Augmented Reality - Display Systems

A survey by Sebastian Klepper
TU Muenchen, 04.07.2007
Sebastian.Klepper@myTUM.de

This article surveys display systems prominently employed in AR applications. In course of this, several applicable display technologies are presented as well as different display types and designs.

Initially, the demands of an AR system are determined with emphasize on enabling the system to give the user the illusion of virtual objects displayed to him being part of his real environment.

As each of the possible display system satisfies a different set of requirements brought along by AR and concurrently bears individual weaknesses, as many aspects of every alternative as possible are treated and compared.

Subsequently, some diverging paradigms in display design are discussed along with challenges and difficulties that come up when you take a closer look on the functionalities of an AR system.

Finally, a short prospect on the capabilities of AR and possible future progression is given, also closely bound to the requirements that efficient AR applications bring along.

Contents

1 Augmented Reality

Augmented Reality (AR) is a variation of Virtual Environments (VE), or Virtual Reality as it is more commonly called. VE technologies completely immerse a user inside a synthetic environment. While immersed, the user cannot see the real world around him. In contrast, AR allows the user to see the real world, with virtual objects superimposed upon or composited with the real world. Therefore, AR supplements reality rather than completely replacing it.

1.1 Definition

A new and major area of current research is into the use of AR outdoors. GPS and orientation sensors enable backpack computing systems to take AR outdoors. AR therefore has clear connections with the Ubiquitous Computing (UC) and wearable computing domains, which in some applications formed the term Ubiquitous Augmented Reality (UAR).

The most salient distinction to be made between AR and UC is that UC does not focus on the disappearance of conscious and intentional interaction with an information system as much as AR does: UC systems such as pervasive computing devices usually maintain the notion of explicit and intentional interaction which often blurs in typical AR work such as Ronald Azuma's work. [8]

As compared to UC, his definition is more focused and covers a subset of AR's original goal, but it has come to be understood as representing the whole domain of AR: Augmented reality is an environment that includes both virtual reality and real-world elements. Azuma defines AR as systems that have the following three characteristics:

1. They combine real and virtual.
2. They are interactive in real time.
3. They are registered in 3-D.

[1, pp. 2-3]

1.2 Requirements

The overall requirements of AR can be summarized by comparing them against the requirements for Virtual Environments (VE or Virtual Reality, VR), for the three basic subsystems that they require.

1. **Scene generator:** Rendering is not currently one of the major problems in AR. VE systems have much higher requirements for realistic images because they completely replace the real world with the virtual environment. In AR, the virtual images only supplement the real world. Therefore, fewer virtual objects need to be drawn, and they do not necessarily have to be realistically rendered in order to serve the purposes of the application. [1, p. 17]

2. **Display device:** The display devices used in AR may have less stringent requirements than VE systems demand, again because AR does not replace the real world. On the one hand, monochrome displays may be enough for AR, wheres nobody would use a VE that is not fully coloured. Also AR displays might get along with lower resolution image data, while this would not be accepted in a VE. [1, p. 17]
3. **Tracking and sensing:** While in the previous two cases AR had lower requirements than VE, that is not the case for tracking and sensing. In this area, the requirements for AR are much stricter than those for VE systems. A major reason for this is the registration problem, which is described in section 5.3.1. [1, p. 17]

2 Types of AR Displays

Most AR applications will need a personal set of earphones, gloves and a display. In the following, the most significant types of AR displays - of which some are already present in everyday life - will be presented.

2.1 Head-Up Displays (HUDs)

HUDs are environmentally fixed displays, for example mounted in the cockpit of a vehicle (e.g. car, aircraft). "Head-Up" means that the user doesn't have to look away to view the desired information – this would be the case with a Head-Down Display (HDD). A HUD in a vehicle typically consist of three primary components:

1. A **computer** which provides the interface between the vehicle's electronic system and the Overhead Projector Unit.
2. The **Overhead Projector Unit** (OPU) is supplied with data by the computer and projects it onto the combiner.
3. A **combiner** is usually made of glass with a special coating that reflects the monochromatic light from the OPU while allowing all other wavelengths of light to pass through, creating a superimposed image.

In airplanes the computer usually is with the other avionics equipment and recieves data from the inertial reference system, flight management system, and other flight guidance systems. A combiner is located in front of the crew member using the HUD and other than in cars the display is here focused to infinity. Tactical military aircraft usually rely on a projection unit incorporated onto the combiner. [9]

HUD systems are also being designed to utilize a synthetic vision system (SVS), which use terrain databases to create a realistic and intuitive view of the outside world. For example, *"The Tunnel In The Sky"* [9] (a perspective 3-D model of the plane's location and orientation) can greatly assist the pilot when more precise flying is required, such as the decreased vertical or horizontal clearance requirements of RNP. Under such conditions the pilot is given a graphical depiction of where the aircraft should be and where it should be going rather than the pilot having to mentally integrate altitude, airspeed, heading, energy and longitude and latitude to correctly fly the aircraft. [9]

Lately, systems like that can also be found in cars where a special area on the windshield is used as the combiner on which real-time data from the speedometer, tachometer or navigation system is projected. Thus the driver is constantly informed about the status of his car without the need of taking his view off the street. However, the driver's eyes have to refocus on the combiner, because other than in aircrafts, the display in automobiles is focused around the distance to the bumper. Thus, the driver cannot read the displayed information without "focusing away" from the street traffic, but due to restricted measurements a HUD focused to infinity is not installable here.

2.2 Head-Mounted Displays (HMDs)

Per eye, a HMD is composed of a modulated light source with drive electronics viewed through an optical system (mirrors and combiner) which, combined with a housing, are worn on a users head via a headband, helmet, or around an eyeglasses frame. [2, p. 1] When mounted to a helmet, e.g. of a pilot or a biker, HMDs are also referred to as Helmet-Mounted Sights (HMSs). It always points in the same direction the user is looking and hence has the capability of displaying virtual objects aligned to the real world in the user's field of vision - of course this does not make tracking and registration redundant.

Military, police and firefighters can use HMDs to display tactical information such as maps or thermal imaging data while viewing the real scene. The i-Vue company make a Night Vision HMD unit with a night vision camera (image intensifier) that feeds imagery to the HMD optics. Engineers and scientists use HMDs to provide stereoscopic views of CAD schematics. Finally, low cost HMD devices are available for use with 3D games and entertainment applications. [10]

Driven by far field and near field applications, the unique distance of the optical images can be set either beyond 6m (i.e. optical infinity), or at about an arms length, respectively. Objects within the optics depth of field at a specific setting will be perceived sharply. Other objects will be perceived blurred. For dual near-far field applications, multifocal planes displays are necessary. [2, p. 2]

Essential to the design of HMDs is an anthropometric measurement: the **Inter-Pupillary Distance** (IPD), which indicates the distance between the two eyes, measured at the pupils. It is used in specifying the size range not only for Head-Mounted Display systems but also for eyeglasses (spectacles), binoculars and other optics.

As every measurement in anthropometry, the IPD spreads statistically. The figures for adults in the USA and in Eurpoe vary from 52mm to 70mm. Of course the exact figure varies from person to person, so depending on age, gender and origin of the target user group, a device's IPD must be adjustable. [10]

2.3 Head Mounted Projection Displays (HMPDs)

A shift in paradigm in HMD design is the replacement of compound eyepieces with projection optics combined with phase conjugate material (e.g. retroreflective optical material), known as head-mounted projection displays (HMPD).

A HMPD consists of a pair of miniature projection lenses, beam splitters, and microdisplays mounted on the head and non-distorting retro-reflective sheeting material placed strategically in the environment. An example would be a so called Augmented Reality Center (ARC), a deployable room coated with retro-reflective material in which a user can interact with a virtual 3-D environment. [2, p. 5]

Projection optics and a retro-reflective screen, instead of eyepiece optics and a diffusing screen, both respectively distinguish the HMPD technology from conventional HMDs and stereoscopic projection systems.

Given a field of view and the internal pupil to the lens which is re-imaged at the eye via the beamsplitter (oriented at 90°from the internal pupil used in conventional HMDs), projection optics can be more easily corrected for optical aberrations like distortion and this on a high level of scalability (an increased field of view does not include the need to adjust the projector optics). [2, p. 5]

2.4 Occlusion Displays

When developing augmented reality applications, scientists have too choose between optical and video see-through mode displays. Video mode is mostly preferred because it is relatively easy to implement occlusions on a pixel-by-pixel basis and despite the fact that they lack quality in terms of resolution (reduced by subsampling through the cameras) and lag (caused by full image pre-processing). Also the viewpoint of the cameras constantly has to match the viewpoint of the eyes.[2, p. 5]

Given these drawbacks, it is desirable to choose optical see-through displays if they can provide occlusion capability: a strong monocular cue to depth perception and may be required for certain applications.[2, p. 6]

Most optical see-through designs will combine computer generated imagery with the real world using a beam splitter which will always transmit some light, regardless of its transmittance and reflection percentages. So unless the image sources are much brighter than the scene it is difficult to achieve opaque display of virtual objects and alternative mechanisms to the conventional head-mounted display designs become necessary.[2, p. 6]

A first approach here could be to uniformly dim the light with liquid crystal shutters under voltage control and combining the modulated output with the image source. Using electrochromic films, light levels can be controlled in a similar way, eliminating the crossed polarizers. Finer grained control over regions within the scene requires masks with multiple pixels.[2, p. 8]

2.5 Eyeglass Displays

A number of factors including aesthetics and social acceptance will push displays targeting daily visual aids towards integration with the eyeglasses form factor. Within this form factor it is extremely challenging to fulfill high-performance optical requirements.

Initial prototypes were working with small light sources mounted directly on the surface of an eyeglass lens or small reflecting mirrors on the lens of the eyeglasses. The latter would result in moving the direct mounting of the light sources away from the lens and being less noticeable and less obstructive to the wearers vision. [2, p. 9]

The eyeglasses lens in a system developed in the late 90's by Spitzer and colleagues [3] has the overall thickness of less than 6.5mm which fits in the commercial eyeglass frame. It features a relay system built into the eyeglass frame to move the microdisplay away from the eyeglasses. The reason for this was that the display necessary for the aimed image measurements based on a practical magnification of a single lens would have been too large for concealment in eyeglasses. [2, p. 9]

This type of display has influence on the user's depth perception (stereopsis). Monocular devices per se impair judgement of distance, speed and size. Also current eyeglass display products are likely to produce vergence lock, a potential health hazard. Therefore they should allow good peripheral vision and should be used in relatively light environments. [5, p. 236] Generally, the effect of viewing computer generated 3-D images on the human visual system depends on both the duration of the viewing and the apparent distance between the viewer and the virtual objects.

3 Display Technologies

The following display technologies are used in various forms of augmented reality display designs, in most of which several of these technologies are usable. Therefore no distinct assignment of one technology to one specific display type is possible.

The suitablility of a technology to be used in an AR application can be determined by reckoning these qualities:

Small Size To be deployable in HMDs and Eyeglass Displays, a display should match a diagonal size of 10-20 mm. Applications that require larger displays (e.g. a HUD) normally use projection techniques with a scalable combiner.

High Brightness & Contrast Ratio These are of special importance when the AR application is to be taken outdoors, because then the display system has to deal with the ambient light to make the virtual objects visible and keep the illusion of them being part of the real world.

Grayscale Performance Besides the contrast ratio, an imaging device's ability to display subtle detail also arises out of its ability to display various levels of gray, so both figures matter.

Short Response Time This parameter is variedly crucial for different display types, as it's measured data reach from microseconds to tens of milliseconds, which may be noticeable for the human eye in evident visual effects, such as the fuzziness of moving objects.

3.1 Cathode Ray Tube (CRT)

The CRT is an evacuated glass envelope containing a source of electrons along with magnetic spools to accelerate and deflect the electrons and a fluorescent screen on which a hit of an electron causes light-emission.

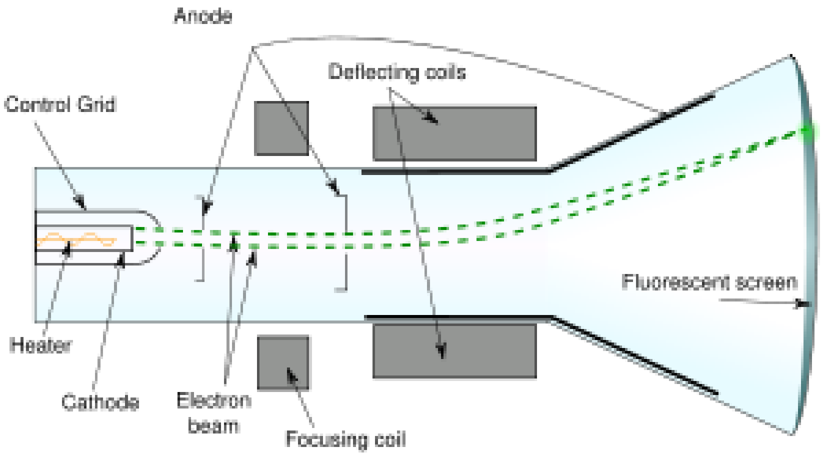

Figure 1: Cathode ray tube employing electromagnetic focus and deflection

CRTs have a pronounced triode characteristic, which results in significant gamma (a nonlinear relationship between beam current and light intensity). In systems where linear response is required, gamma correction needs to be applied.

The evacuated glass envelope required for the electron beam to reach the screen is large, deep, heavy and relatively fragile, also a CRT is very delicate to magnetic influences and bear health dangers in terms of electromagnetic emission, ionizing radiation, toxins used inside the glass envelope and the high voltage required to operate the CRT.

While early HMDs employed monochrome and lateron also color field-sequential CRTs, they have largely been replaced by flat display technologies like LCDs and OLEDs. They were initially preferred due to greater color fidelity and contrast, better resolution when displaying moving images and better view from angles, but improvements in other technologies increasingly alleviate these concerns.

3.2 Liquid Crystal Display (LCD)

A LCD is a thin, flat display device made up of any number of color or monochrome pixels arrayed in front of a light source or reflector.

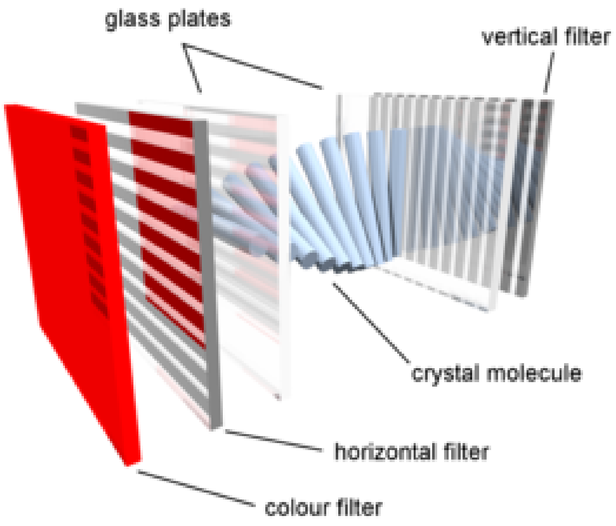

Figure 2: Red subpixel of a reflective twisted nematic liquid crystal display

More precisely each subpixel of the LCD consists of a colour filter, a filter polarizing the light as it enters, twisted nematic liquid crystals, a filter polarized opposite to the first one and thus allowing to block light and a reflective surface to send the light back to the view (in a backlit LCD this layer is replaced with an additional light source). Glass substrate with ITO (indium tin oxide) electrodes can be placed over the LC layer to determine shapes on the overall LCD.

Two active matrix technologies are used in LCDs:

1. **Twisted nematic (TN)** – Twisted nematic displays contain liquid crystal elements which twist and untwist at varying degrees to allow light to pass through. When no voltage is applied to a TN liquid crystal cell, the light is polarized to pass through the cell. In proportion to the voltage applied, the LC cells twist up to 90 degrees changing the polarization and blocking the light's path. By properly adjusting the level of the voltage almost any grey level or transmission can be achieved.
2. **In-plane switching (IPS)** – In-plane switching is an LCD technology which aligns the liquid crystal cells in a horizontal direction. In this method, the electrical field is applied through each end of the crystal, but this requires two transistors for each pixel instead of the one needed for a standard thin-film transistor (TFT) display. This results in blocking more transmission area requiring brighter backlights, which consume more power making this type of display less desirable for mobile devices.

Also, zero power devices have been developed with different approaches. Such devices, e.g. the zenithal bistable device (ZBD) or polymer stabilized cholesteric liquid crystals (ChLCD), are capable of retaining an image without power supply, but yet find no use in augmented reality displays. [11]

3.3 Digital Light Processing (DLP)

DLP uses so called Digital Micromirror Devices (DMDs) to reflect light and modulates the image by tilting the mirrors either into or away from the lens path.

There are two primary methods by which DLP projection systems create a color image:

1. **Single-Chip Projector** – The DMD chip is synchronized with the rotating motion of a color wheel. The red, green, and blue images are thus displayed sequentially at a sufficiently high rate that the observer sees a composite "full color" image. [12]
2. **Three-Chip Projector** – A three-chip DLP projector uses a prism to split light from the lamp, and each primary color of light is then routed to its own DMD chip, then recombined and routed out through the lens. Three-chip DLP projectors can resolve finer gradations of shade and color than one-chip projectors, because each color has a longer time available to be modulated within each video frame; furthermore, they have a reduced potential for flicker and rainbow effect. Like three-tube CRT projectors, the optics for three-chip DLP projectors must be carefully aligned. [12]

3.4 Liquid Crystal on Silicone (LCoS)

LCoS combines the ideas of LCD and DLP. Thus it is a reflective technology using liquid crystals instead of individual mirrors which are applied to a reflective mirror substrate: a silicon chip coated with an aluminized layer. As the liquid crystals open and close, the light is either reflected from the mirror below, or blocked. This modulates the light and creates the image. [13]

Alike the DLP technology, there are two broad categories of LCoS displays:

1. **Three-Panel Design** – One chip is responsible for each display color and the images are combined optically. Similar to DLP devices the light is separated into three components and then cmbined back, additionally the light is polarized and then analyzed, which makes four beam splitters necessary. [13]
2. **Single-Panel Design** – A single display chip shows all RGB components in succession with the observer's eyes relied upon to combine the color stream, As each color is presented, a color wheel or an RGB LED array illuminates the display with only red, green or blue light. While less expensive, this requires high-speed display elements (with a frequency $\geq$ 540Hz) to process all three colors during a single frame time and avoid color breakup (an effect where false colors are briefly perceived when either the image or the observer's eye is in motion). [13]

3.5 Organic Light-Emitting Diode (OLED)

An OLED is any light-emitting diode (LED) whose emissive electroluminescent layer comprises a film of organic compounds. The layer usually contains a polymer substance that allows suitable organic compounds to be deposited. They are deposited in rows and columns onto a flat carrier by a simple "printing" process. The resulting matrix of pixels can emit light of different colors.

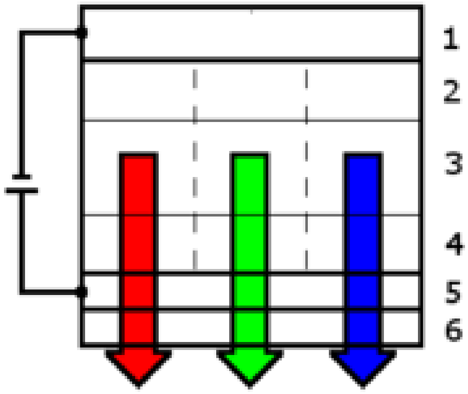

Figure 3: Buildup of a OLED with 3 RGB subpixels

An OLED consists of several organic layers:

Mostly a so called **Hole Transport Layer (HTL, 4)** is placed upon the **Anode (5, e.g. indium tin oxide, ITO)** which sits on top of a **Glass Plate (6)**. Depending on the production process, a layer of PETO/PSS (Poly-3,4-ethylenedioxythiophene polystyrene-sulfonate, a mixture of two ionomeres carrying negative respectively positive charging and forming a macromolecular salt when charged) is applied and used as a conductive polymer. On top of the HTL a **Emitter Layer (EL, 3)** is located which either contains or (rarely) completely consists of colouring. On top of this there is the **Electron Transport Layer (ETL, 2)** and finally the **Cathode (1)** consisting of a metal or alloy with low electron work function. [14] To emitt light, the OLED is charged electrically, leaving electrones in the emissive layer and "holes" in the conductive layer. When these are recombined by electrostatic forces, the energy levels of electrons drop, accompanied by an emission of light.

This happens closer to the emissive layer, because in organic semiconductors holes are more mobile than electrons, (unlike in inorganic semiconductors). The recombination causes a drop in the energy levels of electrons, accompanied by an emission of light.

There are several subtypes of the OLED technology:

PLED Polymer light-emitting diodes involve an electroluminescent conductive polymer that emits light when subjected to an electric current. Developed by Cambridge Display Technology, they are also known as Light-Emitting Polymers (LEP). They are used as a thin film for full-spectrum color displays and require a relatively small amount of power for the light produced. No vacuum is required, and the emissive materials can be applied on the substrate by a technique derived from commercial inkjet printing. The substrate used can be flexible, such as PET. Thus, flexible PLED Displays may be produced inexpensively.

[14]

TOLED Transparent organic light-emitting device uses a proprietary transparent contact to create displays that can be made to be top-only emitting, bottom-only emitting, or both top and bottom emitting (transparent). TOLEDs can greatly improve contrast, making it much easier to view displays in bright sunlight. [14]

SOLED Stacked OLED uses a novel pixel architecture that is based on stacking the red, green, and blue subpixels on top of one another instead of next to one another as is commonly done in CRTs and LCDs. This improves display resolution up to threefold and enhances full-color quality. [14]

3.6 Virtual Retinal Display (VRD)

A VRD, also called Retinal Scanning Display (RSD), consists of the following elements:

Drive electronics – The drive electronics control the acousto-optic modulators that encode the image data into the pulse stream. The color combiner multiplexes the individually-modulated red, green, and blue beams to produce a serial stream of pixels, which is launched into a singlemode optical fiber to propagate to the scanner assembly. The drive electronics receive and process an incoming video signal, provide image compensation, and control image display. [6]

Light source – The light source module contains laser light sources, acousto-optic modulators to create the pulse stream, and a color combiner that multiplexes the pulse streams. To provide sufficient brightness, full-color displays suitable for outdoor, daylight applications incorporate diode, solid-state or gas lasers while systems designed for indoor use can incorporate LEDs. [6]

Scanner assembly – The scanner assembly contains two scanning mirrors. One scanning mirror sweeps the beam horizontally at a high frequency which corresponds to one-half

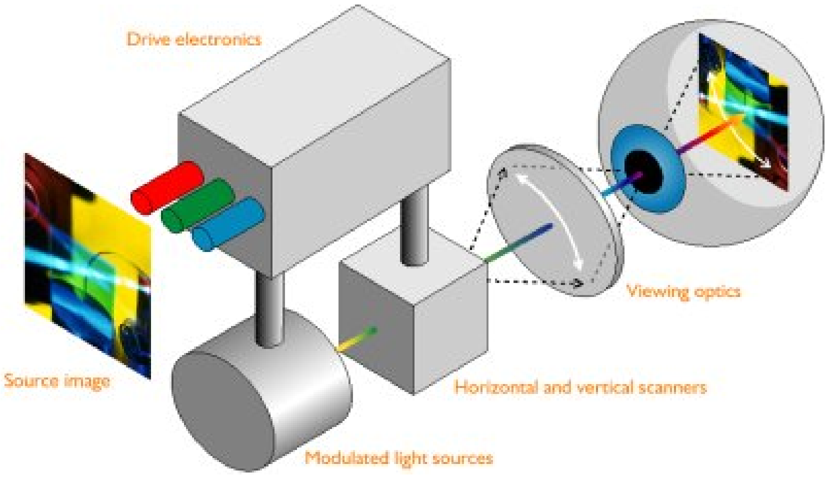

Figure 4: Setup of a Virtual Retinal Display

the VESA monitor-timing standard since the retinal scanning display can process and display pixels bidirectionally. A second scanning mirror sweeps the beam of laser light vertically to complete the raster image. [6]

Pupil expander – Nominally the entire image would be contained in an area of 2mm^2. The exit-pupil expander (not shown in Fig. 4) is an optical device that increases the natural output angle of the image and enlarges it up to 18mm on a side for ease of viewing. The raster image created by the horizontal and vertical scanners passes through the pupil expander and on to the viewer optics. [6]

Viewer optics – The viewer optics relay the scanned raster image to the oculars worn by the user. The optical system varies according to the application. In the case of military applications such as helmet mounted or head mounted display optics, the system incorporates glass and or plastic components; for medical applications such as image-guided surgery, head-mounted plastic optics are used. In industrial or personal displays, the optics might be a simple plastic lens. [6]

VRD devices produce very high resolution images which is only limited by diffraction and optical aberrations in the light source. Due to the brightness of a VRD being able to be adjusted from very high to very dim, the device's contrast ratio is inherently high and in a see-through mode, the VRD can be controlled to allow the user to see an image that matches the brightness levels of the ambient light conditions.

Also its power consumption behaviour is much better than those of CRTs or flat display technologies, which by far cannot reach the scalability of the VRD and its capability to be mass-produced at low cost, because it consists of subsystems that are very simple in their design and largely make use of established optical and electronic technologies.

4 Paradigms in Display Design

When designing displays for AR applications, engineers have to choose from opposed paradigms concerning display method and buildup of the device. The choice of following one paradigm sometimes determines another choice as well, for example building a video display that is see-through doesn't make much sense.

Therefore every design pattern has to fit the exact requirements of the domain the device is going to be used in. In the following, the most prominent choices to make are discussed, in each case taking advantages and disadvantages of both sides into consideration.

4.1 Optical vs. Video

A basic design decision in building an AR system is how to accomplish the combining of real and virtual. Two basic choices are available: optical and video technologies.

Optical see-through HMDs work by placing optical combiners in front of the user's eyes. These combiners are partially transmissive, so that the user can look directly through them to see the real world. The combiners are also partially reflective, so that the user sees virtual images bounced off the combiners from headmounted monitors. [1, p. 10]

Video see-through HMDs work by combining a closed-view HMD with one or two head-mounted video cameras. The video cameras provide the user's view of the real world. Video from these cameras is combined with the graphic images created by the scene generator, blending the real and virtual. The result is sent to the monitors in front of the user's eyes in the closed-view HMD. [1, p. 11]

The following 10 categories are meant to identify particular advantages and disadvantages of both technologies in relevant matters of AR display design.

1. **Simplicity:** Optical blending is simpler and cheaper than video blending. Optical approaches only have to deal with the graphic images while video blending must deal with separate video streams for the real and virtual images. Since video requires cameras and combiners that optical approaches do not need, video will probably be more expensive and complicated to build than optical-based systems. [1, pp. 13-14]

2. **Distortion:** Optical see-through HMDs with narrow field of view combiners offer views of the real world that have little distortion. Video cameras almost always have some amount of distortion that must be compensated for, along with any distortion from the optics in front of the display devices. [1, p. 14]

3. **Resolution:** Video blending limits the resolution of what the user sees, both real and virtual, to the resolution of the display devices. Optical see-through also shows the graphic images at the resolution of the display device, but the user's view of the real world is not degraded. [1, p. 14]

4. **Safety:** Video see-through HMDs are essentially modified closed-view HMDs. If the power is cut off, the user is effectively blind which is a severe safety concern in some

applications. In contrast, when power is removed from an optical seethrough HMD, the user still has a direct view of the real world. [1, p. 14]

5. **Flexibility:** Other than with optical see-through devices, where the virtual objects do not completely obscure the real world objects, because the optical combiners allow light from both virtual and real sources. Video see-through compositors can take the real or the virtual images as fully opaque, or some blend between the two to simulate transparency.

   That is because optical see-through HMDs cannot block out light in specific, dynamically changing areas and thus cannot reach occlusion of the real world behind virtual objects. Also an optical system would have to have two places where the image is in focus: at the user's eye and the point of the hypothetical filter that would selectively block out light. Both is not the case with video composing, since both the real and virtual are here available in digital form. [1, pp. 14-15]

6. **Field of view:** Distortions in optical systems are a function of the radial distance away from the optical axis. The further one looks away from the center of the view, the larger the distortions get. Any distortions of the user's view of the real world must be corrected optically, rather than digitally, because the system has no digitized image of the real world to manipulate.

   On the other side, a digitized image taken through a distorted optical system can be undistorted by applying image processing technique to unwarp the image, provided that the optical distortion is well characterized. This requires significant amounts of computation, but this constraint will be less important in the future as computers become faster. [1, p. 15]

7. **Delay:** Video offers an approach for reducing or avoiding problems caused by temporal mismatches between the real and virtual images. Optical see-through HMDs offer an almost instantaneous view of the real world but a delayed view of the virtual. This temporal mismatch can cause problems. With video approaches, it is possible to delay the video of the real world to match the delay from the virtual image stream. [1, p. 15]

8. **Offset:** With video see-through, the user's view of the real world is provided by the video cameras. In essence, this puts the user's "eyes" where the video cameras are. In most configurations, the cameras are not located exactly where the user's eyes are, creating an offset between the cameras and the real eyes. The distance separating the cameras may also not be exactly the same as the user's interpupillary distance (IPD). This difference between camera locations and eye locations introduces displacements from what the user sees compared to what he expects to see.

   Offset is generally not a difficult design problem for optical see-through displays. While the user's eye can rotate with respect to the position of the HMD, the resulting errors are tiny. Using the eye's center of rotation as the viewpoint in the computer graphics model should eliminate any need for eye tracking in an optical see-through HMD. [1, p. 14]

9. **Brightness:** As said in "5. Flexibility", a video HMD has the advantage of being in charge of both the real and the virtual objects' images. Thus the brightness of both can be aligned exactly, creating a better illusion. Optical systems, however, can make

use of both the ambient light and the display illumination, while video approaches only rely on an artificial light source. [1, pp. 15-16]

10. **Registration:** In optical see-through, the only information the system has about the user's head location comes from the head tracker. Video blending provides another source of information: the digitized image of the real scene. This digitized image means that video approaches can employ additional registration strategies unavailable to optical approaches. [1, p. 15]

4.2 Immersive vs. See-Through

HMD designs may be classified as immersive or see-through. While immersive optics refer to designs that block the direct real-world view, see-through optics refer to designs that allow augmentation of synthetic images onto the real world. Whether immersive or see-through, the optical path may or may not be folded.

Ideally, immersive HMDs target to match the image characteristics of the human visual system. Because it is extremely challenging to design immersive displays to match both the FOV and the visual acuity of human eyes, tradeoffs are made often.

See-through designs more often follow a folded design, particularly optical see-through displays. In such displays, the optical combiner is a key component in distinguishing designs. In folded designs, the center of mass can be moved back more easily. Folded designs however, often indicate optical system complexity.

A large majority of folded designs use a dual combiner, where reflections off a flat plate and a spherical mirror combined are used. The use of a toroid combiner, however, serves to minimize the large amount of astigmatism introduced when tilting a spherical mirror. [2, p. 3]

Balancing Field of View and Resolution:

Three main approaches have been investigated to increase FOV while maintaining high resolution: high-resolution insets, partial binocular overlap and tiling.

4.3 Non-Pupil vs. Pupil-Forming

Three current basic forms of optical design for HMDs are eyepiece, objective-eyepiece combination, and projection optics.

Only the simple eyepiece design is non-pupil forming, because it requires no intermediary image surface conjugate to the microdisplay within the optics. In this case, the eyes pupils serve as the pupil of the HMD.An eyepiece, in addition to creating a virtual image for the human visual system, forms an exit pupil by imaging the pupil of the system prior to it to the image space.

In an objective-eyepiece combination, a objective lens or mirror is placed at the focal point of the objective to magnify the image created by bringing the collected light to focus. The amount of magnification depends on the focal length of the eyepiece. [7]

Fig. 5 shows the difference between projecting a real image (top) and viewing the real image with the eye through an eyepiece (bottom). The idea of image projection is to focus the real image (called the object in the projection system) into another real image (called

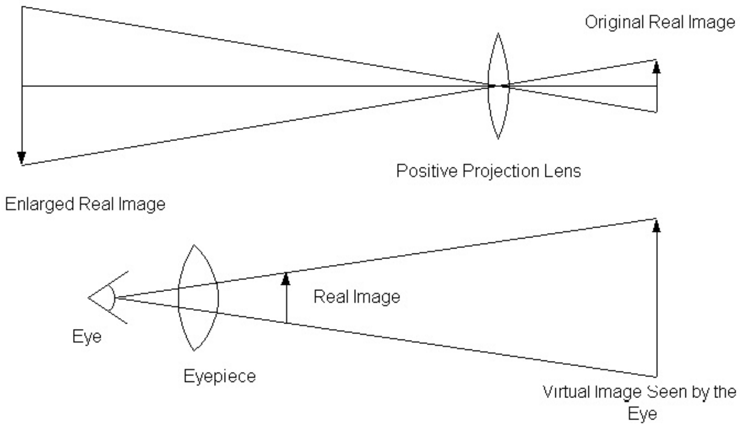

Figure 5: Difference between projection optics and eyepiece

the image) which is a significantly enlarged version of the object. In this case, the object lies outside the focal length of the projection lens by a small amount and the projected image is outside the focal length by quite a bit so it is substantially larger than the original.

For each eye of a user, as long as a possible light path exists between any point on the microdisplay and the eye, the user will see the virtual image of that point. An advantage of non-pupil forming systems is the large eye-location volume provided behind the optics. Their main disadvantage is the difficulty in folding the optical path with a beam splitter or a prism without making a significant trade-off in field-of-view. Unfolded optics prohibits see-through capability and balancing the weight of the optics around the head.

Pupil forming systems on the other hand consist of optics with an internal aperture which is typically conjugated to the eye pupils. A mismatch in conjugates will cause part or the entire virtual image to disappear, and therefore large enough pupils must be designed. The requirements for pupil size should be tightly coupled with the overall weight, ergonomics of the system, field of view, and optomechanical design. [2, pp. 2-3]

Telecentricity Requirement

Whether in object or image space, telecentric optics operates with a pupil at optical infinity in that space. In the telecentric space, the chief rays (i.e. the rays from any point on the microdisplay that pass through the center of the pupil) are parallel to the optical axis. Telecentricity in microdisplay space is desirable to maximize uniform illumination across the visual field, however it is not necessarily true because many microdisplays exhibit asymmetry off-axis. Telecentricity also further imposes that the lens aperture be at least the same size as the microdisplay, which has to be balanced against the weight constraint. A relaxed telecentric condition is often successfully applied in HMD design. [2, p. 3]

5 Difficulties in Display Design

Despite the variety of types and forms of displays usable in AR, every application bears the same design challenges. Therefore, it has to be carefully chosen which display technology is going to be used as this choice has major impact on contrast, brightness and other visual qualities of the final display. Other qualities like focus are determined by design and optics used. Of course, also proper tracking, sensing and registration capabilities must be considered when designing an AR display device.

5.1 Focus

Focus means displaying the virtual objects in the right distance on the one hand, and displaying them in a way that the two lines of sight used to reconstruct the 3-D image have the right angle in respect of real environment on the other.

It can be an issue in both optical and video approaches. Ideally, the focus of virtual objects should match the real. In a video-based system, the combined virtual and real image will be projected at the same distance by the monitor or HMD optics. However, depending on the video camera's depth-of-field and focus settings, parts of the real world may not be in focus. Usually the graphics are rendered with a pinhole model with the result that all objects are in focus. To overcome this, the graphics could be rendered to simulate a limited depth-of-field.

In optical devices the virtual objects are projected in a specific distance away from the user, which may be adjustable, but is often fixed to a reasonable value. So while the distance of real objects to the user is dynamic, all virtual objects are projected in the same distance. If the virtual and real distances are not matched for the particular objects that the user is looking at, it may not be possible to clearly view both simultaneously. [1, p. 16]

5.2 Contrast

Another issue due to its large dynamic range in real environments is contrast. In the range of what the human eye can distinguish, the brightness of the real and virtual objects should be appropriately matched. The eye is a logarithmic detector, where the brightest light that it can handle is about eleven orders of magnitude greater than the smallest, including both dark-adapted and light-adapted eyes. In any one adaptation state, the eye can cover about six orders of magnitude.

A display devices would now be able to match a very large range of brightness levels, which is mostly now even roughly the case. This is a particular problem with optical technologies, because the user has a direct view of the real world and the illusion of the virtual objects being a part of it is diminished.

With video, however, contrast problem are not that severe, because the video cameras themselves have limited dynamic response and everything displayed to the user (real and virtual) must be clipped or compressed into the display's dynamic range. [1, p. 16]

5.3 Brightness

Alternatives to microdisplays like LCDs are laser or laser-diode based scanning displays like the VRD which optically conjugates the pupil of the eyes to the microscanner exit pupil.

Many devices have used a projection device, a screen or an eyepiece magnifier to expand the viewing volume, because the challenge of reaching a small exit pupil (i.e. 1 3mm) within which the eye needed to be located to see the image could be overcome by forming an intermediary image followed by a pupil expander. Controlled angle diffusers have been designed for pupil expansion in HMDs, including diffractive exit-pupil expanders.

Given an intermediary image, the VRD also functions with an equivalent microdisplay in this case formed using scanned laser light. Therefore the VRD can match with other HMD technologies.

A recently developed approach - the optical CRT.15 - uses a single infrared laser diode scanned across a polymer thin plate doped with microcrystals. Optical upconversion is used to have the microcrystal emit light in the red, green, and blue regions of the spectrum while taking the advantage of the laser diode to reduce speckle noise. [2, pp. 3-4]

5.4 Tracking and Registration

Tracking and registration are enabling technologies essential to AR. While tracking is understood as "keeping track" of the user's line of vision, orientation and the position of all modifiable objects in the environment, registration can be understood as alignment or calibration.

One of the most basic problems currently limiting Augmented Reality applications is the registration problem. The objects in the real and virtual worlds must be properly aligned with respect to each other, or the illusion that the two worlds coexist will be compromised. More seriously, many applications demand accurate registration.

Static Errors are the ones that cause registration errors even when the user's viewpoint and the objects in the environment remain completely still. The four main sources of static errors are: optical distortion, errors in the tracking system, mechanical misalignments and incorrect viewing parameters such as center of projection, viewport dimensions, offset (both in translation and orientation) and field of view. [1, pp. 19-22]

Dynamic Errors occur because of system delays (lags) and have no effect until either the viewpoint or the objects begin to move. The time difference between the moment that the tracking system measures the position and orientation of the viewpoint to the moment when the generated images corresponding to that position and orientation appear in the displays is called *end-to-end delay* and values of 100ms are fairly typical on existing systems. The following categories of methods are used to reduce dynamic registration errors: reduce system lag, reduce apparent lag, match temporal streams and predict future locations. [1, pp. 22-25]

Tracker And Sensor Requirements

Registration errors are difficult to adequately control because of the high accuracy requirements and the numerous sources of error. These sources of error can be divided into two types: static and dynamic, which will both be discussed in the upcoming sections. For current HMD-based systems, dynamic errors are by far the largest contributors to registration errors, but static errors cannot be ignored either.

Many systems assume a static viewpoint, static objects, or even both. Even if the viewpoint or objects are allowed to move, they are often restricted in how far they can travel. Registration is shown under controlled circumstances, often with only a small number of real-world objects, or where the objects are already well-known to the system. Ronald Azuma developed a system that shows registration typically within ±5 millimeters from many viewpoints for an object at about arm's length. Closed-loop systems, however, have demonstrated nearly perfect registration, accurate to within a pixel. [1, p. 29]

A possible solution: Vision-Based Systems

Since video-based AR systems have a digitized image of the real environment, it may be possible to detect features in the environment and use those to enforce registration. This is called a "closed-loop" approach, since the digitized image provides a mechanism for bringing feedback into the system, and is not a trivial task. Both detection and matching must run in real time and must be robust. This often requires special hardware and sensors. [1, pp. 25-28]

6 Conclusion

Admittedly, both AR and VE bring along related requirements and thus resembling problems. But in AR the problem domains are slightly altered and hence also the requirements. In some cases (image quality, environmental data, scene generating performance) requirements of AR applications are less stringent, elsewhere (optics, visual complexity, tracking and registration) AR requires by far more from employed technologies than VE.

The reason can be found in the different aims of AR and VE: While VE completely replaces the natural environment and therefore chiefly has to deal with affinity to reality, the intetion of AR lies in leaving the user in his real environment but augmenting it with virtual objects using all kinds of data.

This data can be drawn in real-time from a variety of sensors, most of which serve the purpose of tracking the user's movement and orientation and register his position in the environment. Here AR bears extremely high accuracy and processing performance, because otherwise registration errors, misalignments and lags occur which can render the whole application useless.

These primary problems can be solved better and better by following hybrid approaches, which combine several technologies (e.g. vision-based and prediction) to cover weaknesses and make applications more robust.

Same applies to display systems used in AR: various and constantly improving display technologies are available and engineers can choose out of quite a number of different designs. Each possesses individual advantages in terms of simplicity, image quality, safety, flexibility and scalability.

Here an optimal setup can be achieved by combining methods that turned out to be best for the desired application domain. All components have to be aligned to each other and work seamlessly and synchronously to sustain the user's illusion of the virtual objects being part of his real environment. If this goal is not reached – be it for technical restrictions or design lapses – the application does not provide a realistic experience but confuses the user, sometimes even affecting his ease and health.

When the capabilities of AR are discovered by industrial precursors, as it is more and more the case, research funds that are more than adequate to solve AR's current troubles could be raised. AR has already made its way into several scopes of everyday life: it supports doctors, researchers, engineers, designers, workers, drivers, pilots, gamers and so forth.

As more and more people discover the great potential of AR when it comes to assist miscellaneous tasks, a broad market for AR applications will open up, encouraging improvement of already established technologies and development of further ones.

Future progression can only be estimated and influenced by rational studies of both the technical resources and their human users. But because AR is a quite young and emerging field of science fraught with potency, there cannot be a punch line on where it is heading.

References

[1] Ronald Azuma. A Survey of Augmented Reality. In Presence: Teleoperators and Virtual Environments 6, 4 (August 1997), 355-385.

[2] Jannick Rolland and Ozan Cakmakci. The Past, Present, and Future of Head Mounted Display Designs.

[3] Spitzer M.B. Eyeglass-Based Systems For Wearable Computing. In Proc. First International Symposium on Wearable Computers (ISWC 1997), 13-14 October 1997, Cambridge, Massachusetts, USA. IEEE Computer Society. ISBN 0-8186-8192-6.

[4] http://www.sciam.com/article.cfm?articleID=0006378C-CDE1-1CC6-B4A8809EC588EEDF *(June 17th 2007)*

[5] Erik Geelhoed, Marie Falahee, and Kezzy Latham. Safety and Comfort of Eyeglass Displays. In: Handheld and Ubiquitous Computing: Second International Symposium, HUC 2000, Bristol, UK, September 2000. Proceedings

[6] http://www.cs.nps.navy.mil/people/faculty/capps/4473/projects/fiambolis/vrd/vrd_full.html *(June 22nd 2007)*

[7] J. E. Melzer & K. Moffitt (editors). Head Mounted Displays: Designing for the User. New York, Publisher: McGraw-Hill, 1997.

[8] http://en.wikipedia.org/wiki/Augmented_reality *(June 17th 2007)*

[9] http://en.wikipedia.org/wiki/Head-up_display *(June 17th 2007)*

[10] http://en.wikipedia.org/wiki/Head-mounted_display *(June 17th 2007)*

[11] http://en.wikipedia.org/wiki/Lcd *(June 18th 2007)*

[12] http://en.wikipedia.org/wiki/Digital_Light_Processing *(June 22nd 2007)*

[13] http://en.wikipedia.org/wiki/Lcos *(June 18th 2007)*

[14] http://en.wikipedia.org/wiki/Oled *(June 22nd 2007)*