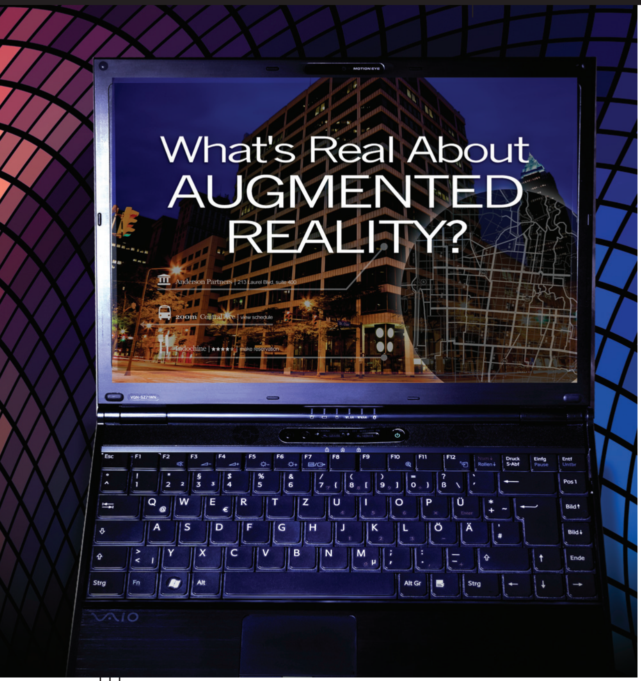


Anywhere Interfaces Using Handheld Augmented Reality

Michael Gervautz, *Qualcomm Research, Vienna*
Dieter Schmalstieg, *Graz University of Technology, Austria*

An investigation of the technology and human factors that drive augmented reality research describes recent developments in handheld AR, concentrating on localization, tracking, interaction, and visualization, and offers several examples illustrating the vast potential and important applications of AR.

We increasingly rely on using computers in our daily activities, but conventional interfaces demand too much attention and using them is often too disruptive. Augmented reality (AR) technology, which superimposes computer-generated images on top of a user's perception of the real world in real time,[1] delivers integrated visual experiences directly related to a place or object that the user views, without any delay.

AR has important applications in fields such as videogaming, interactive marketing and advertising, instructional aids and how-to for use, construction and maintenance, and navigation. However, some major technical obstacles still must be overcome before AR can realize its true potential.

Mobile devices such as tablets and smartphones are rapidly evolving, but they still trail desktop computer adoption and offer only limited capacity for processing and storage. Moreover, mobile devices have ergonomic limitations, such as a small screen size, that make development of satisfactory interfaces difficult. In addition, the need for exact localization of the user limits the circumstances under which AR can be successfully applied. Finally, applications that incorporate AR must not present an overflow of information that confuses rather than informs the user.

HANDHELD AR FORM FACTORS

In the past, AR has been tied to the use of head-mounted displays (HMDs), also known as AR goggles, that include a small screen for presenting computer-generated images to the user. AR uses see-through displays, which combine a live feed from the real world with computer-generated images, either digitally (through a digital video camera and image compositing) or optically (using a half-silvered mirror). Although HMDs leave the hands free, letting users steer the system by simply looking at an object of interest, they have not been commercially successful. This might be due primarily to a lack of attractive, lightweight, or inexpensive devices.

Meanwhile, handheld AR has established itself as an increasingly popular alternative. Tablets and smartphones are lightweight and equipped with high-resolution screens and high-definition cameras. All the necessary hardware is packaged in a convenient, power-efficient system, delivering a video see-through AR experience via a powerful processor and various environment sensors.

Published by the IEEE Computer Society

Handheld AR can be delivered on today's mobile devices as a software-only solution running on hardware that billions of users already own. However, because mobile devices are optimized for low power consumption rather than peak performance, developing software that meets AR's stringent real-time demands on these devices is not easy.

LOCALIZATION AND TRACKING

The most significant technical challenge on mobile devices is measuring the camera's pose (that is, its position and orientation) in relation to relevant objects in the environment. In AR, this measurement is commonly called *tracking*. If the measurement is established in a global rather than a local frame of reference, the procedure is sometimes called *localization*.

Most modern smartphones contain several sensors, including GPS, compasses, linear accelerometers, and gyroscopes. Using these sensors, the device can determine position (from GPS) and orientation (from the compass and by estimating gravity from the accelerometer). However, because of size and power considerations, these sensors offer lower quality than dedicated GPS systems with large antennas. GPS does not work indoors and is only accurate up to an error of tens of meters, even under good conditions. The compass is only exact to tens of degrees and is easily disturbed by metallic objects and electromagnetic interferences. Because accelerometers and gyros work only incrementally, they require constant recalibration from other sources.

A more powerful alternative to using sensors is to apply computer vision methods to determine localization and tracking from the camera image delivered by the AR system. We distinguish several types of computer vision tracking.

Fiducial markers are printed patterns applied to the object of interest. The most popular type of marker is a black-and-white square framing a 2D barcode pattern.[2] This arrangement is easy to detect, and the barcode allows discrimination of multiple markers. *Eye of Judgment*, a Sony PlayStation 3 game, is an example of this type of tracking system.

Natural feature tracking determines the pose relative to a known surface pattern in the environment, as in the advertisement in Figure 1. The device extracts feature points from a video image and compares them to a database that stores known features together with their position in the environment. Given enough successful matches, the device can determine the camera's pose relative to the observed features.

Natural feature tracking requires no special instrumentation of the environment. Given a sufficient degree of surface texture or a known shape, any object can be tracked—for example, a CD cover, a magazine, a statue, or the user's own hand.

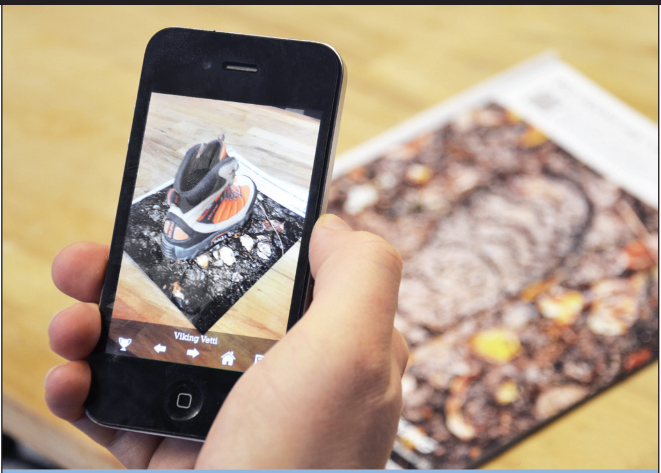


Figure 1. Natural feature tracking. Footwear appears at the top of an advertisement in a magazine.

Global localization systems can determine an object's current position in a large environment, such as a city. Global localization requires that a 3D reconstruction of the environment is available before the user starts the AR application. Such a database can be produced by systematically scanning the environment, as Google, Microsoft, and NAVTEQ currently do. Today, the level of detail required by global localization systems exceeds that of commercially available databases. However, this situation might change quickly.

On-the-fly reconstruction dynamically regenerates the physical environment from a camera stream. These simultaneous localization and mapping (SLAM) systems do not require having a model of the environment available beforehand. This enables new kinds of applications. For example, a user can decorate a living room with virtual furniture that stays put once it is in place. However, given the high computational demands, SLAM is still a large technical challenge for mobile devices and has only been successfully demonstrated for small workspaces.[3]

Using AR in games and object-related applications such as advertising requires deploying computer-vision-based tracking. Tracking purely from sensors is not sufficiently accurate for many interesting applications. For example, sensor-based tracking might identify the street where the user is currently located, but it does not discriminate the items in a shop window, which must be augmented with advertising.

To overcome these difficulties, researchers have combined computer vision techniques with other sensors. A simple approach uses GPS information as a filter to narrow the search area for initialization of vision-based tracking.[4] Given a suitably prepared environment, this approach can even be powerful enough to allow global localization to work on smartphones.[5]

Researchers can also enhance real-time tracking with mathematically advanced sensor fusion based on

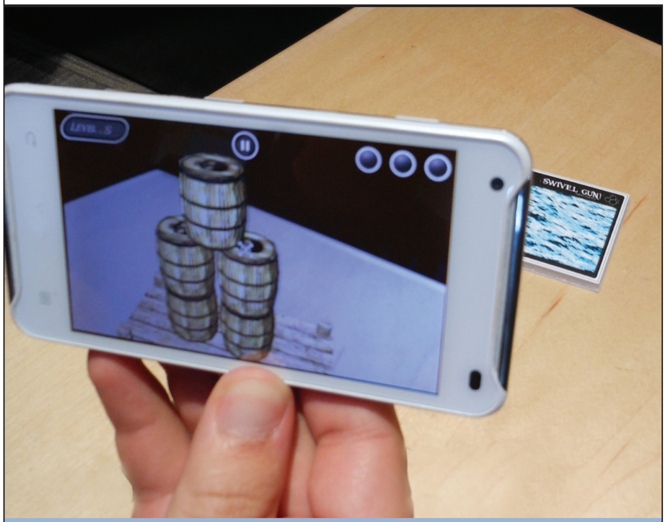


Figure 2. In *Swivel Gun*, a game available on the Android market, users position themselves relative to the target to determine the best shooting position for knocking down all the barrels.

statistics. For example, most computer-vision tracking techniques can be confused by fast rotational movements, which let observed feature points suddenly disappear from the image. This situation can be stabilized by fusing the information from a gyro, which informs the computer-vision-based tracker about the expected rate of rotational movement so that it does not search for features that can no longer be observed.

Another recent idea uses a gravity estimate from a linear accelerometer to determine the orientation of feature points observed in the image.[6] This reduces the degrees of freedom in the search for matching feature points and yields more robust camera pose estimations.

INTERACTION WITH HANDHELD AUGMENTED REALITY

Interaction through handheld devices is sometimes called a *magic lens* because the user observes the physical environment through the device's viewfinder screen, similar to looking through a lens.

Currently, mobile AR uses two kinds of interaction techniques.

Embodied interaction focuses on the device itself (through device movements and the touchscreen) to interact with virtual objects in the scene. Examples include navigation, pan-and-zoom by moving the device relative to the scene, actions triggered by changes in the device's orientation or distance, screen gestures, or tapping on the touchscreen, as Figure 2 shows.

Tangible interaction is based on direct manipulation of known objects—the user reaches into the scene and moves objects that exist in the real world.[7] Actions can be triggered by the appearance or disappearance of objects in the view, a change in an object's position and orientation, the proximity of two or more objects, gestures, or a combination of these. In most use cases, virtual content attached to the objects moves with the object when it is manipulated.

The main challenge is tracking all relevant objects. Typically, objects are identified with fiducial markers or tracked focusing on natural features. Tangible interaction therefore suffers from interobject occlusions and interference from the user's hand. However, some AR applications turn this drawback into a feature by triggering actions if the user's hand occludes a certain spot in the real world.[8]

Ray picking is typically the primary interaction mode. This technique involves selecting an object by casting a virtual ray through a location on the screen into the environment and selecting the first object the ray hits. Although touchscreens support this approach, it represents a level of indirection (users cannot directly grab objects with their hand), and the small screen size of many handheld devices can make it difficult to aim accurately.

Another interaction technique, *layered pie menu*, involves moving the device to select from a hierarchical menu.[9] The user moves the phone closer to or farther away from the display to scroll through menu levels. Rotating or shaking the device opens up further possibilities for setting up commands such as canceling or providing additional parameters.

To observe a physically large environment, it is necessary to move or rotate a handheld device frequently. Ergonomic constraints and the need to keep a line of sight to the display limit the type and number of possible handheld movements. Moving a device by walking is more disruptive than moving or rotating the supportive arm, and it can be difficult to keep the screen in view while physically navigating the environment. Consequently, application designs often aim to minimize physical movement—for example, by only requiring rotation while stationary. However, users might also enjoy the immediacy of physically navigating the environment if it contains interesting physical artifacts that are augmented.

A popular approach for instantaneous AR is the placement of a fiducial marker on a table, which the handheld's camera tracks. Establishing a tracked reference frame also is applicable to multiuser collaboration, in which any manipulation of physical objects is immediately shared. Multiple users can establish a shared space containing common virtual objects, yet still retain an individual view through their personal devices. This shared space is a powerful scenario for collaborative work, such as reviewing 3D architectural design or collaborative (or competitive) games. The availability of ad hoc networking technologies such as Wi-Fi or Bluetooth on mobile devices makes sharing virtual objects relatively straightforward as well.

VISUALIZATION

In addition to localization and tracking, which largely rely on computer vision, AR draws heavily from computer graphics to produce visual output. One important goal of AR is the seamless integration of virtual and real images. In most applications, the scene's virtual objects should be indistinguishable from the real ones. This requires using photorealistic rendering techniques and, more importantly, interaction between the virtual and real objects. Factors such as registration, occlusion, and shadows contribute to successful visual integration.

Registration involves correctly placing and orienting virtual objects in the scene. Registration quality depends primarily on tracking and the camera's calibration relative to the scene.

Correct occlusion between virtual and real objects requires a geometric model of the real objects in the scene, which lets the device determine which virtual objects are visible and which are occluded by real objects. Modeling the real scene in advance can be tedious, but recent commodity depth sensors such as the Microsoft Kinect can determine the necessary geometric information in real time. It seems plausible that future versions of AR-enabled mobile devices will also have depth sensors.

Given a geometric scene model, creating realistic-looking shadows between virtual and real objects is straightforward. Such shadows can be computed on standard graphics hardware using shadow mapping. However, computing shadows also requires the user to estimate the real light sources in the scene. Users must set the positions of light sources manually or determine them using an additional photometric calibration step, which is less desirable. A possible solution is to add light sensors to future devices.

Beyond creating realistic-looking AR scenes, it is also important to determine what augmented information should be shown. In a simple AR application, which just displays additional information related to real objects in the user's environment, the user can be easily overwhelmed with too much information and a cluttered display. To limit the amount of information, an application should consider not only the user's intentions, but also the density of information on the screen.[10]

Labels should be positioned and scaled so they do not obscure important real objects and do not overlap. An algorithm that performs this kind of view management must use a model of the physical environment and its associated virtual content, such as text annotations. Using priorities associated with real and virtual objects, the system searches for an optimal solution to the layout of all virtual objects. To yield real-time performance, this approach exploits temporal coherence—that is, the layout is recomputed only if the viewpoint changes significantly.

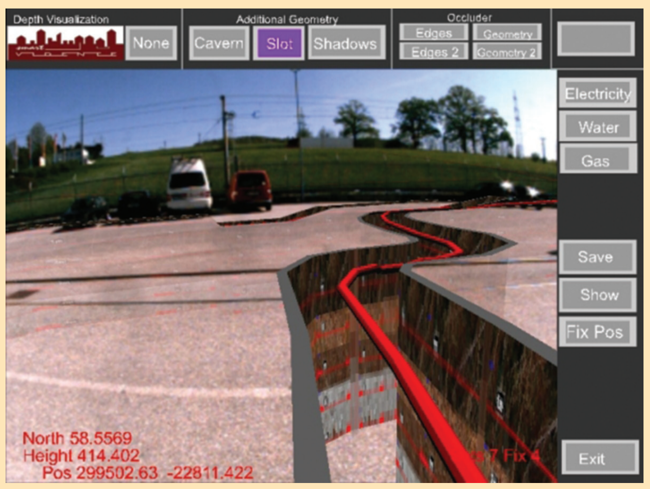


Figure 3. X-ray vision visualization used in the Smart Vidente project to visualize underground water and gas pipes.

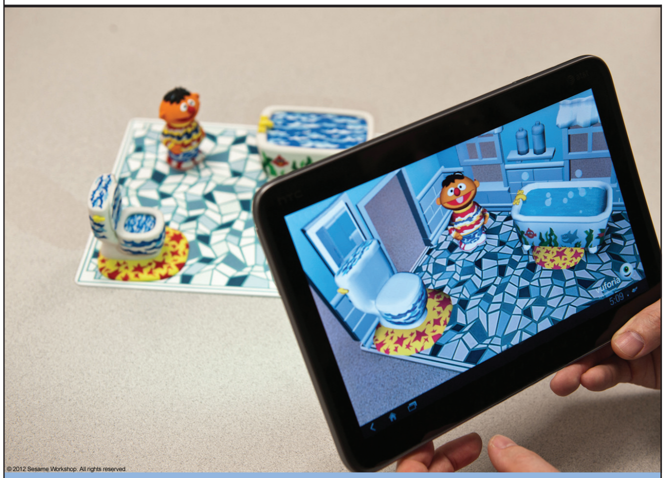
X-ray visualization is a particularly powerful type of AR that simulates looking behind or inside real objects, such as in the application shown in Figure 3. A user can inspect inaccessible locations without changing the viewpoint or otherwise manipulating a real object.

An important issue in x-ray visualization is that the hidden object cannot simply be rendered atop the occluding real object. Such naive rendering would create the impression of the hidden object floating in front of the occluder. To avoid such an unwanted impression, the AR visualization software must synthesize artificial depth cues, which are similar to the cues found in real scenes. For example, a user can add a window-shaped or cut-out region on top of the hidden object to enhance depth perception through partial occlusion. Another approach is to extract strong shape cues, such as contours and strong edges, from the occluding object and impose them atop the hidden object.[11]

APPLICATIONS

Handheld AR technology has great potential for many application areas, such as entertainment, product marketing and sales, education and training, navigation and tourism, and social networking. In the entertainment area, for example, users of AR games benefit from the real-world context and richer and more powerful user interfaces. A tracked mobile device is in fact a six-degrees-of-freedom input device that is already in the user's hands. In gaming, the user can move with this input device to take different physical positions relative to the content—for example, walking around a billiards table, looking out a window, or moving around a target to find the best position from which to throw something.

Consider the mobile game *AR Zombie Gate*, available on Android, which uses a wooden gate graphic as a target. When the user points the mobile device onto the poster,



© 2012 Sesame Workshop. All rights reserved

Figure 4. A prototype playset consisting of individual toy pieces and figurines comes to life when viewed through a tablet, built using Qualcomm's Vuforia platform.

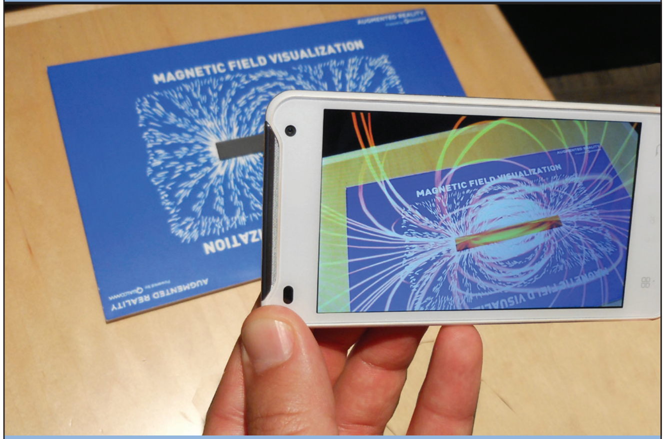


Figure 5. Visitors at San Diego's Reuben H. Fleet Science Center can view and explore magnetic fields in 3D.

the gates open, and a graveyard appears. Zombies start to shuffle from all sides toward the gate. The user must physically shift positions to look into the gate from various angles and shoot the zombies before they reach the gate. This example shows how AR can completely change the user experience. To play the game successfully, users must stand up and move around the target. AR is changing handheld gaming into a Nintendo Wii-like experience.

The toy industry can also benefit from AR, which can make figurines and cards come to life, like the playset in Figure 4. Toy makers can provide directed play and virtual storytelling with their products to enhance learning and skill development.

The largest application opportunity for AR is interactive marketing, advertising, and sales. An AR marketing campaign might include new car models appearing on flyers, cereal boxes featuring games, ketchup bottles displaying recipes, or magazine covers transforming into videos. Different types of AR marketing can add value to the product itself, either in the form of an enhanced product or a more engaging viewer experience. AR lets brands associate their digital content directly with their product or advertising material. Thus, the entire marketing and sales chain can use AR, including advertising material, in-store navigation, in-store experience, product experience, salesperson support, and after-purchase experience.

Moosejaw, a US outdoor clothing company, distributed Moosejaw x-ray, a mobile AR app for iOS and Android that lets mobile users who have downloaded the app see models in their undergarments when pointing their smartphone or tablet at the catalog. Moosejaw reported a 37 percent increase in sales due to the popularity of this app.

In addition to marketing and sales, education and training vertical markets can benefit from the ability to exactly overlay information on real-world objects. Examples range from instruction manuals for appliances, cars, and consumer electronics to education and children's pop-up books. One application provides instructions on how to operate a conference phone by overlaying step-by-step instructions atop the phone. Museums and exhibits can make historic events come to life again and engage the visitor through interactive learning.

The Reuben H. Fleet Science Center in San Diego recently opened the Magnet Field View Finder exhibit, where visitors can learn more about magnetic fields by looking at a magnet through a mobile device. As Figure 5 demonstrates, the screen displays 3D moving magnetic field lines that appear as if they are coming from the magnet. Visitors can move the phone around the magnet to observe the magnetic field lines from various angles, or they can rotate the image and magnet on its turntable.

Navigation and tourist information systems use primarily sensor-based AR approaches, displaying navigation aids and information bubbles on a real-world view. Users benefit from keeping the real world in view (even if they are looking at the navigation screen in a car), getting a much better understanding of where to go than they would by looking on a map. The system displays information about famous buildings or historical landmarks, and users can retrieve more information by pointing the camera at the objects. Technically, tourism information systems are less demanding because in most cases the information does not need to be precisely overlaid and the precision of built-in sensors in today's commercial devices is sufficient. However, many such applications suffer from information cluttering and poor screen estate management.

Although social networking has yet to embrace AR, it could benefit from incorporating the technology to let people leave messages, images, and other content on specific objects or places that others could retrieve later. This content might include reviews on restaurant menus, messages on virtual billboards in public or semipublic spaces, or links to social media pages on conference badges. For such applications, in situ authoring—editing content directly in the real world—is essential. Most mobile devices already have tools for editing images and even videos; however, in situ authoring of 3D objects is still a challenge.

SozializeAR combines social networks with AR. This application was given to each participant at the recent "Emerce e->DAY" conference together with a unique marker that replaced the conference badge. Conference participants assigned their social network address to the marker so others could use the socializeAR app to focus the camera on a badge and see related business information such as business cards or links to social networks. Attendees could take augmented photos of other participants and use the links after the conference.

Although the first commercial AR success stories are now available, much remains to be done in research. For example, tracking works well only in small workspaces; wide area localization and tracking systems are still in the experimental stages. Leading digital map providers are orchestrating massive digitization efforts that will likely change this situation eventually. However, these efforts do not yet involve areas that cannot be accessed by scanning vehicles. Furthermore, they do not cover indoor environments.

Dramatic advances in digitalization technology will be required before the digital models of our environment will allow reliable localization. AR will require not only information on geometry and appearance, but also on the semantics of artifacts in the environment.

Objects and environments that do not have rich surface features are also problematic for localization and tracking. Although vision-based tracking of richly textured surfaces already works well, this is not true for plainly colored objects. Researchers will need to consider other properties, such as shape and contour, to optimize localization and tracking. Thus far, limited progress has been made in this respect.

Finally, even with a flawless AR technical implementation, researchers will still be faced with the problem of insufficient knowledge about AR user interface design. Only a fraction of the user interface design space afforded by 3D interaction with the environment through AR has been explored. We do not yet have a solid understanding of which human factors are important in the design of AR user interfaces. Interaction through a magic lens imposes many constraints on a user's perception and cognition, which must be successfully addressed before AR can become as widespread as the desktop computing metaphor. C

Acknowledgment

This work was partially funded by the Christian Doppler Research Association.

References

1. R.T. Azuma, "A Survey of Augmented Reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, 1997, pp. 355-385.
2. H. Kato and M. Billinghurst, "Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing System," *Proc. Int'l Workshop Augmented Reality* (IWAR 99), IEEE CS, 1999, pp. 85-94.
3. G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces," *Proc. Int'l Symp. Mixed and Augmented Reality* (ISMAR 07), IEEE CS, 2007, pp. 1-10.
4. G. Reitmayr and T. Drummond, "Initialisation for Visual Tracking in Urban Environments," *Proc. Int'l Symp. Mixed and Augmented Reality* (ISMAR 07), IEEE, 2007, pp. 161-172.
5. C. Arth et al., "Real-Time Self-Localization from Panoramic Images on Mobile Devices," *Proc. IEEE Int'l Symp. Mixed and Augmented Reality* (ISMAR 11), IEEE, 2011, pp. 37-46.
6. D. Kurz and S. BenHimane, "Gravity-Aware Handheld Augmented Reality," *Proc. Int'l Symp. Mixed and Augmented Reality* (ISMAR 11), IEEE, 2011, pp. 111-120.
7. I. Poupyrev et al., "Tiles: A Mixed Reality Authoring Interface," *Proc. 7th Conf. Human-Computer Interaction* (INTERACT 01), IFIP, 2001, pp. 334-341.
8. G.A. Lee, M. Billinghurst, and G.J. Kim, "Occlusion-Based Interaction Methods for Tangible Augmented Reality Environments," *Proc. SIGGRAPH Int'l Conf. Virtual Reality Continuum and Its Applications in Industry* (VRCAI 04), ACM, 2004, pp. 419-426.
9. T.R. Hansen, E. Eriksson, and A. Lykke-Olesen, "Mixed Interaction Space—Designing for Camera-Based Interaction with Mobile Devices," *Proc. Conf. Human Factors in Computing Systems, Extended Abstracts* (CHI 05), ACM, 2005, pp. 1933-1936.
10. S. Julier et al., "Information Filtering for Mobile Augmented Reality," *Proc. IEEE 2000 Int'l Symp. Augmented Reality*, IEEE, 2000, pp. 12-15.
11. D. Kalkofen, E. Mendez, and D. Schmalstieg, "Comprehensible Visualization for Augmented Reality," *IEEE Trans. Visualization and Computer Graphics*, vol. 15, no. 2, 2009, pp. 193-204.

__Michael Gervautz__ is a director of business development at Qualcomm Research, Vienna, and a senior lecturer at the University of Technology, Vienna. His research focuses on building up an augmented reality developer network and managing relationships with the universities participating in and contributing to Qualcomm's AR initiatives. Gervautz received a PhD in computer science from the University of Technology, Vienna. Contact him at mgervaut@qualcomm.com.

__Dieter Schmalstieg__ is a full professor and head of the Institute for Computer Graphics and Vision at Graz University of Technology, Austria, where he directs the Studierstube research project on augmented reality. He is also director of the Christian Doppler Laboratory for Handheld Augmented Reality. His research interests include AR, virtual reality, real-time graphics, 3D user interfaces, and ubiquitous computing. Schmalstieg received a Dr. Habilitation degree from Vienna University of Technology. Contact him at schmalstieg@tugraz.at.