Optical AR Glasses

When we talk about AR glasses, we usually mean such OST displays. While the view of reality is always direct and thus immediate, the virtual contents are only optically superimposed. Thus, in contrast to VST displays, there are no limitations in terms of quality and resolution when viewing the real environment. As with VR glasses, a virtual image is generated by a display and enlarged by a lens. However, here it is projected into the user's eye with the help of a beam splitter. A major problem is that the virtual image is projected at a fixed distance, which may differ from the distance of real-world content currently focused by the user.

A general problem of OST-HMDs is the insufficient background contrast ratio in bright environments. Due to the low luminance ratio of the display with respect to ambient light, virtual content is perceived only faintly, i.e., it appears increasingly transparent to reality. In practice, the transparency is reduced accordingly in bright environments to provide the required background contrast ratio for a given front luminance. The different superimposition methods usually lead to a significant reduction of the amount of incident light, so that the surrounding reality appears darkened to the observer. In comparison to viewing without a display, in some cases only about 25% of the light reaches the eye of the observer. This corresponds approximately to sunglasses with a medium protection factor (S2). Due to the low light intensity, most HMDs of this type are not, or are only partially, suitable for sunlight use, even if the transparency is reduced by means of appropriate filters. There are a large number of different designs of OST displays, the most important of which will be briefly explained in the following.

#### Waveguide Optics Glasses

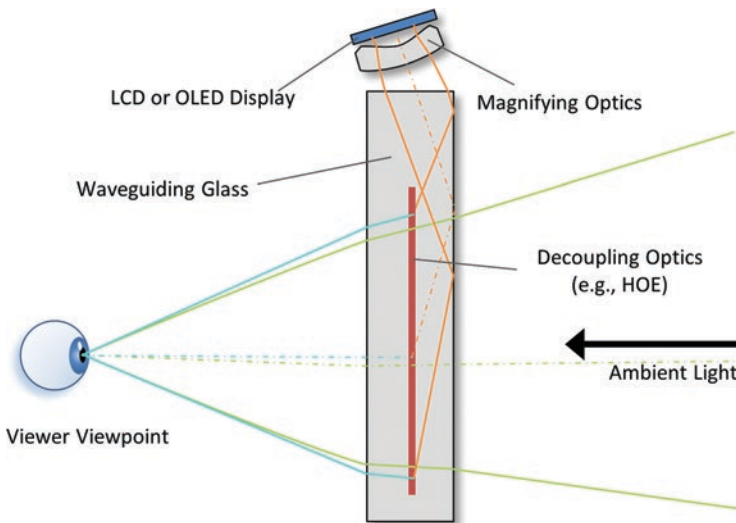
Strictly speaking, this refers to a whole range of different approaches. These approaches have in common that the light is fed into a largely planar glass body, which acts as *waveguide optics*, and then travels through it as in a fiber optic cable by being reflected from the outside of the glass body. The decisive factor now is how the light enters the glass body (coupling-in) and how it exits it (coupling-out).

Special optical elements are used for this purpose. These result in the light being transmitted and radiated in a previously predefined direction. If the light is fed into the waveguide from the side, there is no need for corresponding elements for coupling-in (see Fig. 5.9). The elements for coupling-out are arranged in such a way that the light leaves the waveguide at the appropriate point in the direction towards the eye. In this way HMDs are possible that look more like conventional glasses due to their flat optics. For a full color display, three layers of light guides must be arranged one above the other, since the individual color channels must be transmitted separately due to the dependence of the refraction on the wavelength of the light. Holographic waveguides applying *holographic optical elements* (HOE) are among the best known representatives of this approach (see Fig. 5.9). By analogy with holograms, light beams impinging on the HOEs generate a secondary light beam in a predefined direction, while the ambient light passes through them without interference. Other approaches include diffractive, polarized, and reflective waveguides.

For AR displays like the Microsoft HoloLens (see Fig. 5.10) or HoloLens 2, the Meta 2 or the display from Daqri, waveguides are used. However, another, mostly curved glass is often used before the actual light guide for protection and shading.

### Prism-Based Glasses

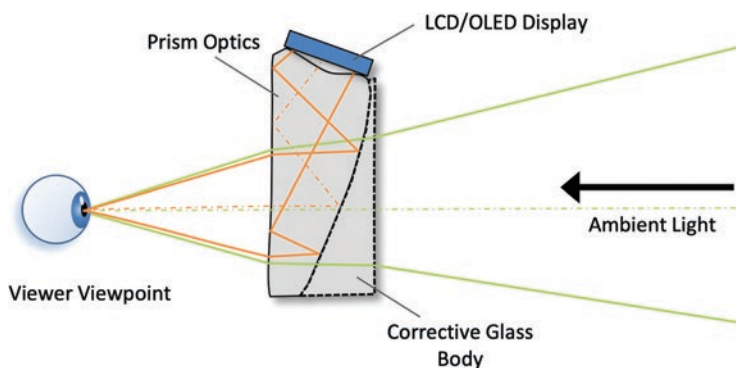
Prism-based glasses, which are also used for VR, enable a relatively high light output with a more compact design when compared to other designs. For use as a see-through display, the prism is complemented by a glass body with parallel outer surfaces, so that the ambient light can pass through the glass without being refracted (see Fig. 5.11).



**Fig. 5.9** Schematic principle of operation of an OST display based on waveguides (here applying holographic optical elements)



**Fig. 5.10** HMD with clearly visible waveguides for the individual color channels (see magnification) (© TU Ilmenau 2019, all rights reserved)

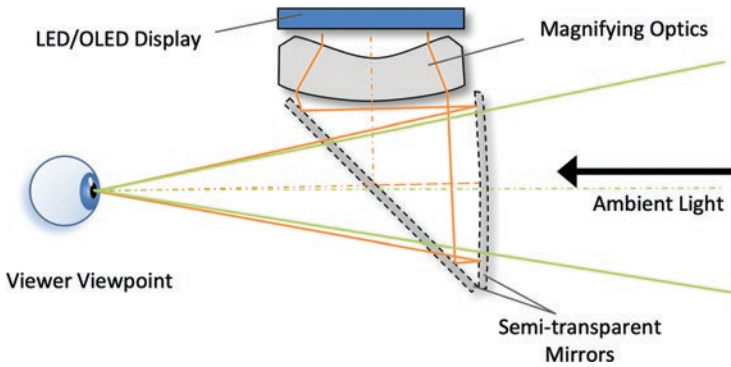


**Fig. 5.11** Schematic principle of a prism-based OST display consisting of two glass bodies. The complement is shown as a dashed line

The best known representative of this type of display so far has been Google Glass, although as a so-called SmartGlass it was not really suitable for AR applications due to its small field of view.

### Mirror-Based Glasses

The use of semi-transparent mirrors has been the preferred design for OST displays for a long time. The content was displayed on an LCD or OLED display and was magnified by means of magnifying optics located directly beneath or above the display. The classic approach used to apply two semi-transparent mirrors, one mounted vertically at the front and a second mirror at a  $45^\circ$  angle right behind it towards the eye. Thus, the ambient light simply passed both mirrors, while the light



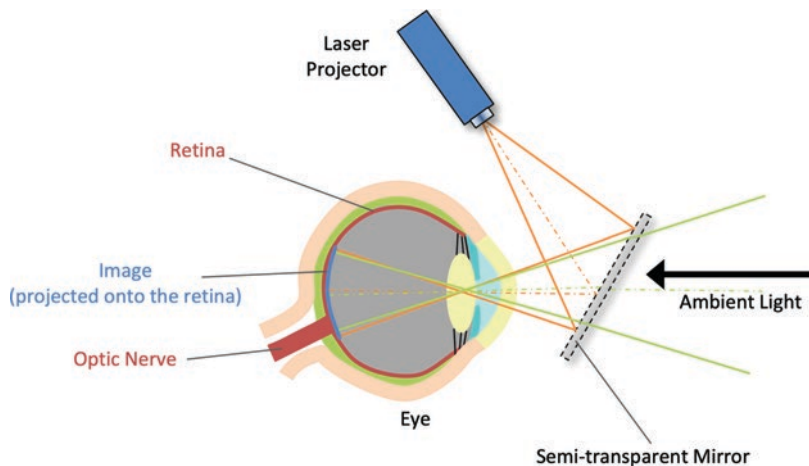
**Fig. 5.12** Schematic structure of a classic mirror-based OST display

from the display was reflected by the diagonal mirror towards the front face, then reflected back towards the eye, passing the diagonal mirror (see Fig. 5.12). A simpler approach would use just a single semi-transparent mirror (with the back mirror flipped by  $90^\circ$ ). While this also allows for better light output, it requires stronger magnification, which typically results in less compact HMDs.

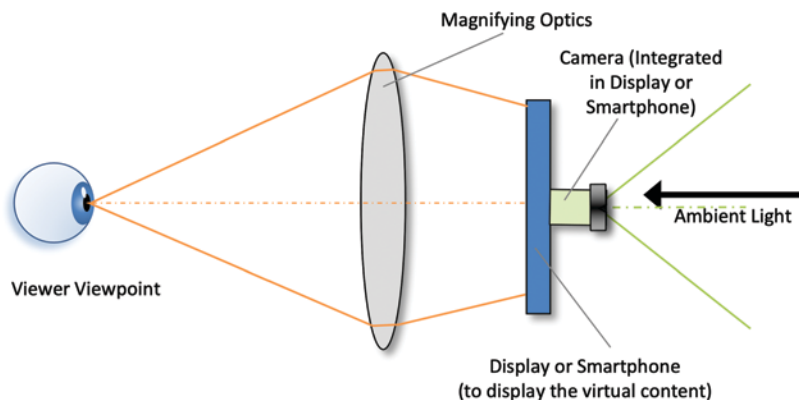
More recently, mirror-based approaches have regained popularity, as they enable the realization of a simple OST display using a smartphone. The smartphone is usually inserted vertically into a holder in the area in front of the user's forehead. A mirror reflects the image downwards into the area in front of the eyes. Here, there is usually also an optical system in the form of one magnifying lens for each eye. There are also models without lenses, but as the eyes then have to focus on the smartphone display, they do not allow for a relaxed viewing position of the virtual image in space. In the area in front of the eyes there is a diagonally aligned semi-transparent mirror, which on the one hand reflects the image of the display towards the eyes and on the other hand allows a view of the environment. In cheap models a perspex panel is used for this purpose. There are also versions where the upper mirror is omitted so that the smartphone is inserted horizontally. Similar models, but without a semi-transparent mirror, use the smartphone camera to realize VST AR glasses.

### Retinal Glasses

*Retinal HMDs* do not have a display in the actual sense, since the content is projected directly onto the retina (see Fig. 5.13). This is also called a *virtual display*. This approach offers two major advantages: on the one hand, a complex optical system is avoided, and on the other hand, despite an extremely compact design, very large fields of view can be generated because no optics in front of the eye have to cover the displayed field of view. Modulated laser light is used as the light source, which is directed into the eye via a semitransparent mirror or prism. Until now, only monochromatic OST glasses of this type have been commercially available. For a full color display three separate lasers (RGB) would be necessary.



**Fig. 5.13** Schematic principle of a retinal virtual OST display (here with mirror)



**Fig. 5.14** Optical principle of recent HMDs with integrated cameras. Since the cameras are mounted right in front of the eyes, their point of view is almost identical

An alternative design approach replaces the laser projector by an RGB light source and the mirror by a DLP (Digital Light Projector) microdisplay.

### Video AR Glasses

Video AR glasses, more precisely Video See Through (VST) displays are basically HMDs as they are used for VR (see Sect. 5.3.1). This means that the user is completely isolated from the environment, at least when the device is completely closed. In contrast to their use for VR, however, a video image of reality is inserted in such a way that the user has the impression that she can look at the world around her through glasses. For this purpose one or two video cameras are attached to the HMD or are directly integrated into it (see Fig. 5.15).



**Fig. 5.15** Example of an HMD with integrated cameras



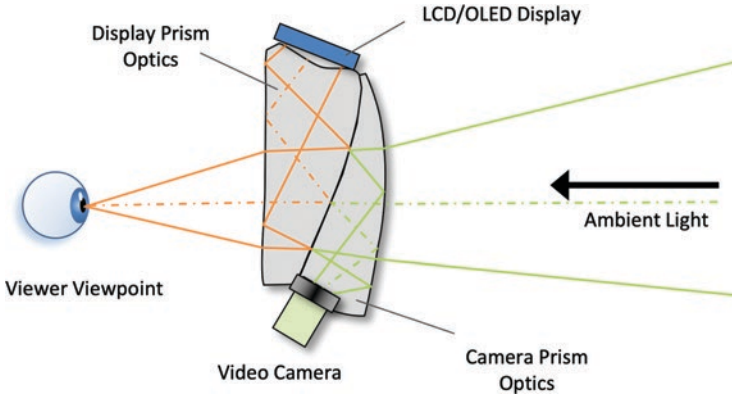
Since the human eye only sees the information projected by the display, the real and virtual content are always in the same focusing plane. Thus, those parts of the real world that are not in focus for the camera cannot be focused by the user either. Furthermore, the perception of the real world is in a reduced resolution and has limited dynamic range due to the camera as well as the display used, when compared to the direct view by the human eye.

The captured video image is correctly inserted as a background image when rendering the scene. Basically, the field of view of the camera must be larger than that of the HMD used. In most cases, it is not possible to position the video cameras directly in the area of the beam path in front of the eyes. Therefore, when correcting the perspective of the camera image, translational and/or rotational offsets often have to be deducted in addition to the rectification and restriction of the viewing angle. Without this, the user will have difficulty in estimating distances, and proportions correctly (at least temporarily until his visual system has adapted). HMDs in which the lens of the camera is positioned directly in front of the eye in the direction of vision, or the light rays arriving there are deflected into the camera, avoid this problem (see Figs. 5.14 and 5.16).

### **5.3.3 General Characteristics and Properties of HMDs**

In this subsection some basic characteristics and properties of HMDs will be reviewed and discussed. Depending on the type of application planned, these can sometimes be decisive for the selection of an HMD to be used.

As already introduced in Sect. 5.3.1 the basic optical principle of VR glasses is that of a magnifying glass. Let us have closer look at its general characteristics according to Melzer and Moffitt (1997). The display, which the user looks at through the lens, is positioned at the distance of the focal length to the lens (see Fig. 5.17).



**Fig. 5.16** Schematic principle of a prism-based video see-through display with cameras for recording in the viewing direction

### Pupil Forming Vs. Non-pupil Forming HMDs

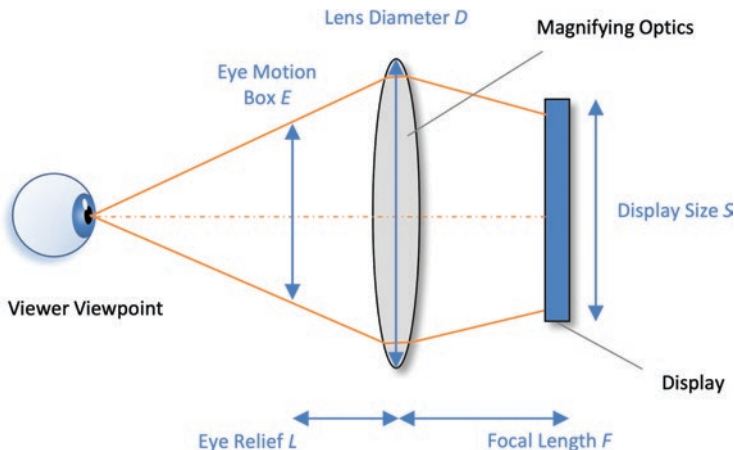
On the optical side, there are two basic approaches to realize an HMD. On the one hand, non-pupil forming HMDs are used, which are based on the principle of a simple magnifying glass. On the other hand, we have pupil forming HMDs, which are based on a projection (Cakmakci and Rolland 2006). An important parameter, which refers to the use of the HMD and is specified for non-pupil forming HMDs, is the so-called *eye motion box* (sometimes also called the head motion box or just eyebox) in its vertical and horizontal dimensions. This is the size of the optical opening of the HMD on the eye side. The larger the eye motion box, the further the position of the HMD can be shifted in relation to the user's eye without restricting the visibility of the virtually projected image. In pupil forming systems, however, a diameter is specified at the optical output of the HMD within which the viewer can see the virtual image. This parameter is called the *exit pupil*. In contrast to the eye motion box, this diameter remains constant regardless of the distance between the user's eye and the HMD optics.

### Field of View (FoV)

Based on the optics shown in Fig. 5.17, the field of view can be calculated for the horizontal, vertical and diagonal by eq. 5.1, where  $F$  represents the focal length of the lens and  $S$  is the size of the display horizontally, vertically or diagonally, respectively.

$$FoV = 2 \arctan \left( \frac{S}{2F} \right) \quad (5.1)$$

Theoretically, the FoV calculated according to eq. 5.1 is independent of the diameter of the lens  $D$ . In practice, however, there is the problem that at a higher distance between the eye and the magnifying lens (the so-called *eye relief*  $L$ ) not all light rays of the display can reach the eye via the lens. In this case, for technical



**Fig. 5.17** Optical principle of VR glasses, where  $F$  represents the focal length,  $S$  the size of the display,  $D$  the diameter of the lens,  $E$  the eye motion box and  $L$  the eye relief

reasons, the lens diameter and the eye relief according to eq. 5.2 determine the maximum possible field of view.

$$FoV_{\max} = 2 \arctan \left( \frac{D}{2L} \right) \tag{5.2}$$

Eq. 5.2 is valid for  $D < L (S/F)$ . HMDs, which optically follow the simple magnifying principle, have an eye motion box  $E$  instead of an exit pupil. The size of the eye motion box for the horizontal, vertical and diagonal direction can be determined according to Eq. 5.3 (Melzer and Moffitt 1997).

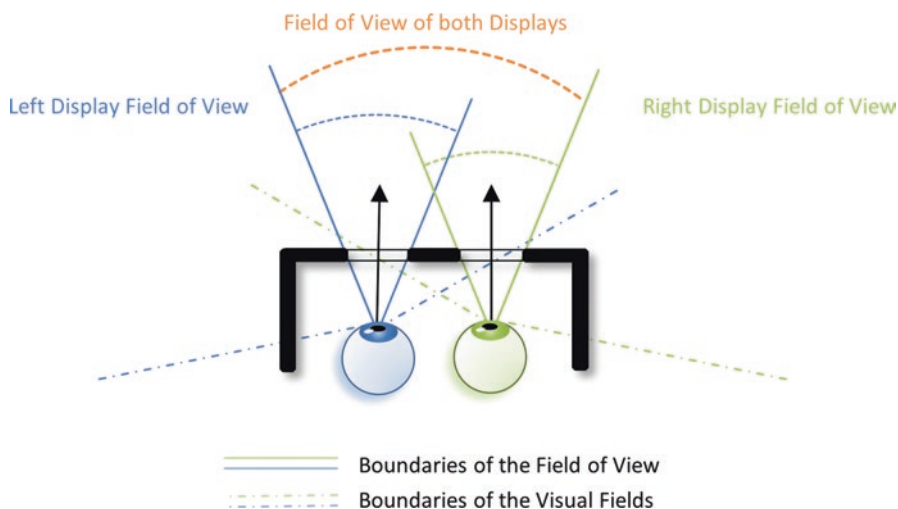
$$E = D - \frac{LS}{F} \tag{5.3}$$

**Accommodation Distance**

The *accommodation distance* indicates the distance from the user’s eye at which the virtual image appears. Most optical see-through HMDs have a virtual image at infinity. For a simple HMD using the magnifying glass principle, the relation between the lens position and the distance of the virtual image  $D_{\text{virt}}$  can be described by eq. 5.4:

$$D_{\text{virt}} = \frac{dF}{F - d} + L \tag{5.4}$$

Here the parameter  $d$  is the distance between the lens and the display. If the display is within the focal length of the lens, as shown in Fig. 5.18, the denominator in



**Fig. 5.18** Field of view of an HMD compared to the user's visual field when using binocular AR-glasses in a closed design

eq. 5.4 becomes zero and the virtual image is at infinity. If  $d$  is smaller than  $F$ , the virtual image is projected enlarged. This means that the projection is larger than the illuminated area of the display. If  $d$  is larger than  $F$ , the virtual image is projected scaled down.

### Interpupillary Distance (IPD)

The interpupillary distance (IPD) is the distance between the two eyes of an observer. It is usually measured from pupil to pupil and in the range of 6 to 8 cm for an adult. Many, though not all, HMDs allow the eye distance to be adjusted to suit the individual user. Otherwise, especially in combination with a small eye motion box, parts of the displayed image may be cut off. The eye distance also has a direct influence on the perception of sizes and distances of the virtual content (see Chap. 2).

### Monocular vs. binocular HMDs.

With HMDs, one can basically distinguish between monocular and binocular variants. *Monocular HMDs* have only one display with associated optics for one eye, while the other eye usually remains free. While this can be useful for certain AR applications, it drastically reduces immersion in VR. *Binocular HMDs* have separate optics for each eye, allowing different content to be viewed. Only this enables stereoscopic perception and thus a spatial impression. In contrast to binocular displays, there are also binocular displays in which both eyes look at the same image through separate optics. However, this does not allow stereoscopic perception. If both eyes look at different areas of one and the same display via separate optics (e.g., in smartphone-based HMDs), they usually see different images nevertheless.

### Open Vs. Closed HMDs

The design of an HMD also affects the perception of the virtual environment (for VR) or the augmented environment (for AR). Basically, one can distinguish between open and closed designs of HMDs. While the closed design limits the visual field of the observer to the field of view of the HMD, the open design allows unrestricted perception of the environment outside the display. Figure 5.18 illustrates an HMD of closed design using OST AR-glasses as an example.

The illustration clearly shows how much the visual field of the observer is restricted by the field of view of the HMD. Stereoscopic vision is only possible in the area where the fields of view of the display for left and right eyes overlap. This is called the stereoscopic or binocular field of view. Its size in VR glasses depends on the distance at which the display appears to the viewer due to the optics. It can therefore vary between 0 and 100% of the individual fields of view.

Small fields of view are problematic for several reasons. With VR glasses as well as with closed AR glasses they lead to tunnel vision and thus to increased cybersickness due to the lack of peripheral perception. An additional complication is that closed AR-glasses shield the viewer from the perception of a large part of his real environment. This is particularly problematic when used in unprotected areas (such as mostly outdoors), since the perception of stairs, cars, bicyclists, etc. occurs much later than normal.

*Monocular HMDs*, i.e., those that only superimpose the vision of one eye, allow an unrestricted view of the surroundings, at least with the other eye. In the field of working environments and military application scenarios, such designs (see Fig. 5.19) are therefore strongly represented, whereby a largely open design is usually used here, so that only the display mounting causes a certain restriction of the visual field.

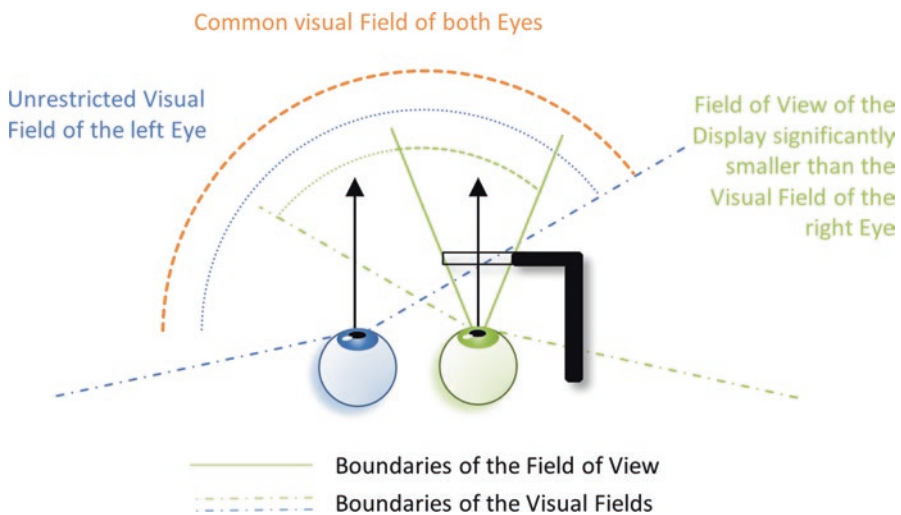
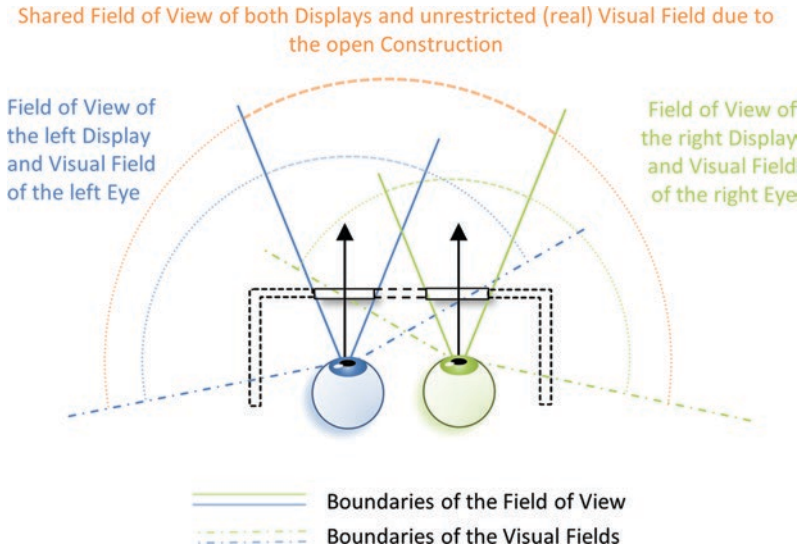


Fig. 5.19 Field of view and visual field for a monocular display (right) in a closed design



**Fig. 5.20** Fields of vision of a binocular HMD with an open design

HMDs in an open design enable users to directly perceive the environment outside the HMD's field of view. Thus, the peripheral vision of the user is not restricted, although virtual content remains limited to the area of the HMD's field of view (see Fig. 5.20). With open AR glasses, it can be disturbing that the area covered by the HMD usually appears significantly darker than the part that is not covered. Furthermore, the limited field of view in comparison to the field of vision causes the problem that virtual objects leaving the field of view of the AR glasses, are only partially displayed at its edges, while the real background remains continuously visible (cf. Figure 5.21). This effect immediately destroys the viewer's impression of a correct registration of the corresponding virtual object in the real world (see frame cancellation, Chap. 2).

### Contrast Ratio

As we have previously learned, the dynamic range or contrast ratio  $CR$  is the ratio between the brightest and darkest representation. For VR glasses, this is the ratio between the luminance of a maximally bright pixel and a completely dark pixel:

$$CR = \frac{L_{\max}}{L_{\min}} \quad (5.5)$$

For OST-AR glasses (see Sect. 5.3.2), however, the contrast ratio between the luminance of the display (the so-called front luminance) and the background of the real environment is of particular interest. The contrast ratio of the background  $CR_{\text{back}}$  is thus the ratio of the front luminance  $L_{\text{front}}$  minus the background luminance  $L_{\text{back}}$  to the background luminance:





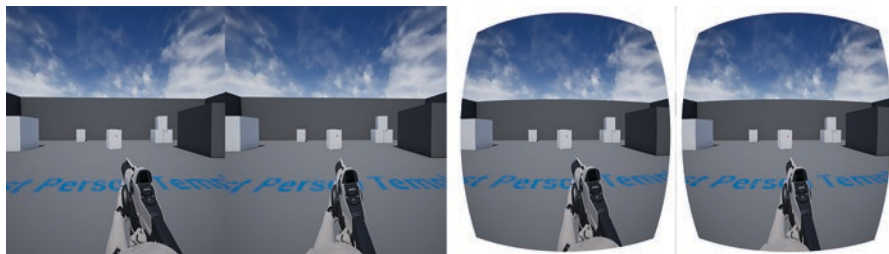
**Fig. 5.21** Problems with the display of virtual objects at the boundaries of the field of view with an open design

$$CR_{back} = \frac{L_{front} - L_{back}}{L_{back}} \quad (5.6)$$

When using OST-AR glasses outdoors, especially in bright sunshine (e.g., on an unclouded day), the brightness of the projected image must be correspondingly high so that the virtual content stands out sufficiently from the background (see also Sect. 8.1.2). Indoors, on the other hand, for example, AR-supported assembly work in a factory building, a significantly lower front luminance may be sufficient to provide the same contrast ratio with respect to the surroundings. With OST-AR glasses, the see-through transparency  $T_{see-through}$  indicates how bright the user can perceive the real environment or by how much the brightness of the environment is reduced by the HMD, similar to sunglasses.

### Distortion

Due to the highly distorting simple magnifying optics used especially in recent consumer VR glasses, the displayed images must be pre-processed (see Fig. 5.22). This is done by applying an appropriately parameterized equirectangular function to the images with the aim of providing the user with an undistorted image after being distorted by the lens. For this purpose, manufacturers often provide corresponding *distortion maps*.



**Fig. 5.22** Display of the two partial images for the left and right eyes of the viewer (left without, right with perspective predistortion for HMD)

In order to be able to judge the image quality of the virtually projected image, the horizontal, vertical and diagonal distortions of the virtual image in relation to the original image are compared. Distortions occur if the virtual image does not have the same projection scale in every area. Distortions become noticeable, for example, when the virtual image has the outer shape of a cushion.

### 5.3.4 *Special HMDs*

#### **Eye Tracking for VR and AR Glasses**

With the availability of VR glasses for the consumer sector, a need to capture where the user is looking in virtual worlds quickly arose. This information can be used for investigations of user behavior in user tests as well as for the fixation of virtual objects for selection and manipulation. A further application area is *Foveated Rendering* (see Sect. 7.1.4), in which different display areas are shown in different detail depending on the retinal area on which they are perceived. While a rigid division of the field of view can lead to disturbing effects when focusing on peripheral areas, in combination with eye-tracking it can be ensured that the rendering always takes place in the center of the current viewing direction at the highest quality.

On the one hand, various commercial eye-tracking systems are now available for direct installation in consumer VR glasses. On the other hand, HMDs are increasingly being delivered directly with integrated eye-tracking for user interaction. Examples are the HoloLens 2, the Magic Leap One (see Fig. 5.23) or the HTC Vive Pro Eye. Commercial systems are usually based on the fact that for each eye several infrared LEDs are arranged mostly in a ring around the HMD's optics. The positions of the reflections of the LEDs are then recorded by a camera, which is also mounted directly next to the HMD's optics. Based on the points identified in the camera image, the direction of vision of the eye can then be calculated (see also Sect. 4.5.5).

**Fig. 5.23** Optical see-through AR glasses with integrated eye tracking (© W. Broll, all rights reserved)



### Multi-Display Glasses

Some high-end systems, like the Varjo glasses, combine multiple displays to achieve very high perceived resolutions (Lang, 2018). The basic idea is to combine a regular, large field-of-view display with a much smaller foveal display that only covers the center of the field-of-view (the fovea). The approach extends *Foveated Rendering* to the usage of high-res foveal displays. The perceived quality is significantly better than one display systems, but the additional effort in design and production results in significantly higher prices. One approach to realize this is the application of a semi-transparent mirror in combination with eye-tracking (see Fig. 5.24).

### Adaptive HMDs

All currently commercially available HMDs have the problem that due to static optical elements the virtual image always has a fixed distance to the eyes of the user. However, the distances of the real objects the user is looking at vary. Since the human eye cannot focus on different distances at the same time, one of them is usually out of focus. One possible solution is an adaptive HMD (Herold et al. 2015). Such adaptive HMDs are based on a liquid lens, which makes it possible to adjust the focal length and thus also the distance of the virtual image to the user.

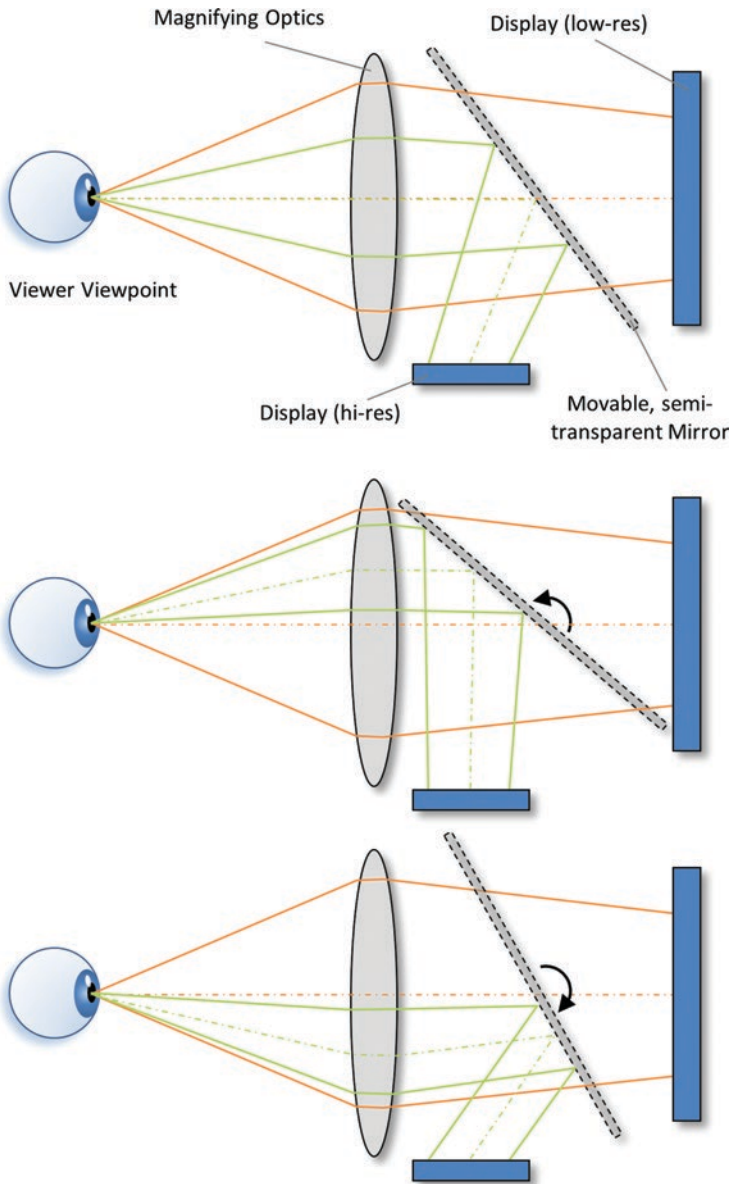


Fig. 5.24 Multi-display system supporting foveated rendering

## 5.4 Stationary VR Systems

Stationary VR systems use one or more mostly vertically oriented (i.e., standing upright) displays (projection screens or large monitors) for visual output. Depending on the type of system, alternatively or additionally horizontally oriented projection surfaces or monitors or even spherical projection surfaces are used. The output is usually stereoscopic. For the correct calculation of perspective, the user's head is usually tracked. The necessity is easily recognized by the following example: if the user bends to the right or left to look past a virtually represented column, the virtual world must be displayed accordingly. This requires an individual calculation of the images shown on the displays from the user's perspective. This is also the reason why even stationary VR systems are still single-user systems almost without exception (and despite the fact that they are often used by several users in parallel). The one exception is new systems that use extremely high-speed projectors that can display a sufficient number of images per second (a typical example would be 360 fps) to display separate stereo pairs for multiple users.

In principle, AR systems can also be stationary. In particular, spatial AR systems such as projection-based AR are usually stationary. While most of the technical aspects discussed here also apply to them, they are dealt with in Sect. 8.4.

### 5.4.1 *Single-Sided Displays*

Many stationary VR systems are simple single-sided displays, i.e., a single projection surface as large as possible is used on which the virtual world is displayed stereoscopically. In the simplest case this can also be just a large monitor.

It is crucial for a high level of user immersion that the display's field of view covers as much of the user's visual field as possible. The larger the field of view (FOV), the less often a virtual object from the user's perspective will reach the edge of the display, destroying the spatial (stereoscopic) impression by frame cancellation (see Sect. 2.4.3). This means that the smaller a display area is, the closer the user has to be in front of it, or the larger the display area, the further away the user can be (cf. Figure 5.25).

#### **Vertical and Horizontal Displays**

Single-sided displays are usually oriented vertically (upright) so that the user(s) can stand or sit in front of the display, comparable to a 3D cinema. Depending on the application, however, horizontal (lying/tabletop/floor) displays are also useful (as an example see *Responsive Workbench*; Krüger and Fröhlich (1994)). In tabletop systems, virtual objects usually appear to lie on the table or hover above it. Users have to stand very close to the display to avoid frame cancellation.

Both single-sided vertically and horizontally arranged displays can be used to view virtual content with multiple users at the same time (see Fig. 5.26). Usually, however, a perspective correct stereo view is only generated for one user. All other users see

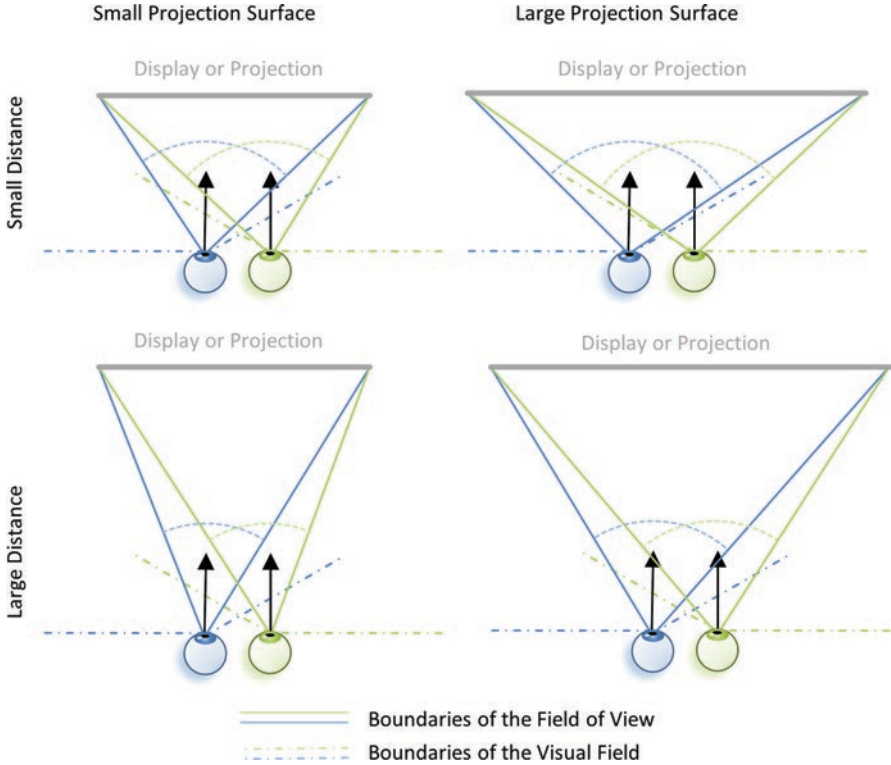


Fig. 5.25 Dependence of the field of view on the size and distance of the display or projection surface



Fig. 5.26 Vertically and horizontally arranged single-sided displays and projections

virtual content stereoscopically, but in a different position. For an interaction with virtual objects in particular, perspective correct stereo presentation is essential.

### Front and Rear Projections

If no monitors but projection systems are used for a display, the projector can basically illuminate the projection surface (screen) from the user's side or from the side opposite to the user. If the projection is made from the same side from which the

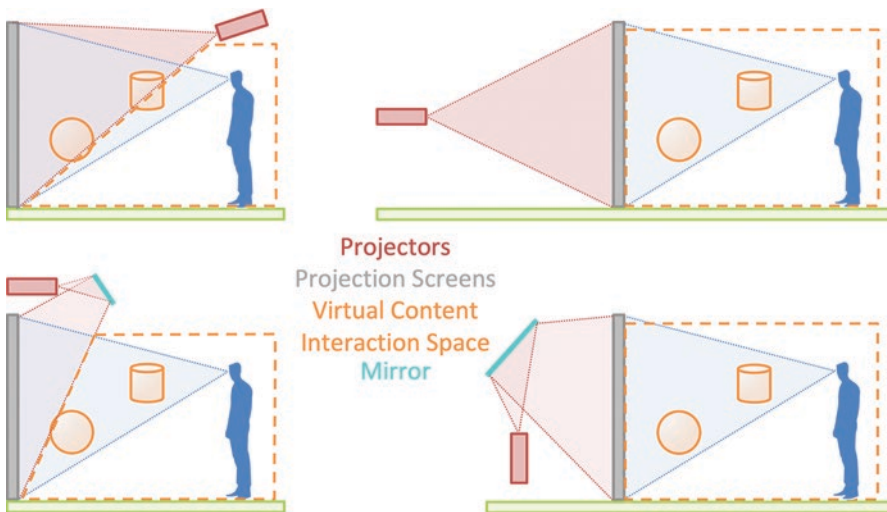


user looks at the projection surface, this is called *front projection*. If, on the other hand, the projection is made from the opposite side, i.e., the rear side, it is called *rear projection*. With front projections, the user must maintain sufficient distance from the projection surface to avoid obstructing the beam path of the projector. The shadows cast on the projection surface can also lead to frame cancellation.

However, if the user has to maintain a greater distance from the display, this inevitably leads to a restriction of her interaction space (see Fig. 5.27) and at the same time increases the risk of frame cancellation. With rear projection systems, these disadvantages are generally avoided, but a correspondingly larger space is required for the beam path of the projector behind the projection surface. Furthermore, specific, usually more costly, canvases must be used for rear projection. By employing mirrors in combination with ultra wide angle projector lenses or ultra short throw (UST) projectors, the space required for both front and rear projections can be significantly reduced, whereby front projections also benefit from an increased interaction space.

### 5.4.2 Multi-Sided Displays

With a single flat display, it is difficult if not impossible to achieve complete coverage of the user's visual field by the field of view and thus a high degree of immersion. Accordingly, there are numerous approaches that combine several display surfaces or realize curved display surfaces. Well-known representatives of the first group are so-called *CAVEs* (*Cave Automatic Virtual Environments*) and *L-Shapes*; the second group particularly includes *spherical displays*.



**Fig. 5.27** While front projections limit the interaction space of the user, rear projections require considerably more space. Usage of mirrors or ultra short throw projectors can significantly reduce space requirements

## L-Shapes

An L-shape uses two displays. One display is usually mounted vertically, while the second display is usually placed horizontally and has an edge directly adjacent to the first display (in side view, the two displays placed next to each other thus resemble the letter L; hence the name). L-Shapes offer the great advantage over single-sided displays, especially in stereoscopic presentation, that the volume for displaying virtual content is significantly larger. Thus, virtual objects can be displayed up to the immediate vicinity of the user, e.g., for hand-based interactions. Frame cancellation, which often occurs in the lower part of a vertically arranged display in single-sided displays, is thus effectively prevented by the second horizontally arranged display (see Fig. 5.28). Similarly, in the case of primarily horizontal displays (e.g., Responsive Workbench; Krüger and Fröhlich (1994)), a second vertically arranged display prevents frame cancellation when viewing virtual objects close to the opposite side of the horizontal display.

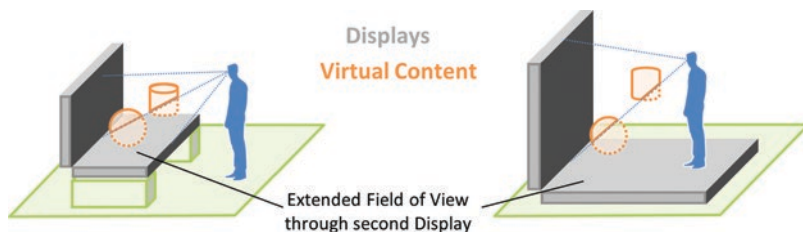
For larger L-shapes it may be necessary for the user to stand on the horizontal display. If this is a monitor or panel, the challenge is that the display must not only have sufficient optical properties but also sufficient static stability to reliably support one or even several users.

## Spherical Displays

Spherical displays or curved screens (also known as *dome projection* when covering 360°) consist of a *curved screen* on which the image is usually displayed with the aid of several projectors (see Fig. 5.29). The projection surface has the shape of a sphere, a cylinder or a cone, or a cutout of these basic shapes. The projector image must be distorted according to the shape of the projection surface. If several projectors are used (see also the next section on tiled displays), their images cannot be projected without any overlapping. The overlapping image areas must therefore always be adjusted accordingly, i.e., masked by software or physical barriers, making these transitions appear seamless to the user.

## Cave

A *CAVE* (*Cave Automatic Virtual Environment*) is a cube-shaped arrangement of displays with the user standing inside the cube (Cruz-Neira et al. 1992). Figures 5.2 and 5.29 show two installations of CAVE-like displays. Depending on how many sides of the cube are designed as displays, we speak of three- to six-sided CAVEs.



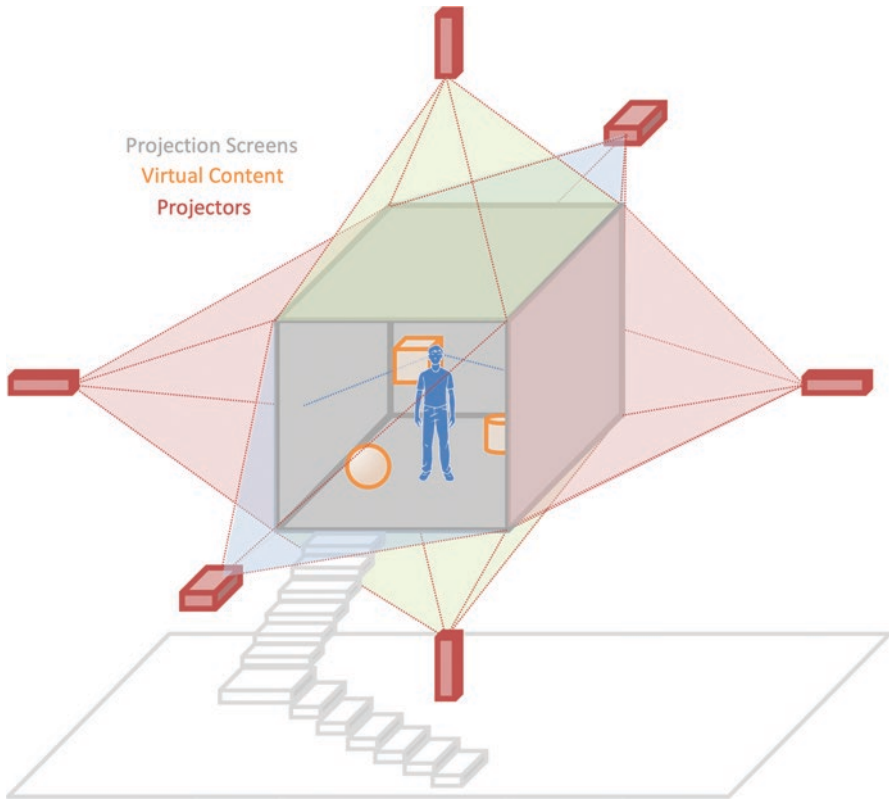
**Fig. 5.28** L-shapes expand the working space available for stereo vision and reduce frame cancellation



**Fig. 5.29** Example of a curved screen projection. (© Fraunhofer IFF 2013. All rights reserved)

In a six-sided CAVE the user is completely surrounded by the virtual world. In this case, only rear projections can be used, which not only requires a sufficiently large space behind each of the projection surfaces, but also, due to the ceiling and floor projections, equally large space above and below the CAVE (see Fig. 5.30). If a projection from above is used for the floor, this first implies that a ceiling projection is no longer possible and second that the users are right in the beam path of the projector. However, such floor shadows are often perceived as less disturbing, since users are used to casting a shadow on the floor in reality. Also, stereoscopic representation is sometimes omitted for floor projection. For stereoscopic representation, CAVEs mostly use active methods, i.e., shutter glasses (see Sect. 5.4.4). Case study 9.7 describes some of the challenges involved in the construction of a CAVE.

An advantage of a CAVE is that the user can move around in it as in reality (at least within the limits given by the surrounding projection surfaces). Another advantage of the CAVE compared to VR glasses is the self-perception of the user's own body. A fundamental problem of CAVEs is that the representation can generally only be calculated correctly only for the position of a single user based on their point of view. For all other users inside a CAVE, this results in a disturbing offset at the boundaries between the projection surfaces (i.e., the edges of the cube). In the best case, this will only lead to frame cancellation if parts of a virtual object extend over several projection surfaces. For many of these users, however, this increases the probability of developing symptoms of cybersickness (see also Sect. 2.4.7). One way to overcome this limitation is to use high-framerate projectors and custom stereo glasses (see also Sect. 5.4.4). In this configuration the projectors can display enough images for multiple (typically two or three) users to provide each user with their own pair of stereo images, in combination with tracking everybody's head



**Fig. 5.30** Layout of a six-sided CAVE

resulting in the correct depth perception for each user. As of today, no CAVEs using these projectors have been built, but several are under construction.

### 5.4.3 Tiled Displays

Stationary VR systems often use displays that are as large as possible. The reasons for this are on the one hand that a large display offers a larger field of view at the same distance from the user and thus results in higher immersion, while on the other hand a larger number of users are able to use such a system at the same time. Since the resolution and brightness of projectors as well as the resolution and size of monitors cannot be increased at will, the size that can be achieved with a single display of a certain quality are limited.

To increase the resolution or to realize large projection and monitor surfaces with high resolution and/or high luminous intensity, a division into several displays (i.e.,

several projection systems or monitors) appears reasonable. We refer to this as *tiled displays*.

The main idea here is to combine several display systems in such a way that the user perceives them as a single, larger system. The idea as such is not new and has been used for a long time in military flight simulators (here to completely cover a dome projection) or for so-called video walls. With this approach, the limitations of a single display can be bypassed to achieve larger sizes and/or higher resolutions.

As the number of individual tiles usually quickly exceeds the number of outputs of a graphics card, tiled display systems usually use a cluster of computers to calculate the output images. Generally, the fewer tiles a computer has to serve, the higher the performance can be. Conversely, the synchronization effort increases with the number of computers used.

Tiled displays can occur with both projection systems and monitors. Both approaches are presented in more detail in the following. Also, tiled displays always have to be calibrated to create the impression of a single display surface. Basic calibration methods for geometric calibration and for achieving brightness and color uniformity are therefore also briefly discussed in the following sections. Various approaches to specific setups exist, e.g., Bajestani (2019) and Okatani and Deguchi (2009).

### Tiled Projections

Figure 5.31 shows the C6, a six-sided CAVE built at Iowa State University in 2006 using *tiled projections*. It was built using 24 projectors with  $4096 \times 2160$  pixels each. A  $2 \times 2$  raster per side with two projectors per tile is used for the stereo display, which combined can display a stereo image with over 100 million pixels. Each



Fig. 5.31 CAVE C6 at the Iowa State University





**Fig. 5.32** Tiled wall using the example of the HEyeWall with 48 projectors. (© Fraunhofer IGD 2013, all rights reserved)

individual pixel is only 0.7 mm in size, a size that is close to the resolution of the human eye at typical viewing distances of 1–5 m.

Instead of fewer very high-resolution and light-intensive projectors, smaller tiles can be used with a correspondingly higher number of projectors, but with lower resolution and light intensity. An early example of such an approach was the HEyeWall shown in Fig. 5.32, a system with 48 standard projectors installed at Fraunhofer IGD in Darmstadt in 2003. Figure 5.33 shows the view behind the screen so that the arrangement of the projectors is visible as a  $6 \times 4$  grid with two projectors per tile.

### Tiled Monitors

Tiled displays consisting of monitors can also be used to realize large display areas with a high resolution. Compared to projectors, monitors have a significantly lower price per pixel and, due to their small installation depth, allow high-resolution systems, even if there is significantly less space available. Figure 5.34 shows the Reality Deck at Stony Brook University, which was built in 2012. The system used 416 standard monitors, each with  $2560 \times 1440$  pixels, so together the whole system can display 1.5 billion pixels simultaneously. As shown by the figure, it is important that the individual monitors have a seamless display. Otherwise, the impression is quickly created that the user is looking through a grid at the virtual world. With stereoscopic displays, a gap between the monitors, which is clearly perceived by the user, very quickly results in frame cancellation. Tiled monitors are suitable for single-sided display surfaces as well as for CAVEs, L-shapes and cylindrical spherical VR systems.

The tile approach is very well suited to overcoming the limitations of individual display systems in terms of resolution, brightness or price. But while the basic idea



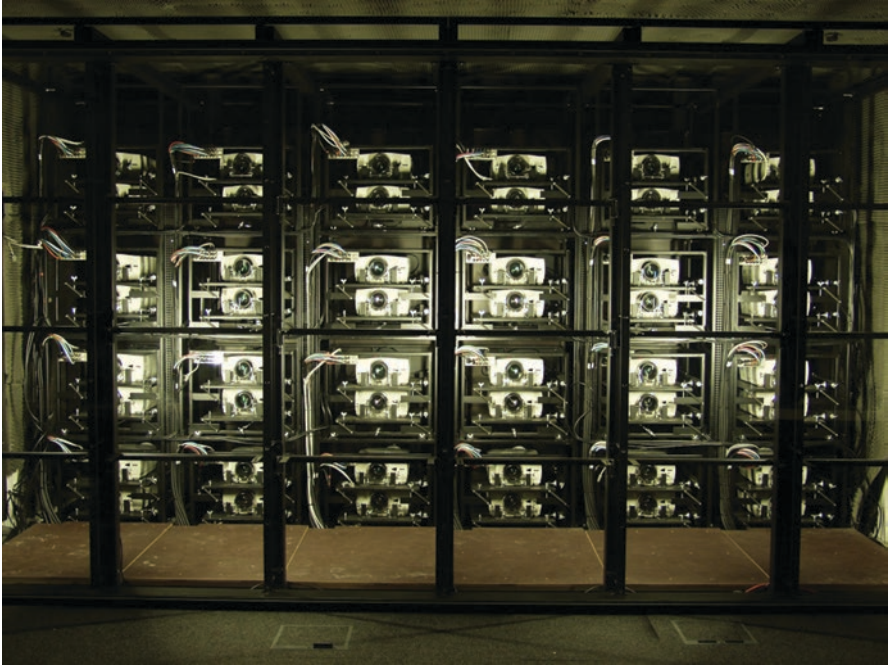


Fig. 5.33 HEyeWall setup. (© Fraunhofer IGD 2013, all rights reserved)



Fig. 5.34 Tiled wall of monitors illustrated by the Reality Deck at Stony Brook University

is very simple, the details require a lot of effort. As a result, the use of tile systems for high-quality applications is either limited or relatively costly. In particular, the calibration of the different display tiles in terms of geometric alignment as well as homogeneity and color representation can very quickly become a significant time and cost factor that is quickly overlooked, or at least underestimated. However, if the method is applied correctly and carefully, extremely impressive display systems can be developed, showing where the journey into virtual worlds may lead.

### **Geometric Calibration**

As soon as several individual displays are to be tiled, geometric consistency is no longer automatically given. A horizontal line that is one pixel wide and runs across all display tiles is not automatically at the same height on each tile. This continuity must be explicitly established.

Under favorable conditions the geometric calibration can be solved purely mechanically. For this purpose, a fixture is used that allows exact mechanical positioning and orientation of the individual display tiles. This requires accuracies in the sub-millimeter range, corresponding to the pixel sizes for high-resolution displays. Obtaining this mechanical accuracy over a large display such as a HEyeWall is a considerable amount of work, which can cancel out a significant part of the price advantage due to installation costs.

This task is further complicated by the inherent assumption that the display tile is geometrically correct in itself. In a conference room it is virtually impossible to see whether the center of the projection is a few pixels higher or lower than the edges, or whether the left edge is a few millimeters larger than the right edge. When several projections are put together, such inaccuracies quickly become obvious. A purely geometric-mechanical calibration cannot always correct such errors, since many variables, such as image border size, squareness and line straightness, depend on each other and cannot be changed independently.

This is especially important if the projection is to be made on an uneven surface (e.g., for spherical displays). A mechanical correction is no longer possible here. The alternative is a correction in the image creation software. There are different approaches possible. The most common is the texture distortion method, in which the image to be displayed is first rendered into a texture and this texture is then displayed on a grid that corrects the geometric inaccuracies of the display. This method is extremely flexible and can correct a wide range of geometric problems. However, it also has some disadvantages. First, the correction must be done within the image creation software, i.e., only software that has knowledge about the display can be used. On the other hand, it involves a (slightly) increased rendering effort, since the image must first be rendered into a texture and then displayed. In many modern systems, however, this is done anyway to produce high-quality images (e.g., in High Dynamic Range Rendering), which is even possible without reducing the refresh rate. Due to the fact that the image is displayed using a texture, however, texture filtering must also be performed, which may result in a certain degree of inaccuracy and image blur.

The biggest challenge, however, is to create an appropriate correction grid. For small systems this can be done manually (and especially in flight simulators this is not uncommon). For larger systems, however, the effort quickly becomes unreasonably high. In such cases, image processing methods that automatically generate corresponding correction grids from test images can help. Nevertheless, this is not a trivial problem and corresponding calibration systems are a price factor that (again) should not be underestimated.

After all these steps the system is now geometrically correct. Straight lines are straight, objects of the same size on all tiles are the same size, etc. Nevertheless, there remain other problems that have to be solved to get a uniform display.

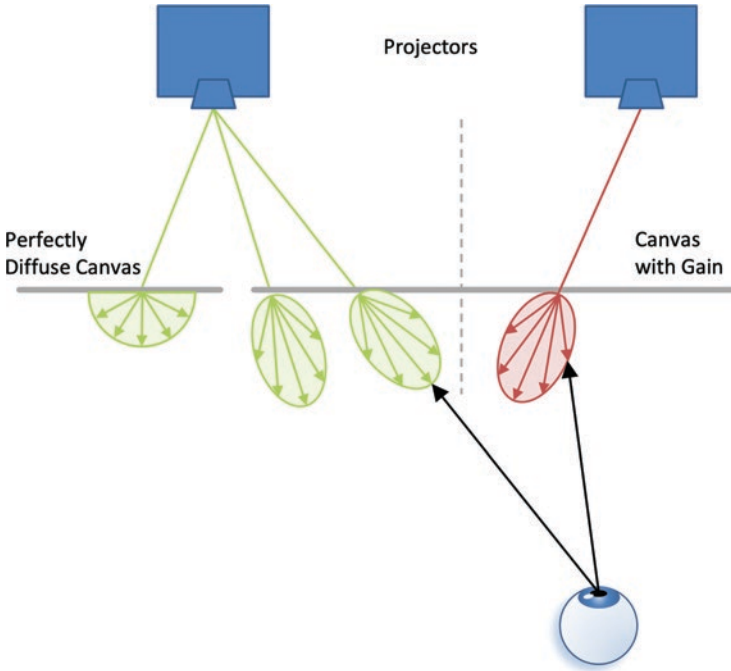
### **Brightness and Color Uniformity**

Besides geometrical problems, projectors also have problems with the *uniformity* of their brightness distribution. These stem from the geometric properties of the light source-lens-screen system, such as vignetting, where the image becomes darker towards the edges. With a single projector, this effect is much less noticeable, since there is no comparison image past the edge of the screen. However, if several tiles are arranged next to each other, the bright-dark-bright transition becomes much more visible. Vignetting is only caused by the projector and lens: it is independent of the viewer's point of view.

Vignetting already occurs with a single projector. When several projectors are used together, production variations in the projectors and especially in the lamps are added. Two identical projectors placed next to each other with the same settings do not necessarily have to be equally bright (and they usually are not). To achieve the impression of uniform brightness, each projector must therefore be individually adjusted. While this is possible using the naked eye, it will not give very accurate results, because the eye can adapt very quickly to different brightness levels. Good results can only be achieved with special light meters.

Another brightness effect comes from the properties of the canvas. Most screens for projections are not perfectly diffuse, i.e., light coming from behind is not emitted uniformly in all directions (see Fig. 5.35). Almost all commercially used screens have a *gain factor* that ensures that more light is emitted to the front than to the sides.

Since the viewer of a normal projection practically never looks very oblique from the side, this arrangement makes sense, because more light reaches the viewer. In the case of tiled projections, however, this ensures that in the transition area between two tiles there are very clear differences in brightness, even if both projectors emit exactly the same amount of light. In Fig. 5.35, the viewer looks directly into the left projector and therefore sees an area of the screen with high gain. The area of the right projector is seen at a much larger angle and therefore in an area of the canvas with low gain. Thus, at the point where the projection areas meet, a clear difference in brightness becomes visible. To make matters worse, this difference is dependent on the angle of viewing: when the viewer moves in front of the screen, one area becomes brighter while the other becomes darker. This makes a uniform image impression practically impossible, the only solution is to use extremely diffuse canvases, which then result in a rather dark projection.



**Fig. 5.35** Brightness discrepancy in the transition area between the images of two projectors due to non-diffuse projection canvas

The transition area between tiles is also critical for another aspect, that of overlapping. There are two alternatives for creating the transition between two tiles: either without overlap (*hard edge*) or with overlap (*soft edge* or *blending*). With hard edge, the projectors are arranged in such a way that the transition from one projector to the next is hard: the last pixel of one projector is immediately followed by the first pixel of the other projector. To make this possible, all components of the system (projectors, projector mounts, canvas, etc.) must be extremely stable. Even the slightest movement in the sub-millimeter range can cause a gap to appear between the two projections, which is clearly visible as a black line, or the projectors can overlap and the result can be seen as a bright line in the image. The HEyeWall (Figs. 5.32 and 5.33) was a hard edge system, so special attention had to be paid to the stability of the screen. For this purpose, precisely adjustable baffles were installed, which made it possible to avoid overlapping.

The alternative is to allow overlapping of the projection areas. This creates an area where both projectors beam onto the canvas. To prevent this area from appearing artificially brighter, the displayed image must be adjusted so that one projector is increasingly faded in and the other is faded out in the overlap area. This adjustment is usually achieved by a blend mask that is placed over the image after the rendering process. The C6 is a soft-edge system in which the two projectors per side overlap by approximately 220 pixels.

The overlap prevents the formation of gaps when the canvas is deformed or moved, and reduces the gain problem. In the transition area, the user no longer sees only the image of a single projector, as the projector images merge seamlessly. The main problem with overlap is when dark images or backgrounds are displayed. Modern LCD or DLP projectors cannot display true black because they rely on filters that attenuate the light from the lamp. These filters are never perfect, so a certain amount of residual light always penetrates. In the overlapping areas a double (at the inner corners a quadruple) residual light is therefore visible. As long as only bright images are displayed, this can be masked, but as soon as darker areas appear at the edges/corners, the overlapping and thus the tiling becomes clearly visible, which considerably disturbs a uniform image impression.

While brightness is only a one-dimensional problem, color uniformity requires three dimensions to be matched. This is already apparent within a single projector. LCD projectors in particular often show significant color differences between different areas of an image. If color differences already occur within an image, it is not surprising that massive color differences often occur between several projectors. To achieve a high quality result, these differences must be compensated. This is a much more complex process than brightness calibration and is practically impossible to do effectively manually.

#### ***5.4.4 Stereo Output Methods***

To support stereoscopic vision (see Sect. 2.2.1) with the goal of making a virtual world stereoscopically experienceable for the user, each eye of a user must be provided with an individual view. While in binocular HMDs this is done by separate optics for each eye, in monitors or projection systems both eyes basically see the same display. Therefore, additional methods for channel separation between the left and right eyes have to be applied. The individual methods used for this purpose are therefore briefly presented below, whereby individual advantages and disadvantages in each case will be highlighted.

##### **Anaglyphs**

Anaglyphs are an approach to stereo imaging in which the two partial images are colored differently for the left and right eye – in the original approach, one in red, the other in green. Both images are then combined into one image by superimposition. Red-green glasses are used for viewing. Here, a red filter is placed in front of one eye and a green filter in front of the other, so that each eye only sees its respective partial image. The two colored partial images complement each other to form a stereoscopic grayscale image. For the observation of a usually colored virtual world the approach is therefore not suitable in this original form.

However, the approach can be extended to color image pairs. In this case, red-cyan glasses are usually used instead of the red-green glasses mentioned above. In a display or projection, each individual pixel usually consists of one red, green and

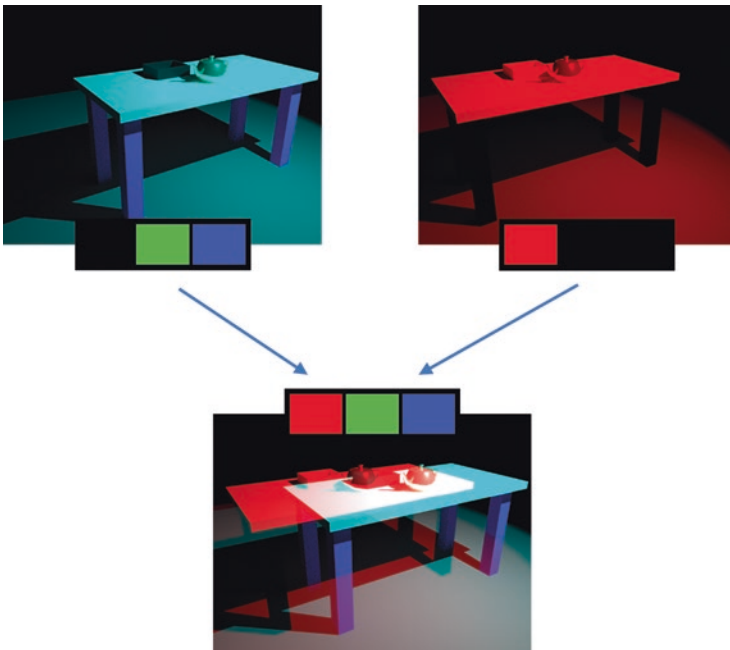


blue subpixel (RGB). For the color anaglyph process, the two images are now divided according to their subpixel assignment (see Fig. 5.36). For one image, only the red channel is used, while the green and blue channels (green + blue = cyan) are used for the other image. The problem is that objects whose color values are only displayed in one subframe cannot be perceived stereoscopically (see also Fig. 5.36). The problem of a unilateral representation can be reduced by a suitable color selection of the objects.

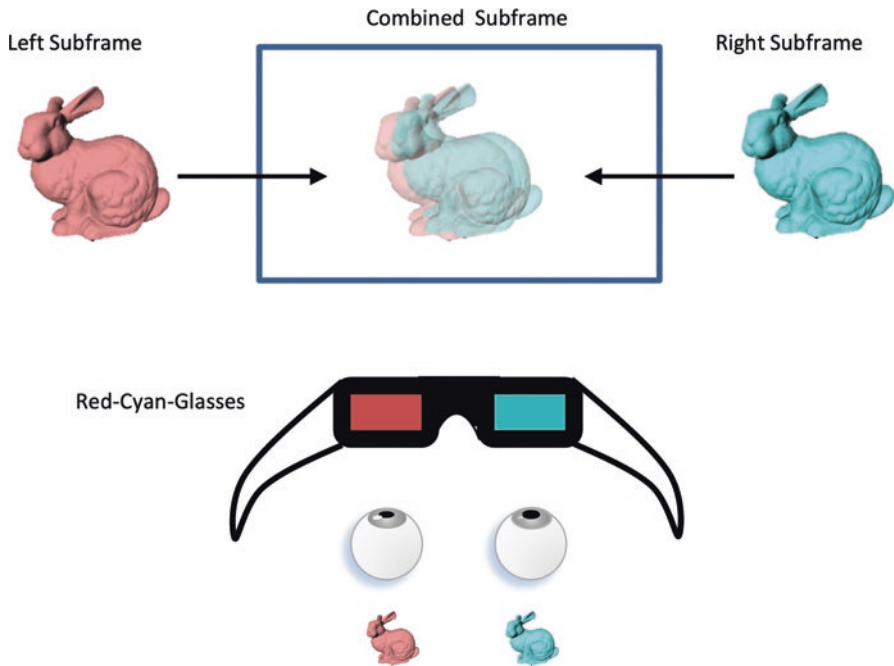
The division does not necessarily have to be along the subpixel boundaries, but a color image can be divided along any complementary colors (e.g., yellow/blue or green/magenta). Of course, the corresponding color filters must then be available for the glasses (see Fig. 5.37). To calculate the partial images, the RGB color value is linked to the respective filter color by a bitwise AND operation with each pixel of the corresponding image.

### Polarization

*Polarization* is a widely used approach to realize stereo vision using channel separation. The method is used in the majority of 3D cinemas. The approach uses the characteristic of light waves to oscillate in different directions. Polarization filters allow only light waves with a certain oscillation direction to pass through. In



**Fig. 5.36** Distribution of pixels along the RGB subpixels (left) and color anaglyph display of a 3D scene with red-cyan channel separation. The problem with this approach can be seen in the dark blue table legs, which are only present in the left subframe. (© Rolf Kruse, FH Erfurt 2019, all rights reserved)



**Fig. 5.37** Superimposition of partial images and channel separation by means of color filters for the anaglyph method

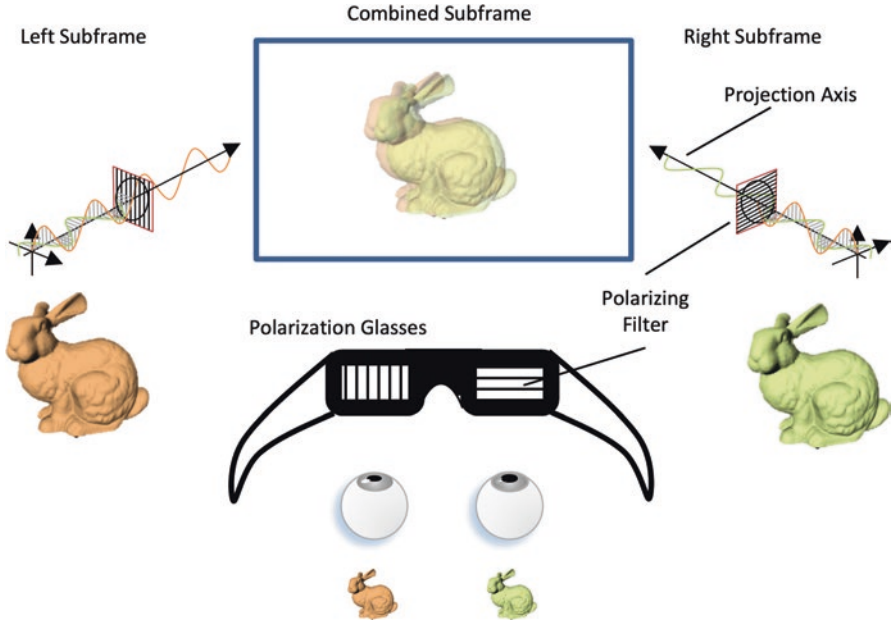
general, there is a distinction between systems that use linear polarization filters and those that use circular polarization filters.

For *channel separation*, typically two projectors (or possibly one projector with two lenses) are required per screen. A polarizing filter is mounted in front of each lens, whereby these are rotated  $90^\circ$  to each other. Thus, one polarizing filter, for example, only allows the horizontally oscillating part of the light to pass, while the second only allows the vertically oscillating part to pass. Since the two partial images overlap on the projection surface, they are perceived simultaneously by the viewer. Polarization glasses are now used to separate the channels of the images for the left and right eyes. Here, the two polarizing filters in front of the eyes are aligned in the same way as those on the lenses. If the polarization axes are exactly the same, each eye only sees the corresponding partial image, enabling stereo vision (see Fig. 5.38).

Due to this approach, however, the procedure is very susceptible to *crossstalk*. If the user tilts the head just a little to the side, the polarization axis changes and a *ghosting* of the other channel results. Instead of a horizontal and vertical alignment, combinations of  $45^\circ/135^\circ$  are often used, which, however, has no advantage with regard to the problems described above.

The use of *circular polarization filters* solves this problem. Here, a distinction is made between left- and right-turning light waves. This is not influenced by the head





**Fig. 5.38** Superimposition of the partial images and channel separation by means of linear polarization filters

tilt, so that crosstalk can be avoided as far as possible. However, circular polarization filters are much more expensive. Therefore, only linear polarization filters are used for “disposable” (cardboard) glasses. A further disadvantage of polarization-based approaches is that the projection surface must retain the polarization. This is only the case with high-quality metal-coated screens, which limits their use and, like the requirement for two projectors per projection surface, increases the costs even further. Since polarization filters generally filter out at least half of the light, only 50% of the light from a projector reaches each eye.

### Wavelength Multiplex

The *wavelength multiplex* method, also known as *interference filter* method, uses dielectric interference filters for channel separation. Each filter is based on several coupled resonators, which filter out three very narrow frequency ranges in the three primary colors red, green and blue (see Fig. 5.39).

By mixing the respective primary colors, full color images can be created. By using different frequencies for the three primary colors, the superimposed partial images can then be separated again into two channels, i.e., one for each eye. Similar to a polarization-based channel separation, a filter pair is used for each of the two projectors (or a projector with two lenses) and an identical filter pair for one pair of glasses. In contrast to polarization methods, the wavelength multiplexing method does not require any special characteristics of the projection surfaces. A further advantage is the low susceptibility to crosstalk. A disadvantage is the color shift

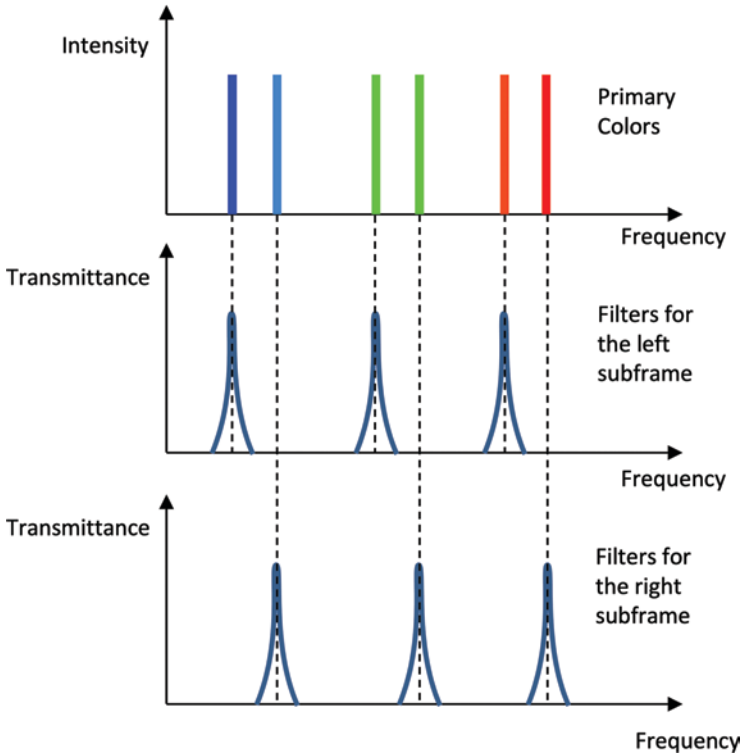


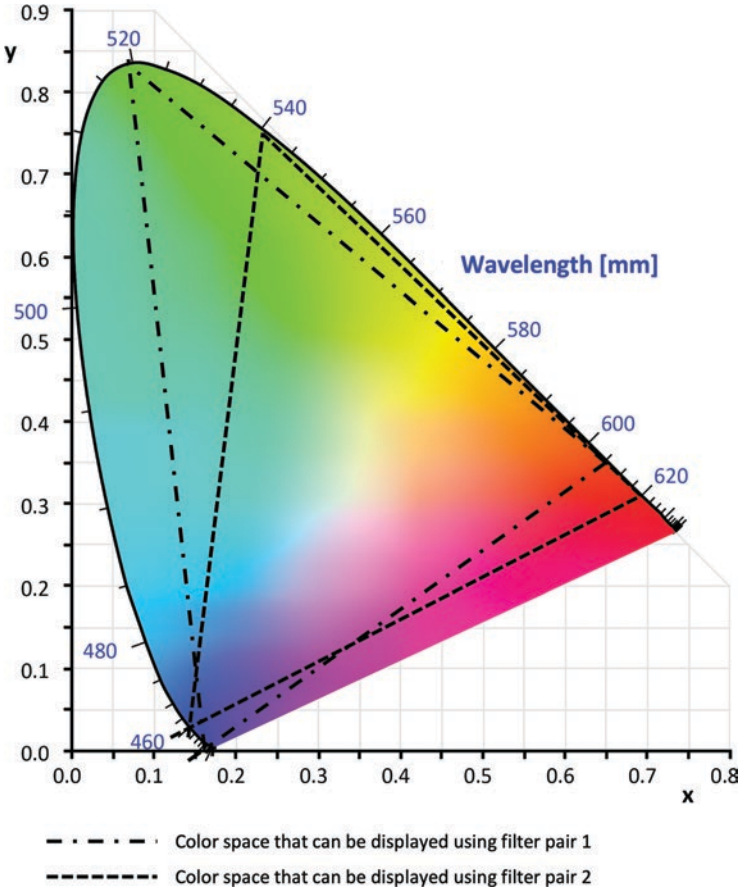
Fig. 5.39 Composition of the two partial images from three different base colors each

between the image for the left and right eyes due to the three different primary colors. To avoid this, the color of the images to be output can be adjusted so that they lie exclusively within the range that can be displayed with both primary color triples (see Fig. 5.40).

However, this further limits the total color space available. The wavelength multiplexing method also allows the independent display of more than two channels. In this case, there are always two channels used for each user, which allows  $n$  users to see a stereo image correctly calculated for their individual point of view. For these  $n$  users,  $2 \times n$  different basic color triples are required and thus a corresponding number of filter types as well as projectors and  $n$  different types of glasses with two different filters each.

### Shutter Glasses

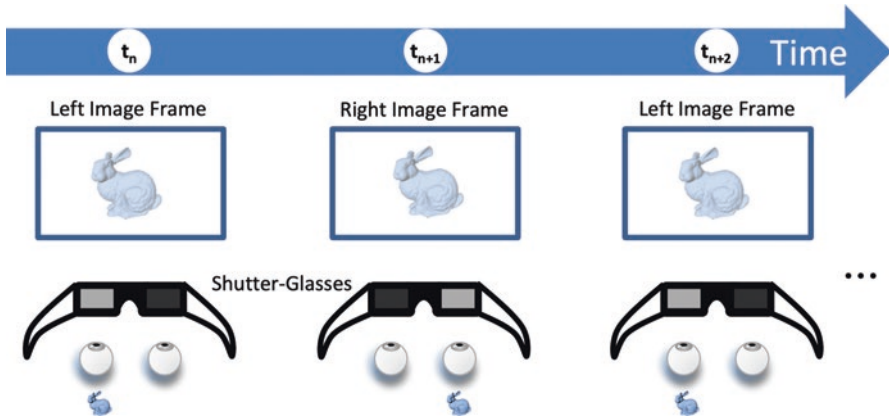
Besides polarization glasses, *shutter glasses* are another widely used method for stereo output. They are also partly used in 3D cinemas and for 3D TV sets. While the methods presented so far were based on superimposition and subsequent channel separation based on filters, shutter glasses display the partial images in time sequence. Here, the left eye sees its partial image for a short time and shortly afterwards the right eye sees its corresponding partial image. Because the change occurs



**Fig. 5.40** CIE standard valence system with the primary colors of two filters using the wavelength multiplexing method

at a high frequency, the brain is still able to fuse the two partial images into one stereoscopic image, even though they are not perceived simultaneously at any time. To ensure that each eye only sees the partial image intended for it, shutter glasses use two LCD shutters (hence the name). This always covers the eye whose partial image is currently not displayed, so that it cannot perceive any image (see Fig. 5.41). Synchronously to the change of the partial image, the corresponding LCD shutter is opened while the shutter of the other eye is closed. Due to the active switching, this is called an *active stereo method* (in contrast to the *passive stereo methods* using filters). Due to the time-sequential display of the two partial images, their frame rate has to be twice as high to achieve the same overall frame rate as with the passive methods.

Proper synchronization between the shutter glasses and the image output is crucial for channel separation. While earlier systems were primarily synchronized via



**Fig. 5.41** Time-sequential display of the left and right partial image synchronously to the alternating opening and closing of the shutter LCD in front of the left and right eye, so that each eye only sees the partial image intended for it

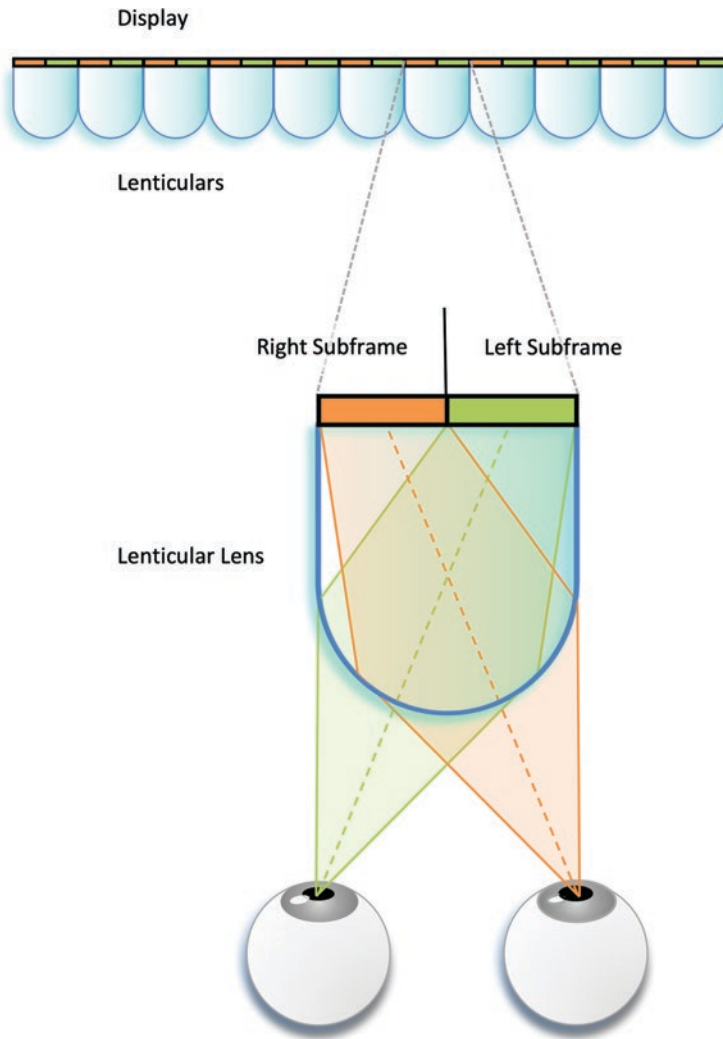
infrared, which is in principle susceptible to occlusion or further interference, nowadays synchronization via radio-based procedures, especially based on Bluetooth, has become generally accepted. An alternative is synchronization via a white flash (known as *DLP link*). Here a very short, completely white image is shown, which is detected by a photo diode attached to the shutter glasses and used for synchronization. The duration is so short that the user is not consciously aware of this white flash.

### Lenticular Lenses

Lenticular lenses are a method to make different (partial) images visible depending on their direction of view. The simplest variant of this are so called “wobble images”, which allow you to view simple animations consisting of very few frames. 3D postcards are based on the same principle. In both cases a prismatic grid consisting of lenticular lenses arranged in vertical rows is used. Each prism covers at least two pixels. Depending on the viewing angle, one or the other pixel becomes visible (see Fig. 5.42).

For stereoscopic output, a prism foil is glued to a screen with pixel accuracy. If the observer is vertically in front of the display at the correct distance, he sees one subframe with one eye and the other subframe with the other eye (see Fig. 5.42).

The advantage of this method is that it does not require any form of glasses, which is why it belongs to the so-called *autostereoscopic methods*. A disadvantage of this method, however, is that the resolution of the display is reduced by half horizontally. Another problem occurs when the user moves to the side or changes their distance to the display. This can lead to the channel separation not working or only working in a limited way (the already mentioned crosstalk). Tilting the head may also cause crosstalk. In principle, the procedure also works for several users at the same time. However, if several users are grouped around such a display, it must be ensured that an individual view is also possible with each eye from other viewing angles. This is achieved by using larger prismatic grids in which each lenticular lens

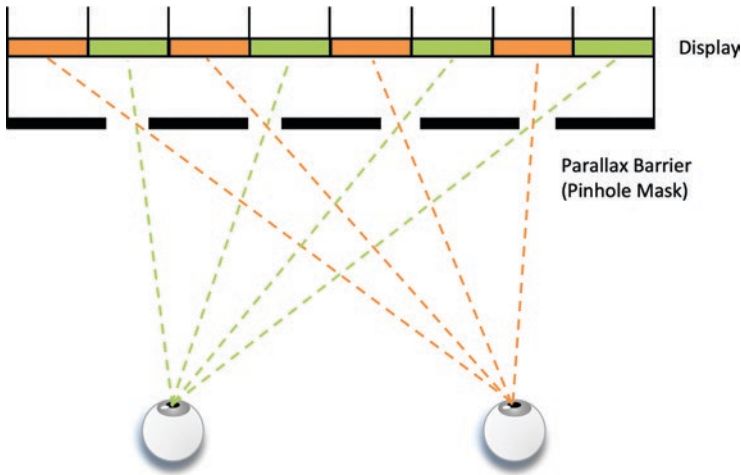


**Fig. 5.42** Prismatic grid of lenticular lenses with detail magnification of a lenticular lens for channel separation

covers more than two pixels. For example, seven different stereo views can be created by using eight pixels. The disadvantage here is that the horizontal resolution is reduced even more (in this case to one eighth!).

### **Parallax Barriers**

*Parallax barriers* represent another autostereoscopic method. Here, a shadow mask is placed in front of the display or the projection surface. Due to the arrangement of the holes in the shadow mask, each eye of the observer sees different pixels, which



**Fig. 5.43** Parallax barrier for channel separation

are then used to display the different subframe for the left and right eyes (see Fig. 5.43).

While rigid parallax barriers also only work within a certain distance from the display, movable parallax barriers allow for adjustment to the distance of the viewer. Either two slit masks are mechanically moved against each other so that the position and size of the holes change accordingly, or an additional LCD layer is used for this purpose. Parallax barriers also reduce the resolution by at least a factor of two and are susceptible to head tilting. In LCD-based systems, if detected, this can be solved by software. Software parallax barriers are used as masks to generate the two subframes, so that ultimately only the pixels visible to the respective eye from each perspective are included in the overall image to be displayed. In principle, the approach can also be extended to more than one user, as Ye et al. (2010) have shown with a display in which they used randomly distributed holes in well known locations.

## 5.5 Audio Output Devices

The goal of acoustic or audible output is to reproduce the sounds and tones of the virtual world in such a way that the user can perceive them in the same way as in the real world. Even though the spatial resolution of human audio perception is lower compared to the visual sense (see Sect. 2.3.1), it clearly supports spatial orientation. A simple audio system at least is also important for the temporal assignment of events that happen in the virtual world: for example, the user can be given audible feedback when selecting objects or controlling a menu.

### **Stereo Speakers**

A simple spatial audio model, such as the one used in X3D, only influences the volume of the output for the left and right ear depending on the distance of the sound source to the virtual position of the ears (whereby this is only taken into account due to the position and orientation of the virtual camera, i.e., the position of the loudspeakers to the real ear has no influence here). For more realistic effects, the individual signal delay due to the distance to the sound source can be included. With this form it is not possible to distinguish between audio sources in front of or behind the user, which is usually not sufficient for VR or AR. Also sound sources above or below the user cannot be determined from their direction.

### **Multi-Channel Systems**

For better spatial orientation the use of more complex audio installations is necessary. Often, *multi-channel audio systems* are sufficient to provide orientation for the user of a virtual world. In multi-channel systems one or more main loudspeakers are usually available as the actual sound source, while several additional loudspeakers are used to support the spatial effects. When installing multi-channel audio systems, it is important to ensure that appropriate loudspeakers are also installed behind the user. The disadvantage of multi-channel audio systems is that spatial perception is really good only in a small area (the so-called sweet spot). If the user is able to move, the limits of such a system are quickly reached. Furthermore, although the horizontal direction of a sound source can be well simulated, the height of the sound source usually cannot be reproduced.

### **Binaural Sound**

One way to achieve a more realistic audio impression is *binaural sound*. This is an attempt to imitate natural, spatial hearing. The output is only possible via headphones. For an optimal hearing impression, the *Head Related Transfer Function (HRTF)* of the user must be known. If this is not known, the HRTF of a standard head is usually used, which can provide very good or even bad results depending on the individual user. The advantage of binaural sound is that with correct HRTF not only audio sources in front of and behind the user, but also below and above the user, can be clearly identified with regard to their direction.

### **Ambisonics**

*Ambisonics* usually uses four channels to record and play back three-dimensional sound sources in the form of a sound field. Although the technology is over 50 years old, it has only recently gained some popularity through its use in conjunction with 360° video and VR. In the meantime, relatively inexpensive commercial microphones and software for mixing ambisonic recordings are available. The four channels represent the sound pressure gradients in the X, Y and Z directions and the sound pressure. For these second-order ambisonics, eight capsule microphones are used. Higher-order ambisonics are rarely used.

### **Wave Field Synthesis**

Another way to create more realistic spatial sound is *wave field synthesis* (Bertino and Ferrari 1998; Brandenburg 2006). The goal of wave field synthesis is to record



the wave field of a real event (e.g., the playing of an orchestra) and to be able to reproduce it at any time as a synthetic wave field. Thus it is possible to position sound sources freely, within certain physical limits. For this purpose, the wave field is generated by a large number of loudspeakers that have to be arranged around the playback area. These loudspeakers are operated by a central computer, controlling the reproduction of the sounds together with their positioning.

### **Application to Stationary VR Systems**

Often it seems to make sense to position the speakers of an audio system behind the display. In principle, this is possible for projection systems with permeable screens. However, with multi-sided projection systems such as CAVES, the sound is partially reflected by the projection surfaces, reducing the quality of the audio simulation. A further problem with loudspeakers located behind the projection surface is that the sound causes the canvas to vibrate. This can sometimes have a negative effect on the quality of the visual impression. When using glass panels as projection surfaces and in the case of monitor-based solutions (see Fig. 5.32), it may be necessary to place the loudspeakers under, above or next to the displays.

A comprehensive overview of sound, especially in the context of VR, is given by Vorländer (2008).

## **5.6 Haptic Output Devices**

Haptic output devices make virtual objects tangible for the user by means of mechanical, pneumatic or electrical stimuli, vibration or the application of force. Haptic output is mostly integrated into input devices like gloves (see also Sect. 4.4) and mechanical input devices (see Sect. 4.6.2).

Generally, haptic output devices are divided into those that generate *tactile feedback* and those with *force feedback*. Tactile feedback generates a haptic sensation for the user when touching a virtual object, without necessarily corresponding to the sensation when touching a similar real object. Force feedback usually requires an external skeleton structure (*exoskeleton*), which restricts the freedom of movement of fingers or other limbs.

### **Haptic Loop**

While haptic output basically uses the same information about the virtual world as graphical output, haptic rendering differs significantly from normal (graphical) rendering. A haptic output device is usually combined with an input device, since the haptic output is typically the result of a movement of the user. Of course, a virtual object can also move towards the user and thus initiate a haptic output. An example would be a virtual bullet in a VR game. If the initiation is done by the user, this results in a so-called *haptic loop* (or *haptic rendering loop*) (see Fig. 5.44). For the user's movement to result in a haptic output, it must first trigger a collision between the user's representation (i.e., their avatar) and an object of the virtual world. This collision is detected by collision detection (see Chap. 7). The resulting collision

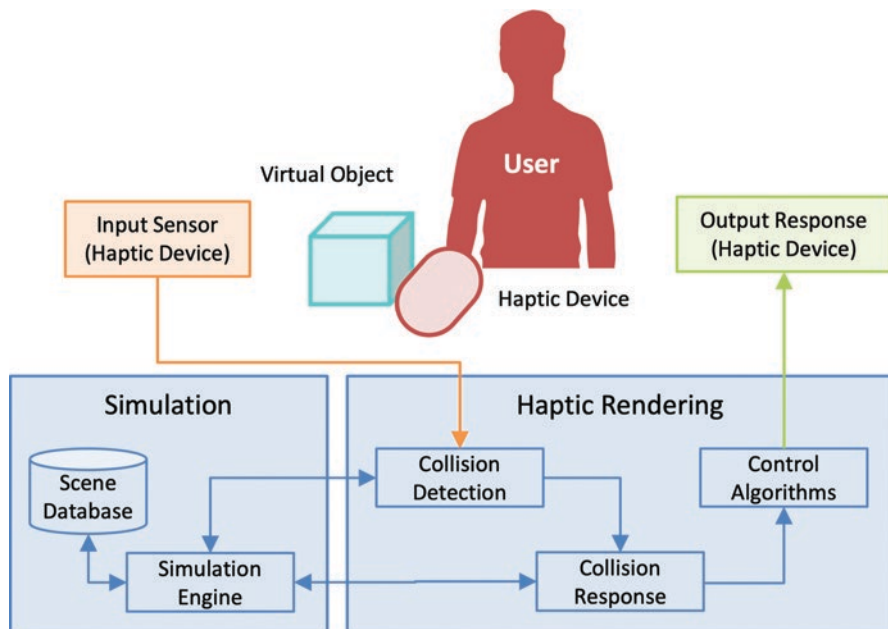


Fig. 5.44 The haptic rendering loop

response induces on the one hand a modification of the VR simulation, while on the other hand the ideal force feedback is calculated as response. The extent to which this can now be (re)transmitted to the user depends in particular on the specific haptic output device used. Appropriate control algorithms are used to convert the ideal force into an actual output reaction. Correspondingly, a massive resistance may end up in a rather soft tactile perception. In contrast to the graphics render loop, which should reach at least 60 frames per second, the haptic render loop typically operates at a much higher frequency. 1000 Hz (or 1000 fps) is not uncommon at this point. An overview of haptic rendering can be found in Salisbury et al. (2004).

### Data Gloves with Tactile Feedback or Force Feedback

While *data gloves* are primarily input devices, a number of models have been developed over time that involve a haptic output component. However, most of them are limited to simple tactile feedback. For example, small vibration motors are placed on the fingertips, the fingertips are mechanically contracted by small bands or electrical impulses are generated. However, some also use external skeletal structures to actually restrict the freedom of movement of the fingers. The *HGlove* uses an exoskeleton for the thumb, index finger and middle finger. The *HaptX* glove (see Fig. 5.45), for example, uses a combination of a miniaturized pneumatic actuator consisting of 12 elements at each fingertip and an exoskeleton.



**Fig. 5.45** Glove with pneumatic finger actuators and external skeletal structure. (© HaptX Inc. 2021. All rights reserved)

### Air and Ultrasound-Based Systems

*Vortex rings* use air over a certain distance to create a short haptic stimulus. If a grid of actuators is used to generate the stimuli, this allows the simulation of even more complex virtual objects.

Similar in effect are haptic output devices based on ultrasound. Those devices use up to several hundred sound generators arranged in arrays. At update rates of up to 40 kHz, a haptic output can be generated at a distance of up to 70 cm.

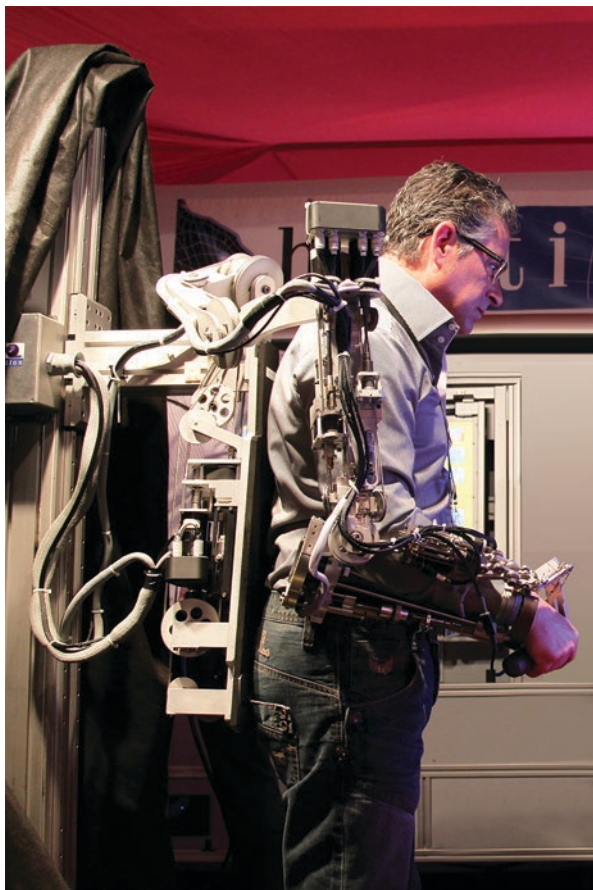
### Vests and Suits

In games and military simulations especially, vests and whole suits for haptic output have been developed. The *Teslasuit* full body suit is an example of an individual, haptic output device that uses electrical impulses to perform *transcutaneous electrical nerve stimulation* and *electrical muscle stimulation*. Furthermore, the suit can influence the temperature sensation of the user. A motion capture system is also integrated for input. Haption's *Able* is another example, providing an exoskeleton for shoulders, arms and hands (see Fig. 5.46).

### End Effector Displays

An *end effector* is a device that is typically mounted on a robot arm (see Fig. 5.47). *End effector displays* are haptic output devices for tactile stimulation, which a user can grasp or otherwise manipulate with hands or feet. In contrast to an input device, an end effector display is not passive, but reacts actively through resistance or force feedback. The best-known representative of this device class is the *Phantom Omni* as a desktop device. Larger versions can easily cover volumes of several cubic meters.

**Fig. 5.46** Example of a vest-based system. (© Haption SA/Laval Virtual, 2020. All rights reserved)



**Fig. 5.47** Example of an end effector display. (© Haption SA, 2020. All rights reserved)

## 5.7 Summary and Questions

VR/AR output devices are used to present the virtual world or the augmented world to the user via appropriate stimulus generation, i.e., to convert the virtual content into something that can be experienced by the user's senses. The visual output can be done with HMDs, monitors or projection systems. To cover the field of vision of stationary VR systems as much as possible, display systems are often composed of several individual displays, which are arranged in different forms: examples are walls, L-shapes, curved screens or CAVEs. The tiling of displays (in the form of tiled displays) is especially used to improve resolution and luminous intensity, but leads to a considerably higher calibration effort. Using active or passive stereoscopy methods, monitors and projection systems can also spatially represent virtual worlds. Although the optical sense is the most important one, a high degree of immersion can only be achieved by addressing additional senses. While there are almost no VR or AR systems without acoustic output and therefore the question of the quality of the effects to be produced is primarily concerned here, haptic output devices are much less common.

Check your understanding of the chapter by answering the following questions:

- What difficulties can occur when using tiled displays and what are the solutions?
- What advantages do multi-sided displays offer compared to single-sided displays?
- When should optical AR glasses (OST) be used for an AR application and when should video AR glasses (VST) be used?
- Explain the basic differences between active, passive and autostereoscopic stereo methods.
- Which audio technologies are suitable for realistic surround sound?
- Explain the difference between tactile and force feedback using a simple example

## Recommended Reading

- Bajestani SA, Pourreza H, Nalbandian S (2019) Scalable and view-independent calibration of multi-projector display for arbitrary uneven surfaces. *Machine Vision and Applications*, 7–8, Springer.
- Burdea GC, Coiffet P (2003) *Virtual reality technology*, John Wiley & Sons, Hoboken, New Jersey.
- Okatani, T., Deguchi, K. (2009) Easy calibration of a multi-projector display system. *International Journal of Computer Vision*, 85, 1–18. <https://doi.org/10.1007/s11263-009-0242-0>
- Sherman W, Craig A (2019) *Understanding virtual reality*, second edn. Morgan Kaufmann.
- Vorländer M (2008) *Auralization – fundamentals of acoustics, simulation, algorithms and acoustic virtual reality*. Springer.

## References

- Bajestani SA, Pourreza H, Nalbandian S (2019) Scalable and view-independent calibration of multi-projector display for arbitrary uneven surfaces. *Machine Vision and Applications*, 7–8, Springer
- Bertino E, Ferrari E (1998) Temporal synchronization models for multimedia data. *TKDE* 10(4):612–631
- Brandenburg K (2006) Digital entertainment: Media technologies for the future. In *Second International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution (AXMEDIS '06)*, 4–5
- Cakmakci O, Rolland J (2006) Head-worn displays: a review. *J Disp Technol* 2(3):199–216. <https://doi.org/10.1109/jdt.2006.879846>
- Cruz-Neira C, Sandin DJ, DeFanti TA, Kenyon RV, Hart JC (1992) The cave: audio visual experience automatic virtual environment. *Commun ACM* 35(6):64–72
- Herold R, Weidenmüller F, Penzel M, Ebert M (2015) Data-glasses for improved user interaction in 3D. In: *Proceedings of SID 2015 International Symposium*, pp. 189–191
- Krüger W, Fröhlich B (1994) The responsive workbench. *IEEE Comput Graph Appl* 14(3):12–15
- Lang B (2018) The key technology behind Varjo's high-res 'Bionic Display' headset. <https://www.roadtovr.com/graphic-illustrates-key-technology-behind-varjos-high-res-bionic-display/>. Accessed 28 Mar 2021
- Melzer J, Moffitt K (1997) *Head-mounted displays: designing for the users*. McGraw Hill, New York
- Okatani T, Deguchi K (2009) Easy calibration of a multi-projector display system. *Int J Comput Vis* 85:1–18. <https://doi.org/10.1007/s11263-009-0242-0>
- Salisbury K, Conti F, Barbagli F (2004) Haptic rendering: introductory concepts. *IEEE Computer Graphics & Applications*, Jan/Feb, pp 24–32
- Vorländer M (2008) *Auralization – fundamentals of acoustics, simulation, algorithms and acoustic virtual reality*. Springer
- Ye G, State A, Fuchs H (2010) A practical multi-viewer tabletop autostereoscopic display. *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR) 2010*. Seoul, South Korea, October 13–16