

Chapter 1

AUGMENTED REALITY: A PROBLEM IN NEED OF MANY COMPUTER VISION-BASED SOLUTIONS

Gudrun Klinker

Moos 2, D-85614 Kirchseeon, Germany

Klinker@in.tum.de

Abstract Augmented reality (AR) is a technology by which a user's view of the real world is augmented with additional information from a computer model. It constitutes a very promising new user interface concept for many applications. Yet, AR applications require fast and accurate solutions to several very complex problems, such as user and real object tracking, occlusion and reflection handling, as well as virtual user motion. Currently, computer vision based solutions are considered to be among the most promising approaches towards solving these issues. This paper discusses several such AR issues and potential solutions.

1. INTRODUCTION

Augmented reality (AR) is a technology by which a user's view of the real world is augmented with additional information from a computer model [4, 26]. Users can work with and examine real 3D objects while receiving additional information about those objects or the task at hand. Rather than pulling the user into the computer's virtual world, AR brings information into the user's real world, thereby building upon people's visual and spatial skills.

AR constitutes a very promising new user interface concept for many applications, e.g., in medicine [6, 13, 15, 34], exterior construction [22], interior design [1, 32], the assembly, maintenance and repair of complex technical objects [8, 11, 28], and games [21, 32]. With the increasing availability of virtual prototypes, industries can benefit from AR during all phases of the life cycle of a product, integrating the computer-generated information with the physical environment.

Yet, AR applications require fast and accurate solutions to several very complex problems, such as user and real object tracking, occlusion and reflection handling, as well as virtual user motion. Currently, computer vision based solutions are considered to be among the most promising approaches towards solving these issues. This paper discusses several such AR issues and potential solutions.

2. TYPICAL AR CONFIGURATIONS

Several different arrangements of display and tracking devices are currently in use depending on the purposes and constraints of different applications.

- *Head-mounted see-through AR*

Head-mounted, see-through displays are the prototypical setup that people envision first when talking about AR [20]. Users wear a head-mounted semi-transparent display through which they can see the world like through a set of sunglasses. Shown inside the glasses are three-dimensional objects. They are rendered according to the current vantage point of the user such that the virtual objects seem to co-exist side-by-side with real objects in the scene. When users move their heads, the virtual objects maintain their position in the world. Magnetic or acoustical trackers, gyroscopes, or mini-cameras [20] are attached to the HMD in order to track its position in the scene.

- *Head-mounted video-feed-through AR*

Video-feed-through HMDs are used in a fashion very similar to see-through displays, except that the display is opaque in this case. The real scene is recorded by one or two (stereo) video cameras attached to the HMD. The video signal is displayed inside the HMD, thereby showing users the real world around them. Compared to the see-through solution, precise augmentations of such video-based reality are easier to achieve since the augmentations can be inserted into a suitable recent video image whenever the results become available. Yet, implementations tend to lack the immediacy of a truly immersive see-through setup due to the offset between the cameras and the users eyes, the limited resolution of the video signal, and the time lag between recording, processing, augmenting and displaying the video signal. Nevertheless, if the geometric offset and time lag can be kept small, video-feed-through AR is very successful. It is thus a very common setup in demonstrations.

- *Monitor-based AR*

In monitor-based setups [20], users view video-based augmentations on a monitor rather than in an opaque HMD. The video images are recorded by a mobile camera which could be anywhere – on the users head, in his hands, on a tripod, on a wall, on a moving robot, or with a collaborator. By decoupling the display coordinate system from the camera coordinate system, AR can thus be related to collaborative and telepresence concepts.

- *Portable monitors*

Flat, portable AR monitors [29], provide hand-held video augmentations. A camera on the back of the device records the scene behind the screen. The picture is augmented and displayed on the monitor, generating the illusion of a picture frame inside which the real world is augmented with virtual objects.

- *Combinations of various wearable devices*

In realistic AR applications, it is likely that no one display modality will be used by itself. Rather, users will prefer having a choice between several display devices each for its own purpose. Feiner et al. are combining the immersive 3D display qualities of a HMD with better 2D resolutions of a portable monitor, allowing users to drag pieces of information involving extended amounts of text from the HMD to a Netscape-based browser in their hand [11].

3. USER TRACKING

Precise user tracking is one of the key issues in AR since it determines the immersive quality and credibility of the augmentations. Virtual objects have to be rendered from a virtual camera perspective that is identical with the current vantage point of the user.

Various carefully calibrated sensing devices have been used for this purpose in the past [2, 5, 33]. Commercial tracking devices such as magnetic trackers and active LED-systems can be used [11, 33]. But the precision and the working range of such devices are insufficient for most AR applications. Thus, research is now focussing on computer vision based methods which promise untethered, higher-precision applicability of AR.

However, the vision-based approaches cannot be arbitrarily sophisticated and complex. They have to perform in real-time and they have to be very robust, degrading gracefully and recovering fast when they fail. Users wearing a HMD cannot be expected to restrain their natural head motions severely while interacting with the system, neither can they be expected to stay motionless while the system is trying to recover from

failures. Thus, vision-based approaches have started from very simple, limited setups. They are now progressing towards more sophisticated approaches.

3.1 TRACKING OPTICAL MARKERS AT KNOWN 3D POSITIONS

Currently, the most successful vision-based approaches track optical markers in indoor laboratory demonstrations (Figure 1) [23, 24, 27, 30, 31, 34] sometimes combined with information from other tracking modalities [22, 27, 30]. Optical markers provide enough simplifying assumptions to achieve real-time tracking performance. Typically, the markers are specially designed to be easily recognizable. Examples are polka dots of varying sizes placed in pre-designed patterns on large form boards for Boeing’s wire harness assembly application [8], large dark rectangles with unique identification labels at Fraunhofer-IGDs doorlock installation demonstration [20, 28, 31], and concentric, multi-colored rings at Neumanns lab at USC [27] and in Fuch’s lab at UNC [30]. The placement of such markers is quite a severe restriction to the overall applicability of the system. It is tedious to install a significant number of them necessary to warrant a robust operation of the system. In some applications (especially outdoors), such an approach is completely impractical.

Yet, the current demonstrators are a good starting point to begin experimenting with more general concepts.

3.2 TRACKING NATURAL FEATURES AT KNOWN 3D POSITIONS

As a first step towards generalizing vision-based, real-time user tracking setups, several approaches are now tracking features that occur naturally at known positions in the scene [27, 31]. Stricker et al. track three-dimensional lines, such as the edges of doors and walls as well as strongly bent edges on car bodies which tend to show up as slim bright lines due to specular reflections [28, 31]. The 3D positions and orientations of these lines are measured ahead of time and provided to the tracker in a data file. After initialization, the system predicts the position of such known lines in each image according to the most recently known camera position and searches for them in a local image area. From several non-collinear lines, it then computes the new camera position. The approach works very well when provided with an initial calibration. Yet, it requires more detailed information at startup time, such as black squares with unique identification marks.

For outdoor exterior construction applications, Klinker et al. use a different approach [19, 20]. It is not possible to install special markers ubiquitously in vast outdoor environments. Besides, the markers would have to be huge to be visible from long distances. Instead, natural landmarks must be used as calibration marks. Klinker et al currently select easily visible landmarks (such as church steeples, tips of power line poles, bridge pillars and river banks) off-line and determine their three-dimensional position from other sources. When calibrating and augmenting a video sequence of the area, they interactively indicate the location of such landmarks in the initial image. The landmarks are then tracked in subsequent images using normalized correlation such that the video sequence can be calibrated with minimal user intervention.

Although these approaches dont require scene modifications, they still depend on complicated setup procedures involving the three-dimensional measurement of natural landmarks and their initial identification in an initial image. Model-based computer vision approaches are needed in order to automatically associate large sets of known 3D landmarks with currently visible image features using suitable feature properties. Such object recognition has to work in real-time since the initialization phase of an AR system cannot require the user to remain motionless for extended periods of time whenever the system loses track and requires a re-initialization. Alternatively, several hybrid approaches address the initialization problem by combining optical trackers with other technology, such as magnetic trackers [30], GPS [22], or gyroscopes [27].

3.3 TRACKING NATURAL FEATURES AT UNKNOWN 3D POSITIONS

The logical next step involves striving for optical tracking solutions that do not require the off-line three-dimensional measurement of landmarks. Mendelsohn et al. use an uncalibrated stereo vision approach for indoor applications [25]. They do place special, easily identifiable targets (black pentagons) in the scene, but they do not determine their 3D location. Using the constraint that each target pentagon lies in a plane, they are able to compute a highly accurate, metric scene reconstruction after tracking the targets for a few images, gaining increasingly precise estimates the longer the tracking proceeds. The system currently runs off-line on prerecorded sequences involving up to 1200 images (40 seconds).

Neumann et al. are presenting first steps towards tracking unknown landmarks outdoors [27]. They track both interesting feature points and entire regions using differential-based local optical flow estimation.

A multi-scale estimation strategy iteratively fits region and point motion estimates to an affine motion model until they agree. The approach is able to dramatically improve the precision of a hybrid inertial (3 DOF) and vision tracker. Treating the fusion of inertial and image tracks as a 2D image stabilization problem, the hybrid tracker is able to annotate three-dimensional objects such as cars and entrance gates to parking lots very reliably after the intrinsic camera parameters and an initial camera position have been determined off-line. So far, the algorithm is used off-line since it does not yet operate in real-time.

3.4 MODELING OF USER MOTION

Due to numerical instabilities of the camera calibration involving a large number of parameters in a system of non-linear equations, it is not advisable to calibrate incoming pictures independently of each other one at a time. Calibrations are likely to jump back and forth between solutions at different local maxima of the parameter space. To introduce a temporal constraint, physical motion models have to be added to the system.

It is very customary in computer vision applications to use Newtonian motion models which describe the velocity and acceleration of a point in space over time. Such models have been applied to the problem of tracking user head motions [3, 23]. Yet, due to the randomness of user head motion, it is hard to predict future head motion from history looking merely at velocity and acceleration. Impuls seems to play a significant role. Sophisticated motion prediction models [23] so far have encountered serious problems tracking a user-held or user-worn camera. The prediction of 3D user motion is too slow on todays computers to quickly react to a users head rotations (e.g., during a quick glance to the side or a head shaking motion), generating an effect of "swimming" off track. If the head motion is very abrupt, the system never recovers from its "detour".

Thus, current approaches tend to be much more pragmatic, ignoring 3D motion constraints altogether and simplifying motion analysis down to a 2D, image-based analysis of local pixel motion coupled with numerical initialization heuristics to ensure that calibrations of consecutive images are likely to settle at that same local extremum of the parameter space. Such approaches are surprisingly successful since they are able to perform at close-to-real-time speeds on current computer systems [20, 31].

4. 3D SCENE MODELLING FOR OCCLUSION AND REFLECTION HANDLING

In order to augment real worlds with virtual objects, the virtual objects need to be integrated seamlessly into the environment. They have to behave in physically plausible manners: they occlude or are occluded by real objects, they are not able to move through other objects, and they cast shadows on other objects. To this end, AR applications require a very accurate model of the real environment (a reality model).

4.1 OCCLUSION HANDLING BETWEEN REAL AND VIRTUAL OBJECTS

Occlusion relationships between real and virtual objects can be computed efficiently by the geometric rendering hardware of today's graphics workstations. By first rendering the AR reality model transparently, the z-buffer of the rendering engine is initialized to account for the distances of real objects from the user. Since the model is drawn transparently, it remains invisible - and thus does not obstruct the view onto the real world (or its video picture). When the virtual objects are rendered subsequently, only those are drawn that are closer than any real objects to the user.

4.2 GEOMETRIC REALITY MODELS

Occlusion handling in AR applications requires geometrically precise descriptions of the real world. Similar descriptions are also used in Virtual Reality (VR) applications. Yet, for AR, such reality models generally don't need to be as complex as those used for VR. VR models are expected to synthetically provide a realistic immersive impression of the simulated environment. Thus, the descriptions of surface textures are crucial. AR, on the other hand, can rely on live optical input (or a live view of the real world) to provide a very high sense of realism. For occlusion handling, the AR reality model only needs to describe the surface shape.

However, AR reality models have to be much more precise than VR models. Since an immersive VR system cuts users off from reality, users can only gain a qualitative impression whether or not the objects are modelled correctly. In AR, on the other hand, users have an immediate quantitative appreciation of the extent of disagreements between the reality model and the real world, since virtual objects then won't integrate into it seamlessly.

Reality models for AR applications are currently acquired in many different ways depending on the requirements and data sources of a particular application. In many cases, 3D scene information is provided manually by the user, e.g., in simple text files describing the locations of markers on planar laboratory walls. In other cases, pre-existing CAD models of man-made objects can be used [28]. Sometimes, maps and geodesic measurements of landscapes and cityscapes are available [19]. Most of the time, however, current AR systems ignore issues related to occlusion handling altogether because appropriate models of cluttered scenes are hard to generate. The automatic construction of scene models by computer vision techniques is a promising way to help alleviate this problem.

The construction of 3D scene descriptions is a long-standing issue in computer vision research. Its application to AR applications is being demonstrated increasingly. Semi-automatic approaches towards generating architectural models from images have been reported by Debevec et al. [9] and by Faugeras [10]. Kanade et al. have built a setup at Carnegie Mellon involving 51 calibrated video cameras that are arranged on the periphery of a "3D Dome" [17]. Dynamic scenes inside the dome such as multi-person ballgames are recorded by all cameras. Applying a trinocular vision approach to each group of three cameras, the system then computes, off-line, a fairly accurate dynamic 3D model of the game, a "virtualized reality". Hirose and Tanikawa are using a van with 8 cameras, a GPS sensor, a terrestrial magnetism sensor and a 3-axis angle sensor on its roof to drive through the streets of Tokyo, recording image sequences and positional data [14]. Back in the laboratory, they invite visitors to a virtual walk through the recorded parts of the city using image-based rendering techniques.

4.3 REFLECTION HANDLING BETWEEN REAL AND VIRTUAL OBJECTS

In addition to handling occlusions between virtual and real objects, AR-systems also need to be able to handle photometric relationships. If the positions of all light sources and the reflective material properties of all real objects are known, inter-reflections and shadows from real objects onto virtual objects can be taken into account when rendering the virtual ones. Furthermore, the video pixels showing the real objects can be modified to simulate the influence of virtual light sources as well as inter-reflections and shadows from virtual objects [12, 30].

4.4 PHOTOMETRIC REALITY MODELS

Quintessential to reflection handling in AR applications is the availability of photometric reality models. The models need to indicate the location of all direct and indirect light sources in the real scene. To this end, Ikeuchi et al. have developed a photometric model-based rendering method [16]. From calibrated input images of real objects and a previously obtained geometric reality model, it obtains reflectance parameters of the real objects by tracking individual small spots on the 3D surface through the image sequence and relating their color changes to user motion: in principle, pixels related to a particular surface area of a 3D object should remain constant when the user moves through the scene. Using a color histogram analysis, Ikeuchi et al. attribute color variations to sensor noise and to occasional specular reflections. After discounting such effects, they can attribute an intrinsic reflective color property to each small surface area of real objects.

4.5 PHYSICAL INTERACTIONS BETWEEN REAL AND VIRTUAL OBJECTS

For an augmented world to be realistic the virtual objects not only have to interact optically with the real world, but also mechanically. This applies to virtual objects when animated or manipulated by the user. For example, a virtual chair shouldn't go through walls when it is moved, and it should exhibit gravitational forces [7]. Given a reality model, this behavior can be achieved using collision detection and avoidance systems that are known from Virtual Reality systems [36].

These two laws make up the most important physical constraints. A full physical simulation including more aspects of the interaction between real and virtual objects, such as elastic behavior and friction, would be desirable. For off-line applications this is possible if enough information about the virtual objects and a complete enough reality model is available. For real-time applications most simulation systems are not fast enough. Yet, even simple implementations of the above rules will make the system much more realistic.

5. DIMINISHED REALITY

Many AR applications require that existing structures be removed before new objects are added to the scene. For example, interior design or refurbishment typically doesn't start from an empty room. Rather, the area is cluttered with all kinds of furniture or structure that needs to be moved or removed. Similarly, many exterior construction projects

require the removal of old buildings before new ones are put into their place. Medical and machine repair applications as well typically require x-ray vision skills, allowing the user to ignore current structures in the foreground to focus on what is behind it. Thus, just as important as augmenting reality is technology to diminish it.

The removal of foreground objects requires a model of the objects behind them, i.e., a three-dimensional reality model of the entire area. Modelling currently unseen parts of the scene is a very complex issue. There is no general solution to this problem since we cannot know what a dynamically changing world looks like behind an object at any specific instant in time - unless another camera can see the occluded area. Yet, some heuristics can be used to solve the problem for various realistic scenarios. We can use morphological operators to extrapolate properties from surrounding "intact" areas into the occluded region (e.g: in a cloudy sky) [22]. Furthermore, when a building is to be removed from a densely populated area in a city, particular static snapshots of the buildings behind it could be taken and integrated into the reality model to be mapped as textures into the appropriate spaces of the current image.

For video sequences, computer vision techniques can be used to suitably merge older image data with the new image. Faugeras et al. have shown that soccer players can be erased from video footage when they occlude advertisement banners: For a static camera, changes of individual pixels can be analyzed over time, determining their statistical dependence on camera noise. When significant changes (due to a mobile person occluding the static background) are detected, "historic" pixel data can replace the current values [37].

In more general schemes involving mobile cameras, such techniques can lead towards incremental techniques to diminish reality. While moving about in the scene, users and cameras see parts of the background objects. When properly remembered and integrated into a three-dimensional model of the scene, such "old" image data can be reused to diminish newer images, thus increasingly effacing outdated objects from the scene as the user moves about. Stricker uses geometric constraints to compute pixelwise correspondences between regions in several images that outline a particular object [19]. From such correspondances, he traces specific points on the object across all images and then decides in which images it is visible or occluded. Accordingly, occluded pixels can be replaced by visible ones, effectively removing the occluding object from the image.

6. MOBILE OBJECT TRACKING

In realistic AR applications, real objects in the scene cannot be expected to remain stationary throughout the entire course to the application. Actions or instructions issued by the computer as augmentations in the HMD or on the monitor cause the user to perform actions changing the real world - which, in turn, prompt the computer to generate new, different augmentations. For example, a machine repair task typically includes a partial disassembly of the machine, followed by a replacement of some parts and a subsequent reassembly.

To maintain correct virtual-real occlusion relationships, the AR application has to keep track of all moving objects in real-time and update the reality model accordingly. Furthermore, the AR-system has to understand the meaning of the users actions such that it can react and propose the next step of a repair procedure or indicate that a mistake has been made.

Several prototypes of two-way human-computer interaction involving limited degrees of reality tracking with non-optical means have been demonstrated. In Feiner et al.'s space frame construction system [11], selected new struts are recognized via a bar code reader, triggering the computer to update its visualizations. In a mechanical repair demonstration system, Breen et al. use a magnetically tracked pointing device to ask for specific augmentations regarding information on specific components of a motor [7].

6.1 DETECTION AND TRACKING OF OBJECTS WITH SPECIAL MARKERS

Klinker et al. currently use rather pragmatic, simple solutions towards reality tracking that can be run approximately in real-time [21]. In a "mixed mockup" demonstration involving the insertion of virtual buildings in a miniaturized scene they attach special markers to mobile objects, such as toy buildings. Both virtual and real buildings can be moved to experiment with different house arrangements.

The current approach assumes that unique markers are attached to all mobile real and virtual objects and that they are manipulated on a set of known surfaces. The marks can then automatically be identified and their 3D position and orientation can be determined by intersecting the rays defined by the positions of the squares in the image with the three-dimensional surfaces on which they lie. If the markers are manipulated in mid-air rather than on a known surface, more sophisticated approaches are needed, such as stereo vision or the computation of the 3D target

location from its projected size and shape. The current system does not apply the latter concepts due to real-time and robustness considerations.

Attaching markers to a few real objects is an elegant way of keeping track of objects even when both the camera and the objects move. The objects can have arbitrary textures that don't even have to contrast well against the background - as long as the markers can be easily detected. Yet, the markers take up space in the scene; they must not be occluded by other objects unless the attached object becomes invisible as well. Furthermore, this approach requires a planned modification of the scene which generally cannot be arranged for arbitrarily many objects. Thus it works best when only a few, well-defined objects are expected to move.

6.2 DETECTION OF OBJECTS USING OBJECT MODELS

Klinker et al. show more general reality tracking schemes in the context of an augmented Tic Tac Toe game [21]. The user and the computer alternate placing real and virtual stones on the board. When the user has finished a move, he waves his hand past a 3D "Go" button. The computer then scans the image area containing the board. If it finds a new stone, it plans its own move and places a new virtual cross on the board. If it could not find a new stone or if it found more than one, it asks the user to correct his placement of stones.

The Tic Tac Toe system uses model-based object recognition principles to find new pieces on the board. Due to the image calibration the location of the game board in the image is known, as well as all nine valid positions for pieces to be placed. Furthermore, the system has maintained a history of the game. It thus knows which positions have already been filled by the user or by its own virtual pieces. It also knows that the game is played on a white board and that the user's stones are red. It thus can check very quickly and robustly which tiles of the board are covered with a stone, i.e. which tiles have a significant number of pixels that are red rather than white. Error handling can consider cases in which users have placed no new stone or more than one new stone - or whether they have placed their stones on top of one of the computer's virtual stones.

Using a model-based object recognition approach is a more general approach than the one based on special markers since it does not require scene modifications. Yet, the detection of sophisticated objects with complex shape and texture has been a long standing research problem in computer vision, consuming significant amounts of processing

power. Real-time solutions for arbitrarily complex scenes still need to be developed.

Thus, the appropriate choice of algorithm depends on the requirements of an application scenario. In many cases, hybrid approaches including further information sources such as stationary overhead surveillance cameras that track mobile objects are most likely to succeed.

6.3 PEOPLE AND HAND TRACKING

During a repair procedure, the user typically manipulates real objects with his hands or with special tools. Potentially, he is also assisted by coworkers. Thus, for large parts of an AR application, moving hands and people will be visible within the scene. Human hand-eye coordination requires that the users hands are integrated particularly well into the augmented world: when the user touches a virtual object or some virtual positioning aid like a pointer or a virtual yard stick on the floor, users have to receive immediate and precise feedback as to where their hands are in relationship to the virtual objects. Thus occlusion handling has to work well. To this end, hands and people have to be tracked in real-time.

Yokoya et al have developed a stereo-based vision system which uses two camera on an HMD [35]. From optical markers at unknown scene locations, the system tracks user head motion. In addition, it also tracks the motion of the users hands, determining their current position by stereo triangulation on an SGI Onyx2 IR. The system is able to perform the stereo approximately in real-time due to its heuristics for quickly detecting skin color in each image and thereby pruning the time-consuming stereo matching process significantly. Kanade et al have reported real-time stereo vision performance for arbitrary objects using a stereo vision machine based upon special-purpose hardware [18]. Kanades "3D Dome" demonstrations of constructing a dynamic 3D model of a several people playing a ball game together is another example of a reality tracking system [17] - yet, it currently still has to rely on 51 precoded video sequences and doesnt run in real-time yet.

7. VIRTUAL USER MOTION AND TELEPRESENCE

In addition to analyzing real user motion and real and virtual object motion, AR applications are likely to also be confronted with requirements of virtual user motion. While looking at an augmented scene and working in it in a reality-based coordinate frame, users may want to temporarily take a side-step from reality to look at the world from a different perspective. For example, while discussing a planned new building at a

construction site, users might be interested in getting a birds-eye view of the location. Similarly, a mechanic may want to temporarily step into his colleagues shoes (or HMD) during a complex repair effort of a big machine or look at details of the machine through a magnifying glass.

In this sense, augmented reality and virtual reality are not two discrete alternatives but rather part of a spectrum of mixed realities [26] with full virtual reality on one end and full physical reality on the other. Augmented Reality is in the middle, combining the best of both worlds. But sometimes it might be desirable to lean more in one direction or the other.

To leave reality behind without getting lost, users need a smooth transition path from their current position to virtual places and back. Whenever possible, available real data should be integrated into the virtual presentations. To this end, 3D scene descriptions are essential. As discussed in the previous sections, computer vision techniques lend themselves to generating and dynamically updating such descriptions from the live image data being obtained while the user moves about.

8. SUMMARY

Augmented reality is an exciting new technology with the potential of becoming a "killer application", combining many aspects of computer science into well-designed and well-tuned systems. One of the most essential ingredients of such a system is intelligent sensor analysis technology, such as provided by computer vision research. This paper has listed many areas in which augmented reality systems can benefit greatly from concepts and approaches that are common within the computer vision community.

Acknowledgments

Most of this work was conceived and developed while the author was at the Fraunhofer Project Group for Augmented Reality at ZGDV. Close collaborators were Didier Stricker, Dirk Reiners, Eric Rose, and Dieter Koller.

References

- [1] K.H. Ahlers, A. Kramer, D.E. Breen, P.-Y. Chevalier, C. Crampton, E. Rose, M. Tuceryan, R.T. Whitaker, and D. Greer. Distributed augmented reality for collaborative design applications. In *Proceedings of Eurographics '95*, 1995. Also available as technical report ECRC-95-03 (<http://www.ecrc.de>).
- [2] R. Azuma and G. Bishop. Improving static and dynamic registration in an optical see-through HMD. In *Proc. Siggraph'94*, pages 197–204, Orlando, FL, July 1994. ACM.
- [3] R. Azuma and G. Bishop. A frequency-domain analysis of head-motion prediction. In *Proc. Siggraph'95*, pages 401–408, Los Angeles, CA, August 1995. ACM.
- [4] R.T. Azuma. A survey of augmented reality. *Presence, Special Issue on Augmented Reality*, 6(4):355–385, August 1997.
- [5] R.T. Azuma. The challenge of making augmented reality work outdoors. In *Proc. ISMR'99 (1. International Symposium on Mixed Reality)*, pages 379–390, Yokohama, Japan, March 1999.
- [6] J.W. Berger and D.S. Shin. Computer-vision-enabled ophthalmic augmented reality: A pc-based prototype. In *Proc. IEEE and ACM IWAR'98 (1. International Workshop on Augmented Reality)*, pages 19–30, San Francisco, November 1998. AK Peters.
- [7] D.E. Breen, E. Rose, and R.T. Whitaker. Interactive occlusion and collision of real and virtual objects in augmented reality. Technical Report ECRC-95-02, ECRC, Arabellastr. 17, D-81925 Munich, <http://www.ecrc.de>, 1995.
- [8] D. Curtis, D. Mizell, P. Gruenbaum, and A. Janin. Several devils in the details: Making an ar app work in the airplane factory.

- In *Proc. IEEE and ACM IWAR'98 (1. International Workshop on Augmented Reality)*, pages 47–60, San Francisco, November 1998. AK Peters.
- [9] P.E. Debevec, C.J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *Proc. SIGGRAPH*, pages 11–20, New Orleans, August 4-9 1996. ACM.
- [10] O. Faugeras, S. Laveau, L. Robert, G. Csurka, and C. Zeller. 3D reconstruction of urban scenes from sequences of images. In A. Gruen, O. Kuebler, and P. Agouris, editors, *Automatic Extraction of Man-Made Objects from Aerial and Space Images*. Birkhauser, 1995.
- [11] S. Feiner, B. MacIntyre, and T. Hoellerer. Wearing it out: First steps toward mobile augmented reality systems. In *Proc. ISMR'99 (1. International Symposium on Mixed Reality)*, pages 363–377, Yokohama, Japan, March 1999.
- [12] A. Fournier. Illumination problems in Computer Augmented Reality. *Journee Analyse/Synthese d'Images (JASI)*, pages 1–21, January 1994.
- [13] W.E.L. Grimson, G.J. Ettinger, S.J. White, P.L. Gleason, T. Lozano-Perez, W.M. Wells III, and R. Kikinis. Evaluating and validating an automated registration system for enhanced reality visualization in surgery. In *Proc. of Computer Vision, Virtual Reality and Robotics in Medicine (CVRMed '95)*, pages 3–12, Nice, France, April 1995. IEEE.
- [14] M. Hirose, T. Tanikawa, and T. Endo. Building a virtual world from the real world. In *Proc. ISMR'99 (1. International Symposium on Mixed Reality)*, pages 183–197, Yokohama, Japan, March 1999.
- [15] W.A. Hoff. Fusion of data from head-mounted and fixed sensors. In *Proc. IEEE and ACM IWAR'98 (1. International Workshop on Augmented Reality)*, pages 167–182, San Francisco, November 1998. AK Peters.
- [16] K. Ikeuchi, Y. Sato, K. Nishino, and I. Sato. Photometric modeling for mixed reality. In *Proc. ISMR'99 (1. International Symposium on Mixed Reality)*, pages 147–163, Yokohama, Japan, March 1999.
- [17] T. Kanade, P. Rander, S. Vedula, and H. Saito. Virtualized reality: Digitizing a 3D time varying event as is and in real time. In *Proc.*

- ISMR'99 (1. International Symposium on Mixed Reality)*, pages 41–57, Yokohama, Japan, March 1999.
- [18] T. Kanade, A. Yoshida, K. Oda, H. Kano, and M. Tanaka. A stereo machine for video-rate dense depth mapping and its new applications. In *Proc. 15th Computer Vision and Pattern Recognition Conference (CVPR'96)*, San Francisco, June 18-20 1996. IEEE.
- [19] G. Klinker, D. Stricker, and D. Reiners. The use of reality models in augmented reality applications. In *European Workshop on 3D Structure from Multiple Images of Large-scale Environments (SMILE)*, Freiburg, Germany, June 6-7 1998.
- [20] G. Klinker, D. Stricker, and D. Reiners. Augmented reality: A balancing act between high quality and real-time constraints. In *Proc. ISMR'99 (1. International Symposium on Mixed Reality)*, pages 325–346, Yokohama, Japan, March 1999.
- [21] G. Klinker, D. Stricker, and D. Reiners. An optically based direct manipulation interface for human-computer interaction in an augmented world. In *Proc. Eurographics Workshop on Virtual Environments (EGVE)*, Vienna, Austria, June 1999.
- [22] G. Klinker, D. Stricker, and D. Reiners. Augmented reality for exterior construction applications. In W. Barfield and T. Caudell, editors, *Augmented Reality and Wearable Computers*. Lawrence Erlbaum Press, 2000.
- [23] D. Koller, G. Klinker, E. Rose, D. Breen, R. Whitaker, and M. Tuceryan. Automated camera calibration and 3d egomotion estimation for augmented reality applications. In *7th Int'l Conf. on Computer Analysis of Images and Patterns (CAIP-97)*, Kiel, Germany, September 10–12, 1997, G. Sommer, K. Daniilidis, and J. Pauli (eds.), Lecture Notes in Computer Science **1296**, Springer-Verlag, Berlin, Heidelberg, New York, 1997.
- [24] J.P. Mellor. Enhanced reality visualization in a surgical environment. Master's thesis, Dept. of Electrical Engineering, MIT, 1995.
- [25] J. Mendelsohn, K. Daniilidis, and R. Bajcsy. Constrained self-calibration for augmented reality registration. In *Proc. IEEE and ACM IWAR'98 (1. International Workshop on Augmented Reality)*, pages 201–208, San Francisco, November 1998. AK Peters.
- [26] P. Milgram and Jr H. Colquhoun. A taxonomy of real and virtual world display integration. In *Proc. 1st International Symposium*

- on *Mixed Reality (ISMR'99)*, pages 5–30, Yokohama, Japan, March 1999.
- [27] U. Neumann, S. You, Y. cho, J. Lee, and J. Park. Augmented reality tracking in natural environments. In *Proc. 1rst International Symposium on Mixed Reality (ISMR'99)*, pages 101–130, Yokohama, Japan, March 1999.
- [28] D. Reiners, D. Stricker, G. Klinker, and S. Mueller. Augmented reality for construction tasks: Doorlock assembly. In *Proc. IEEE and ACM IWAR'98 (1. International Workshop on Augmented Reality)*, pages 31–46, San Francisco, November 1998. AK Peters.
- [29] J. Rekimoto. Navicam: A magnifying glass approach to augmented reality. *Presence, Special Issue on Augmented Reality*, 6(4):399–412, August 1997.
- [30] A. State, G. Hirota, D.T. Chen, W.F. Garrett, and M.A. Livingston. Superior augmented reality registration by integrating landmark tracking and magnetic tracking. In *Proc. SIGGRAPH*, pages 429–438, New Orleans, Aug 4-9 1996. ACM.
- [31] D. Stricker, G. Klinker, and D. Reiners. A fast and robust line-based optical tracker for augmented reality applications. In *submitted to 1. International Workshop on Augmented Reality (IWAR'98)*, pages 129–145, San Francisco, November 1998. AK Peters.
- [32] H. Tamura, H. Yamamoto, and A. Katayama. Steps toward seamless mixed realitz. In *Proc. ISMR'99 (1. International Symposium on Mixed Reality)*, pages 59–59, Yokohama, Japan, March 1999.
- [33] M. Tuceryan, D. Greer, R. Whitaker, D. Breen, C. Crampton, E. Rose, and K. Ahlers. Calibration requirements and procedures for a monitor-based augmented reality system. *IEEE Transactions on Visualization and Computer Graphics*, 1, September 1995.
- [34] M. Uenohara and T. Kanade. Vision-based object registration for real-time image overlay. In *Proc. of Computer Vision, Virtual Reality and Robotics in Medicine (CVRMed '95)*, pages 13–22, Nice, France, April 1995. IEEE.
- [35] N. Yokoya, H. Takemura, T. Okuma, and M. Kanbara. Stereo vision based video see-through mixed reality. In *Proc. ISMR'99 (1. International Symposium on Mixed Reality)*, pages 131–145, Yokohama, Japan, March 1999.

- [36] G. Zachmann. Real-time and exact collision detection for interactive virtual prototyping. In *Proc. of the 1997 ASME Design Engineering Technical Conferences*, Sacramento, CA, Sept 14–17 1997. ASME. CIE-4306.
- [37] I. Zoghliami, O. Faugeras, and R. Deriche. Traitement des occlusions pour la modification d'objet plan dans une sequence d'image. Private communication; see also <http://www.inria.fr/robotvis/personnel/zimad/Orasis6/Orasis6/html>, 1996.