

————— Technical Report No. 154 —————

Selected Technical and Perceptual Aspects of Virtual Reality Displays

Bernhard E. Riecke¹, Hans-Günther Nusseck², & Jörg Schulte-Pelkum³

————— October 2006 —————

¹ Department Bühlhoff, E–mail: bernhard.riecke@tuebingen.mpg.de

² Department Bühlhoff, E–mail: hg.nusseck@tuebingen.mpg.de

³ Department Bühlhoff, E–mail: joerg.s-p@tuebingen.mpg.de

Selected Technical and Perceptual Aspects of Virtual Reality Displays

Bernhard E. Riecke, Hans-Günther Nusseck, Jörg Schulte-Pelkum

There is an increasing amount of different presentation techniques available for producing visual Virtual Reality (VR) scenes. The purpose of this chapter is to give a brief and introductory overview of existing VR presentation techniques and to highlight advantages and disadvantages of each technique, depending on the specific applications. This should enable the reader to design and/or improve their VR visualization setup in terms of both the perceptual aspects and the effectiveness for a given task .

In this overview, we relate the different types of presentation techniques to aspects of human physiology of visual perception which have important implications for VR setups. This will, by no means, be a complete overview of all physiological aspects. For a detailed overview and introduction, see, e.g., Goldstein (2002).

The aim of a visual simulation is to achieve a convincing and perceptually realistic presentation of the simulated environment. Ideally, the user should feel present in the virtual environment and not be able to tell whether it is real or simulated. The human visual system uses several cues to form a percept of the surrounding environment. We will have a closer look at some of these cues in the first section, as they are of crucial importance when looking at simulated scenes. The remaining sections are concerned with possible technical implementations and how these relate to the perceptual aspects and effectiveness for a given task.

1 Basic principles of visual perception

Visual perception is only one part of our sensory experience. There is also auditory, vestibular, kinesthetic, proprioceptive, haptic, gustatory, and olfactory information. Only the combination of all these senses forms our overall natural perception of the environment, and of our own position in it. While truly immersive VR setups should include all of these sensory modalities, this will not be accomplished in the near future, since multi-modal research on Presence is only in its infancy. The POEMS project is one of the contributors in this new field: It combines visual, auditory, and vibratory cues in an ego-motion simulator. In this chapter, we will focus only on the visual modality, since this topic is already quite complex. Vision is often considered the most important or dominant among the different modalities, since most of our everyday actions like walking, reaching, etc. are visually guided. Furthermore, vision is able to pick up vast amounts of information very accurately and quickly. The more convincingly this visual information is presented in a simulation, the more perceptual realism and possibility for spatial presence are created. This chapter provides an overview of important perceptual as well as technical aspects of a visual VR simulation that can help to build a powerful visual simulation setup and improve its effectiveness from both a technical and a perceptual side.

1.1 Field of view (FOV)

The visual field of view is the part of our outside world that can be viewed by both eyes. Our FOV is limited by the facial boundary around the eyes.

1.1.1 Size and shape of the FOV

The normal FOV of one eye has an approximately oval shape. It extends about 90° horizontally to both sides along the line of sight and about 120° vertically, specifically, 50° upwards and 70° downwards. With both eyes together, the horizontal FOV is a bit larger than a semi-circle. The central visual field is viewed by both eyes. Only within this binocular overlapping zone of about 100° to 120° can we perceive depth through stereopsis. Our visual field is further extended by the fact that we can rotate our eyes in their orbits. With the head fixed, the eyes can maximally cover an area of approximately 290° horizontally and 190° vertically.

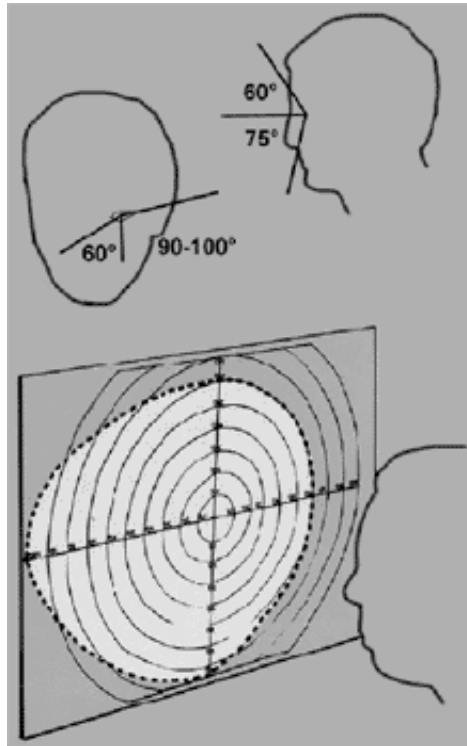


Figure 1: Size and shape of the human field-of view (FOV), shown for the left eye.

1.2 Perception of brightness and color

The human eye contains over 100 million light-sensitive cells, called photo-receptors. There are two types of photo-receptors: rods and cones. Rods are sensitive to brightness while cones are sensitive to colors. Rods are more sensitive to light than cones by a factor of 1000. Since they only register brightness, they produce only black and white images. Their high sensitivity enables us to see at night or under conditions of very low light intensity, but also causes them to quickly saturate (e.g. in twilight). The cones then take over and allow us to see in bright environments.

We distinguish between three situations:

- **Photopic Vision** (normal daylight vision – cones active [$> 3,4\text{cd/m}^2$])
- **Scotopic Vision** (night vision – only rods active – no color [0 to $0,034\text{cd/m}^2$])
- **Mesopic Vision** (intermediate range – cones and rods active)

Images that are presented in a visual simulation are viewed with mesopic vision because of limitations in current display technology. This fact affects the perception and interpretation of simulated environments. Standard monitors have a dynamic range of about 500:1, meaning that the ratio between the brightest and darkest portion of an image they can display is 500. New technologies, notably High Dynamic Range displays (HDR) are able to produce much higher ratios (60.000:1 for current HDR monitors). They are, however, still far from the real world dynamic range, and are so costly that only a handful of research labs currently possess them.

1.2.1 Color vision

We can see colors because the human eye is able to perceive a whole range of photonic wavelengths (400-750nm). This comprises all colors between, but not including, infrared and ultraviolet.

To detect different colors, roughly 6 million cones are present in each eye. They can be divided into three categories: cones sensitive to red (64%), green (32%) and blue (2%) (see Figure 2).

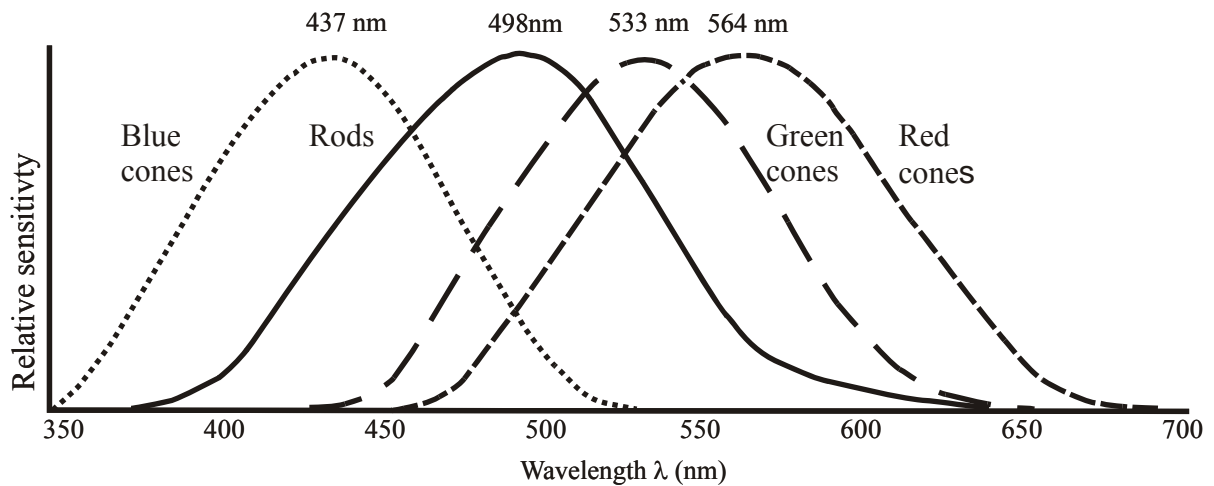


Figure 2: Normalized spectral sensitivity of retinal rod and cone cells of the human eye (adopted from Dowling, 1987)

The distribution of the rods and cones on the retina is not homogenous (see Figure 3). In the far periphery of the retina, the rod-type receptors are the majority and no colors are perceived in the far peripheral field. We do not notice this because our eyes are constantly in motion. Only in the central visual field are colors seen in high resolution. This is due to the fact that cones are only present at high density in the central part of the retina, called the fovea.

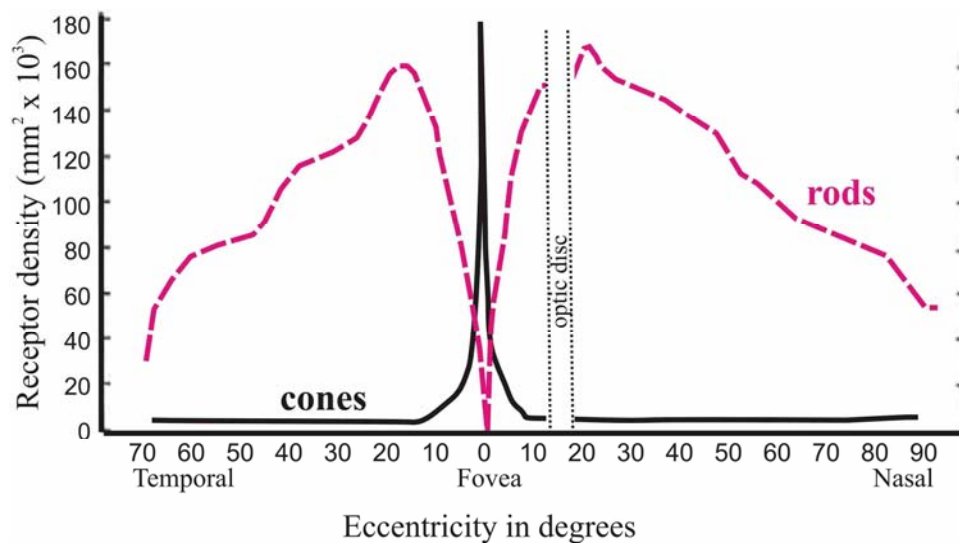


Figure 3: Receptor density of rod and cone cells on the retina.

1.2.2 Image resolution and motion detection sensitivity

Image resolution is reduced towards the visual periphery. Only in foveal vision, which covers only 2° around the line of sight, do we see sharp images. Up to 18° around the line of sight, objects can be seen at sufficient resolution. At 27° around the line of sight, objects appear with a clearly noticeable blurriness.

In contrast, motion detection sensitivity increases towards the visual periphery. Also, flicker is perceived more clearly in the periphery than in the fovea.

Subjectively, we do not notice at all that we can only see very blurry contours in the periphery, since we constantly move our eyes very quickly towards points of interest, making them fall on the fovea so that they appear to us in high resolution.

1.3 Space perception

When perceiving our environment, it is important to be able to estimate the size of objects, their distance to one another, as well as their distance to the observer. Here, we discuss only a few of the essential related physiological aspects that are important for VR simulations.

1.3.1 Binocular disparity

Stereoscopic vision is made possible by the fact that we see objects with two eyes, from slightly different angles. The eyes are positioned 60 to 70mm apart, and the images collected from each eye are combined into a single three-dimensional image. The human brain uses the horizontal disparity to compute the third dimension from two two-dimensional retinal images. In section 4, technical solutions for 3d-visualizations in VR will be addressed.

1.3.2 Accommodation and vergence

When we fixate an object, the eyes converge to it, and the lenses automatically focus on it (accommodation). In natural viewing conditions, this mechanism always works hand in hand with stereopsis, i.e., binocular disparity, vergence, and accommodation all allow us to perceive the distance from the viewed object to the observer. In most VR applications, however, this is not the case. For example, in head-mounted displays (HMDs), optical lenses are used to make the eyes accommodate at a comfortable distance. This, combined with vergence and disparity, produces a conflict. It has been suggested that this conflict may be the cause for eye-strain that users of VR visualization setups frequently experience.

Further information and overviews on 3D-vision can be found for example at www.webvision.med.utah.edu, webexhibits.org/colorart/cones.html, and webexhibits.org/colorart/ag.html.

2 From the virtual model to the displayed scene

In this section, some technical details on the implementation of virtual environments are discussed and related to the physiological aspects described above.

The first step to generate a simulation of a virtual environment (e.g., an outdoor scene like an open place or an indoor scene like a room) is to build a 3D model of the environment on the computer using specialized software. This computer model consists of a description of the geometry (e.g., the shape of the room and the objects contained therein) and the surface properties (typically, color, reflection properties, and texture) of the environments to be simulated. To be able to move through such a virtual environment (VE), a virtual camera needs to be positioned in the model to specify the current point of observation. Furthermore, light sources can be positioned in the model to enable a convincing natural illumination. The process by which the position of the camera and light sources is used to generate an image of the simulated scene is called rendering. That is, rendering is the process that converts the high-level representation of the scene, camera, and light sources into an array of pixels. Modern graphics cards have highly specialized chips (so-called graphical processing units or GPUs) that are optimized for rendering and are, for this purpose, much more powerful than the main processor of the computer.

2.1 Rendering

Graphic cards produce individual rendered images of the simulated scene at fixed time intervals (frame rate). This frame rate is typically 60Hz for most digital projectors and TFT/LCD displays. For cathode ray tube (CRT) monitors and projectors, where each pixel is illuminated only for a few milliseconds, a higher frequency is chosen whenever possible to be above the perceivable visual flicker threshold. This is not required for liquid crystal displays (LCD), thin film transistors (TFT), digital light processing projector (DLP), or liquid crystal on silicon projector (LCoS) systems because they display the image at almost constant brightness during the whole duration of each frame, thus effectively eliminating perceived flicker (see section 3.1 for more details). One of the most serious constraints on VR rendering systems is its real-time capability, meaning that each new image has to be rendered within one frame (e.g., 16ms = 1/60Hz). For very large and complex models, and when a high rendering quality is required, even the rendering power of modern graphics cards might not be sufficient. This rendering of a new image then requires more than one frame, which results in jerky image motion that disturbs the smooth flow of images and can be particularly annoying when fast motions are displayed. For example, when simulating a car ride at high speed, the frame rate can drop below the desired value (the refresh rate of the display device, e.g., 60Hz) when objects of high geometrical complexity (complex buildings, large number of trees etc.) enter the image. Such jerky motion is most disturbing when self-motion is being simulated, as not only individual objects but the whole scene appears jerky.

2.2 Aliasing

A computer-generated image consists of an array of individual square pixels (picture elements) for the commonly used raster graphics. A subsequent drawback is that lines that are neither exactly horizontal nor vertical cannot be displayed perfectly. They have to be approximated by a succession of horizontal and vertical line elements. This results in jagged lines and stair-step-like artifacts (“**aliasing**”), as illustrated in Figure 4 (top). To avoid these stair-stepping artifacts induced by the raster graphics, graphics cards include a number of so-called **anti-aliasing algorithms** that essentially blend the border of the lines into the background color. This is illustrated in Figure 4 (bottom).

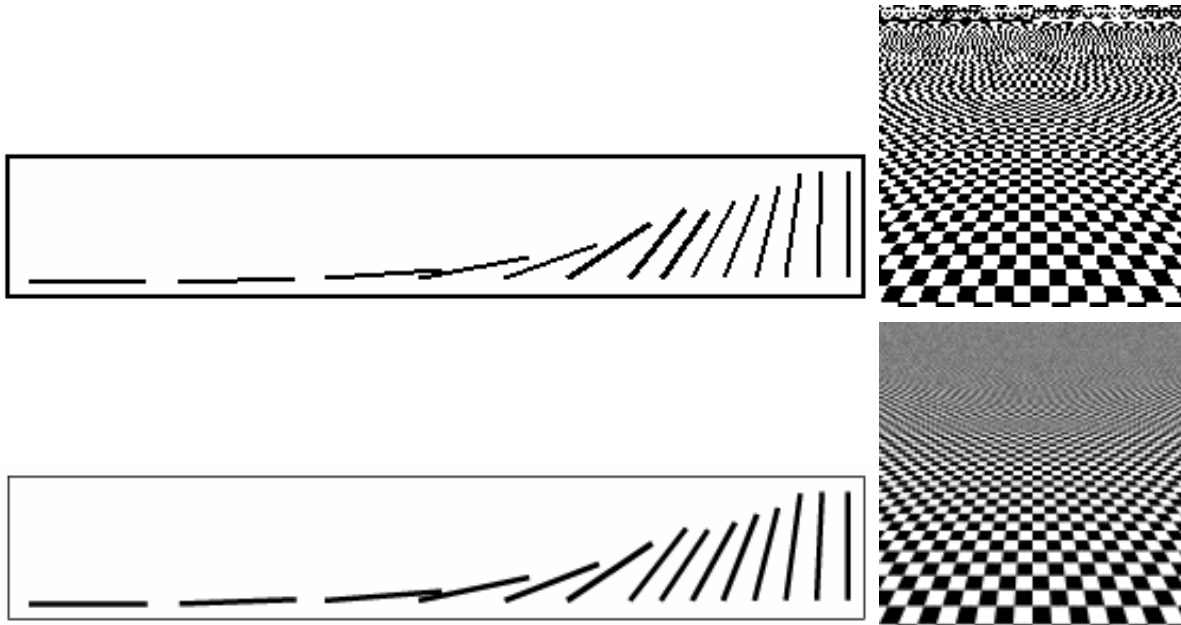


Figure 4: Illustration of aliasing artifacts for oblique lines due to the pixel rastering (top figure). The bottom figure shows an antialiased version of the same images.

One often used antialiasing approach is to simply render the image several times with slightly offset camera position (below one pixel offset), and then to “mix” together these individual images (“**supersampling**”). As each of the images has to be rendered separately, however, the supersampling method results in a considerable reduction of the overall rendering speed, which can in turn lead to jerky motions as discussed above. For moving stimuli, aliasing can lead to moving stair-stepping artifacts. These moving aliasing artifacts result from the stair-steps occurring at a slightly offset position when an object or the camera is moved. These moving aliasing artifacts can be rather disturbing especially for slow motions.

2.3 Texture filtering

To add colors to objects made of geometrical surfaces, the most commonly used method is called texture mapping. It consists in wrapping the objects with one or several images, very much like wallpaper is laid on walls, or tin cans are wrapped with a picture of their content. This technique often allows for the use of simpler geometry. For instance, from a distance, a cube wrapped with the pictures of the different faces of a house can be used as a fairly convincing model of a complex looking house but uses very little geometry.

Due to the raster graphics paradigm, details of the texture become more and more blurred for increasing distances from the camera. For close objects, highly detailed textures are needed. However, using these highly detailed textures for far objects makes little sense as most details will not be visible. Hence, to decrease computational load, the same texture is made available in different resolutions in the graphics card memory.

This technique known as MIP-mapping is based on creating a set of lower resolution, pre-filtered versions of the original texture. (The acronym “MIP” originates from the Latin phrase *multum in parvo*, meaning “much in a small space”). During the rendering process, the appropriate resolution is chosen according to the size of the object in the rendered image. This has the advantage that the texture pixels (texels) have previously been filtered, which can considerably improve the rendering speed compared to supersampling approaches to antialiasing. Additional filtering is needed to allow for smooth transitions between the different texture resolutions. Several filtering methods are built into modern graphics cards, and allow excellent results in terms of level of visible detail and sharpness of distant textures. Some of these filtering algorithms are computationally

costly and can decrease the rendering power of graphics cards. Without any filtering or with insufficient filtering, however, simulated scenes can appear “washed-out” and hence unnatural.

2.4 Realtime rendering issues (framerate, latency, double-buffering)

In most VR simulations, it is essential to have only a minimal delay between the motion of the simulated camera or objects in the scene and the visualization of the corresponding image on the display device. This is of particular importance for interactive applications like driving or flight simulations, where even small delays can lead to noticeable performance decreases. Many steps are involved between the motion of the camera/object and the corresponding change in the image, and delays may occur in several of these steps.

The first delay occurs through the rendering process itself: While one image is shown on the display device, the next image is being rendered into the back-buffer of the graphics card (**back-buffer rendering** or **double-buffering**). This back buffer needs to be completely rendered before it can become the new primary or **front buffer**. To avoid artifacts, this swapping of the front and back buffer (**swap-buffer**) process does not occur until the next frame. When a frame rate of 60Hz is used, a new frame is displayed every 16ms, implying that the buffers can be swapped at most every 16ms (or multiples thereof). Hence, this swap-buffer process already induces a delay of at least 16ms between the time when the new camera/object position is being set and the moment when the graphics card starts displaying the new result. Adding time-consuming methods for anti-aliasing or texture filtering can further increase the time necessary to generate new images.

If the image is directly outputted from the graphics card to a CRT monitor or projector, the time until the image is displayed is extremely small (less than a tenth of a millisecond). If further image processing equipment is added before the final display device (e.g., distortion, switching, color correction, or signal conversion), more delay would be accumulated. Using LCD, LCoS, or DLP display devices adds even more delay (see Section 3.1). Overall, the delay between the update of the virtual scene and the update of the display can easily reach values in the range of 30 – 60ms, which is a perceptually noticeable delay.

2.5 Framesync & tearing

As described above, graphics cards output individual images at a fixed frame rate. The computation of the individual images can occur at a different frequency, though, and actually be faster than the frame rate. To avoid such asynchronies, the graphics card can be set to **frame synchronization mode**, in which the swapping of the front and back buffer is synchronized with the frequency of the image output. If the synchronization is not activated, buffer swapping can occur as soon as the final image is ready. This can, for example, occur while the current image is being outputted, resulting in the device displaying the top of the previous image and the bottom of the new one (**frametear**), with a clear horizontal separation line. This effect is not visible when the scene is static, but can be rather disturbing for moving observers or objects, where the upper part appears horizontally offset from the lower part of the image.

2.6 Graphic-Cluster, framelock, & genlock

For multi-channel projection setups or when high quality rendering is needed, the image generation can be distributed to several graphics cards. While earlier rendering clusters typically used graphics-supercomputers, nowadays more and more PC clusters are used due to their reduced cost and the fast evolution of relatively low cost graphics processors. In these graphics clusters, each node renders only a part of the complete image. These individual parts are then combined to a high-resolution composite image using either specialized hardware (**compositors**) or several projectors displaying each sub-image separately. Either way, it is important that the graphics cards of the individual cluster nodes are synchronized to avoid artifacts in the composite image. This is done by either synchronizing the beginning of the rendering of each image (framesync) between the graphics cards (so-called **framelock**) or by synchronizing the rendering of each individual pixel (**genlock**).

3 Display techniques suitable for VR setups

This section provides an overview on the most commonly used visualization techniques suitable for VR setups, and we discuss specific advantages and disadvantages of each technique from a perceptual and effectiveness point of view.

3.1 Projection Systems

3.1.1 Cathode ray tube (CRT) projectors

Cathode ray tube technology has long been used for TV-like displays and was also the first technology to be used for video projectors. In general, a CRT system consists of a cathode ray (i.e., a focused electron beam in a vacuum tube) that points to a phosphorescent screen and scans it in a raster-like fashion, line by line, from top left to bottom right. The brightness of each pixel is controlled by the acceleration voltage of the electron beam and the resulting amount of energy that excites the phosphor. The actual time during which each pixel emits light is quite short due to the fast decay of the luminescence of the phosphor. This fast decay of the phosphorescent pixels is responsible for the impression of flicker that can sometimes be observed on CRT-based projection systems, as each pixel of the presented image is indeed shown only very briefly. This short presentation time is also responsible for the relatively low brightness achievable with CRT projectors. It allows, however, for a very sharp presentation of moving objects or scenes, which is important for many VR applications. This can be an important advantage over LCD, DLP, or LCoS-based projections, where the whole image is displayed for the entire duration of a frame, which can result in smeared-out impressions for fast moving stimuli.

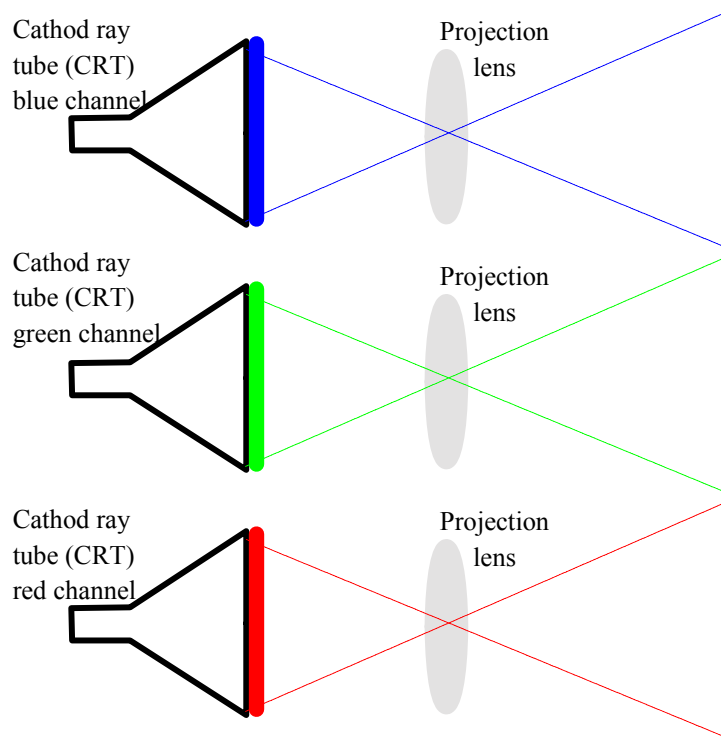


Figure 5: Schematic view of a CRT projector, showing the cathode ray tube screen being projected onto a screen using three separate CRTs and lens systems for each color channels.

For TV-like CRT displays, colors are generated by subdividing each pixel into three phosphorescent subpixels (red, green, and blue). Mixing these three colors with different intensities re-creates any color of the visible spectrum. CRT-based projectors, however, almost exclusively use three separate cathode ray tubes (one for each color) to generate color images. Calibrating such a system can be quite tedious, as each tube uses its own optics and needs to be adjusted separately in terms of focusing and positioning (convergence).

3.1.2 Liquid Crystal Display (LCD) projectors

LCD projectors typically use one or three liquid crystal glass panels to modulate the light generated by a lamp. Brightness is controlled by modulating the amount of light transmitted through each of the LCD pixels. The LCD panel thus acts like a large array of Venetian blinds. Colors are generated by using either a single panel with red, green, and blue subpixels very close to each other, similar to TFT monitors, or by using separate LCD panels for each of the three primary colors. Currently most available LCD projectors use the 3-panel setup, as it provides higher brightness and better image resolutions.

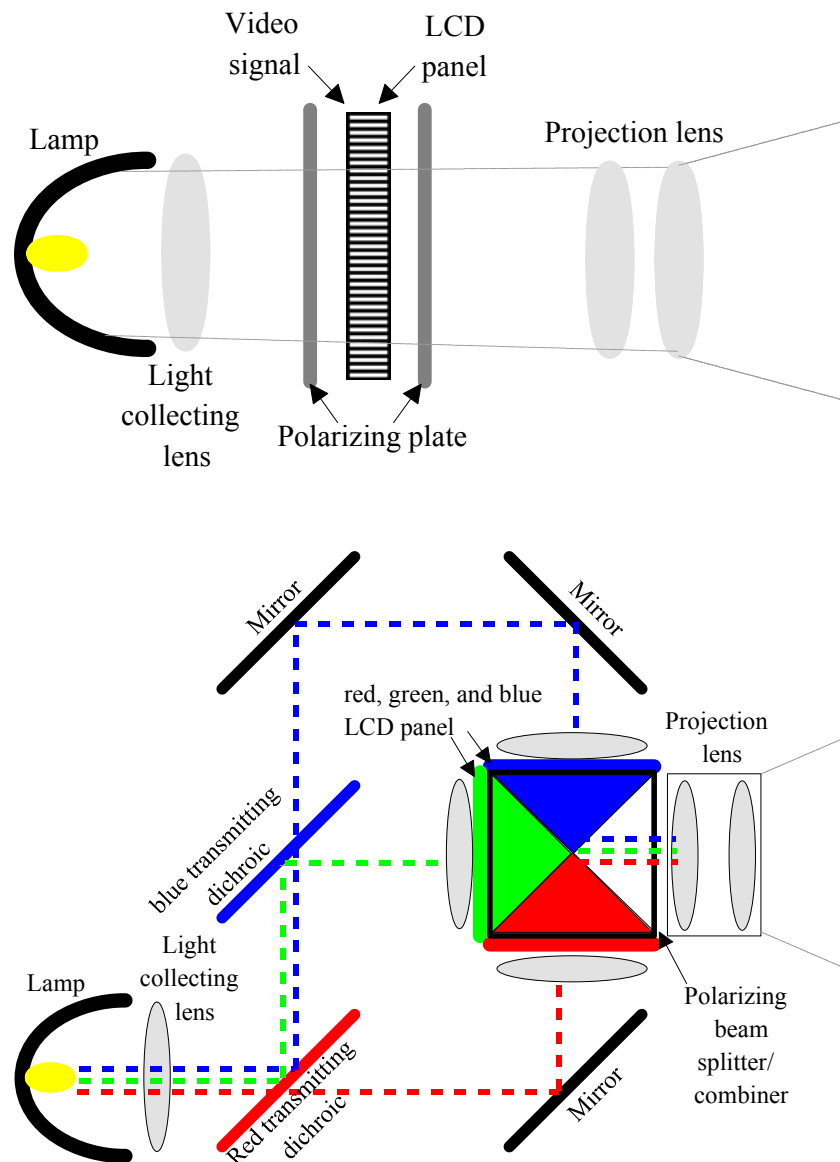


Figure 6: Schematic view of a single-chip LCD projector (top) and a 3-chip LCD projector (bottom).

3.1.3 Digital Light Processing (DLP) projectors

While transmissive technology is used in LCD projectors, the Digital Light Processing (DLP) system uses a reflective display technology. In DLP projectors, the image is created by reflecting light off a microscopically small array of mirrors (Digital Micro-Mirror Device or DMD) mounted directly on a semiconductor chip. Each mirror corresponds to one pixel in the projected image, and the intensity of each pixel is modulated by tilting the micro-mirrors to reflect the light either into the optical path (“on”) or away from it onto a heatsink (“off”). The micro-mirrors are vibrated in a stochastic fashion, and the relative amount of “on” states for each frame determines the brightness for a given pixel. The size of each mirror is only about 14 micrometer, which allows them to move very fast.

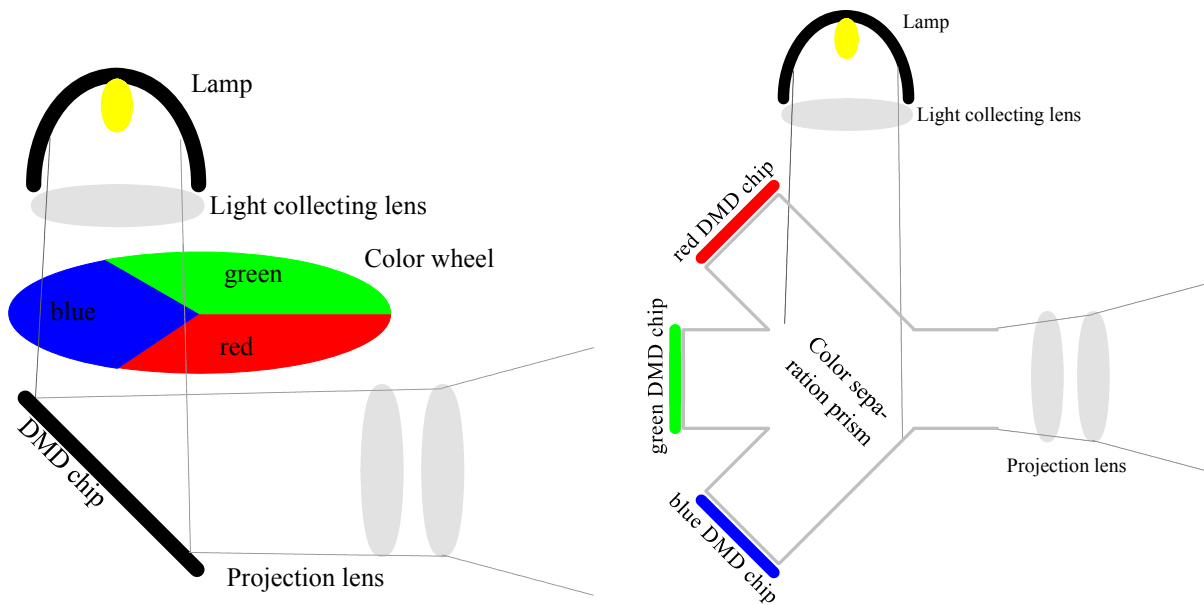


Figure 7: Schematic view of 1-chip DLP projection unit (left) and 3-chip DLP projection unit (right).

Colors are created by using either a color wheel to present the colors sequentially (1-chip and 2-chip DLPs – the latter are hardly used, though) or three separate DMD chips (3-chip DLPs). For **1-chip DLP projectors**, a color wheel is placed between the lamp and the DMD where the light is reflected through the optics (see Figure 7, left). The color wheel is subdivided into several sectors, one or two for each of the three basic colors (red, green, and blue). Sometimes, white (i.e., transparent) sectors are also used to increase brightness. The color wheel typically spins with one or two revolutions per frame (i.e., typically 60Hz or 120Hz). **3-chip DLP projectors** avoid the usage of a color wheel by using a prism to split the light beam originating from the lamp into the primary colors, which are then directed to three different DMD chips (see Figure 7, right).

3.1.4 Liquid Crystal on Silicon (LCoS) projectors

A more recent LCD-based projector design is the LCoS (Liquid Crystal on Silicon) or D-ILA (Direct Drive Image Light Amplifier) technology. It is a reflective projection technology similar to DLP projectors, but uses one or three panels of reflective liquid crystal instead of a micro-mirror array. LCoS projectors might thus be seen as a hybrid between the reflective DLP technology and normal LCD projectors, which uses transmissive LCD panels. In LCoS chips, a liquid crystal array is mounted on a reflective mirror substrate. The image is created by individual liquid crystals switched to transparent or opaque, thus either reflecting the light from the mirror below or not. Most LCoS projectors use three separate chips, one to modulate the light of each of the red, green, and blue channel. LCoS chips are typically built directly onto silicon wafers, much like for memory or processor chips used in computers. This provides extremely high pixel densities and resolutions, thus compact and cost-effective designs. Compared to LCD projectors, the reflective design of LCoS projectors allows for smooth images without the pixelation present in LCD projectors, and allows for higher contrasts and more compact designs (just like DLPs). Currently available LCoS projectors are, however, still costly, need rather expensive lamps that often have only little more than 1000-2000h lifetime, and do not (yet) allow for image contrasts comparable to DLP projectors.

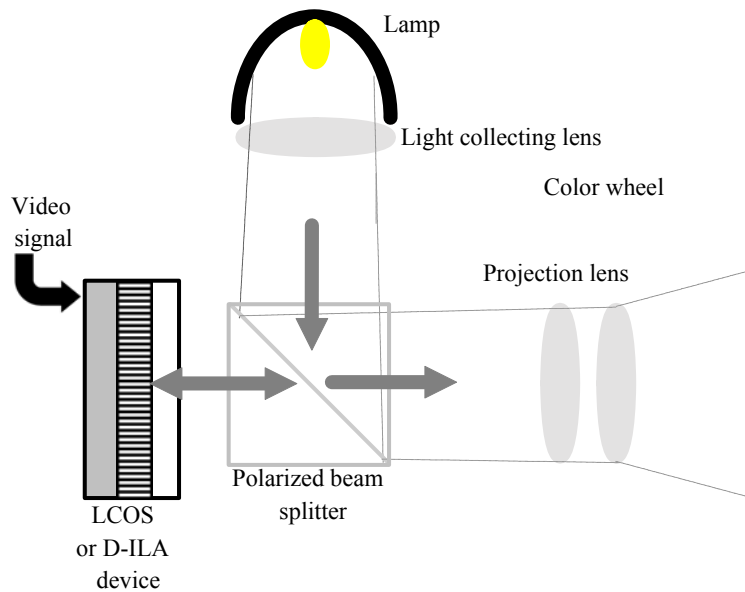


Figure 8: Schematic setup of LCoS or D-ILA projection unit using reflective LCD panels.

3.1.5 Comparing the different projector types

Even though DLP projectors have many advantages, there are a few drawbacks associated with them that might not be visible at first glance. Hence, we will describe them in a bit more detail in the following.

Generally, 1-chip DLP projectors use a color-wheel for color separation. That means that colors are presented sequentially and not simultaneously as for 3-chip DLP projectors. Typically, this color wheel consists of 4 segments (red, green, blue and white) and rotates at a frequency of 120Hz for a resulting 60Hz framerate. Therefore, each color is presented twice per frame. This leads to a noticeable "double burst" in the output signal of DLP projectors. This can be seen in Figure 9, which shows the response of a 1-chip DLP projector to a one-frame flicker signal (one frame on, one frame off; notice that the input signal is inverted in the figure).

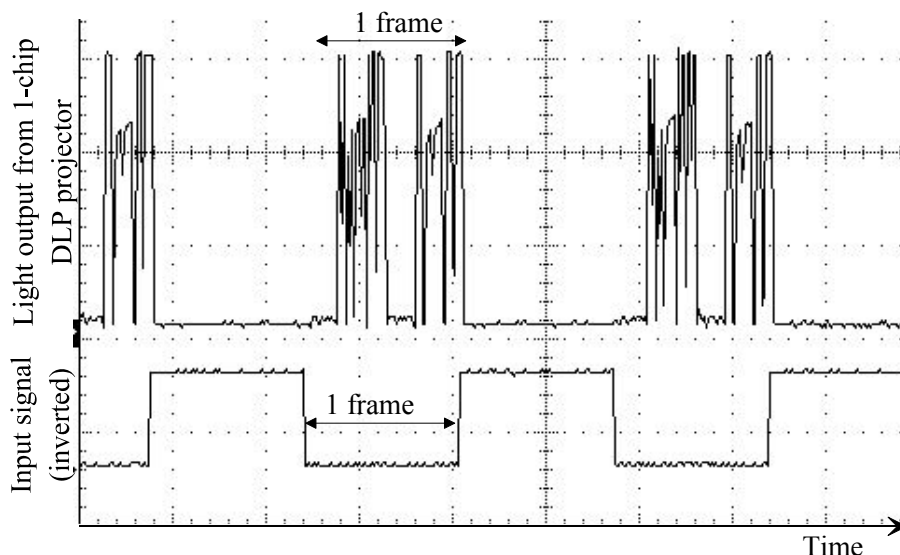


Figure 9: Resulting output of a 1-chip DLP-Projector (top signal) for a one frame flicker input signal (square wave at the bottom). Note that each color is presented twice per frame, as the color wheel rotates at twice the frame rate (here: 120Hz rotation for 60Hz framerate). This results in a noticeable double burst in the output signal.

This combination of sequential color separation and double burst is fine for static stimuli and does not pose any problems for displaying dynamic stimuli as long as the user fixates a point on the screen. For normal, unrestricted viewing conditions, however, our eyes typically follow the displayed moving stimulus smoothly (“smooth pursuit”). This results in a spatial segmentation of the different colors (so-called “rainbow effect”): When the eyes follow, for example, a line that moves quickly across a black background, the line splits up into several individual lines (one for each segment of the color wheel times the number of revolutions per frame). That is, 1-chip DLP systems can result in eye-motion based spatial color variations. This can be quite disturbing, especially when showing moving high-contrast stimuli. Finally, in the case of saccadic eye movement or blinking, one can see spots of individual color in areas of high contrast. This occurs both for static and dynamic stimuli. This effect is difficult to describe but very easy to see. 3-chip projectors do not suffer from the “rainbow effect”, since all three colors are presented synchronously, not sequentially.

DLP projectors (both the 1-chip and the 3-chip versions) manipulate the luminance levels of each color channel by controlling the individual mirrors in a stochastic fashion (**temporal flickering**). This results in temporally unstable images (i.e., there is some scintillation). This is especially noticeable at low luminance levels. Figure 10 illustrates this scintillation in the response of a 3-chip DLP projector for a dark grey flicker signal. Note that 1-chip DLP projectors have a similar response characteristic.

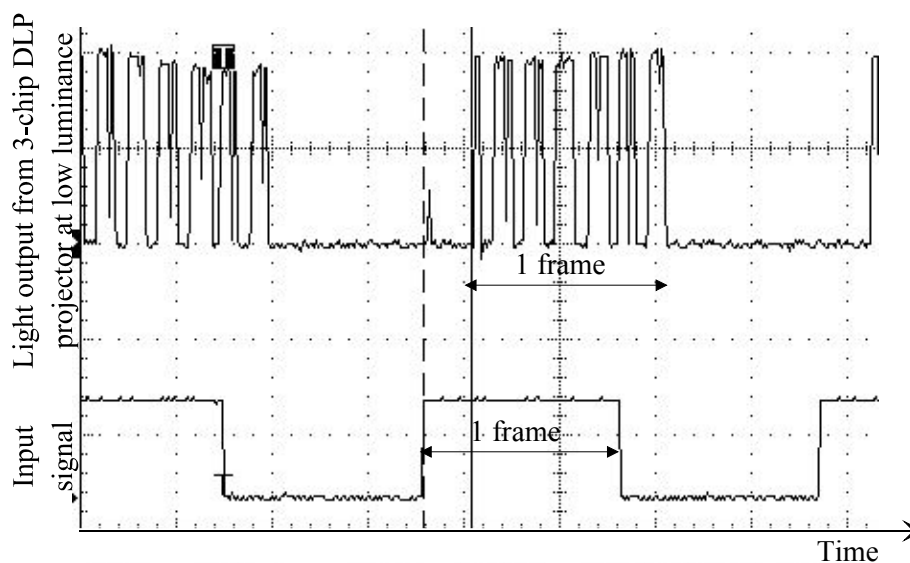


Figure 10: Brightness signal of a three chip DLP-Projector (top signal) for a one frame dark gray flicker signal of a square wave input (bottom). Note the temporal flickering for DLP-based projectors due to the stochastic motion of the individual mirrors.

All of the above-mentioned drawbacks associated with DLP projectors do not exist for 3-chip LCD or LCoS-based projectors, which present all three colors simultaneously and with relatively constant amplitude for the full duration of a given frame. Hence, even though 1-chip DLP projectors are becoming cheaper and are suitable for many general-purpose applications, their usefulness for high-quality VR presentations with requirements for high stimulus control needs to be carefully evaluated. Often, LCD and LCoS projectors are more suitable as they involve fewer artifacts. The table below provides a rough comparison between the different projection types, and can be helpful to decide which type of projector is optimal for a given task.

See also <http://www.kybervision.net/en/TechRev/projectors.html> for an overview of using different projector types for vision research.

Table 1: Comparison of different projector types used for VR applications.

	CRT projectors	LCD projectors	DLP projectors	LCoS/D-ILA projectors
Advantages	<ul style="list-style-type: none"> ● High contrast ratio ● Excellent deep black levels ● Excellent color and image quality ● No pixelation visible ● High resolutions possible, as no grid mask or fixed pixels necessary ● Moving objects are sharp, not blurred during smooth pursuit ● Minimal delay for displaying rendered image ● High frame rate ● Long lifetime ● Quiet 	<ul style="list-style-type: none"> ● Excellent color saturation and reproduction ● Precisely focused (sharp) image at a given resolution 	<ul style="list-style-type: none"> ● High contrast ratio ● Good color saturation ● Reduced pixelation ● Lightweight and compact ● High light efficiency ● High response speed ● Long lifetime ● Fully digital displays without analogue conversion 	<ul style="list-style-type: none"> ● Good black level ● Excellent color and image quality ● Reduced pixelation ● High resolutions possible ● High pixel density (7 μm pitch) ● Long lifetime ● Easier to manufacture and more compact than LCD panels ● Fully digital displays without analogue conversion ● High frame rate
Disadvantages	<ul style="list-style-type: none"> ● Large, bulky, and heavy ● Adjusting convergence and color can be tedious and needs to be re-adjusted at times ● Only moderate brightness ● White is sometimes not really white ● Drift possible that requires recalibration ● Problems possible for interlaced images ● No zoom available ● Rather expensive 	<ul style="list-style-type: none"> ● Visible pixelation (Screendoor effect) ● Poor black level ● Low contrast ratio ● Often more bulky than DLP projectors, as they need more internal components. ● Possibility of “dead pixels”, which are permanently on or off. ● Contrast is somewhat lower than for DLPs/LCoS ● Relatively low brightness, especially for higher resolutions 	<ul style="list-style-type: none"> ● Rainbow effect for single chip projectors ● Light leakage or “halo” effect possible: Stray light reflected off the edges of the micro-mirrors can produce a gray band around the projected image, especially for older models. ● Fan and color wheel in 1-chip DLPs can be noisy ● Noticeable flickering at low brightness levels ● Very expensive for 3-chip DLPs 	<ul style="list-style-type: none"> ● Still rather expensive

3.2 Head-mounted display (HMD)

Probably the most commonly used display devices for VR are head-mounted displays (HMD). They consist of two small displays, one for each eye, worn like a helmet or heavy glasses; and they allow for stereoscopic vision. Special optics are used to present a sharp image at a comfortable viewing distance (typically 0.5-1m), even though the displays are only a few centimeters away from the users' eyes.

Similar to the projection systems discussed above, different display technologies are employed for HMDs. Earlier HMDs typically used LCD or CRT-based displays, whereas more recent models often use LCoS and sometimes organic LED technology. Compared to LCoS-based projection systems, where three panels are typically used to generate images, (with one panel for each of the three primary colors), HMDs often use a **color multiplexing** system similar to the color wheel in 1-chip DLP projectors. Instead of using one lamp and a color wheel, three different LED-based light sources are used which can be separately switched on or off. This can lead to color artifacts similar to the ones discussed above for 1-chip DLP projectors.

The FOV of HMDs depends on the model and currently ranges from about 20° - 120° horizontally and 16° - 67° vertically. HMDs have a fixed angular resolution, since they are pixel-based display systems and are presented at a fixed viewing distance. The angular resolution varies considerably between models, with values ranging from 1.4 to 10 arcmin/pixel. For most HMDs, the region of the image that is seen simultaneously by both eyes (thus in stereo) covers 100% of the simulated FOV ("100% stereo overlap"), but some models use only a partial overlap (as low as 35%) by shifting the displays horizontally. This way, the horizontal FOV can effectively be enlarged by the part that is only seen monocularly. Nevertheless, even the most advanced HMDs are still far from approaching the full human visual FOV (see Section 1.1). Interestingly, a binocular overlap of only 20° is typically sufficient to allow for convincing depth perception.

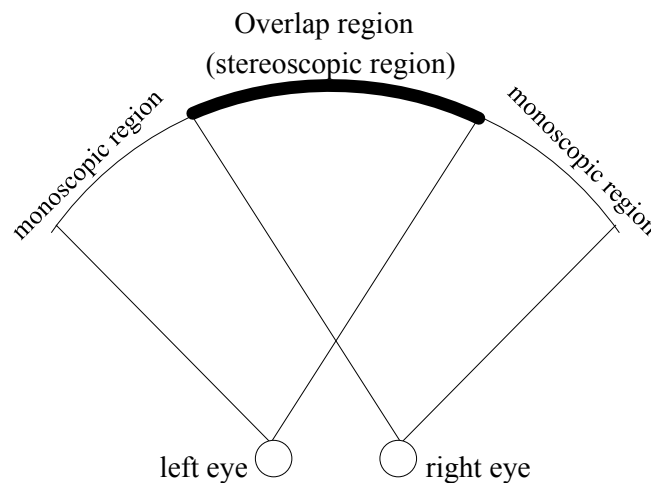


Figure 11: Region of stereoscopic overlap in the human.

Since the presented image of the simulated environment is fixed relatively to the head in HMDs, it is important to track the position of the observer's head using a real-time tracking system and move the virtual camera in the virtual scene accordingly. Without such a tracking, a head motion to the right would, for example, not result in the expected change in the simulated viewing direction. The change in head position and orientation should ideally result in an immediate change in the presented image, just like in the real world. This is technically impossible in principle, as the image is only presented with a limited update rate (typically 60Hz), which already induces a delay of 16ms. The tracking system itself adds a further delay of at least several milliseconds. Hence, care should be taken to minimize additional delays between head motion and resulting changes of the displayed image. Large delays result in images that noticeably lag behind what is expected for fast head motions. This is mostly noticeable during rotations, and can easily lead to simulator sickness symptoms such as eye strain, dizziness, headaches, and general discomfort.

Further information and overview can be found for example at www.genreality.com/comparison.html, www.stereo3d.com/hmd.htm#chart,

http://vresources.jump-gate.com/articles/vre_articles/analyhmd/analysis.htm.

4 Stereo projection techniques

The benefit of stereoscopic image presentation for an increased sense of Presence experienced in a virtual environment has been shown experimentally (e.g., IJsselstein, 2004). To achieve a convincing simulation of spatial depth and distances, a binocular projection is crucial. HMDs make this possible by providing binocular cues, since each eye looks at a different LCD display. For projection systems, several techniques exist, with different advantages and disadvantages, as will be discussed in the following. The main idea behind each of the technologies is that a different image has to be presented to each eye, just like for the HMD. Monocular depth cues are currently not used for VR applications, even though there are promising evaluations of monitor-based experimental setups using several depth layers (Akeley, Watt, Girshick, & Banks, 2004).

4.1 Active stereo using a single projector

With active stereo setups, the user is equipped with special shutter glasses that alternatively block the vision of the left and right eye, and are synchronized with a projector that alternatively projects the image destined to the right and left eye accordingly. That is, when the projector shows the right eye image, the left eye shutter is closed, and vice versa. Hence, each eye sees an image only every other frame. This can result in perceived flicker if the update rate of the projector is not high enough – which poses a serious constraint for most low-cost projectors. Furthermore, the flicker can increase eye strain and cause headaches, especially for longer presentation times.

4.2 Passive Stereo

Passive stereo setups do not require the user to wear active shutter glasses that switches between the left and right eye. Instead, the user is equipped with glasses that act as light filters. A similar pair of filters is used in front of the projector to give light properties that will allow the glasses to let only either the left or the right image through. For single projector setups, this can be achieved by using a filter that switches between two modes. For two projector setups, one projector always displays the left image, and the other the right image. This has the advantage of permitting both projectors to operate at the standard refresh rate, effectively reducing flicker. Furthermore, the brightness is increased, as each eye sees the full image the whole time.

4.2.1 Passive stereo using polarization filters

Light is an electromagnetic transversal wave that can be decomposed into two orthogonal polarization directions. Most commonly, linear polarization filters are used to decompose the unpolarized light into two polarization directions (e.g., vertical and horizontal). By arranging the polarization of the light oppositely for the left and right eye, it is possible to send different information to the left and right eye, thus enabling binocular depth perception. Note that humans are largely insensitive to polarization and are thus unaware of the manipulation. Using linear polarization filters is the most cost-effective and commonly used solution to achieve good binocular stereo projections and works well as long as the head tilt of the user is limited – note that users perceive double images when tilting their head more than a few degrees. A more costly solution is to use circular polarization filters, which work well independent of head tilt.

4.2.2 Passive stereo using wavelength multiplex imaging (Infitec™ color bandpass filters)

A recent alternative to polarization filters is to use Infitec (Interference Filter Technique) color bandpass filters to present two different images to the two eyes simultaneously. Each of the two optical filters splits the color spectrum into a wavelength triplet that is slightly offset one from another (see Figure 12). By mounting two different filters in front of the two projectors, it is possible to display two images that are complementary in terms of their color spectrum. Hence, by using the same pair of color bandpass filters in the glasses worn by the user, it is possible to present stereo information with minimal cross talk, independent of head tilt or viewing direction, and without the need for special projection screens that maintain the polarization direction of the projected light. As humans have only three different types of color receptors (see Section 1.2.1), the bandpass filtering remains unnoticed as long as the color for each projector is carefully calibrated. Thus, wavelength multiplex is an interesting alternative to the most commonly used polarization or shutter-glass based stereo projection techniques, as it circumvents many of their drawbacks.

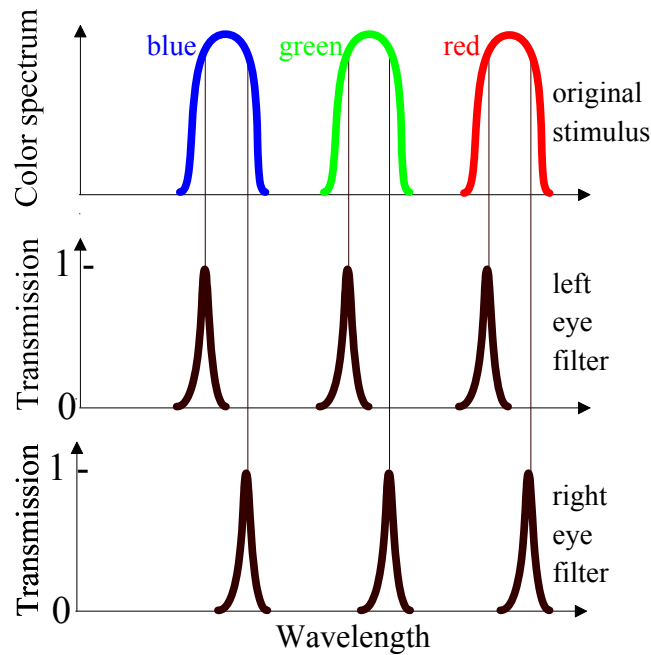


Figure 12: Schematic illustration of the functioning of Infitec color bandpass filters depicting the separation of the original color spectrum (top graph) into the left and right eye image using two different filters.

4.2.3 Light efficiency of different stereo projection techniques

Whenever stereo projection is used, creating a separate image for each eye results in a loss of effective light that reaches the eyes. Due to the increased brightness of recent projectors, this is typically not a problem for smaller projection setups and/or when the room can be completely darkened. Whenever larger projection screens are needed and/or ambient light cannot be excluded, however, effective stereo brightness (stereo lumen) can become critical and a deciding factor between the different possible stereo projection paradigms.

The light efficiency of the different stereo projection techniques obviously depends a lot on the individual projection setups and filters used, but we will nevertheless try to give some general rule-of-thumb here. More detailed information is available from the respective companies, and a good overview on this issue can be found under <http://www.barco.com/VirtualReality/en/steroscopic/lumens.asp>.

Active stereo using a single projector and shutter glasses results in the highest loss in light effectiveness. Presenting the images sequentially to the left and right eye already reduces the maximum achievable brightness by at least 50%. At least another 50% of the remaining brightness is lost due to the polarization in the active shutter glasses. This leads to a theoretical maximum in achievable brightness of $50\% \times 50\% = 25\%$. Due to technical limitations, however, the overall effectiveness hardly exceeds 16%.

Light effectiveness can, however, be increased when using **two projectors in a passive stereo** setup. Using **polarizing filters** in front of the projector and the eye reduces the brightness by at least 50%, and limitations in the effectiveness of polarizing filters currently limits the effective stereo brightness to about 38%. Using wavelength multiplex imaging with **Infitec™ filters** results in a slightly decreased effective brightness (maximum efficiency of about 27%) due to practical limitations in the color filtering. A maximum light efficiency of about 60% can be obtained when using highly specialized LCD projectors that use an optimized internal polarization.

5 Conclusions

The goal of this chapter was to relate human perception and artificial stimulation of our different senses using current technology. As this is a vast domain, we have restricted ourselves to visual stimulation, as this is probably the cue that gives us the most information about our immediate surrounding environment. We have reviewed a few low-level features of the extremely complex human visual system, and how it uses them to create a representation of the surrounding environment, notably how it captures colors, brightness and how it allows the brain to recreate an accurate 3 dimensional representation of the surroundings. This relates directly to the way we need to stimulate our visual system to allow us to build a convincing mental representation of a

virtual environment. We reviewed a few of the current technologies available for this purpose, and, for each of them, described their advantages and drawbacks, from both the human perception and purely technical points of view. Although technology progresses very fast towards generating more realistic looking visual displays, it is still very far from being able to stimulate our visual system in the same manner as the real world does. There are also problems in terms of practicality (comfort, size, physical constraints ...), and cost is often a limiting factor. However, for a given task and application, the hardware can already give satisfactory results, provided it is carefully chosen for its properties. Spending a lot of money to buy the latest, most advanced technology is not always necessary. Producing enough relevant information for the brain is what should be considered. For instance, stereoscopic vision can be crucial to train surgeons, in order for them to have a more accurate idea of the relative position of different organs. However, although it is known that stereoscopic projection increases the sense of presence, monocular vision is often sufficient for architectural walk-throughs. Hence, carefully choosing hardware depending on the purpose of the task can greatly reduce the costs while producing equivalent results as the most advanced technology in terms of human perception. Even though technology progresses at a tremendous rate, finding the appropriate cost/efficiency compromise will remain an important factor when designing virtual reality systems for the next years, and probably decades. Before setting up a new simulation system, experimental evaluations should be performed in order to optimize task related performance while keeping the costs within a reasonable range.

6 References

- Akeley, K., Watt, S. J., Girshick, A. R., & Banks, M. S. (2004). A stereo display prototype with multiple focal distances. *ACM Transactions on Graphics*, **23**, 1804-813.
- Goldstein, E. B. (2002). *Sensation and perception*, 6th edition. Belmont, CA: Wadsworth Publishing Co.
- IJsselstein, W.A. (2004). *Presence in depth*. PhD thesis, Technische Universiteit Eindhoven.