

AN INTRODUCTION TO AUGMENTED REALITY

ALEX OLWAL

www.olwal.com



*Adapted from: Olwal, A. Unobtrusive Augmentation of Physical Environments: Interaction Techniques, Spatial Displays and Ubiquitous Sensing. Doctoral Thesis, KTH, Department of Numerical Analysis and Computer Science, Trita-CSC-A, ISSN 1653-5723; 2009:09.
<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-10439>*

1 Introduction

1.1 Augmented and Mixed Reality

The goal of Augmented Reality (AR) is to improve and enhance our perception of the surroundings by combining sensing, computing and display technologies.

Most AR research addresses human vision, as it is generally considered to be our most important sense. Visual systems are also the focus in this overview, but it is worth noting that other stimuli, such as feedback from auditory, tactile or olfactory displays, may be equally or even more important, depending on the specific scenario and individual.

The characteristics of these systems can be further understood from three classical and widely used criteria for AR systems [Azuma 1997]:

- 1 “Combines virtual and real”
AR requires display technology that allows the user to simultaneously see virtual and real information in a combined view. Traditional displays can show only computer-generated images and are thus insufficient for AR.
- 2 “Registered in 3-D”
AR relies on an intimate coupling between the virtual and the real that is based on their geometrical relationship. This makes it possible to render the virtual content with the right placement and 3D perspective with respect to the real.
- 3 “Interactive in real time”
The AR system must run at interactive frame rates, such that it can superimpose information in real-time and allow user interaction.



Figure 1. The Sonic Flashlight uses a see-through display to overlay real-time ultrasound images over a patient's body parts. (Images courtesy of George Stetten [Stetten et al. 2001].)

The fundamental idea of AR is to combine, or mix, the view of the real environment with additional, virtual content that is presented through computer graphics. Its convincing effect is achieved by ensuring that the virtual content is aligned and registered with the real objects. As a person moves in an environment and their perspective view of real objects changes, the virtual content should also be presented from the same perspective.

The Reality-Virtuality Continuum [Milgram and Kishino 1994] spans the space between *reality*, where everything is physical, and *virtual reality*, where virtual and synthesized computer graphics replace the physical surroundings. *Mixed reality* is located between them, and includes AR and augmented virtuality.

AR adds *virtual content* to a predominantly *real environment*, whereas augmented virtuality adds *real content* to a predominantly *virtual environment*. Although both AR and augmented virtuality are subsets of mixed reality by definition, most of the research in the area focuses on AR, and this term is therefore often used interchangeably with mixed reality.

AR techniques exploit the spatial relationships between the user, the digital information, and the real environment, to enable intuitive and interactive presentations of data. An AR system can, for example, achieve medical see-through vision, by using a special display in which the images displayed by the computer are seen overlaid on the patient [State et al. 1996, Stetten et al. 2001], as shown in Figure 1 and Figure 2.

Such configurations rely on the proper acquisition and registration of internal medical imagery for the relevant perspective, and careful calibration to establish the geometrical relationship between the display, the viewer, and the patient, to ensure that the correct image is accurately presented.

1.2 Research challenges

The user experience for an AR system is primarily affected by the display type, the system's sensing capabilities, and the means for interaction. The display and sensing techniques determine the effectiveness and realism possible in the blending of the two realities, but may at the same time have ergonomic and social consequences.

It may, in particular, be desirable to achieve *walk-up-and-use* scenarios that support spontaneous interaction with minimal user preparation [Encarnacao et al. 2000]. Unencumbering technology can also be emphasized, avoiding setups that rely on user-worn equipment [Kaiser et al. 2003, Olwal et al. 2003], such as head-worn displays [Cakmakci and Rolland 2006] or motion sensors [Welch and Foxlin 2002]. It can also be useful to, to the greatest extent possible, preserve the qualities of the real space, while augmenting and assisting the user with unmediated view and control. The excessive use of artificial elements, such as visual reference patterns used for tracking, may, for example, have negative side-effects by cluttering or occluding the real environment that the system is meant to augment. Some display technologies may also result in significantly reduced visual quality due to optical properties, or the use of a downsampled view of the real environment.



Figure 2. The visualization of a 3D ultrasound scan is registered with a pregnant woman's body, to allow the physician to "look into the body" for a view of the fetus. (Images courtesy of Henry Fuchs and Department of Computer Science, UNC-Chapel Hill [State et al. 1996].)

An *unobtrusive AR system* [Olwal 2009] emphasizes unencumbering techniques and avoids changes to the appearance of the physical environment. An unobtrusive AR system must address issues in three areas:

- 1 Display systems
The system should merge the real and virtual, while preserving a clear and direct view of the real, physical environment, and avoiding visual modifications to it.
- 2 Sensing and registration
The system should present perspective-correct imagery without user-worn equipment or sensors.
- 3 Interaction techniques
The system should support direct manipulation, while avoiding encumbering technologies.

2 Fundamental technologies

This chapter provides an overview of work in three areas that are fundamental to the development of unobtrusive AR technology: display systems, sensing and registration, and interaction techniques.

Display systems merge the view of the real and virtual, while *sensing and registration* makes it possible to render graphics from the right perspective. Direct manipulation and user interface control are enabled through *interaction techniques*.

2.1 Display systems

Combining graphics and the real world

A fundamental characteristic of AR systems is that they allow the user to see a combined view of virtual imagery and real objects.

The display hardware used in these systems can be head-worn (retinal displays, miniature displays, and projectors), handheld (displays and projectors) or spatial (displays or projectors in the environment) [Bimber and Raskar 2005].

The following sections focus on display technology for handheld and spatial AR systems. The first three sections discuss classical AR display technologies in this context, where optical see-through, video see-through and direct projection display systems make it possible to visually merge the real and the virtual. The last section discusses spatially aware handheld displays, which use a tracked display to provide a virtual view of data associated with the real environment. We are particularly interested in the four approaches described in these sections, since they can be used in configurations that avoid encumbering technology and visual modifications to the environment.

2.1.1 Optical see-through displays

Optical see-through capabilities are achieved by using an optical combiner, such as a half-silvered mirror or a holographic material.

The role of the combiner is to provide an *optically direct* view of the environment, with a simultaneous presentation of computer-generated imagery. The combiner is typically able to transmit light from the environment, while also reflecting light from a computer display. The combined light reaches the user's eyes. (See Figure 3.)

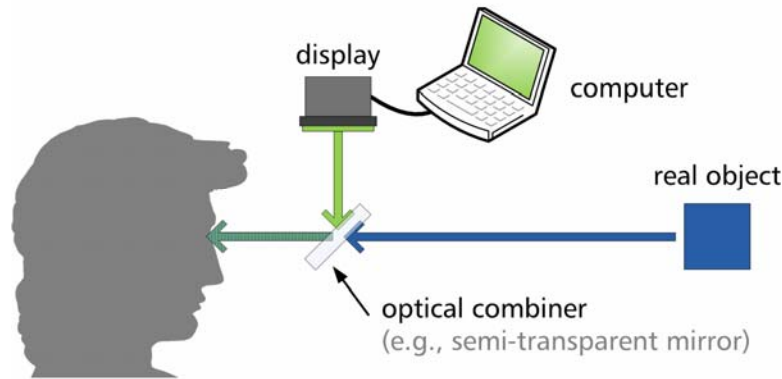


Figure 3. The optical path in an optical see-through display system. The light from the real environment passes through a transparent combiner that simultaneously reflects computer-generated images from a display. The *optically combined* light reaches the user's eyes. (Other optical elements, such as lenses that control the focal distance, are not shown in this illustration.)

Spatial optical see-through systems use situated transparent displays that are integrated with the environment [Bimber and Raskar 2005]. They free the user from worn equipment, but require the display to not only be calibrated with the environment, but also registered with the user to ensure perspective-correct imagery. (See Figure 4.)

Advantages

- + Direct view of the real physical environment

The main advantage of optical see-through techniques is the directness with which the real environment is viewed. Factors that limit the quality of computer-generated imagery, such as resolution, image quality or system responsiveness, do not affect the view of the real environment. This property may be critical for many applications where a small delay or reduced visibility may not only be distracting, but could even be hazardous [Navab 2003].

Disadvantages

- Reduced brightness

The view of the real environment will suffer a decrease in brightness, depending on the type of optical combiner used.



Figure 4. The Virtual Showcase augments physical objects inside showcases with dynamic annotations or related media. The Virtual Showcase is covered with half-silvered mirrors that optically combine the view of the objects with computer graphics reflected from the rear-projected table, on which the showcase is placed. (Images courtesy of Oliver Bimber [Bimber et al. 2003, Bimber et al. 2001].)

- Lack of occlusion
The blending of the two light sources means that the computer graphics can become transparent, making it difficult to achieve occlusion effects in many optical see-through systems. Occlusion may however be solved through special hardware that can selectively block the user's view of the physical world [Cakmakci et al. 2004, Kiyokawa et al. 2003, Kurz et al. 2008].
- Need for advanced calibration and/or tracking
Precise calibration and/or tracking are necessary to ensure that the virtual content is correctly registered with the direct view of the real environment.
- Multiple focal planes
The virtual graphics are typically presented in a fixed image plane, unlike the real environment, which has objects at varying distances. The larger the distance between the image plane and a real object, the harder it is to properly accommodate and comfortably perceive a combined view. The user could either shift focus between the two, or choose to have either the virtual imagery or the real objects out of focus. Head-up displays [Wood and Howells 2006], which are often used in aircraft, employ a technique where the virtual image plane is placed at the focal point of the optical system to achieve infinite focus. Recent work has demonstrated the use of liquid lenses to achieve variable focal planes in head-worn displays [Liu et al. 2008].

2.1.2 Video see-through displays

A popular AR technique is based on a camera that acquires the view of the environment, a computer that adds virtual content, and an ordinary video display that presents the combined view to the user. (See Figure 5.)

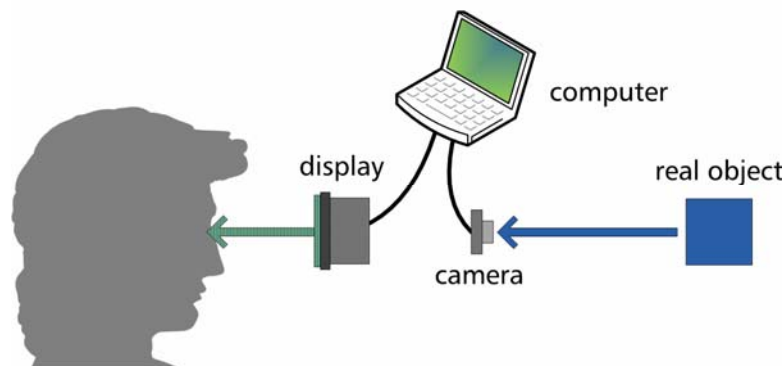


Figure 5. The optical path in a video see-through display system. The view of the real environment is acquired by a camera and is combined with virtual imagery by a computer. The combined video is presented to the user on a computer display. (Other optical elements, such as lenses that control the focal distance, are not shown in this illustration.)

Head-worn displays can use video see-through techniques by placing cameras close to the eye positions. Ideally, two cameras should be used to acquire a stereo view, with one perspective for each eye, but monoscopic single-camera systems are common and easier to design and implement [Takagi et al. 2000].



Figure 6. The NaviCam project illustrates how the camera on a handheld display can be used to recognize features in the environment, such that annotations can be overlaid onto the video feed. (Images courtesy of Jun Rekimoto [Rekimoto 1997, Rekimoto and Nagao 1995].)

Some video see-through displays use a camera to capture the scene, but present the combined view on a regular, typically handheld, computer display. A window-like effect, often referred to as a “magic lens,” is achieved if the camera is attached on the back of the display, creating the illusion of see-through [Rekimoto 1997, Rekimoto and Nagao 1995], as shown in Figure 6.

It is becoming increasingly popular to directly implement video see-through on mobile devices with built-in cameras, as illustrated in Figure 7. Camera-equipped mobile phones are particularly attractive devices for AR, given their widespread use, connectivity, portable form factor, and rapidly evolving processing and graphics capabilities (e.g., [Henrysson and Ollila 2004, Mohring et al. 2004, Takacs et al. 2008, Wagner et al. 2008a, Wagner et al. 2008b, Wagner and Schmalstieg 2003]).

Advantages

- + Controlled combination of real and virtual

Rather than optically combining the light from the real and virtual scenes, everything is merged in the computer, since the view of the real world is continuously captured by the camera. This provides greater compositional flexibility, allowing the virtual content to be rendered more realistically with respect to the real environment [Klein and Murray 2008]. The virtual content could, for example, occlude the real environment. The system may also synchronize the real and virtual content, to compensate for delays in the pipeline.
- + Integrated image-acquisition, calibration and tracking

The closed loop of the video see-through system makes it advantageous to use the camera to not only acquire the real scene, but also track features in the environment. Registering the camera’s pose relative to the environment greatly simplifies the introduction of perspective-correct virtual imagery in the combined image shown to



Figure 7. Video see-through AR can be achieved using commercial camera phones. The camera on the back of the device captures video of the real environment, which is used by software on the device to recover the phone’s pose relative to tracked features in the environment. This makes it possible to render 3D objects that are registered with the real environment, overlaid on the video that is shown on the phone’s display [Henrysson 2007].

the user [Kato and Billinghurst 1999]. Image-based techniques may also be exploited to improve registration of the virtual content with the video of the real scene [DiVerdi and Höllerer 2006].

Disadvantages

- Reduced quality and fidelity of the real environment
The major drawback of video see-through systems is their sampling of the real environment at the camera's video resolution, and their dependency on the quality of the image sensor.
The computational load associated with intensive image processing means that most video see-through systems today handle less than 1 megapixel video resolution at interactive frame rates (often $640 \times 480 \approx 0.3$ megapixels, at 30 frames per second). The quality can of course be improved with better hardware, but this requires additional computational power to avoid reduced performance, and will still result in a significant downsampling of the real environment.
The camera's image sensor not only has less resolution than the human eye, but also differs in sensitivity. This can, for example, affect a video see-through system's performance in low or bright lighting conditions and impact its ability to acquire and present scenes with high dynamic range.
The downsampled, indirect view of the camera plays an important role in the level of immersion and realism a user experiences.
- Potential perspective problems due to camera offset
Perceptual problems may arise if the cameras are not effectively placed so that their views correspond to those of the user's eyes [State et al. 2005, Takagi et al. 2000].
- Single focal plane
The combination of real and virtual content into a single image, with a single focal plane, eliminates the focal cues of real-world objects. While stereoscopic systems simulate the disparity between the left and right eye views, accommodation remains inconsistent because everything the user sees is focused at the same distance [Bimber and Raskar 2005]. The typical stereoscopic optical see-through system only suffers from this drawback for virtual content.
- Sensitivity to system delay
The video see-through imaging pipeline may introduce system delays, which may be unacceptable, or even dangerous, for some applications. In contrast, optical see-through systems always provide a direct view of the real environment.
- Dependency on camera operation
Video see-through systems rely heavily on the camera's ability to acquire the real scene, and their visual pathway will fail if the cameras, for example, are blocked, overexposed, or lose power.

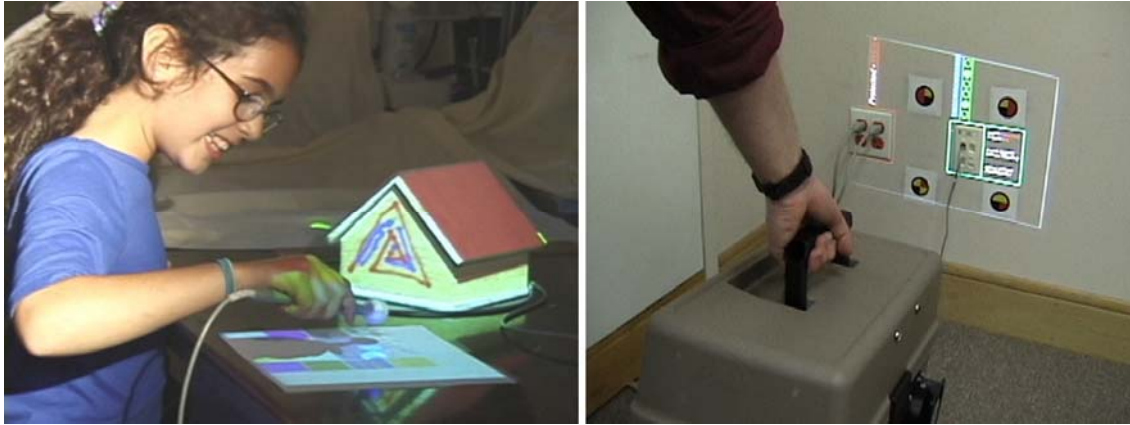


Figure 8. Left: A child uses a tracked brush to apply virtual paint, which is projected onto physical objects. (Image courtesy of Ramesh Raskar [Bandyopadhyay et al. 2001].)

Right: A handheld projector is combined with a camera that identifies elements of interest in the environment and augments them with projected light. In this example, a network socket is augmented with visualizations of network status and traffic. (Image courtesy of Ramesh Raskar [Raskar et al. 2003].)

2.1.3 Direct projection

Augmentation can also be achieved by directly projecting graphics onto the real environment. Figure 8 and Figure 9 show examples of how the real world can be modified through controlled light that alters its appearance [Bandyopadhyay et al. 2001, Pinhanez et al. 2003, Pinhanez and Pingali 2004, Pinhanez and Podlaseck 2005, Pinhanez 2001, Raskar et al. 2004, Raskar et al. 2003, Zaeh and Vogl 2006].

Advantages

- + Direct integration of the virtual with the real
Direct projection eliminates the need for an optical or video-based combiner, since the light is directly added to the real environment. The integration of real with virtual material takes place on the surfaces of the real environment and their combination is naturally perceived by the user.

Disadvantages

- Dependence on environmental conditions
Direct projection depends on the availability of suitable projection surfaces, and on their geometry and appearance. It also relies on non-conflicting environmental lighting and the absence of objects in the optical path from the projector to the



Figure 9. The Everywhere Displays project uses “steerable displays”, where the system’s projector can create augmentations on different surfaces in the environment, while the camera senses the user’s interaction. (Images courtesy of Claudio Pinhanez [Pinhanez et al. 2003, Pinhanez and Pingali 2004, Pinhanez and Podlaseck 2005, Pinhanez 2001].)

surface, to avoid occlusions and shadows.

- Dependence on projector properties

Parts of a projected image may end up distorted or out of focus if the projection surface is not perpendicular to the direction of projection. Focusing issues can also occur if the surface is too close or too far away from the projector. In some cases, these issues can be corrected in software [Bimber and Emmerling 2006, Bimber et al. 2005, Grossberg et al. 2004, Raskar et al. 1999a, Raskar et al. 1998, Raskar et al. 1999b, Wetzstein and Bimber 2007] or through the use of special hardware, such as short-throw or laser projectors [Zach and Vogl 2006].

2.1.4 Spatially aware handheld displays

An alternative approach to the combination of computer graphics and the real environment is to use a tracked handheld display without optical or video see-through capabilities. This approach can provide useful context-sensitive information by leveraging proximity and relative position to real objects, despite not matching Azuma's criteria for AR systems (see Section 1.1), due to the lack of a see-through display, and less strict requirements for the registration between real and virtual.

Fitzmaurice presented an early prototype of how a tracked display could be moved over a physical paper map, causing dynamically updated information to be presented about the different map areas [Fitzmaurice 1993], as shown in Figure 10.

Advantages

- + Simplified tracking and calibration

Less accuracy is generally needed, since the real and virtual are not visually registered with each other. The tracking resolution is, of course, dependent on the requirements of the application and how densely the real objects are positioned.

- + Tangible interaction on the surface

Tracked displays may compensate for their lack of see-through by supporting placement and manipulation directly on surfaces [Fitzmaurice et al. 1995, Ishii and Ullmer 1997] where optical or video see-through capabilities may not always be useful. This applies especially to video-see through systems, since the camera would be either blocked or out-of-focus, when it is placed on, or very close to, the surface.

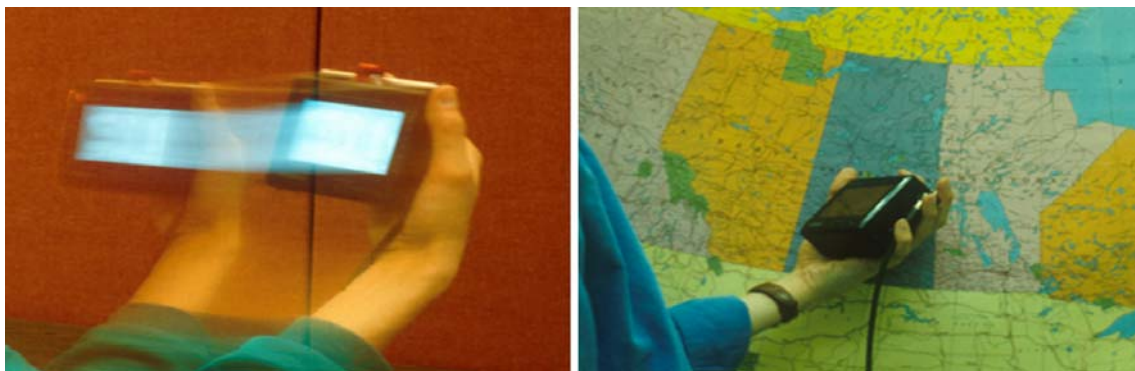


Figure 10. The Chameleon system illustrates how dynamically updated graphics in a tracked display can provide context-sensitive information about real objects underneath. (Images courtesy of George Fitzmaurice [Fitzmaurice 1993].)

Disadvantages

- Less realism
The absent visual combination of real and virtual affects the level of realism and can weaken the link between the real objects and their associated digital content.
- Display occludes the real objects
As the display is moved in the environment, since it is not see-through, it may occlude real elements of interest.

2.2 Sensing and registration

Perspective-correct imagery

The geometrical relationship between the user's viewpoint, the display, and the environment needs to be known for accurate and properly aligned combinations of the real and virtual.

There are many different technologies that can be used to recover the position and/or orientation of users, displays and objects in AR systems. Common techniques use ultrasonic, electromagnetic, optical, or mechanical sensing, and can be deployed in an environment to sense and recover the properties of beacons, fiducials, markers and other features, or serve as an infrastructural frame of reference for sensors in the space. *Exocentric* approaches rely on technology in the environment, while *egocentric* approaches are self-contained and independently perform the necessary sensing and computation [Welch and Foxlin 2002].

All approaches to AR typically share a need for the display to be registered with the environment. A static display, such as the Virtual Showcase [Bimber et al. 2003, Bimber et al. 2001], may require only initial calibration. In contrast, a movable display, such as NaviCam [Rekimoto 1997, Rekimoto and Nagao 1995], needs continuous tracking to recover its relative pose to the environment.

Camera-based techniques, as used in NaviCam, have become increasingly popular due to their cost-effectiveness and availability in commercial products (such as portable computers and mobile phones), where they can serve a dual purpose as both egocentric tracking device and video see-through combiner. Their use in AR was further facilitated through the development of ARToolKit [Kato and Billinghurst 1999], an open source framework for video see-through AR [ARToolKit 2009]. ARToolKit uses camera-based tracking to recover the position and orientation of the camera relative to special fiducial markers in the environment, such that perspective renderings of 3D graphics can be registered with the environment and overlaid on the live video stream. The visual patterns greatly simplify the tracking and reduce the required computational power. Their obtrusiveness can however be avoided thanks to rapid advances in techniques such as SLAM (simultaneous location and mapping) and natural feature tracking [Klein and Murray 2007, Neumann and You 1999, Reitmayr et al. 2007, Takacs et al. 2008, Wagner et al. 2008b]. (See Figure 11.)



Figure 11. Natural feature tracking techniques make video see-through AR possible without the need for fiducial markers in the environment. Here, a mobile device tracks features on a printed page and overlays registered 3D graphics at interactive frame rates. (Images courtesy of Daniel Wagner and Graz University of Technology [Wagner et al. 2008b].)

The recovery of the user's pose relative to the display may also be necessary for situations where the viewpoint is not fixed. This is primarily an issue for spatial optical see-through displays, where the relationship between the user, the rendered content on the display, and the real environment behind the display change, as the user moves.

The need for user tracking is, however, dependent on many factors, such as how and where the virtual imagery is created and the display's placement relative to the user.

Head-worn displays [Cakmakci and Rolland 2006], for example, are typically assumed to be rigidly positioned relative to the user's eyes, such that the user's viewpoint can be directly derived from the tracked position of the display. Video see-through systems present the combined view in the plane of the video display, updating it only when the relative positions of the camera and the environment change. The same view of the real environment is thus shown on the screen if the user moves, while the camera and display remain stationary. Dedicated user tracking is also not necessary for systems that eliminate dependencies on the user's perspective through direct augmentation of the real environment's surfaces with projected 2D graphics [Bandyopadhyay et al. 2001, Pinhanez et al. 2003, Pinhanez 2001, Raskar et al. 2004, Raskar et al. 2003, Zaeh and Vogl 2006]. The same principle applies to optical see-through displays that are used on the surfaces of real objects to present 2D overlays [Stetten et al. 2001].

Perspective-correct imagery may also be provided without user tracking, through the use of a *multi-view* display system (e.g., [Jones et al. 2007]). Such systems render multiple views, where the principle of parallax can limit their respective visibility to certain perspectives. The user will thus only see the appropriate view, when viewing the display from the right position and/or direction, which implicitly avoids the need for tracking. This approach may however result in an increased computational load for the simultaneous generation of multiple views, while also limiting the user to a discrete number of fixed viewpoints [Halle 1997].

Interactive tabletops and surfaces are closely related to AR technologies in their augmentation of the user's interaction with physical objects. Object sensing has been explored using a wide variety of exocentric techniques, where the main challenges lie in the detection, identification and tracking of unique artifacts. Most projects rely on either electronic tags or visual markers for remote identification and tracking [Patten et al. 2001, Reilly et al. 2006, Rekimoto and Saitoh 1999, Rekimoto et al. 2001, Want et al. 1999, Wellner 1993, Wilson 2005].

2.3 Interaction techniques

Direct and unencumbered manipulation

An AR system's interactive capabilities often play a significant role in complementing the display capabilities that help to augment real-world tasks. While the literature has many compelling examples of techniques for exploring interaction in AR scenarios, many of these have unintentional side effects that influence the user's *interaction with the real world*. Problems may be related to ergonomic factors, such as head-worn displays that limit visibility and peripheral vision, systems tethered with cables that constrain movement, or other wearable devices that cover parts of the user's body. Social aspects may also be important, as user-worn technology can be inappropriate in many real-world scenarios.

It may therefore be advantageous to emphasize interactive technologies that avoid such potential conflicts and minimize the negative effects an AR system may have on the user's normal interaction with the real world.

A number of different technologies that can enable interaction, while avoiding encumbering and user-worn equipment, are described in the following sections.



Figure 12. An acoustic tap tracker allows touch-screen interaction in public spaces by sensing the position of knocking actions on glass surfaces. (Images courtesy of Joseph Paradiso [Paradiso 2003, Paradiso et al. 2000].)

2.3.1 Touch

The direct-manipulative property of touch-sensitive surfaces is often viewed as natural, intuitive and easy to use. Finger- or stylus-based pointing, dragging, tapping and gesturing can be directly applied to graphical objects of interest, without the need for special-purpose devices that users may need to operate in other forms of human-computer interaction [Han 2005, Matsushita and Rekimoto 1997, Paradiso et al. 2000, Rekimoto et al. 1998, Selker 2008]. (See Figure 12.)

Numerous characteristics of the surface can affect the user's experience and performance in finger-based interaction. Improper calibration, parallax between the touch-sensitive layer and display element, low display resolution, low sensing resolution, and occlusion by the user's finger can prohibit accurate and precise interaction. Solid surfaces also typically lack the tactile feedback provided through texture, actuation, physical properties and mechanics of special-purpose input controls. This can be considered to be one of the most serious issues in interaction on touch devices, since technology that provides varying levels of passive or active tactile feedback (e.g., [Harrison and Hudson 2009, Poupyrev and Maruyama 2003, Poupyrev et al. 2002]) is still rare.

2.3.2 Gesture and pose

It may be beneficial to *remotely* sense the user's motion as a complement or alternative to direct touch [Wilson 2004], for example, if a system's particular hardware configuration is incompatible with touch-enabling technology. Public installations can depend on issues such as material characteristics of the window glass, or the requirement for robust technology that is protected from users. Touch-based interfaces may also be inappropriate, or even prohibitive, for certain user scenarios. These could include laboratory work or medical procedures, where the user's hands are occupied or wearing gloves, or the task requires aseptic operation.

The controlled, or semi-controlled, measurement of light is a popular remote sensing approach for interactive applications. The methods vary, for example, based on the type of



Figure 13. VIDEOPLACE was among the first systems to successfully demonstrate telepresence, gestural control, interactive computer graphics and other important concepts. (Images courtesy of Myron Krueger [Krueger 1977, Krueger et al. 1985].)



Figure 14. The ALIVE system tracks a user’s body parts and supports full body interaction with virtual creatures. (Images courtesy of Alex Pentland, Bruce Blumberg and Trevor Darrell [Maes et al. 1995].)

hardware used, the type of light that is sensed, and whether the scene is actively illuminated by the system.

Basic scene changes may, for example, be detected with an ordinary camera that captures visible light. Such remote sensing approaches can be used for tracking body parts, such as the head or hands, where the resulting detection of postures and gestures may be mapped to explicit interface control [Krueger 1977, Krueger et al. 1985, Maes et al. 1995], as shown in Figure 13 and Figure 14.

An eye tracker, on the other hand, can recover the 3D positions of the user’s eyes by illuminating the scene with multiple infrared light sources and capturing the reflecting light from the pupils on an image sensor that senses only infrared light [Tobii 2009]. TouchLight [Wilson 2004], shown in Figure 15, uses two infrared cameras and stereo computer vision to detect the position of the user’s hands near the surface of a transparent display.

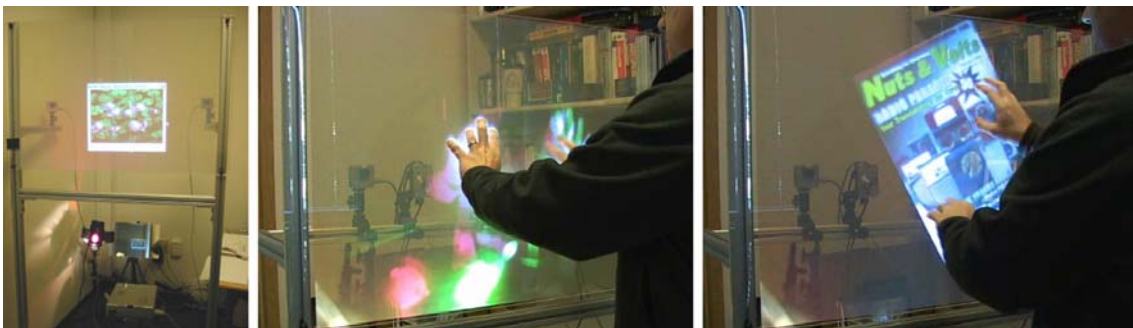


Figure 15. TouchLight uses computer vision to track the user’s gestures in front of a transparent display. (Images courtesy of Andrew Wilson [Wilson 2004].)

Gesture-based interfaces are often regarded as natural, intuitive, futuristic or “magical”, but there are numerous issues that can complicate their usability. The lack of physical support may, for instance, cause fatigue due to lengthy interactions, unnatural poses, or the recognizer’s need for exaggerated motion. The potential difficulty of detecting the start and end of an action is another problem that is often referred to as the “Midas touch problem” in eye tracking [Jacob 1991], which can complicate the distinction between movement and actuation in the user interface.

It is worth noting that sensing a user’s movement, gaze or other actions can also be used *implicitly* to increase the level of immersion, through subtle changes or responses in the user interface.

2.3.3 Handheld devices

There is a distinction between handheld devices, which can be seen as analogous to tools, and user-worn devices, which often serve as accessories that extend our capabilities. How much a device encumbers the user depends, of course, on the specific activity and the device’s properties. A wristwatch, for example, may not be disturbing in an office environment, but may be uncomfortable to wear at night and, in contrast to a pocket watch,

requires some time and effort for removal. A diving watch is similarly a useful tool underwater, in contrast to the potential problems that could be caused by the use of an ordinary watch.

A handheld device has many benefits that leverage our inherent capabilities to manipulate physical objects with our hands [Fitzmaurice et al. 1995, Ishii and Ullmer 1997, Yee 2003]. Changing grips for better comfort is a natural instinct, and it is often easy to, for example, put down a device, or switch to the other hand, if required. The device's onboard controls can also support a different set of input and output functionality than what is possible with touch or gesture alone. Physical buttons, scroll wheels, joysticks, trackballs, touch-screens, accelerometers, microphones and hardware for vibro-tactile feedback are examples of standard technology in many of today's handheld devices. On-board components may also assist in sensing user activity, as in the case of using accelerometer data for motion sensing [Rekimoto 1996, Wang et al. 2006].

Other benefits are found in multi-user environments, where the devices provide an implicit indication of which persons are in control at the moment. The act of passing a physical device to another person is also a strong metaphor for delegation of control.

The physical properties of handheld devices can, on the other hand, become a problem, as the hardware may not only be costly but also introduces potential issues of damage, wear, loss and theft. Requirements for regular service, such as battery charging, software updates or repair, could have an impact on practical use. Some scenarios might additionally not benefit from limiting the number of simultaneous users by availability and support for specific hardware devices.

Handheld devices can include everything from classical remote controls to advanced mobile devices, with simple behavior that transmits button presses, to complex input/output functionality and spatial awareness.

2.3.4 Speech input

Speech interfaces typically employ a command-based interaction style, where the user executes actions through verbal commands. Interactivity can however also be achieved through analysis of non-verbal features, for example, to provide continuous parameter control [Harada et al. 2006, Igarashi and Hughes 2001, Olwal and Feiner 2005, Patel and Abowd 2007]. Delays in audio processing and the inherently limited bandwidth of speech are factors that need to be taken into account for each scenario, as they can affect the system's perceived performance, responsiveness and interactivity.

The issue of ambient noise and conflicting audio is a major challenge in speech interfaces, where the system's ability to focus on the relevant sound source often determines its success in accurately interpreting the user. This has led to various hardware strategies that spatially narrow the system's capture of audio, ranging from array microphones that remotely sense direction, to user-worn microphones that clip onto clothing for advantages of proximity, to body-attached microphones that exploit bone conduction to isolate the user's voice from the surroundings [Basu et al. 2000, Xu et al. 2004].

References

- ARToolKit (2009). <http://www.hitl.washington.edu/artoolkit/>. Accessed May 12, 2009.
- Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385.
- Bandyopadhyay, D., Raskar, R., and Fuchs, H. (2001). Dynamic shader lamps: Painting on movable objects. *Proc. ISMAR '01 (IEEE and ACM International Symposium on Augmented Reality)*, 207–216.
- Basu, S., Schwartz, S., and Pentland, A. (2000). Wearable phased arrays for sound localization and enhancement. *Proc. ISWC '00 (International Symposium on Wearable Computers)*, 103–110.
- Bimber, O. and Emmerling, A. (2006). Multifocal projection: A multiprojector technique for increasing focal depth. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):658–667.
- Bimber, O., Encarnacao, L. M., and Schmalstieg, D. (2003). The virtual showcase as a new platform for augmented reality digital storytelling. *Proc. EGVE '03 (Workshop on Virtual environments)*, 87–95.
- Bimber, O., Frohlich, B., Schmalstieg, D., and Encarnacao, L. M. (2001). The virtual showcase. *IEEE Computer Graphics and Applications*, 21(6):48–55.
- Bimber, O. and Raskar, R. (2005). *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A K Peters, Ltd. ISBN 1-56881-230-2.
- Bimber, O., Wetzstein, G., Emmerling, A., and Nitschke, C. (2005). Enabling view-dependent stereoscopic projection in real environments. *Proc. ISMAR '05 (IEEE and ACM International Symposium on Mixed and Augmented Reality)*, 14–23.
- Cakmakci, O., Ha, Y., and Rolland, J. P. (2004). A compact optical see-through head-worn display with occlusion support. *Proc. ISMAR '04 (IEEE and ACM International Symposium on Mixed and Augmented Reality)*, 16–25.
- Cakmakci, O. and Rolland, J. (2006). Head-worn displays: A review. *Display Technology*, 2(3):199–216.
- DiVerdi, S. and Höllerer, T. (2006). Image-space correction of AR registration errors using graphics hardware. *Proc. VR '06 (IEEE Conference on Virtual Reality)*, 241–244.
- Dodgson, N. A. (2004). Variation and extrema of human interpupillary distance. *Stereoscopic Displays and Virtual Reality Systems XI*, volume 5291 of *Proc. SPIE*, 36–46.
- Encarnacao, I. M., Barton, R. J., I., Bimber, O., and Schmalstieg, D. (2000). Walk-up VR: Virtual reality beyond projection screens. *IEEE Computer Graphics and Applications*, 20(6):19–23.
- Fitzmaurice, G. W. (1993). Situated information spaces and spatially aware palmtop computers. *Communications of the ACM*, 36(7):39–49.
- Fitzmaurice, G. W., Ishii, H., and Buxton, W. A. S. (1995). Bricks: Laying the foundations for graspable user interfaces. *Proc. CHI '95 (SIGCHI conference on Human factors in computing systems)*, 442–449.
- Grossberg, M., Peri, H., Nayar, S., and Belhumeur, P. (2004). Making one object look like another: Controlling appearance using a projector-camera system. *Proc. CVPR '04 (IEEE Conference on Computer Vision and Pattern Recognition)*, volume 1, 1–452–1–459 Vol.1.
- Gustafsson, J. and Lindfors, C. (2004). Development of a 3D interaction table. *Stereoscopic Displays and Virtual Reality Systems XI*, volume 5291 of *Proc. SPIE*, 509–516.
- Gustafsson, J., Lindfors, C., Mattsson, L., and Kjellberg, T. (2005). Large-format 3D interaction table. *Stereoscopic Displays and Virtual Reality Systems XII*, volume 5664 of *Proc. SPIE*, 589–595.
- Halle, M. (1997). Autostereoscopic displays and computer graphics. *ACM SIGGRAPH Computer Graphics*, 31(2):58–62.

- Han, J. Y. (2005). Low-cost multi-touch sensing through frustrated total internal reflection. *Proc. UIST '05 (ACM symposium on User interface software and technology)*, 115–118.
- Harada, S., Landay, J. A., Malkin, J., Li, X., and Bilmes, J. A. (2006). The vocal joystick: evaluation of voice-based cursor control techniques. *Assets '06 (ACM SIGACCESS conference on Computers and accessibility)*, 197–204.
- Harrison, C. and Hudson, S. E. (2009). Providing dynamically changeable physical buttons on a visual display. *Proc. CHI '09 (International conference on Human factors in computing systems)*, 299–308.
- Henrysson, A. (2007). *Bringing Augmented Reality to Mobile Phones*. PhD thesis, Linköping University.
- Henrysson, A. and Ollila, M. (2004). UMAR: Ubiquitous mobile augmented reality. *Proc. MUM '04 (International conference on Mobile and ubiquitous multimedia)*, 41–45.
- Igarashi, T. and Hughes, J. F. (2001). Voice as sound: Using non-verbal voice input for interactive control. *Proc. UIST '01 (ACM symposium on User interface software and technology)*, 155–156.
- Ishii, H. and Ullmer, B. (1997). Tangible bits: towards seamless interfaces between people, bits and atoms. *Proc. CHI '97 (SIGCHI conference on Human factors in computing systems)*, 234–241.
- Jacob, R. J. K. (1991). The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Transactions on Information Systems*, 9(2):152–169.
- Jones, A., McDowall, I., Yamada, H., Bolas, M., and Debevec, P. (2007). Rendering for an interactive 360° light field display. *Proc. SIGGRAPH '07 (ACM International Conference on Computer Graphics and Interactive Techniques)*, Article No. 40.
- Kaiser, E., Olwal, A., McGee, D., Benko, H., Corradini, A., Li, X., Cohen, P., and Feiner, S. (2003). Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality. *Proc. ICMI '03 (International conference on Multimodal interfaces)*, 12–19.
- Kato, H. and Billinghurst, M. (1999). Marker tracking and HMD calibration for a video-based augmented reality conferencing system. *Proc. IWAR '99 (ACM International Workshop on Augmented Reality)*, 85–94.
- Kim, S., Ishii, M., Koike, Y., and Sato, M. (2000). Development of tension based haptic interface and possibility of its application to virtual reality. *Proc. VRST '00 (ACM symposium on Virtual reality software and technology)*, 199–205.
- Kiyokawa, K., Billinghurst, M., Campbell, B., and Woods, E. (2003). An occlusion-capable optical see-through head mount display for supporting co-located collaboration. *Proc. ISMAR '03 (IEEE and ACM International Symposium on Mixed and Augmented Reality)*, 133.
- Klein, G. and Murray, D. (2007). Parallel tracking and mapping for small ar workspaces. *Proc. ISMAR '07 (IEEE and ACM International Symposium on Mixed and Augmented Reality)*, 1–10.
- Klein, G. and Murray, D. (2008). Compositing for small cameras. *Proc. ISMAR '08 (IEEE and ACM International Symposium on Mixed and Augmented Reality)*, 57–60.
- Krueger, M. W. (1977). Responsive environments. *AFIPS '77: Proceedings of the June 13-16, 1977, national computer conference*, 423–433.
- Krueger, M. W., Gionfriddo, T., and Hinrichsen, K. (1985). VIDEOPLACE—an artificial reality. *Proc. CHI '85 (SIGCHI conference on Human factors in computing systems)*, 35–40.
- Kurz, D., Kiyokawa, K., and Takemura, H. (2008). Mutual occlusions on table-top displays in mixed reality applications. *Proc. VRST '08 (ACM symposium on Virtual reality software and technology)*, 227–230.
- Liu, S., Cheng, D., and Hua, H. (2008). An optical see-through head mounted display with addressable focal planes. *Proc. ISMAR '08 (IEEE and ACM International Symposium on Mixed and Augmented Reality)*, 33–42.
- Maes, P., Darrell, T., Blumberg, B., and Pentland, A. (1995). The ALIVE system: Full-body interaction with autonomous agents. *Proc. Computer Animation '95*, 11–18, 209.
- Matsushita, N., Ayatsuka, Y., and Rekimoto, J. (2000). Dual touch: a two-handed interface for pen-based PDAs. *Proc. UIST '00 (ACM symposium on User interface software and technology)*, 211–212.
- Matsushita, N. and Rekimoto, J. (1997). Holowall: designing a finger, hand, body, and object sensitive wall. *Proc. UIST '97 (ACM symposium on User interface software and technology)*, 209–210.
- Milgram, P. and Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems (Special Issue on Networked Reality)*, volume E77-D.
- Mohring, M., Lessig, C., and Bimber, O. (2004). Video see-through AR on consumer cell-phones. *Proc. ISMAR '04 (IEEE and ACM International Symposium on Mixed and Augmented Reality)*, 252–253.
- Navab, N. (2003). Industrial augmented reality (IAR): Challenges in design and commercialization of killer apps. *Proc. ISMAR '03 (IEEE and ACM International Symposium on Mixed and Augmented Reality)*, 2.
- Neumann, U. and You, S. (1999). Natural feature tracking for augmented reality. *IEEE Transactions on Multimedia*, 1(1):53–64.
- NFC Forum (2009). <http://www.nfc-forum.org/>. Accessed May 12, 2009.
- Olwal, A., Benko, H., and Feiner, S. (2003). Senseshapes: Using statistical geometry for object selection in a multimodal augmented reality. *Proc. ISMAR '03 (IEEE and ACM International Symposium on Mixed and Augmented Reality)*, 300–301.

- Olwal, A. and Feiner, S. (2005). Interaction techniques using prosodic features of speech and audio localization. *Proc. IUI '05 (International conference on Intelligent user interfaces)*, 284–286.
- Olwal, A. (2009). *Unobtrusive Augmentation of Physical Environments: Interaction Techniques, Spatial Displays and Ubiquitous Sensing*. PhD thesis, KTH.
- Paradiso, J. A. (2003). Tracking contact and free gesture across large interactive surfaces. *Communications of the ACM*, 46(7):62–69.
- Paradiso, J. A., Hsiao, K., Strickon, J., Lifton, J., and Adler, A. (2000). *IBM Systems Journal*, 39(3-4):89–914.
- Patel, S. N. and Abowd, G. D. (2007). BLUI: Low-cost localized blowable user interfaces. *Proc. UIST '07 (ACM symposium on User interface software and technology)*, 217–220.
- Patten, J., Ishii, H., Hines, J., and Pangaro, G. (2001). Sensetable: A wireless object tracking platform for tangible user interfaces. *Proc. CHI '01 (SIGCHI conference on Human factors in computing systems)*, 253–260.
- Pinhanez, C., Kjeldsen, R., Tang, L., Levas, A., Podlaseck, M., Sukaviriya, N., and Pingali, G. (2003). Creating touch-screens anywhere with interactive projected displays. *Proc. MULTIMEDIA '03 (ACM international conference on Multimedia)*, 460–461.
- Pinhanez, C. and Pingali, G. (2004). Projector-camera systems for telepresence. *ETP '04: Proceedings of the 2004 ACM SIGMM workshop on Effective telepresence*, 63–66.
- Pinhanez, C. and Podlaseck, M. (2005). To frame or not to frame: The role and design of frameless displays in ubiquitous applications. *Proc. of Ubicomp '05 (International conference on Ubiquitous computing)*.
- Pinhanez, C. S. (2001). The everywhere displays projector: A device to create ubiquitous graphical interfaces. *UBICOMP '01 (International Conference on Ubiquitous Computing)*, 315–331.
- Poupyrev, I. and Maruyama, S. (2003). Tactile interfaces for small touch screens. *Proc. UIST '03 (ACM symposium on User interface software and technology)*, 217–220.
- Poupyrev, I., Maruyama, S., and Rekimoto, J. (2002). Ambient touch: Designing tactile interfaces for handheld devices. *Proc. UIST '02 (ACM symposium on User interface software and technology)*, 51–60.
- Rakkolainen, I. and Palovuori, K. (2002). Walk-thru screen. *Projection Displays VIII*, volume 4657 of *Proc. SPIE*, 17–22.
- Raskar, R., Beardsley, P., van Baar, J., Wang, Y., Dietz, P., Lee, J., Leigh, D., and Willwacher, T. (2004). RFIG lamps: interacting with a self-describing world via photosensing wireless tags and projectors. *Proc. SIGGRAPH '04 (Conference on Computer graphics and interactive techniques)*, 406–415.
- Raskar, R., Brown, M. S., Yang, R., Chen, W.-C., Welch, G., Towles, H., Seales, B., and Fuchs, H. (1999a). Multi-projector displays using camera-based registration. *Proc. VIS '99 (Conference on Visualization 1999)*, 161–168.
- Raskar, R., van Baar, J., Beardsley, P., Willwacher, T., Rao, S., and Forlines, C. (2003). iLamps: Geometrically aware and self-configuring projectors. *Proc. SIGGRAPH '03 (Conference on Computer graphics and interactive techniques)*, 809–818.
- Raskar, R., Welch, G., and Fuchs, H. (1998). Seamless projection overlaps using image warping and intensity blending. *Proc. VSMM '98 (International Conference on Virtual Systems and Multimedia)*.
- Raskar, R., Welch, G., and Fuchs, H. (1999b). Spatially augmented reality. *Proc. IWAR '98 (International workshop on Augmented reality)*, 63–72.
- Reilly, D., Rodgers, M., Argue, R., Nunes, M., and Inkpen, K. (2006). Marked-up maps: combining paper maps and electronic information resources. *Personal and Ubiquitous Computing*, 10(4):215–226.
- Reitmayr, G., Eade, E., and Drummond, T. W. (2007). Semi-automatic annotations in unknown environments. *Proc. ISMAR '07 (IEEE and ACM International Symposium on Mixed and Augmented Reality)*, 67–70.
- Rekimoto, J. (1996). Tilting operations for small screen interfaces. *Proc. UIST '96 (ACM symposium on User interface software and technology)*, 167–168.
- Rekimoto, J. (1997). Navicam: A magnifying glass approach to augmented reality systems. *Presence: Teleoperators and Virtual Environments*, 6(4):399–412.
- Rekimoto, J. and Nagao, K. (1995). The world through the computer: computer augmented interaction with real world environments. *Proc. UIST '95 (ACM symposium on User interface and software technology)*, 29–36.
- Rekimoto, J., Oka, M., Matsushita, N., and Koike, H. (1998). Holowall: interactive digital surfaces. *SIGGRAPH '98: ACM SIGGRAPH 98 Conference abstracts and applications*, 108.
- Rekimoto, J. and Saitoh, M. (1999). Augmented surfaces: a spatially continuous work space for hybrid computing environments. *Proc. CHI '99 (SIGCHI conference on Human factors in computing systems)*, 378–385.
- Rekimoto, J., Ullmer, B., and Oba, H. (2001). Datatiles: a modular platform for mixed physical and graphical interactions. *Proc. CHI '01 (SIGCHI conference on Human factors in computing systems)*, 269–276.
- Schmandt, C. (1983). Spatial input/display correspondence in a stereoscopic computer graphic work station. *Proc. SIGGRAPH '83 (Conference on Computer graphics and interactive techniques)*, 253–261.
- Selker, T. (2008). Touching the future. *Communications of the ACM*, 51(12):14–16.
- State, A., Keller, K. P., and Fuchs, H. (2005). Simulation-based design and rapid prototyping of a parallax-free, orthoscopic video see-through head-mounted display. *Proc. ISMAR '05 (IEEE and ACM International Symposium on Mixed and Augmented Reality)*, 28–31.

- State, A., Livingston, M. A., Garrett, W. F., Hirota, G., Whitton, M. C., Pisano, E. D., and Fuchs, H. (1996). Technologies for augmented reality systems: realizing ultrasound-guided needle biopsies. *Proc. SIGGRAPH '96 (Conference on Computer graphics and interactive techniques)*, 439–446.
- Stetten, G., Chib, V., Hildebrand, D., and Bursee, J. (2001). Real time tomographic reflection: phantoms for calibration and biopsy. *Proc. ISAR '01 (IEEE and ACM International Symposium on Augmented Reality)*, 11–19.
- Takacs, G., Chandrasekhar, V., Gelfand, N., Xiong, Y., Chen, W.-C., Bismpiagiannis, T., Grzeszczuk, R., Pulli, K., and Girod, B. (2008). Outdoors augmented reality on mobile phone using loxel-based visual feature organization. *Proc. MIR '08 (International conference on Multimedia information retrieval)*, 427–434.
- Takagi, A., Yamazaki, S., Saito, Y., and Taniguchi, N. (2000). Development of a stereo video see-through hmd for ar systems. *Proc. ISAR '00 (IEEE and ACM International Symposium on Augmented Reality)*, 68–77.
- Tobii (2009). <http://www.tobii.com/>. Accessed May 12, 2009.
- Wagner, D., Langlotz, T., and Schmalstieg, D. (2008a). Robust and unobtrusive marker tracking on mobile phones. *Proc. ISMAR '08 (IEEE and ACM International Symposium on Mixed and Augmented Reality)*, 121–124.
- Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., and Schmalstieg, D. (2008b). Pose tracking from natural features on mobile phones. *Proc. ISMAR '08 (IEEE and ACM International Symposium on Mixed and Augmented Reality)*, 125–134.
- Wagner, D. and Schmalstieg, D. (2003). First steps towards handheld augmented reality. *Proc. ISWC '03 (IEEE International Symposium on Wearable Computers)*, 127.
- Wang, J., Zhai, S., and Canny, J. (2006). Camera phone based motion sensing: interaction techniques, applications and performance study. *Proc. UIST '06 (ACM symposium on User interface software and technology)*, 101–110.
- Want, R. (2006). An introduction to RFID technology. *IEEE Pervasive Computing*, 5(1):25–33.
- Want, R., Fishkin, K. P., Gujar, A., and Harrison, B. L. (1999). Bridging physical and virtual worlds with electronic tags. *Proc. CHI '99 (SIGCHI conference on Human factors in computing systems)*, 370–377.
- Welch, G. and Foxlin, E. (2002). Motion tracking: no silver bullet, but a respectable arsenal. *IEEE Computer Graphics and Applications*, 22(6):24–38.
- Wellner, P. (1993). Interacting with paper on the digitaldesk. *Communications of the ACM*, 36(7):87–96.
- Wetzstein, G. and Bimber, O. (2007). Radiometric compensation through inverse light transport. *Proc. PG '07 (Pacific Conference on Computer Graphics and Applications)*, 391–399.
- Wilson, A. D. (2004). Touchlight: An imaging touch screen and display for gesture-based interaction. *Proc. ICMI '04 (International conference on Multimodal interfaces)*, 69–76.
- Wilson, A. D. (2005). Playanywhere: A compact interactive tabletop projection-vision system. *Proc. UIST '05 (ACM symposium on User interface software and technology)*, 83–92.
- Wood, R. B. and Howells, P. J. (2006). *The Avionics Handbook*, chapter 7. Head-up displays, 7:1–7:24. CRC Press. ISBN-10 0849384389, ISBN-13 978-0849384387.
- Xu, Y., Yang, M., Yan, Y., and Chen, J. (2004). Wearable microphone array as user interface. *Proc. AUIC '04 (Conference on Australasian user interface)*, 123–126.
- Yee, K.-P. (2003). Peephole displays: pen interaction on spatially aware handheld computers. *Proc. CHI '03 (SIGCHI conference on Human factors in computing systems)*, 1–8.
- Zaeh, M. and Vogl, W. (2006). Interactive laser-projection for programming industrial robots. *Proc. ISMAR '06 (IEEE and ACM International Symposium on Mixed and Augmented Reality)*, 125–128.
- ZCam (2009). <http://www.3dvsystems.com/>. Accessed May 12, 2009.