Nada är en gemensam institution mellan
Kungliga Tekniska högskolan och Stockholms universitet.

Unobtrusive Augmentation of Physical Environments

Interaction Techniques, Spatial Displays and Ubiquitous Sensing

ALEX OLWAL

Avhandling som med tillstånd av Kungliga Tekniska högskolan
framlägges till offentlig granskning för avläggande av teknologie doktorsexamen
fredagen den 5 juni 2009 kl 13.00
i sal E1, Lindstedtsvägen 3,
Kungliga Tekniska högskolan, Stockholm.

TRITA-CSC-A 2009:09
ISSN-1653-5723
ISRN-KTH/CSC/A--09/09-SE
ISBN: 978-91-7415-339-2

Abstract

The fundamental idea of Augmented Reality (AR) is to improve and enhance our perception of the surroundings, through the use of sensing, computing and display systems that make it possible to augment the physical environment with virtual computer graphics. AR is, however, often associated with user-worn equipment, whose current complexity and lack of comfort limit its applicability in many scenarios.

The goal of this work has been to develop systems and techniques for uncomplicated AR experiences that support sporadic and spontaneous interaction with minimal preparation on the user's part.

This dissertation defines a new concept, *Unobtrusive AR*, which emphasizes an optically direct view of a visually unaltered physical environment, the avoidance of user-worn technology, and the preference for unencumbering techniques.

The first part of the work focuses on the design and development of two new AR display systems. They illustrate how AR experiences can be achieved through transparent see-through displays that are positioned in front of the physical environment to be augmented. The second part presents two novel sensing techniques for AR, which employ an instrumented surface for unobtrusive tracking of active and passive objects. These techniques have no visible sensing technology or markers, and are suitable for deployment in scenarios where it is important to maintain the visual qualities of the real environment. The third part of the work discusses a set of new interaction techniques for spatially aware handheld displays, public 3D displays, touch screens, and immaterial displays (which are not constrained by solid surfaces or enclosures). Many of the techniques are also applicable to human-computer interaction in general, as indicated by the accompanying qualitative and quantitative insights from user evaluations.

The thesis contributes a set of novel display systems, sensing technologies, and interaction techniques to the field of human-computer interaction, and brings new perspectives to the enhancement of real environments through computer graphics.

Acknowledgments

I would like to first and foremost thank Professor Steven Feiner at Columbia University for his guidance, engagement, and generous support of this work. The opportunity of collaborating with such an inspiring advisor has been a true pleasure and a highly appreciated privilege.

I thank Professor Lennart Johnsson who financially supported my work at KTH and Professor Yngve Sundblad for the much appreciated discussions, joint proposals, and shared interests in visualization and interaction.

The main part of this work was based at PDC at KTH in Stockholm, with large parts conducted at a number of institutions around the world, where I enjoyed many stimulating challenges and enriching interactions with new colleagues. I would like to express my gratitude towards the following institutions, for hosting and supporting the work:

- PDC, Department of Computer Science & Communication, KTH
- Computer Graphics and User Interfaces Laboratory, Columbia University
- Department of Production Engineering, KTH
- Imaging, Interaction and Innovative Interfaces Laboratory, UC Santa Barbara
- Adaptive Systems and Interaction Group, Microsoft Research, Redmond
- VITA group, Linköping University

I would like to thank the many colleagues and collaborators that I have been fortunate to interact with over the years, including: Patrick Baudisch, Andy Beall, Mats Bejhem, Blaine Bell, Hrvoje Benko, Gábor Blaskó, Nicola Candussi, Phil Cohen, Andrea Corradini, Ed Cutrell, Stephen DiVerdi, Arvid Engström, Stefan Evensen, Rogerio Feris, Elias Gagas, Jonny Gustafsson, Sinem Güven, Drexel Hallaway, Anders Henrysson, Susanna Heyman, Ken Hinckley, Hans von Holst, Tobias Höllerer, Edward Ishak, Jacob Jacobson, Ed Kaiser, Maryam Kamvar, Torsten Kjellberg, Lars Kjelldahl, Xiaoguang Li, Christoffer Lindfors, Simon Lok, Lars Mattson, David McGee, Gary Ngai, Axel Nordberg, Mihai Nicolescu, Takashi Okuma, Matthias Pusch, Ismo Rakkolainen, Sajid Sadi, Christian Sandor, Raman Sarin, Jan Stamer, Gert Svensson, Nick Temiyabutr, Matthew Turk, Jan Wikander, Andrew Wilson, Ryuji Yamamoto, Tiantian Zhou, and many others. Lars Winkler Petterson, Otmar Hilliges and Oskar Rönnberg kindly took the time to provide their much appreciated perspectives on drafts of this dissertation.

I have also had the rewarding experience of supervising numerous ambitious M.Sc. thesis, B.Sc. thesis, and project students, who I thank for their hard work and enthusiasm.

Last but certainly not least, I would like to very much thank my family and friends for supporting my interests and explorations in the sciences. I would like to, in particular and especially, emphasize my gratitude towards my mother Lena, and my grandmother Valentina, for their endless inspiration from my early days.

The research in this dissertation was funded in part by the Sweden–America Foundation, VINNOVA, and NSF Grant IIS-01-21239.

Generous support was provided by Ericsson, Sony Ericsson, Nokia, Microsoft Research, Mitsubishi, 3DV Systems and the Swedish National Maritime Museums.

Table of Contents

Publications in the thesis

Paper I ASTOR: An Autostereoscopic Optical See-through Augmented Reality System
Olwal, A., Lindfors, C., Gustafsson, J., Kjellberg, T., and Mattson, L.
Proceedings of ISMAR 2005
IEEE and ACM International Symposium on Mixed and Augmented Reality
Vienna, Austria, Oct 5–8, 2005

Paper II Spatial Augmented Reality on Industrial CNC-Machines
Olwal, A., Gustafsson, J., and Lindfors, C.
Proceedings of SPIE 2008 Electronic Imaging, Volume 6804
The Engineering Reality of Virtual Reality 2008
San Jose, CA, January 27–31, 2008

Paper III POLAR: Portable, Optical see-through, Low-cost Augmented Reality
Olwal, A. and Höllerer, T.
Proceedings of VRST 2005
ACM Symposium on Virtual Reality and Software Technology
Monterey, CA, Nov 7–9, 2005

Paper IV LightSense: Enabling Spatially Aware Handheld Interaction Devices
Olwal, A.
Proceedings of ISMAR 2006
IEEE and ACM International Symposium on Mixed and Augmented Reality
Santa Barbara, CA, Oct 22–25, 2006

Paper V LUMAR: A Hybrid Spatial Display System for 2D and 3D Handheld Augmented Reality
Olwal, A. and Henrysson, A.
Proceedings of ICAT 2007
International Conference on Artificial Reality and Teleexistence
Esbjerg, Denmark, Nov 28–30, 2007

Paper VI SurfaceFusion: Unobtrusive Tracking of Everyday Objects in Tangible User Interfaces
Olwal, A., and Wilson, A.
Proceedings of GI 2008
Graphics Interface
Windsor, Ontario, May 28–30, 2008

Paper VII Rubbing and Tapping for Precise and Rapid Selection on Touch-Screen Displays
Olwal, A., Feiner, S. and Heyman, S.
Proceedings of CHI 2008
SIGCHI Conference on Human Factors in Computing Systems
Florence, Italy, April 5–10, 2008

Paper VIII Unencumbered 3D Interaction with See-through Displays
Olwal, A.
Proceedings of NordiCHI 2008
Nordic Conference on Human–Computer Interaction
Lund, Sweden, October 20–22, 2008

Paper IX An Immaterial Pseudo-3D Display System with 3D Interaction
DiVerdi, S., Olwal, A., Rakkolainen, I., and Höllerer, T.
In Three-Dimensional Television: Capture, Transmission, and Display
ISBN 978-3-540-72531-2. Editors: Haldun M. Ozaktas and Levent Onural
Springer, 2008

Paper X Consigalo: Multi-user, Face-to-face Interaction with Adaptive Audio, on an Immaterial Display
Olwal, A., DiVerdi, S., Rakkolainen, I., and Höllerer, T.
Proceedings of INTETAIN 2008
International Conference on Intelligent Technologies for Interactive Entertainment
Cancun, Mexico, January 8–10, 2008

Paper XI Spatially Aware Handhelds for High-Precision Tangible Interaction with Large Displays
Olwal, A., and Feiner, S.
Proceedings of TEI 2009
International Conference on Tangible and Embedded Interaction
Cambridge, UK, February 16–18, 2009

Other related publications

Rubbing the Fisheye: Precise Touch-Screen Interaction with Gestures and Fisheye Views
Olwal, A., and Feiner, S.
Conference Supplement of UIST 2003
ACM Symposium on User Interface Software and Technology
Vancouver, BC, November 2-5, 2003

An Autostereoscopic Optical See-through Display for Augmented Reality
Olwal, A., Lindfors, C., and Gustafsson, J.
SIGGRAPH 2004 Sketches
International Conference on Computer Graphics and Interactive Techniques
Los Angeles, CA, Aug 8-12, 2004

A Novel Walk-through 3D Display
DiVerdi, S., Rakkolainen, I., Höllerer, T., and Olwal, A.
Proceedings of SPIE 2006 Electronic Imaging, Volume 6055
Stereoscopic Displays and Virtual Reality Systems XIII
San José, CA, Jan 15-18, 2006

An Immaterial, Dual-sided Display System with 3D Interaction
Olwal, A., DiVerdi, S., Candussi, N., Rakkolainen, I., and Höllerer, T.
Proceedings of IEEE VR 2006
IEEE Virtual Reality Conference 2006
Alexandria, VA, Mar 25-29, 2006

The Audiator: A Device-Independent Active Marker for Spatially Aware Displays
Olwal, A.
SIGGRAPH 2007 Posters
International Conference on Computer Graphics and Interactive Techniques
San Diego, CA, August 5-9, 2007

My roles in the publications in this dissertation

I am the main contributor to the work in Papers I–IV, VI–VIII, and X–XI.

I initiated the collaboration between KTH and the VITA group at Linköping University, where joint work with the second author led to the results in Paper V.

The work described in Paper IX combined the different authors' expertise and interests in display technology, computer graphics, interaction techniques, and tracking technologies. I am the main contributor to two of the projects described in Paper IX, in which I developed "Consigalo" and "Elastic Surface Deformer". I constructed the custom input device hardware, which included a miniature version of an active infrared LED marker for the tracking system, and two joysticks, which I extended for wireless operation and position tracking. Paper X describes the "Consigalo" system and interaction concepts in more detail.

Papers I and II describe the results from collaborative work on a novel autostereoscopic spatial AR system. I initiated and was responsible for the collaboration with the Department of Production Engineering to develop an AR system based on their display hardware. We collaborated on concept development and project scenarios, as well as on practical issues like hardware configurations and setup. They provided the industrial domain knowledge, expertise in optics and the HOE technology. I developed the system, which integrated with the machine, extracted real-time data, rendered 3D visualizations, and supported interaction. I am the main contributor to Papers I and II, in which Jonny Gustafsson provided the expertise on the display technology, and the co-authors contributed feedback.

The work described in Papers III, IX and X was hosted by University of California, Santa Barbara, while the work in Paper VI was hosted by Microsoft Research in Redmond.

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 dissertation, 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.

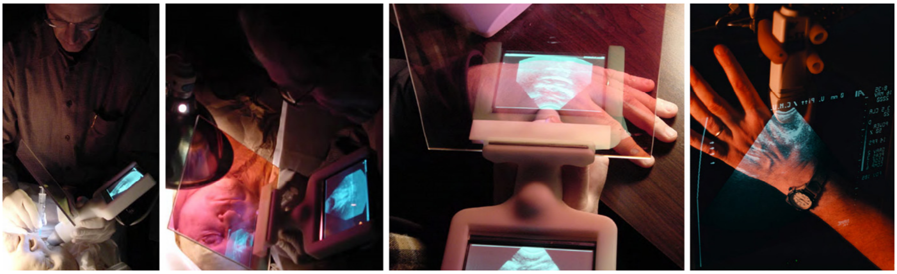


Figure 1. The Sonic Flashlight uses a see-through display to overlay real-time ultrasound images over a patient's body parts. (Images courtesy of George Stetten [Stetten et al. 2001].)

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

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].)

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.

The goal of the work described in this dissertation is to enable unobtrusive AR in *walk-up-and-use* scenarios that support spontaneous interaction with minimal user preparation [Encarnacao et al. 2000]. Unencumbering technology is 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]. The system should also, 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.

This work defines an *unobtrusive* AR system to be one that 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.

1.3 Summary of contributions

This thesis makes contributions in each of the three areas related to unobtrusive AR: display systems, sensing and interaction techniques.

1.3.1 Spatial display systems for unencumbered AR

The first part of the work (Chapter 3) contributes the design and development of ASTOR and POLAR, two spatial AR display systems that do not rely on user-worn equipment. In contrast to head-worn and handheld displays, *Spatial AR displays* either use large optical elements that are positioned relative to the real environment, or use digital projectors for direct augmentation of real-world surfaces [Bimber and Raskar 2005].

ASTOR and POLAR illustrate the feasibility of AR experiences through see-through displays that are positioned in front of the real environment to be augmented.

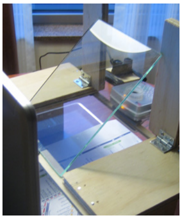
ASTOR

The ASTOR (AutoSTereoscopic Optical see-thRough) system is the first autostereoscopic spatial AR system based on a holographic optical element (HOE). It was developed in collaboration with the Department of Production Engineering at KTH.

Paper I and Paper II describe the display system and how it is used to augment the user's view of the operation of an industrial lathe. The display component is integrated with the machine's existing safety glass, and data from the control computer is directly overlaid as 3D visualizations on machine components.

ASTOR demonstrates how the view of a real process can be efficiently combined with computer-monitored information, and shows the directness of presenting this real-time information as overlaid, in-situ, 3D visualizations.

POLAR



Paper III describes the POLAR (Portable, Optical see-through, Low-cost Augmented Reality) system, where portability, low-cost, and commercial off-the-shelf components are emphasized. The work introduces a foldable display system and a hybrid tracking approach that fuses computer vision and infrared distance sensing to enable perspective adjustments based on the user's viewpoint.

POLAR's hybrid tracking technique and modular physical design illustrate how 2D AR workspaces can be created with a minimal set of compact and affordable hardware.

1.3.2 Unobtrusive sensing of devices and objects

The contributions of the second part of the work (Chapter 4) are two novel sensing techniques that employ an instrumented surface for hidden tracking of active or passive objects. These techniques have no visible sensing technology or markers, and are thus suitable for deployment in scenarios where it is critical to maintain the original visual qualities of the real objects and environments.

LightSense and LUMAR

Paper IV presents the LightSense system, which uses two approaches for tracking an unmodified mobile device using light sensors underneath a surface. External tracking of the mobile device's activated camera light allows the device to be moved on the surface, while its display shows dynamic, context-sensitive information about surface items. LightSense enables natural, tangible interaction with unmodified mobile phones, on visually unaltered surfaces.

Paper V describes LUMAR, where the LightSense system is combined with an egocentric marker-based technique, to form a three-layered information space. These layers consist of static printed content, a dynamic 2D space where the device is moved on the surface, and a dynamic 3D space where the device is moved above the surface.

SurfaceFusion

The tracking of physical objects on digital displays often involves custom sensing hardware, visual markers or exotic modifications. In the future, however, many real-world objects can be expected to contain passive electronic tags, based on potentially unobtrusive technology such as Radio Frequency Identification (RFID) [Want 2006, Want et al. 1999]. This work investigates unobtrusive techniques for detecting and tracking such objects.

Paper VI presents a hybrid approach in which the fusion of RFID sensing and computer vision enables the identification and tracking of passive, visually unaltered objects on a digital surface. The paper proposes a set of minimal image processing operations that efficiently and robustly classify changes on the surface, and describes a fusion pipeline that synchronizes activities in the radio-frequency domain with activities detected by the camera.

1.3.3 Interaction techniques for direct manipulation

The third part of the work (Chapter 5) contributes a set of new interaction techniques that were designed to fulfill the previously discussed criteria for unobtrusive AR (Section 1.2). These techniques not only complement the specific display systems and sensing techniques introduced in this work, but most of them are also widely applicable to human-computer interaction in general.

Touch-based interaction

Touch-based interaction is commonly regarded as more direct and intuitive than traditional mouse and keyboard interfaces. User experience for touch-based interaction may, however, greatly vary with the system's sensing capabilities and hardware configuration. Paper VII introduces Rubbing and Tapping, two families of interaction techniques for touch-sensitive surfaces, which use minimal gestures to expand the expressiveness of stylus and finger input. A formal evaluation shows their quantitative and qualitative advantages over existing state-of-the-art techniques for target selection on touch-screen displays.

Interaction with public 3D-displays

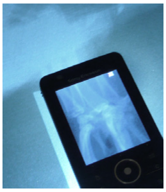
Paper VIII provides a qualitative evaluation of techniques that were designed and implemented for novice user interaction with public 3D-displays. The techniques include touch, head tracking, gesture tracking, and control with a mobile device. A qualitative user study indicates an advantage in terms of intuitiveness, ease-of-use and comfort, for both large and handheld touch-screen displays.

Immaterial displays

Paper IX and Paper X describe development and interaction explorations on an *immaterial* display that is not constrained by physical surfaces. This new type of display system uses vaporized water to form a projection screen that allows users to walk through and reach through interactive visuals that appear in mid-air.

An existing immaterial display system was extended for dual-sided projection, allowing individual, but synchronized, content to be displayed on each side of the vapor screen. The display system was combined with a 3D tracking system, and custom input devices were built, to enable the exploration of new interaction metaphors. Two applications were developed that exploited the unique capabilities of the vapor display for face-to-face interactions, simulated touch-based interactions, and full 3D interaction. Informal feedback was gathered during a one-week demonstration of the system in the Emerging Technologies exhibit at the *ACM SIGGRAPH* Conference in 2005.

Spatially aware handheld displays



Paper XI introduces a framework based on the previous work on spatially aware handheld displays (Paper IV and Paper V) and touch-screen interaction (Paper VII and Paper VIII). It presents a set of techniques that exploit the higher visual and input resolution of a tracked handheld display to improve interaction with a display of larger size, but less dense resolution.

The quantitative and qualitative results from a formal user study show the potential of complementing finger-based interaction on a large touch-screen display with techniques performed on a handheld device.

1.4 Thesis overview

This first chapter presented a brief introduction to AR, the specific challenges in achieving AR without encumbering technology or user-worn equipment, and a summary of the contributions of the thesis. Chapter 2 describes related work in three areas that are fundamental to an unobtrusive AR system: display systems, sensing for perspective-correct imagery, and interaction.

The contributions of the thesis are discussed in Chapters 3–5, which summarize and complement the work in Paper I–Paper XI. Chapter 3 introduces two spatial display systems that augment the environment behind them without encumbering equipment. Chapter 4 presents sensing techniques that can be integrated into an environment to support unobtrusive tracking and detection of devices and objects. Chapter 5 describes four projects that investigate interaction techniques for direct manipulation in unobtrusive AR scenarios.

Conclusions and future work are presented in Chapter 6.

2 Fundamental technologies

This chapter provides an overview of and describes the related 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.)

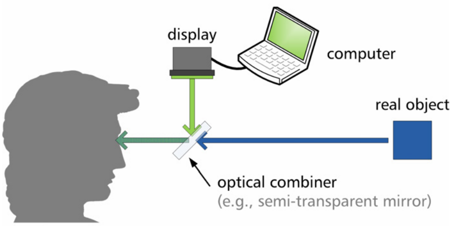


Figure 3. The optical path in an optical see-through display system. The light from the real environment passes through a transparent combiner that simultaneously reflects computer-generated images from a display. The *optically combined* light reaches the user's eyes. (Other optical elements, such as 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.)

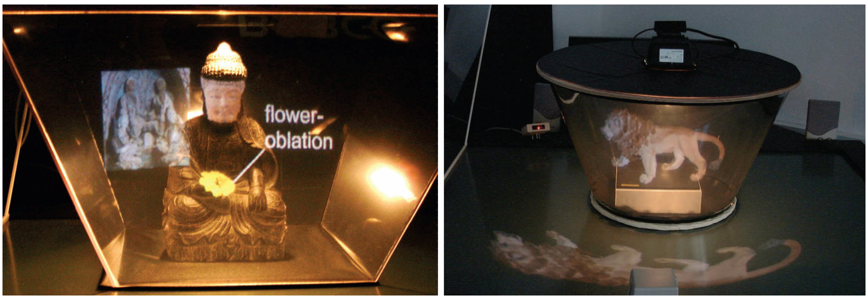


Figure 4. The Virtual Showcase augments physical objects inside showcases with dynamic annotations or related media. The Virtual Showcase is covered with half-silvered mirrors that optically combine the view of the objects with computer graphics reflected from the rear-projected table, on which the showcase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

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.

– **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.)

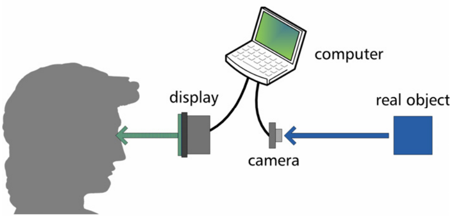


Figure 5. The optical path in a video see-through display system. The view of the real environment is acquired by a camera and is combined with virtual imagery by a computer. The combined video is presented to the user on a computer display. (Other optical elements, such as lenses that control the focal 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].

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]).

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].)

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].

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 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 640×480 ≈ 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.

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].

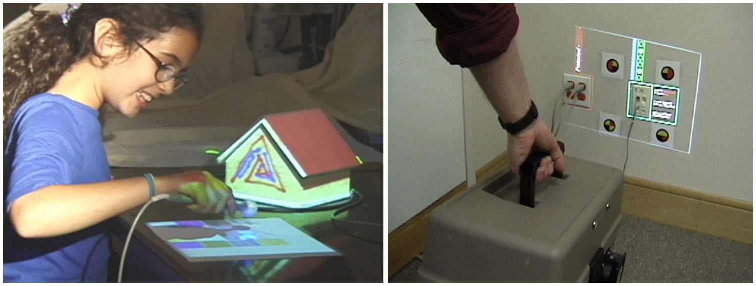


Figure 8. Left: A child uses a tracked brush to apply virtual paint, which is projected onto physical objects. (Image courtesy of Ramesh Raskar [Bandyopadhyay et al. 2001].)

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].)

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 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.

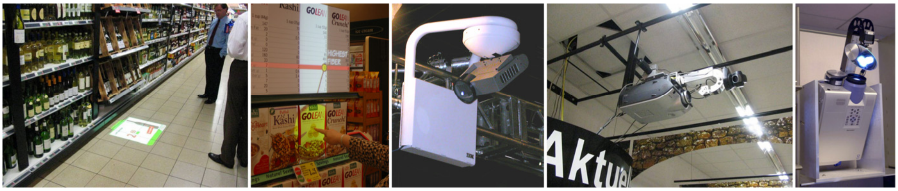


Figure 9. The Everywhere Displays project uses "steerable displays", where the system's projector can create augmentations on different surfaces in the environment, while the camera senses the user's interaction. (Images courtesy of Claudio Pinhanez [Pinhanez et al. 2003, Pinhanez and Pingali 2004, Pinhanez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

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 [Zaeh 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

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].)

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.

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

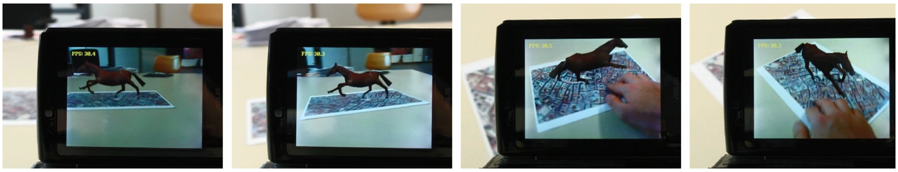


Figure 11. Natural feature tracking techniques make video see-through AR possible without the need for fiducial markers in the environment. Here, a mobile device tracks features on a printed page and overlays registered 3D graphics at interactive frame rates. (Images courtesy of Daniel Wagner and Graz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

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.)

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.

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.)

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].)

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

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].)

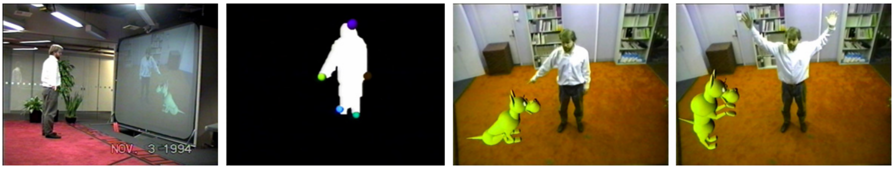


Figure 14. The ALIVE system tracks a user's body parts and supports full body interaction with virtual creatures. (Images courtesy of Alex Pentland, Bruce Blumberg and Trevor Darrell [Maes et al. 1995].)

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.

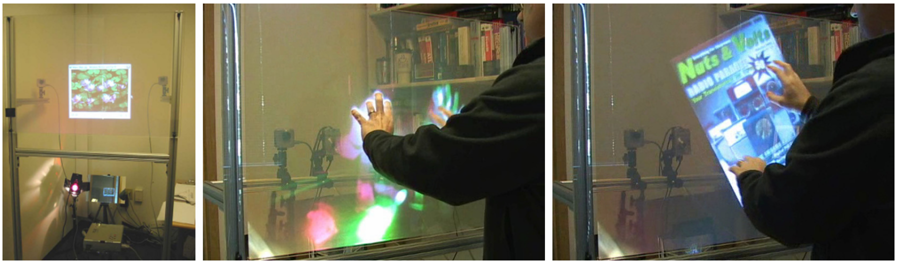


Figure 15. TouchLight uses computer vision to track the user's gestures in front of a transparent display. (Images courtesy of Andrew Wilson [Wilson 2004].)

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].

3 Spatial display systems for unencumbered AR

The display system is responsible for providing a combined view of the real and virtual, and thus is an essential component in AR technology. The many possible design choices depend on the system's desired characteristics and potential limitations, as discussed in Section 2.1. Scenarios may, for example, vary in how they prioritize properties, such as display configuration, optical qualities, visual resolution, user experience, immersion and ease-of-use.

This chapter describes ASTOR and POLAR, two spatial optical see-through display systems developed for unencumbered optically direct views of the real world with superimposed computer graphics.

3.1 ASTOR

An autostereoscopic optical see-through 3D display system

The goal of the ASTOR project was to enable unobtrusive user experience enhancements by implementing an optical see-through system for AR, in which 3D graphics and a view of the real world could be seen simultaneously, without requiring user-worn equipment [Paper I, Paper II].

Multi-view displays are a class of autostereoscopic displays, where parallax can be used to control which images are seen from a certain viewpoint [Halle 1997]. *Autostereoscopy* is achieved by presenting the user's left and right eyes with individual images of the 3D content from the corresponding two perspectives. Those two images form a stereo pair that is fused by the brain into a perceived 3D image, resulting in the illusion of depth.

Commercial multi-view displays typically employ optics in front of the display to limit the visibility of certain pixels to specific directions, such that different images are presented to each eye, from the same display surface. Such techniques, which, for example, can be based on a parallax barrier or lenticular sheet [Halle 1997], rely on content that is rendered in a specific format where multiple perspectives are combined into a single image and the resulting resolution often decreases with the number of unique perspectives. They are not directly appropriate for optical see-through AR, as their optics would distort the view of the real world behind the display. A half-silvered mirror could however combine the reflection of such multi-view displays with the view of the real world.

ASTOR instead employs a holographic optical element (HOE) that overcomes many of the previously discussed limitations. The optical properties of an HOE, which allow great flexibility in how light is diffused and reflected, are specified during the holographic manufacturing process. The HOE was developed by colleagues in the Department of Production Engineering, based on the group's previous work on an "Interaction Table" for tabletop Virtual Reality [Gustafsson and Lindfors 2004, Gustafsson et al. 2005].

The HOE used in ASTOR was manufactured to reflect light from multiple projectors, such that individual real images of thin vertical slices are formed, floating in front of the HOE plate. The image from each projector can only be seen from within the corresponding vertical slice, and the system limits the visibility of each of these views to a single eye at a time by using a slice width of 50 mm (which is less than the average interpupillary distance for an adult [Dodgson 2004]). Two horizontally aligned projectors can thus create two separate off-axis projections for autostereoscopy, where each eye sees a uniquely rendered image from the corresponding perspective, as shown in Figure 16.

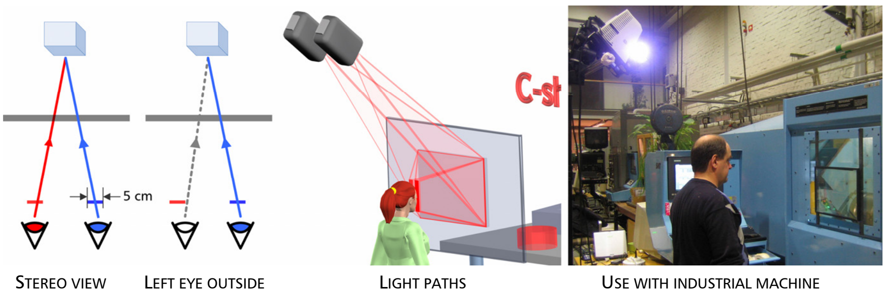


Figure 16. Stereo view: The user sees two individual images when looking through the 50 mm wide vertical slices. Images from the right and left projector are seen through the left and right slices, respectively (colored red and blue for clarity).

Left eye outside: The user has to look from the right direction and through the slice to see an image. Here, the left eye is outside the viewport and thus does not see an image.

Light paths: The light travels from each of the two projectors, reflects off the HOE and reaches the eyes, which are looking through the slices.

Use with industrial machine: The user views the augmented workspace through the HOE, which is installed in the center of the industrial lathe's safety glass.

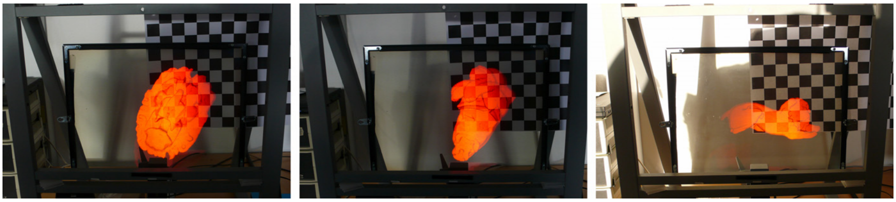


Figure 17. The HOE used in ASTOR combines bright computer graphics with high transparency. The HOE's properties allow it to very efficiently direct most of the light from the projectors to the user's eyes. The direct view of the real environment is slightly brighter than the view through the HOE, as indicated by the attached checkerboard pattern. The images were shot through one of the viewing slices under varying lighting conditions.

The setup can be extended with additional projectors, to expand the viewing volume for a single viewer through more stereo perspectives, and to support multiple simultaneous users.

One of the main advantages of using an HOE, compared to a half-silvered mirror, for example, is its superior brightness and contrast [Gustafsson et al. 2005]. It is able to effectively direct a large fraction of the projected light to the user's eyes, which yields high-quality imagery even under significant ambient illumination that would disturb many other displays. The quality of the reflected light is less dependent of the transmitted light, which makes it possible to display bright computer graphics while simultaneously maintaining a highly transparent see-through view of a real environment [Gustafsson et al. 2005]. (See Figure 17.)

The versatility of holographic techniques allowed the HOE used in ASTOR to be manufactured for optimal viewing configurations. In contrast to half-silvered mirrors, an HOE does not require that the light's angle of incidence be equal to the angle of reflection. ASTOR can thus use a custom HOE, where light projected from a 35° angle above the surface normal is reflected out of the display along the normal. This avoids potentially disturbing reflections from the projectors, while allowing the user to comfortably look straight into the real environment behind the display. (See Figure 18.)

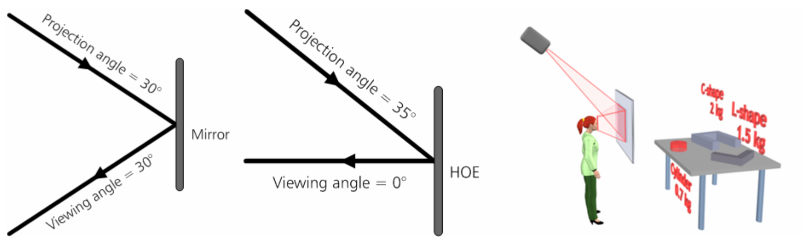


Figure 18. Left: A mirror's angle of incidence and angle of reflection are the same.

Center: The optical properties of an HOE are specified when it is manufactured, allowing flexible configurations that can be adapted to the specific requirements of a scenario.

Right: The HOE used in ASTOR was manufactured for an optimal projection and viewing configuration, where the projectors are placed in a convenient position and the user looks straight through the display.

Figure 19. Left: The operator of an industrial machine has to simultaneously focus on both the important real-time sensor measurement provided from a control computer, and the view of the workspace.

Right: ASTOR integrates the real-time sensor data as intuitive and natural 3D visualizations, which are overlaid on the corresponding parts of the machine.

The currently monochromatic HOE can also be extended to full color, which would be especially relevant for rendering realistic virtual objects [Gustafsson et al. 2005]. It is worth noting however, that simple monochromatic 3D overlays, which are optically merged with the direct view of the real world, may be perfectly sufficient for augmentations and enhancements in many application scenarios.

The system was applied to an industrial machine, where the operator typically has to focus simultaneously on both the workspace and a computer monitor that shows important real-time measurements from a cutting process. Figure 19 and Figure 20 illustrate how ASTOR can be used to instead integrate the presentation of these measurements with the workspace through real-time in-situ visualizations.

The ASTOR system was developed to receive continuous updates from the industrial machine of important data, such as tool position, RPM, temperature, and cutting

Figure 20. Left: The projectors behind the operator illuminate the HOE with autostereoscopic 3D imagery, which is optically combined with the view of the machine, as seen through the safety glass.

Right: The ASTOR system receives real-time measurements from the industrial machine which are presented as stereoscopic 3D graphics, optically merged with the workspace. The image shows how the cutting forces are rendered as 3D vectors at the tip of the tool (the three components and the resultant of the force).

forces at the tip of the tool. Figure 20 shows how ASTOR renders the force data as 3D vectors, which are overlaid in 3D at the current tool position to provide the operator with a direct and intuitive representation of the machine's state.

Key features and insights

- ### Direct view of the real workspace

  ASTOR's optical see-through display system preserves a direct view of the real workspace, which can be critical for many real-world tasks, including the industrial scenario presented here.

- ### High brightness and high transparency

  The HOE is more effective than half-silvered mirrors in showing bright computer graphics, while maintaining a clear view of the environment behind the display.

- ### Autostereoscopic 3D

  The system overlays 3D visualizations on the relevant machine parts, such that the user, for example, can follow the tool position in 3D, even if the real tool is cutting inside or behind the workpiece. Autostereoscopy makes it possible to see the 3D visualization at the correct depth, without the need for eyewear.

- ### Multi-view

  The use of additional projectors would increase the number of possible views, which could both provide individual visualizations for multiple people, and more stereo pairs. The view-dependent perspectives and the lack of user-worn equipment make it easy to step to the side for an unaugmented view of the workspace.

- ### Resolution not dependent on the number of perspectives

  Many autostereoscopic systems encode all views in a single image and thus trade display resolution for additional perspectives [Halle 1997]. In contrast, the HOE in ASTOR provides full resolution for each perspective.

- ### Optimal display configuration

  The possibility of specifying projection and viewing angles at the time of manufacture for the HOE allows comfortable viewing directions, avoids reflection from the projectors, and greatly simplifies alignment and integration with existing surfaces. In contrast, half-silvered mirrors require that the angle of incidence is equal to the angle of reflection, which limits the possibilities of alignment between the user, the mirror, and the display element. A half-silvered mirror must always be mounted at an angle with respect to the user to support see-through, and would thus integrate poorly with the industrial machine in the discussed setup.

- **Tested in an industrial environment**

  The ASTOR system was successfully integrated with a fully operational lathe in an industrial environment, which further demonstrates its flexible software and hardware combination that can complement existing work processes without the need to alter them. Only an initial one-time calibration is needed after installation of the rigid setup. Paper I and Paper II discuss numerous additional features that could provide deeper integration with the machine to further amplify the user's understanding of its operation.

Issues

- **Limited number of fixed views**

  Using the fixed views for an extended period of time could result in fatigue, due to the requirement for the user to stand in a specified position. The slices in ASTOR's HOE are, however, intentionally made sufficiently wide to support a bit of motion and movement, and no negative feedback has been communicated from the different people who have tried the system so far. It would be valuable to formally investigate the ergonomic aspects of the system in future work.

- **Scalability**

  Multiple perspectives must be simultaneously generated and rendered in parallel, which affects the system's scalability due to limitations in display hardware and computational power. Emerging display technologies and new generations of powerful graphics processors are examples of technologies that may help address this.

- **Monochromatic**

  The current version of ASTOR provides monochromatic imagery only, but the underlying HOE technology can be extended to full color.

- **Horizontal-parallax-only**

  The HOE limits the view of each perspective in the horizontal direction. If the user moves inwards or outwards (while viewing the HOE through the slices) the same images will be seen, which can result in misregistration of the virtual content with the real world behind the display. Remote tracking of the user's head or eyes is a well-known approach to address such issues. Paper VIII demonstrates how a 3D camera and remote eye tracker can provide interaction in this type of display system. It would also be straightforward to use them for the tracking of the user's viewpoint to support continuous adjustments of the rendered perspectives. Such technologies do, however, typically limit the number of simultaneous users.

- Different focus for real and virtual

  The projected images are focused on the HOE's surface, and the further away the system places the stereoscopically rendered 3D graphics, the more problematic the mismatch between vergence and accommodation will become. This affects how far away real-world objects can be placed, if co-located augmentations are desired [Bimber and Raskar 2005, Schmandt 1983]. Future work includes the investigation of optical techniques that can address this limitation.

- Custom manufacturing process

  The HOE is currently commercially unavailable, and the complicated manufacturing process requires access to special material, a holographic laboratory, and optical expertise.

- Stationary

  The rigid setup with a number of carefully aligned hardware components makes the system suitable primarily for stationary installations.

3.2 POLAR

A portable, low-cost, optical see-through 2D display system

The POLAR system emphasizes low-cost and portability, and is based on a foldable 2D display system [Paper III]. It uses a half-silvered mirror as the optical combiner to merge the virtual computer graphics with the real environment. The half-silvered mirror in this classical optical see-through configuration reflects images that are generated on a computer display, while it also allows light to pass through, for a simultaneous view of the real environment [Schmandt 1983], as shown in Figure 21.

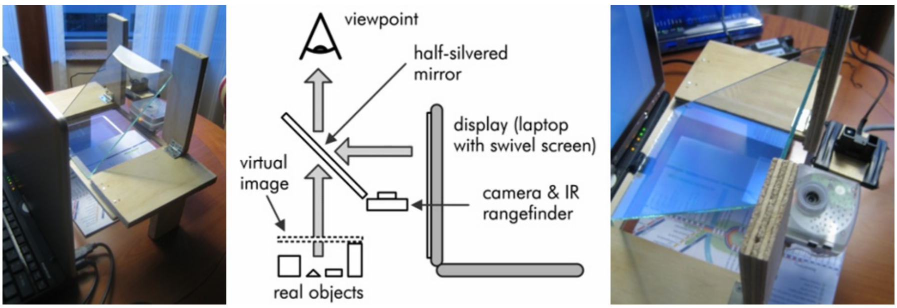


Figure 21. The POLAR system is a compact, portable and low-cost system for spatial AR. It provides overlays in a small workspace underneath the half-silvered mirror.

Figure 22. Example augmentations (highlighted here with white outline for clarity).
Left: POLAR is used to label components on equipment placed in the workspace.
Right: POLAR provides dynamic augmentations of a paper map, with colored highlights of areas of interest to the user.

The user looks down through the half-silvered mirror, which is mounted at a 45° angle with respect to a vertically placed digital display. The reflected content from the display will appear as a virtual image in a 2D plane below the mirror, superimposed on the view of the real objects, as shown in Figure 22. The height and location of the virtual image plane can be adjusted by varying the distance between the display and the mirror.

This display configuration is sufficient for AR if the virtual image plane is co-located with the real objects, for example, as in the case of the Sonic Flashlight [Stetten et al. 2001].

It is, however, necessary to track the user to enable general scenarios, where the virtual image plane might be placed at a distance from real-world objects. That makes it possible to continuously adapt the rendering of the overlaid graphics to the user's viewpoint if the virtual image plane is located above or below real-world objects. This might, for example, be useful in the case of multiple objects at different heights.

POLAR avoids user-worn equipment and exotic hardware through a hybrid technique that recovers the 3D position of the user's eyes. The system uses an infrared rangefinder to establish the distance to the user's face, which is derived from a non-linear mapping of the sensor's voltages to distance values. POLAR simultaneously tracks the 2D position of the user's eyes in video captured by a standard PC camera, using a software-based eye tracker. The 3D position can then be recovered through trigonometry by combining the distance, the 2D position in the video image, and the known field-of-view of the camera, as shown in Figure 23.

The POLAR system is thus able to provide the user with a perspective-correct overlay, where 2D graphics are superimposed on real objects in the workspace seen through the half-silvered mirror.

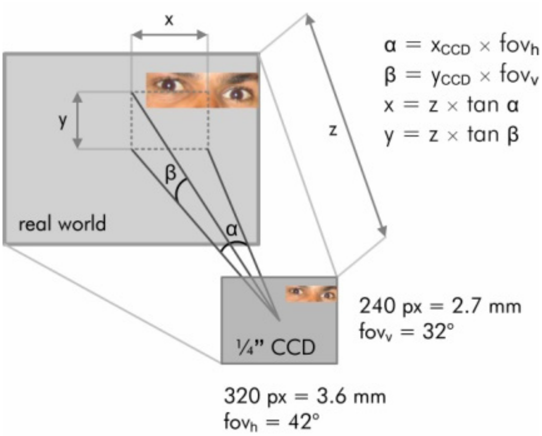


Figure 23. POLAR uses a hybrid tracking technique based on a digital camera and IR distance sensor. The position of the user's eye in the camera image, which are recovered using eye tracking software, are combined with the user's distance (z) to the camera, to establish the 3D position (x, y, z) of the user's eyes in the real world.

Key features and insights

- ### Direct view of the real workspace

  As in ASTOR, a direct optical view of the real workspace is provided, which is a desirable characteristic for many application scenarios.

- ### Color

  Color is supported through the use of an ordinary digital display and a half-silvered mirror, in contrast to the monochromatic graphics on ASTOR's HOE.

- ### Continuous views based on the user's 3D position

  The hybrid technique tracks the user's position unobtrusively, to support perspective-correct overlays during movement, without user-worn sensors.

- ### Adjustable focus plane

  In contrast to ASTOR, where the user focuses on the display surface, POLAR has an adjustable virtual image plane. Like all configurations based on half-silvered mirrors, the location of the reflected virtual image is determined by the distance to the display. It makes it possible to adjust the plane, such that it is as close as possible to the objects of interest, to minimize the conflict between vergence and accommodation [Bimber and Raskar 2005, Schmandt 1983].

- ### Portable

  The compact and foldable physical design, which is easy to assemble and disassemble, simplifies transportation and mobility.

- Commercially available low-cost hardware

  POLAR emphasizes the advantage in cost and availability of its components, as that expands possible applications and scenarios.

Issues

- 2D overlays

  The system can only show 2D imagery, but could be extended to reflect imagery from autostereoscopic displays. These are, however, rare in portable configurations.

- Reduced brightness and contrast

  Half-silvered mirrors have an inevitable trade-off in the ratio of transmitted and reflected light, and are also sensitive to ambient light.

- Need for tracking

  Ordinary displays can not present different perspectives based on the user's viewpoint, without the use of tracking or additional optics.

- Single-user and limited range

  The system's tracking technique supports only one person at a time, and is constrained by the tracking range of the distance sensor, and the field of view of the camera.

4 Unobtrusive sensing of devices and objects

The ability to sense physical objects is very useful in AR environments, as it allows the systems to extend beyond solely virtual imagery, to include interaction with physical devices and real-world artifacts. Passive and active objects can be sensed through a wide variety of techniques [Welch and Foxlin 2002].

LightSense and SurfaceFusion are two new exocentric sensing techniques that are introduced in this work's spirit of unobtrusive systems. They share an emphasis on preserving the appearance of the real environment and avoiding visual modification of the objects to be sensed, which makes them suitable for ubiquitous deployments. LUMAR combines LightSense with an egocentric 3D tracking technique for the exploration of hybrid interaction modes that include handheld video see-through AR.

4.1 LightSense

Exocentric sensing of spatially aware handheld displays

In LightSense, a surface with embedded sensing capabilities enables unobtrusive, exocentric tracking of a mobile device. The mobile device can then act as a spatially aware handheld display that provides information about areas underneath [Paper IV]. In contrast to egocentric approaches, all computation and sensing is offloaded to the environment, leaving only the need to communicate the sensed position to the device.

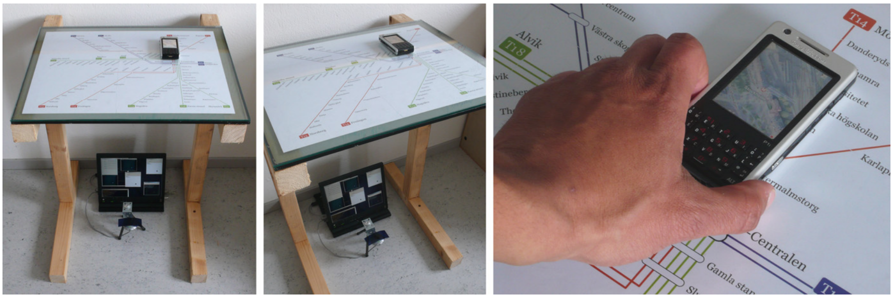


Figure 24. Left and center: The setup in the PC based version of LightSense. A PC tracks the position and distance of the mobile device's camera light to the surface using an attached camera. The device's position is communicated wirelessly to the device, which allows it to display context-sensitive information about the areas on the surface.

Right: As the mobile device is moved over a route map for the Stockholm subway, it is able to show geographical maps over the area for each subway stop. The user can control the zoom level by moving the device closer to or further away from the surface.

The built-in LEDs that are becoming increasingly common in modern camera phones are used as active markers in LightSense, which makes it possible to track the device without any modifications to it.

Embedded sensors are used to detect the activated LED (camera light or focus light) on the mobile phone. The light sensors in the surface are connected to a processing unit that continuously informs the device of its position through a wireless connection.

4.1.1 Sensing with a PC

Figure 24 shows how a flexible tracking solution is achieved using a PC that establishes the position and size of the light spot as seen from a camera underneath the surface. The data is then streamed wirelessly to the mobile device, for example, over Bluetooth or Wireless LAN. A chain of linked image processing filters isolates the light spot in the camera image. The center of the spot is used for the position, while the distance to the surface is estimated by the size of the spot. (See Figure 25.)

Figure 25. The camera's view of the mobile device's LED light from underneath the surface. The three images show the size of the light spot at different distances to the surface. From left to right: 0 cm, ~10 cm, and ~20 cm.

4.1.2 Sensing with a microcontroller

A compact, low-cost alternative to camera tracking on a PC is the use of embedded photo sensors that are attached to a microcontroller equipped with a wireless radio. The mobile device's light can, for example, be sensed by light-dependent resistors, where the microcontroller measures and detects their changed resistance on incident light, and wirelessly communicates the information to the mobile device over Bluetooth. (See Figure 26.)

The hardware enables a flat form factor, but with lower sensing resolution than a digital camera's image sensor and without estimation of distance. The setup is thus more appropriate for detecting the mobile device over a discrete number of points, and is also limited in its ability to sense distance. The RFIG lamps project explored a similar approach for projection-based AR [Raskar et al. 2004].

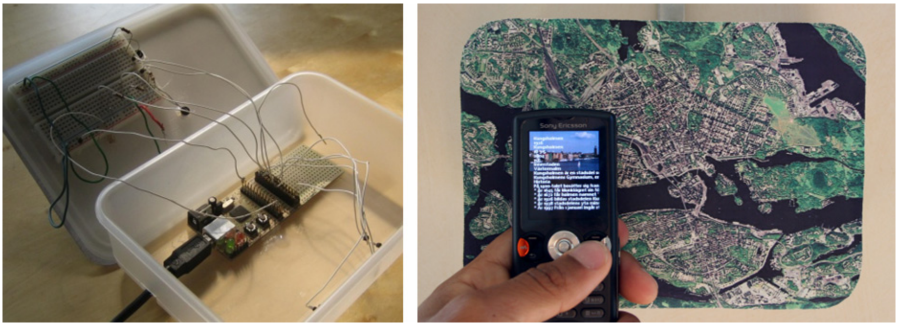


Figure 26. Left: A microcontroller with attached LDR sensors can sense the mobile device's light in discrete positions.

Right: A thin form factor makes a lightweight design possible. A wall-mounted version of LightSense allows the mobile device to provide dynamically updated information about the different boroughs in Stockholm.

Key features and insights

- **Preserves visual qualities**

  The system complements 2D graphics on a static map with a tracking technique that avoids visual alteration, fiducial markers and artifacts on the surface.

- **Tracking on, and near, the surface**

  The tracking technique works well on the actual surface, in contrast to typical egocentric camera-based approaches. This may have advantages in collaborative scenarios, where the supporting surface enables tangible interaction for multiple users. The possibility for moving the device on the surface may also avoid the fatigue that can result from lengthy interactions with devices held in mid-air, which could be a potential issue in handheld video see-through applications.

- Offloads computation to the environment

  The tracking technique enables spatial awareness on any modern mobile device that supports wireless communication with the tracking system. It is independent of the device's on-board computing power and performance, as all the computation and processing is offloaded to the environment. It can thus co-exist with resource-intensive third-party applications or run in parallel with other tracking techniques, as demonstrated in the LUMAR project. (See Section 4.2.)

- Robustness and protection of the technology

  The surface physically protects the equipment from damage, and at the same time provides robust tracking in controlled conditions, by minimizing the influence of ambient light.

- Unmodified devices

  The system tracks commercially available, unmodified mobile devices with built-in LED lights.

Issues

- Infrastructure and LED tracking

  LightSense depends on active technology in the environment. The alternative sensing technique, based on a microcontroller and light sensors, demonstrates how compact, low-cost equipment, could be used.

  The approach of sensing light also limits the system to mobile devices with LEDs, and to devices that provide software control over it.

  RFID sensing may become an interesting alternative in the future, as it uses flat, passive tags, which can be placed inside objects and surfaces. The Near Field Communication (NFC) initiative specifies the functionality of such sensing for mobile devices [NFC Forum 2009]. This emerging technology could support similar interaction, if it becomes widely available on mobile hardware.

- Limited tracking range and sensed information

  The tracking technique supports sensing of the mobile device's 2D position and an estimate of its distance to the surface. Data can also be encoded in modulated light, which, for example, could allow identification of the device being tracked. Egocentric approaches, in contrast, typically provide full 3D position and orientation, as used in LUMAR (Section 4.2).

- Non–see-through vs. see-through

  The current system was not designed for see-through operation, and as such, inherits the previously discussed limitations of spatially aware handheld displays. It does, however, illustrate that see-through capabilities are not always critical for enhancing a physical environment. This is especially relevant in scenarios that provide complementary representations of data associated with a physical location, but where geometrical registrations are of little use. The benefits of registering an abstract map with a geographical map may, for example, not necessarily be important, as in the implemented scenario shown in Figure 24.

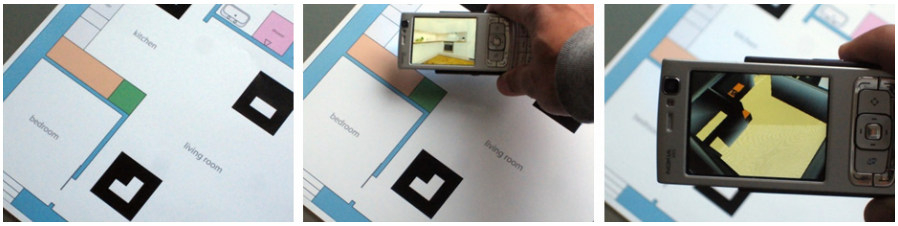
4.2 LUMAR

A hybrid egocentric/exocentric system for handheld AR

LightSense was combined with an egocentric tracking framework to expand the possible interactions of the mobile device. LUMAR [Paper V] uses the UMAR system [Henrysson and Ollila 2004] to provide a 3D video see-through mode when the device is held above the surface and out of range for LightSense. The LightSense framework is used to support tangible interaction with the device on, or near, the surface, where UMAR's camera is blocked, out-of-focus, or loses tracking for other reasons.

UMAR (Ubiquitous Mobile Augmented Reality) brings ARToolKit marker tracking [Kato and Billinghurst 1999] to mobile devices, allowing them to recover their position and orientation with respect to printed reference patterns, such that 3D models can be superimposed and registered with the environment in the live video from the device's camera.

LUMAR (LightSense + UMAR) illustrates the concept of layered information spaces through its different interaction modes, which are shown in Figure 27. The static layer consists of printed media, which can benefit from superior resolution, color



STATIC PRINT | DYNAMIC 2D MEDIA | DYNAMIC 3D MEDIA

Figure 27. LUMAR provides three types of information layers.

Static print: Although only able to present static information, a high-resolution print has many benefits compared to digital displays, such as its higher quality, low cost and passive nature.

Dynamic 2D media: LUMAR's exocentric tracking makes it possible to complement the map with 2D media when the tracked mobile device is on the surface (where egocentric camera-based tracking fails due to a blocked or out-of-focus camera).

Dynamic 3D media: LUMAR's egocentric framework provides 3D pose tracking and video see-through when the mobile device is held above the surface.

depth, viewing angles and texture, along with other aspects that influence the perceived quality and fidelity compared to existing digital displays. Its static nature is complemented by one or more dynamic 2D layers, which are enabled through the movement of a spatially aware handheld display on the surface. The information space can be further enhanced through video see-through AR, by superimposing layers with registered 3D data.

Key features and insights

- **Interaction on, near, and above the surface**

  LUMAR leverages the best of egocentric and exocentric techniques by combining the advantages of spatially aware handheld displays (LightSense) with the benefits of handheld video see-through AR (UMAR). The hybrid system fuses the interaction spaces provided by the two techniques, with increased flexibility as a result.

- **Multiple information layers**

  The work presents the possibilities of multiple information and interaction layers, where the different data can be accessed and represented in its most optimal form. Text and detailed graphics, for example, may be well-suited for static print, whereas video and linked content will likely benefit from presentation in dynamic 2D. The video see-through mode complements these with 3D graphics registered with the physical space.

- **Intuitive transitions**

  The user transitions between the different layers through intuitive single-handed movement of the handheld device.

Issues

- **Markers in the environment**

  The egocentric camera-based tracking in LUMAR relies on fiducial markers to recover its position and pose relative to the printed material. The use of these black and white rectangular patterns in the environment conflicts with the desired properties of unobtrusive AR, but was the best technical solution at the time to achieve real-time video see-through AR on mobile devices. The future work in Paper III discusses the extension of this work to techniques based on natural feature tracking, where no fiducial markers are needed. Advances in tracking algorithms and mobile computing have recently made such techniques possible [Wagner et al. 2008b].

- Limited interaction

  The work focuses on the potential of multiple information layers, but only provides control over perspective and content through device motion.

  Paper XI indicates how touch-screen techniques can provide fast, precise and direct manipulation in these scenarios. Future work would also include investigating complex tasks that require sophisticated interaction techniques, which may span different layers and dimensions.

4.3 SurfaceFusion

Unobtrusive sensing of everyday objects

It is useful to sense passive, everyday objects in AR scenarios, especially with the increased interest in interactive surfaces and displays. Traditional object sensing approaches often rely on complex pattern recognition techniques, active electronics in the objects, or the use of fiducials that alter visual qualities (as in LUMAR, Section 4.2). We can, however, expect that many physical objects will be manufactured with embedded electronics for identification in the near future. Passive RFID was mentioned in Section 4.1, and is an example of such technology, where cheap and small tags can be made invisible to the user, while providing robust identification.

Figure 28. SurfaceFusion's unobtrusive hybrid sensing enables tangible interaction with visually unaltered everyday objects.

RFID technology can only sense the presence of tags in the volume being scanned, but does not inherently recover their position. Computer vision, on the other hand, is well-suited to track shapes and movement with a camera, but relies on complex, computationally expensive algorithms for the recognition of arbitrary objects, where the performance may heavily depend on the changing nature of environmental lighting conditions.

SurfaceFusion introduces an approach, based on activity sensing, that leverages the capabilities of these respective modalities [Paper VI]. A *frame difference algebra* was developed as a means to *visually track* shapes that appear, disappear or move on the surface with minimal assumptions about object appearance, while RFID *identifies* added or removed objects. A fusion pipeline merges the geometry recovered from computer vision with RFID presence, such that arbitrary RFID-tagged objects can be tracked and identified without requiring visual markers. (See Figure 28.)

Key features and insights

- Identification and tracking of everyday objects

  SurfaceFusion illustrates the power of fusing simple techniques, which allows it to unobtrusively identify and track the position of arbitrary passive objects that contain embedded RFID tags.

- Preserves object appearance

  By complementing computer vision with RFID, many negative aspects of fiducial-based tracking are avoided. RFID greatly simplifies the process of identification, while eliminating the need for visual tags.

- Uniqueness and on-board storage

  An RFID tag can store large amounts of data, which, for example, can be used to identify far more objects than possible with 2D barcodes of equivalent size.

- Generally applicable techniques

  The image processing techniques in Paper VI make very few assumptions about object appearance and are therefore applicable to other activity detection applications.

  Certain scenarios also find unobtrusiveness less important and may instead require, or even benefit, from the use of visual markers. The introduced techniques can in such cases complement the identification and tracking with increased robustness and performance.

Issues

- Ambiguous actions

  The current system may encounter ambiguous information from certain event sequences, for example, if two objects are introduced to the system at the exact same time. The sensitivity to this timing is dependent on the system's ability to limit object detection to the surface only, and on the update rate of the RFID reader and the computer vision system.

  These issues may in part be addressed by a global fusion technique, which would resolve ambiguities over a history of events in an integrated on-line probabilistic process, as discussed in Paper VI .

- Limitations of RFID hardware

  Paper VI discusses several additional methods for extending activity sensing by exploiting properties in the RFID reader and RFID tags. The RFID reader and antennas in the project were however quite limited in their capabilities, and while a set of complementary methods, hardware, modifications and techniques were explored, their practical use was limited to detecting tag presence only. This made it possible, however, to demonstrate the power of the hybrid approach, through the use of basic RFID technology, image processing and the fusion pipeline.

  The modularity of the approach is important as it would be interesting to not only improve existing RFID sensing and computer vision techniques, but also to complement them with other sensing modalities.

5 Interaction techniques for direct manipulation

This chapter presents a number of interactive methods that complement the displays, view-dependent imagery and sensing methods described earlier. A rich set of techniques provides the user with direct mechanisms for manipulating content through touch, gestures and user sensing.

Section 5.1 introduces novel general techniques that can improve touch-screen interaction, while Section 5.2 specifically focuses on 3D interaction for public see-through displays. Section 5.3 explores the challenges and opportunities made possible by a new type of display system that forms images on water vapor. Spatially aware handheld displays are discussed in Section 5.4 as tools that can complement a larger digital display with high visual fidelity and precise interaction.

5.1 Rubbing and Tapping

Expanding the expressiveness of touch interaction

Traditional touch-screens are limited in their sensing capabilities, which typically allow them to sense only the location of the user's touch, or the lack of contact altogether. The constraint of being able to detect only two states, combined with the device's limited precision, results in fewer input options compared to other pointing devices. A computer mouse can, for example, have several buttons and a clickable scroll wheel, in addition to the precise adjustments that are possible through its relative position sensor.

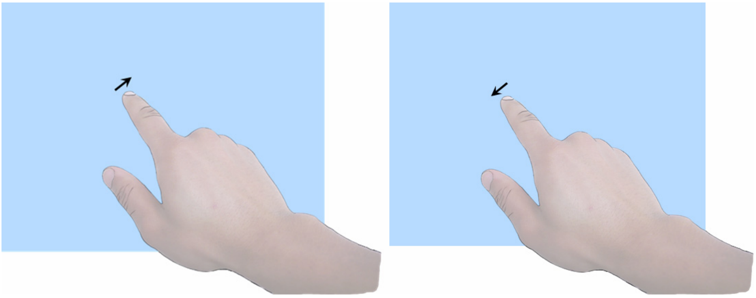


Figure 29. The *Rubbing* interaction technique uses small diagonal rubbing strokes to expand the user's expressiveness on touch-screen displays. The Rubbing gesture provides quick and convenient access to additional functionality, while the user is moving the finger on the surface. *Rubbing* is detected as at least two, short and fast, consecutive diagonal strokes in roughly opposite directions. The arrow illustrates the directions of the two strokes.

Two families of techniques were developed to address limitations in touch-screen interaction [Paper VII]. Both techniques default to regular touch-screen pointing, where their power lies in the ability to expand the expressiveness of interaction through simple gestures. The gestures can be mapped to different actions, which are integrated with standard pointing and dragging, to ensure fluid interaction.

Rubbing is a single-handed technique that uses a series of short and fast, diagonal strokes in roughly opposite directions, to trigger actions. (See Figure 29.)

Consider the three successive cursor position samples: $P_0(x_0, y_0)$, $P_1(x_1, y_1)$, and $P_2(x_2, y_2)$. If $\mathrm{sign}(x_1-x_0) \neq \mathrm{sign}(x_2-x_1)$ AND $\mathrm{sign}(y_1-y_0) \neq \mathrm{sign}(y_2-y_1)$, then (x_1, y_1) ends the previous stroke and begins the next. By classifying the gestures based on the slope of the strokes, two unique actions can be distinguished. The strokes roughly along the lower-left-to-upper-right diagonal (positive slope) are referred to as "Rubbing in", whereas the strokes roughly along the lower-right-to-upper-left diagonal (negative slope) are referred to as "Rubbing out". (See Figure 30.)

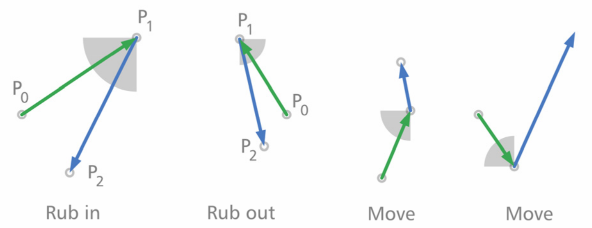


Figure 30. *Rubbing* also requires that the lengths of the strokes (here P_0–P_1 and P_1–P_2) are within the limits of preset minimum and maximum stroke lengths, and are executed sufficiently quickly and in roughly diagonally opposite directions. It is possible to distinguish between two types of gestures based on the slope of the strokes. The two leftmost images illustrate how "Rubbing in"/"Rubbing out" is performed along the diagonal with the positive/negative slope. The second stroke must lie within the 90° sector. The two rightmost images show how ordinary movement (dragging) of the finger on the surface can be supported in the same interaction mode, due to differences in speed, direction, and stroke lengths compared to Rubbing.

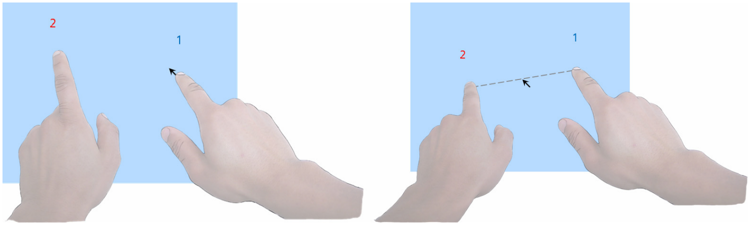


Figure 31. The *Tapping* interaction technique simulates aspects of multi-touch functionality on single-touch displays. The cursor jump that occurs when a second finger touches the screen on many single-touch displays can be detected in software, and mapped to different actions.

Left: The first (1) finger is touching the screen, while the second (2) finger is above the surface. The first (1) finger controls the cursor position (black arrow).

Right: The touch screen reports an average cursor position that lies on the vector between the two touch points when both the first (1) and second (2) fingers touch the screen.

Tapping is a bimanual technique that can simulate certain aspects of multi-touch functionality on many single-touch displays. It exploits the fact that many single-touch displays report an average cursor position when the screen is touched in two locations. The cursor jump that results from a second touch can be detected in software, if the distance between the two touches is sufficiently large [Matsushita et al. 2000]. It is both possible to establish the presence of a second touch, and its relative location to the first touch, since the average position will be located along the vector between the two touch points. The possibility of detecting this event is used for a quick touch-and-release action with the second finger, while the first finger is touching the screen. (See Figure 31.)

Five interaction techniques were designed, using combinations of Rubbing and Tapping to allow the integration of pointing and zooming into a single mode. They were evaluated against two well-known baseline techniques in a target selection task on a desktop touch screen. The results from a formal within-subject user study with twenty participants showed that the best of the introduced techniques were faster, resulted in fewer errors and had a higher user preference [Paper VII], compared to the two baselines.

The techniques were also used in two applications that were developed in this work. Figure 32 shows *TouchView*, an image browsing application, where the Rubbing and Tapping gestures avoid on-screen widgets and make it possible to support fluid interaction through integrated pointing and zooming. *TouchDaemon* is a background application, which was developed to make the techniques available in third-party applications, such as a map application in an unmodified web browser. (See Figure 33.)

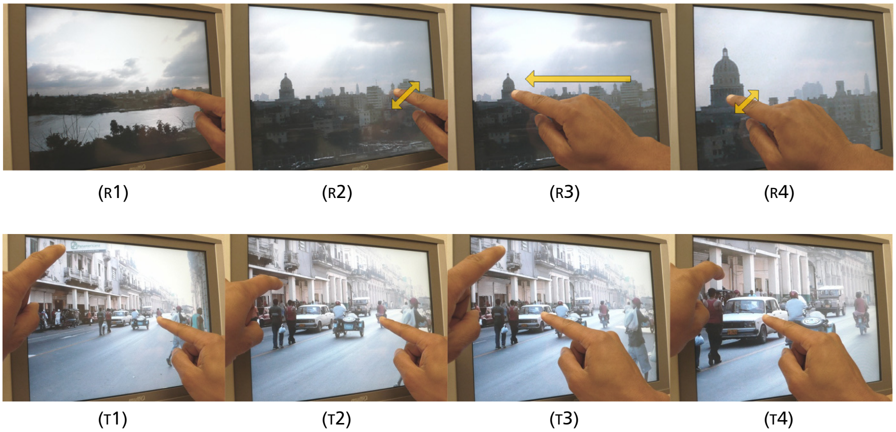


Figure 32. The *TouchView* application employs Rubbing and Tapping to support zooming and panning on images, without requiring menus or toolbars.

Top: (R1) Finger touches screen. (R2) Rubbing actions zoom in on a set of buildings. (R3) Finger is moved to a building of interest. (R4) Additional rubbing gestures zoom in on the building. (Arrows added to indicate finger motion.)

Bottom: (T1) First finger touches screen. (T2) A tap action with the second finger zooms in on a person. (T3) Second finger is released and first finger is moved to the car. (T4) A second tap zooms in again.

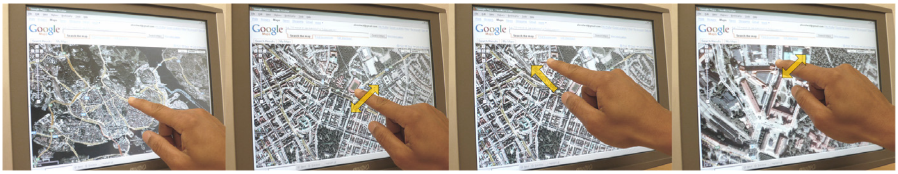


Figure 33. The *TouchDaemon* is an application that runs in the background and listens to the touch-screen input. It can detect Rubbing and Tapping gestures and makes it possible to trigger specific functionality through these interaction techniques in most third-party applications. In this example, TouchDaemon is used to provide gesture-based zoom through Rubbing in an unmodified map application that is running in an unmodified web browser. (Arrows added to indicate finger motion.)

Key features and insights

- ### Fast, precise and consistent performance

  Statistically significant results from a quantitative evaluation show numerous advantages over existing baselines.

- ### Higher user preference

  The majority of users preferred the introduced techniques over baseline techniques.

- ### Expands expressiveness with simple gestures

  The introduced techniques make it possible to access additional functionality without relying on mode-switches, widgets or user interface elements. The work emphasizes the increased precision and performance possible through simple gestures that can be performed at any time during the interaction on the display.

- ### Integrated pointing and zooming

  Pointing and zooming actions are integrated in a single mode, where the user is always pointing to the target. The two-finger stretching gesture that is popular on multi-touch displays, instead require that the user alternates between pointing and zooming. It also requires that the user actively adjusts to keep the target in the center of the vector formed by the fingers, since the two fingers actually move away from the target in that technique.

- ### Unlimited repetitions

  The techniques make it possible to quickly perform large scale changes, since they were designed for repeated execution, either through repeated Rubbing gestures, or through numerous taps with the second finger, while the user is pointing to the target.

- ### Rubbing: compact, single-finger interaction

  The Rubbing gestures require little space, and are thus also suitable for small screens. They work well near edges and corners, where many other gestural techniques would require an initial translation of the content to the center of the screen. A mobile version of Rubbing for stylus-based operation was included in the study presented in Paper XI.

- ### Compatible with existing hardware

  Rubbing works with any single-touch technology, and does thus not rely on special hardware. It can both be used on finger-sensing displays, as well as with stylus-based technologies.

  Tapping works on technology that reports average positions for two touches, such as today's common resistive single-touch sensors.

- ### Preserves standard behavior

  The techniques do not interfere with or change the standard way touch screens are operated. Instead, their additional functionality is invoked on demand, which allows them to be integrated in most existing touch-screen interfaces.

Issues

- ### Relative vs. absolute mapping

  The two-finger stretching gestures supported on multi-touch hardware can have an advantage in their absolute, 1:1 mapping of finger movement to geometrical transformations, despite the limitations in how fast they can be performed. The proprioceptive cues are, however, lost in the relative actions of the introduced Rubbing and Tapping techniques.

  It would be interesting to, in future studies, compare the impact of the differences between absolute and relative mappings for various scenarios.

- ### Simplicity vs. guessability

  The techniques were designed to be as simple as possible, in order to provide fast and convenient access to additional functionality. While their mechanisms do not directly map to real-world actions, experiences from the studies and from various demonstrations, indicate that their simplicity allows most people to understand and use them after a brief explanation.

- ### Discrete steps

  The current implementations trigger actions in discrete steps. It would be useful to explore their extension for continuous scale changes, which is technically possible in both Rubbing and Tapping.

- ### Comparison with multi-touch techniques

  The work focuses on general techniques that can be supported on the vast number of existing single-touch displays, as well as in multi-touch systems.

  Important future work includes a formal evaluation of our techniques in comparison with multi-touch gestures.

- ### Tapping: 2 fingers required

  Some users disliked the need for two fingers in Tapping, and this is an interesting issue, where a study comparing single-touch and multi-touch techniques may be able to provide more insight.

- ### Rubbing: risk for fatigue

  Certain touch-screens require hard touches, or use surface material that causes friction during dragging gestures. The resistive touch-screen display used in the study, caused fatigue for some users, who felt skin sensations after the large number of performed trials. It is unclear whether this would also be a problem under normal circumstances, where these gestures would be performed much less frequently.

Friction is also less of an issue for touch screens that use a glass surface, such as many capacitive or acoustic sensors.

The mobile version of Rubbing, which is used in Paper XI, avoids this problem altogether through the use of a stylus.

5.2 3D manipulation on public see-through displays

Touch, gestures, head movement and remote control

Spatial optical see-through AR systems, such as ASTOR, can integrate the display element with the environment, and allow walk-up-and-use interaction. The potential for damage, breakage and general wear in unstaffed public spaces is, however, often best addressed by making the installed equipment physically inaccessible to users. This means that such setups must provide input devices that work on, or through, the surface, while avoiding the use of external mechanical parts and interference with the view.

Seven interaction techniques were designed, implemented and evaluated in their support of novice user interaction with 3D content on a public 3D display [Paper VIII]. The techniques were designed to allow the user's exploration of different 3D models through translation, zoom and rotation actions.

Two touch-screen techniques were based on gestural manipulation on a transparent single-touch overlay (Figure 34, left; Figure 35, left). Remote-controlled operation was provided in two other techniques through the controls on a wirelessly connected mobile device (Figure 34, center and right; Figure 35, right).

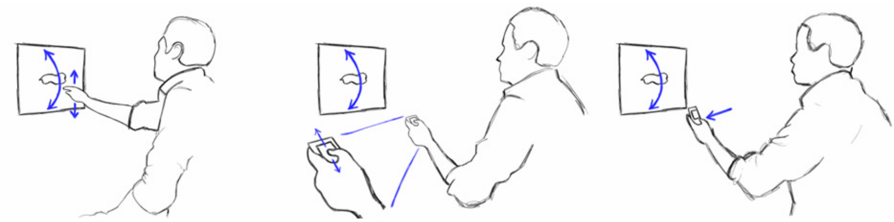


Figure 34. Techniques for interacting with 3D models on a public display.

Left: Touch-screen interaction is supported with on-screen buttons and rubbing gestures. On-screen buttons allow the user to browse different models, and change the mapping of dragging gestures between zoom, translation and rotation. The user can also, at any time, use the rubbing gesture to change the zoom level, as it does not conflict with dragging gestures.

Center: A connected mobile device's touch-screen is used as a remote touchpad for rotation of the model.

Right: All operations can also be accessed using the device's buttons.

Figure 35. Left: Interaction with a 3D model using techniques on the large touch-screen overlay.
Right: Manipulation through the touch-screen of a wirelessly connected mobile device.

In three additional techniques, the horizontal and vertical movement of a tracked mobile device were used as input (Figure 36, left), an eye tracker [Tobii 2009] followed head movement (Figure 36, center), and a 3D camera [ZCam 2009] tracked hand gestures (Figure 36, right).

The first four interfaces supported full manipulation (rotation, movement, zooming, cycling through models) of a displayed 3D model, while the last three only supported rotation about the vertical and horizontal rotation (pitch and yaw). Figure 37 shows the technology used in the study.

A focus group interview was conducted with three representatives from the Vasa Museum. A brainstorming session and discussions of how this technology could be applied to interactive exhibits provided interesting insights and affected the design of the techniques to be tested. The Vasa Museum also provided 3D-scans of sculptures and artifacts, which were used as the 3D models in the study.

A formal within-subject qualitative study with twelve participants was conducted, where each participant explored the possibilities for manipulation and interaction with the 3D models, using all seven techniques.

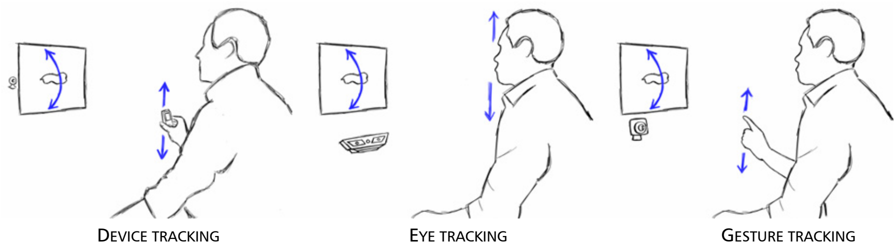


Figure 36. Sketches illustrating techniques for interacting with 3D models on a public display.
Left: Remote sensing of the mobile device makes it possible to rotate the content by moving the device vertically and horizontally with respect to the display.
Center: An eye tracker recovers the 3D position of the user's eyes to support rotation through head movement.
Right: A 3D camera is used to track hand gestures, where hand motion controls rotation.

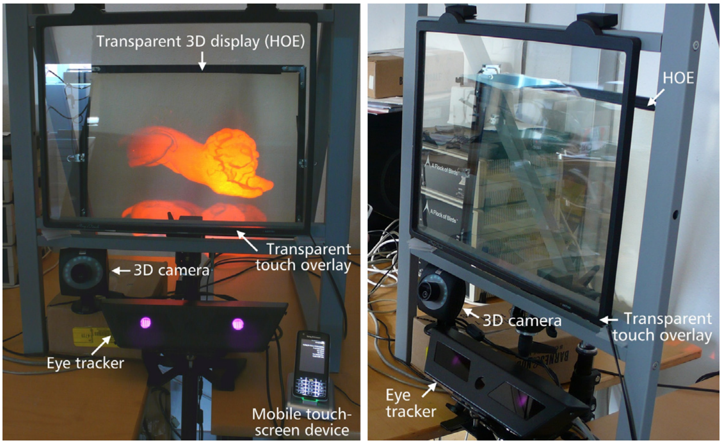


Figure 37. The technology used in the setup consists of a prototype 3D camera, a remote eye tracker, a resistive touch overlay, a mobile touch-screen device, and the HOE from the ASTOR system to provide autostereoscopic 3D imagery.

The results from the study indicated a preference for regular and mobile (remote) touch-screen interaction, which may be attributed to the directness of the tactile and graphical feedback, robust sensing, and users' familiarity with touch-screen technology. Remote-controlled operation using the mobile device's buttons or touch-screen was appreciated, as each individual could choose their own preferred pose for comfortable interaction.

Key features and insights

- ### Interviews with museum staff

  The museum staff provided valuable input regarding use cases, possible scenarios for digital showcases and the importance of unencumbered interaction in public installations.

- ### Potential of unencumbering technology

  The work demonstrates the potential of four technologies (eye tracker, 3D camera, touch sensor, and mobile device) in their ability to support unencumbered 3D manipulation in a public scenario, through the seven implemented interaction techniques.

- ### Advantages of touch-screen interaction

  A qualitative study shows that fluid interaction on the large touch-screen, as well as on the mobile device, was particularly preferred by the participants in the study.

- Compatible technologies for redundancy or functionality

  While the technologies were explored individually, most of them are compatible with simultaneous operation, and can thus be used in parallel to either support redundant interaction, or to provide additional functionality.

Issues

- Simple interaction

  The work emphasizes interaction for novice users in public scenarios. It would be interesting to address more complex scenarios that require sophisticated interaction techniques, which may be based on combinations of technologies.

- Quantitative study

  This study focuses on the qualitative aspects of the technologies for causal exploration. These insights could now be complemented with a quantitative evaluation that would measure the performance in execution of specific, well-defined tasks. It would be especially beneficial to develop such tasks based on actual requirements, which could come from the museum representatives in the focus group.

- Hardware limitations

  Some users were negatively affected in their interaction by tracking limitations of the eye tracker and 3D camera. Newer hardware with increased robustness, range and performance would likely address this.

  The current HOE was manufactured for front-projection (to allow integration with the industrial machine, see Section 3.1), which requires equipment on the user's side of the display surface, and also creates potential occlusion problem with the hand during touch-screen interaction. It is, however, possible to manufacture an HOE for rear-projection that would address these issues.

5.3 Immaterial displays

Multi-user, face-to-face, and reach-through interaction

Immaterial displays introduce new challenges and possibilities for human-computer-interaction. They represent a class of output devices in which the display system forms images without relying on physical surfaces or enclosures. The display system in this work is based on the FogScreen device, which produces a diffusion screen through a controlled flow of vaporized water [Rakkolainen and Palovuori 2002]. A hanging ceiling unit emits vaporized water, while simultaneous lamination by continuous air flow on both sides minimizes turbulence and makes it possible to establish a rectangular projection surface. (See Figure 38, left.)

Figure 38. Left: The flow of vaporized water is trapped between two layers of stabilizing air flow, making it possible to create projection screens that users can walk or reach through.

Center and Right: The water particles primarily disperse light in the forward direction, enabling dual-sided projections, where each side can show individual rear-projected content. In this example, a user sees the face of a 3D model from one side, and can walk through the screen to see the back of the head from the other side.

The display's resulting permeability removes many limitations of typical digital displays. Multiple users can casually *walk through* and *reach through* the screen, exchange physical objects from one side to the other, and freely communicate through the projection surface while maintaining visual contact.

Another advantageous aspect for novel interaction is the vaporized water's property of primarily diffusing light in the forward-direction, which means that rear-projected graphics will dominate the imagery on the surface. This allows the use of dual-sided rendering, such that different, but synchronized, rear-projected content can be shown on each of the display's two sides, as illustrated in Figure 38 (center and right). The work explores 2D and 3D dual-sided interaction for immaterial display systems, and resulted in a number of interface prototypes [Paper IX, Paper X]. A commercial 3D tracking system was used to explore isomorphic 3D interaction, as well as simulated touch-screen behavior, which were complemented with visual and auditory feedback. Custom input device hardware was also developed to miniaturize the tracked devices for touch-screen emulation, and to provide a comfortable form factor, with richer input and actuation, for the 3D interface.

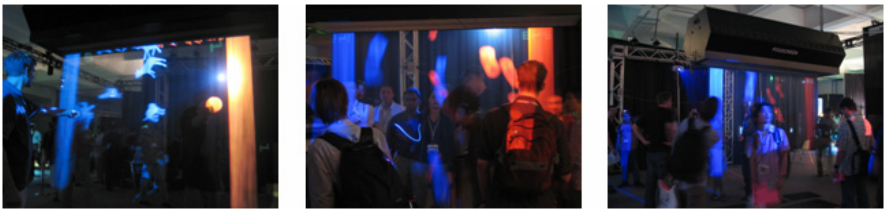


Figure 39. Interactive demonstrations for immaterial displays were demonstrated during a week at the Emerging Technologies exhibit at ACM SIGGRAPH 2005. One of the four applications shown was "Consigalo", a multi-user game that specifically focuses on simulated touch-screen behavior, face-to-face interaction, and auditory feedback.

The prototypes illustrate how engaging interaction can be supported by bringing computer graphics into the physical space, while allowing unrestricted user movement and the support for multi-user, face-to-face interaction without physical limitations, as shown in Figure 39.

Key features and insights

- **Interactive computer graphics in mid-air**

  The system introduces technology that allows interactive computer graphics to seemingly appear in mid-air.

- **Unhindered reach-through and walk-through**

  The system illustrates how new interaction techniques can be designed for unbreakable displays that provide reach-through and walk-through capabilities. The display’s immaterial properties simplify collaboration and communication around the content, as no physical barriers constrain verbal communication or movement.

  This can, for example, be advantageous in simulators, where it may be desirable to allow a user’s body parts to pass through the display surface.

- **Face-to-face interaction**

  The two-sided immaterial display allows two people to maintain face contact, while discussing synchronized, individual views of computer graphics that appear in-between them.

- **Accumulated experience for immaterial displays**

  The work provides insights from experiences in integrating several 2D and 3D tracking technologies for user tracking and interface control. It also reports on experiments with 2D and 3D rendering approaches and the use of auditory feedback.

- **Tested with large numbers of users**

  The applications were demonstrated during a week at ACM SIGGRAPH 2005, which made it possible to informally evaluate the ease-of-use and intuitiveness of the developed techniques, by observing the large number of users that were able to successfully use the system with no or minimal instruction.

Issues

- **Display quality**

  The display’s limitations in fidelity and field-of-view can only be addressed by improvements in the underlying display technology.

- Formal evaluation

  Future work would include a formal user study that can provide qualitative and quantitative results from interaction techniques specifically designed to exploit the system's immaterial properties.

- Form factor

  The considerable weight and size of the FogScreen device, the two projectors, and the tracking system, are currently limiting factors for widespread adoption of the system.

- No tactile feedback

  Auditory feedback was implemented as a means to partially provide a sense of touching a physical surface. Unconventional string-based haptic devices [Kim et al. 2000] could however be integrated with the system, since there is no physical display surface that would interfere. Such approaches do, on the other hand, have the drawback of limiting the number of simultaneous users.

5.4 Spatially aware handheld displays

Enhancing large touch-screens for precise interaction

Several previously discussed issues, such as parallax, occlusion, calibration and imprecision in finger-pointing, may affect interaction on touch-sensitive surfaces. (See Sections 2.3.1 and 5.1.) Apart from these, it is also worth noting that surface quality, input resolution and robustness can greatly *vary* across different touch technologies.

This work introduces a framework to overcome such issues, by leveraging the capabilities of small, coarsely tracked mobile touch-screen devices [Paper XI]. These

Figure 40. A spatially aware handheld display can complement interactions with a large digital display, by overcoming its limitations in visual output and touch input. In this example, the smaller device has over ten times the visual and input spatial resolution of the rear-projected surface, thanks to its denser arrangement of smaller pixels and touch-sensor elements. Besides improved resolution, it is also clearly shown how the handheld display also has significantly better brightness and contrast. This scenario illustrates how detailed focus can be achieved with the handheld, while the larger screen provides the context for interaction.

RUB CIRCLING GESTURES DRAG BUTTONS

Figure 41. Four different techniques to improve zooming and target selection on a large digital surface by interacting *through* the mobile device. The first three techniques use the mobile device's touch screen, where three simple gestures provide means for zooming, whereas the last technique relies solely on the mobile device's buttons.

devices typically have considerably higher visual and input resolution than large displays, due to a denser arrangement of the smaller pixel and touch-sensor elements, which are complemented by a rich set of additional on-board sensors and input controls. (See Figure 40.)

The mobile device is tracked using an extended version of the LightSense framework (described in Section 4.1 and Paper IV), where rear-projected imagery and touch sensing replace the static print. Finger-based interaction can thus be performed on a surface with dynamic content, in addition to the mobile interaction.

Four mobile interaction techniques were designed for interaction with the large digital display through the mobile display, which provides a high-resolution view of the content. Three of these techniques were implemented for gestural selection and zooming using the mobile device's touch-screen, while one technique instead used discrete buttons presses, as shown in Figure 41.

The techniques were evaluated in a formal within-subject user study with twelve participants, where two baseline techniques used direct finger-interaction on the large rear-projected screen.

The mobile techniques generally had consistent performance, smaller variations among participants, and a higher user preference, than the finger-based techniques for the large display. However, the characteristics of the hardware configuration and the touch overlay in our setup seem to have negatively affected the performance of the finger-based techniques. It is not always feasible to provide a setup that is ergonomically optimal for all possible users, and the choice of hardware configurations for installed touch screens is likely to be subject to many environmental constraints. Certain scenarios may even prohibit the use of touch altogether, for example, due to installation sites that do not allow exposed display surfaces, or the use of protective glass layers that are incompatible with available touch technology.

The study indicates that a coarsely tracked mobile device can avoid such problems, through predictable high-resolution input and output that support personalization and user-specific customization of its behavior.

Key features and insights

- **Spatial awareness for precise interaction in larger context**

  The system preserves the original view of the large digital display, while making it possible to achieve more precise interaction on a smaller screen, which has higher visual and touch resolution.

- **Personalized device — personalized interaction techniques**

  The majority of participants in the study selected one of the mobile techniques as their preferred, but the technique of choice varied. Several of the techniques can coexist in the same mode, similarly to other work in this dissertation (see Sections 5.1 and 5.2), and support either redundant functionality, or different functionality. An important difference compared with the other systems, is that these interaction techniques are running on the user's personal device. It would thus not be unreasonable for the user to customize a preferred technique of choice, just like the device is personalized with ringtones, background images, or text input methods.

- **The power of a coarsely tracked high-resolution viewport**

  The device in the study was tracked with almost a fifth of the resolution (352×288) compared to the finger-based pointing (800×600). The device was, however, able to overcome this limitation through its densely placed pixels and touch sensor elements. The higher quality screen might also have contributed through its better contrast and brightness.

Issues

- **Limited performance on resistive single-touch overlay**

  To address the limitations and issues with the resistive single-touch overlay, it would be valuable in future work to explore similar complementary interactions on multi-touch surfaces.

- **Refined and combined techniques**

  The interaction techniques in this scenario could be combined to support more advanced tasks, and also be extended with possibilities for continuous and elastic control.

The study identified an easily fixable design mistake, which affected one of the techniques (MobileDrag), where the system's misinterpretation of quick strokes led to high error rates for selections of small targets.

- Hybrid techniques

  Apart from the use of different techniques to trigger the same action (increased redundancy), or support multiple actions in a single mode (increased functionality), it is also interesting to explore techniques that are simultaneously or sequentially performed on multiple displays.

6 Conclusions and future work

Unobtrusive AR makes it possible to augment our view of the surroundings through walk-up-and-use scenarios that require minimal effort or preparation on the user's part.

The dissertation argues that the strategy of leveraging equipment that is integrated in the environment can, for many scenarios, eliminate the need for encumbering, user-worn technology. It also shows how techniques can be adapted to avoid intrusive modifications to the devices, objects or environment with which the user interacts. The many hybrid techniques described here illustrate how the combination of complementary methods can efficiently surpass their individual capabilities.

The display systems, sensing approaches, and interaction techniques that were designed, developed and evaluated in this dissertation, demonstrate the potential of unobtrusive AR to enable widespread and seamless interaction with real and virtual content.

6.1 Displays, sensing and registration

AR depends on display systems that combine computer-generated virtual imagery with the user's view of a real environment. Unobtrusive AR refines this requirement through its preference for an optically direct view of the real environment, and the avoidance of head-worn displays.

The directness with which the environment is perceived varies across display technologies and desired qualities. This work identifies spatial and handheld optical see-through, direct projection, and spatially aware handheld displays as particularly interesting display technologies for unobtrusive AR systems.

The camera's sampling of the real-world in video see-through systems, on the other hand, may have numerous negative consequences for the perception of the physical environment. This is, however, primarily an issue for head-worn video see-through displays, which tend to completely *replace* the user's view with augmented video. Handheld video see-through systems instead have the potential to act as viewports that *complement* an optically direct view, given their portability and less encumbering physical form factor.

Besides achieving the visual combination of real and virtual, AR systems also need to register the virtual content with the real, and provide perspective-correct imagery to the user. This thesis introduces combinations of display systems, hardware configurations, and sensing techniques that accomplish this purpose.

The HOE in ASTOR [Paper I, Paper II], is used both as an advanced optical combiner and as the enabling component for the system's multi-view capabilities. The unique characteristics of the system make it possible to simultaneously support flexible display configurations, a high-quality see-through view, and bright computer-generated imagery with high contrast. Its multi-view capabilities allow it to both present autostereoscopic 3D graphics, and support view-dependent imagery without user tracking. The clarity of the optical see-through display is particularly important, since the complementary nature of the monochromatic AR overlays relies on a high-quality view of the real environment. Like most multi-view displays, ASTOR also has limited scalability, due to the computational load and image generation that would be required to support a large number of simultaneous perspectives.

The POLAR project [Paper III] explores mobile scenarios by investigating the use of inexpensive, commercially available and portable technology as an alternative to ASTOR's custom-manufactured HOE. The system compensates for the lack of multi-view capabilities with remote sensing that makes it possible to continuously update the perspective as the user moves. The two interaction techniques in Paper VIII, which were implemented with an eye tracker and a 3D camera, illustrate how classical user tracking could be used to address the fixed number of views that limit the original ASTOR system.

LightSense also exploits remote sensing, but for the unobtrusive registration and tracking of mobile phones [Paper IV]. The environment senses and wirelessly informs the phone of its position on a surface, allowing context-sensitive graphics to be shown in its display. LUMAR combines the exocentric tracking of LightSense with egocentric marker-based tracking, which makes video see-through AR possible when the device is held above the surface [Paper V]. LightSense and LUMAR do not provide an optically direct view through their displays, but preserve instant access to an unmediated view of the real environment, either by moving the device out of the way, or by looking to the side of the device's display.

In SurfaceFusion, sensing is used to identify and track arbitrary passive real-world objects without the need for changes to their appearance [Paper VI]. This makes it possible to present associated digital media next to physical objects, which blends the manipulation of real and virtual through touch and tangible interaction.

6.2 Interaction techniques

The potential of perspective-correct augmentation, made possible by AR displays and sensing technologies, becomes even more interesting when extended with intuitive user interaction. The detection of touch, gesture, pose, and speech, or the use of handheld devices, are examples of interaction technology that can support direct manipulation, while avoiding user-worn equipment. These technologies are particularly interesting for unobtrusive AR systems, and led to the implementation of a set of diverse interaction techniques, which were designed for 2D displays, 3D displays, immaterial displays, and hybrid display configurations.

Two new techniques for touch-screen interaction are introduced in Paper VII. The Rubbing and Tapping techniques were specifically designed for precise and fluid interaction. A controlled user study shows that these techniques could allow faster interaction with fewer errors when selecting objects on a desktop display. Rubbing, in particular, was designed to work with any touch-screen display, independent of its size and technology, and was implemented in a number of different systems.

The manipulation of content by novice users on 3D displays is discussed in Paper VIII, with a focus on public installations, in which requirements of unencumbered interaction are similar to those of unobtrusive AR. A qualitative study compared seven interaction techniques for basic manipulations of a 3D model through rotation, zoom and translation actions, which can be achieved without pixel-precise target selection. The techniques were based on touch, remote control with a handheld device, eye tracking, gesture tracking, and device tracking. The study indicates a preference among subjects for touch-screen interaction, due to ease-of-use, directness, tactile feedback, and robustness of the technology.

The experimental design of techniques and applications for a new display type resulted in a number of interesting insights [Paper IX, Paper X]. An immaterial projection screen, which was extended for the support of individual content on each side, enabled a large number of people to simultaneously engage in face-to-face interactions while moving freely in front of, or *through* the display. A series of experiments led to a final design, where a 3D tracking system was used for simulated touch-screen behavior, using custom-built miniature markers that could be comfortably held between two fingers. Continuous auditory feedback was designed to vary with the user's distance to the screen and the actions on the surface, to compensate for the lack of tactile feedback. The interface was successfully used by a large number of users of different ages, with no or minimal instruction, during a one-week demonstration at the *ACM SIGGRAPH* Conference in 2005.

Paper XI introduces a framework where a mobile device is used as a precision tool, to enhance both the visual quality of a small portion of a larger digital display, and perhaps even more interestingly, the accuracy with which a user can interact. The work illustrates how limitations of a stationary display can be avoided by extending its capabilities with complementary interaction techniques performed through a tracked mobile touch-screen device.

6.3 Summary of conclusions

This thesis focuses on unobtrusive AR and introduces a number of techniques that address important issues in this area. Many of the presented techniques provide interesting contributions on their own, and can also be generalized to other types of applications and scenarios, as discussed in this thesis, and in full detail in the appended publications.

This work has resulted in five important insights:

1 *Novel, unencumbering display technology can effectively merge virtual imagery with real environments.*

ASTOR (Section 3.1, Paper I and Paper II) describes how autostereoscopic 3D graphics can be merged with an optically direct view of a workspace through the use of a multi-view HOE. It is used to augment machine operation with real-time process data in the discussed examples. POLAR (Section 3.2 and Paper III) enhances small workspaces with 2D graphics and avoids user-worn equipment through remote tracking. Techniques for immaterial displays (Section 5.3, Paper IX and Paper X) show new possibilities for bringing interactive computer graphics into a physical environment.

2 *Exocentric techniques can be advantageous in their ability to sense unmodified devices and visually unaltered objects, while preserving the appearance of the real environment.*

LightSense (Section 4.1 and Paper IV) and SurfaceFusion (Section 4.3 and Paper VI) introduce techniques that enable surfaces to track mobile devices and everyday objects, while preserving their appearance and avoiding modifications.

3 *Hybrid approaches can expand the possibilities for interaction and remove the need for intrusive technology by combining sensing methods, display technologies and interaction techniques.*

POLAR fuses computer vision and basic distance sensing to achieve low-cost 3D tracking of the user (Section 3.2 and Paper III), while SurfaceFusion combines RFID sensing and computer vision to track visually unaltered objects (Section 4.3 and Paper VI). LUMAR leverages the best of egocentric and exocentric techniques by using a hybrid approach (Section 4.2 and Paper V). Section 5.4 and Paper XI describe how a spatially aware handheld display can complement the interaction on a large digital display, with more precise input and improved visual fidelity.

4 *Many different display configurations can benefit from important qualities of touch-based interaction, such as directness, robustness, predictability and feedback, while avoiding the need for user-worn equipment.*

The benefits of touch-screen interaction are discussed in Chapter 5. The results from the study in Paper VIII (Section 5.2) indicate an advantage of touch-screen interaction for simple 3D manipulations on public see-through displays, compared to gestures, head motion and device movement. The positive user feedback from the immaterial display system's simulated touch-screen behavior is also encouraging (Section 5.3, Paper IX and Paper X). Paper XI (Section 5.4) describes how techniques that use touch or button input on a mobile device can overcome the limitations of a large touch-screen. The dissertation also introduces two new techniques, Rubbing and Tapping, and demonstrates how they can expand the expressiveness of general touch-screen interaction (Section 5.1 and Paper VII).

5 *Unobtrusive methods can coexist with other interactive techniques, while emphasizing minimal interference with the real-world procedures that they augment.*

The systems introduced in this dissertation minimize interference with the user's view of the display (Chapter 3), avoid changes to the visual appearance of the physical environment (Chapter 4), and emphasize unencumbered means for interaction (Chapter 5). Their unobtrusiveness also simplifies integration with other techniques to achieve redundant or complementary functionality.

6.4 Future work

We can expect important technological advances in the near future that will have significant impact on the areas that this thesis has discussed.

Developments in imaging technologies will continue to generate smaller, less expensive and better performing components for digital projectors, displays and optics, with direct benefit to the display systems that are introduced in this work.

Increasingly efficient image generation will enable multi-view systems, like ASTOR, that are more compact, support a large number of simultaneous viewers, and exhibit full-color capabilities that improve the quality of virtual content.

The rapid improvements in computational power and camera hardware on mobile devices already allow the use of sophisticated computer vision for sensing, detection and tracking of real-world events, making an even more mobile and low-cost version of POLAR feasible. Advances in performance are also critical to the shift from marker-based techniques, as used in LUMAR, to natural feature tracking [Takacs et al. 2008, Wagner et al. 2008b], for mobile video see-through AR. An increase in sensors that are ubiquitously deployed in the surrounding environment will provide the infrastructure for the type of interaction made possible by LightSense. It is also likely that a wide range of new sensing technologies, such as NFC [NFC Forum 2009], will find their way into the mobile devices, which will lead to improved spatial awareness, while new output mechanisms, such as embedded projectors and tactile displays, will increase their expressiveness.

SurfaceFusion was developed with an expectation that the vast majority of everyday objects will contain embedded passive electronics in the near future, where the increased sophistication in the surfaces that sense them will enable radical new forms of interaction. Many surfaces will also embed spatial optical see-through displays, like ASTOR and POLAR, for intuitive augmentation of real-world processes behind them. Examples could range from laboratory and industrial workspaces to shopping windows and museum showcases [Bimber et al. 2003, Bimber et al. 2001]. Immaterial displays have the potential of bringing virtual images into the real environment, where the unrestricted movement around the content they present provides unique possibilities for collaboration and natural communication.

This research has demonstrated the potential of techniques for visual augmentations and interaction that maintain real-world qualities, preserve the mechanisms of current methods, and leverage the advantages in existing procedures.

The presented techniques have been illustrated with prototypes based on examples from industrial scenarios [Papers I and II], field work [Paper III], map navigation [Papers IV and V], information exchange [Paper VI], archeological visualization [Paper VIII], medical imaging [Paper XI], and various forms of media manipulation [Papers IV–XI]. It is desirable to now apply the insights gathered in this thesis, to develop and extend these prototypes and concepts for real-world use.

The general characteristics of these emerging techniques allow them to be applied to a wide range of real-world problems, and many scenarios are expected to benefit from continued research in combining, developing and refining the techniques introduced in this thesis for even more complex and advanced interactions.

7 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., Schmalsteig, 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, I–452–I–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.

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., Bismpigiannis, 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.