Ronald T. Azuma
Hughes Research Laboratories
3011 Malibu Canyon Road, MS RL96
Malibu, CA 90265
azuma@isl.hrl.hac.com
http://www.cs.unc.edu/~azuma

A Survey of Augmented Reality

Abstract

This paper surveys the field of augmented reality (AR), in which 3D virtual objects are integrated into a 3D real environment in real time. It describes the medical, manufacturing, visualization, path planning, entertainment, and military applications that have been explored. This paper describes the characteristics of augmented reality systems, including a detailed discussion of the tradeoffs between optical and video blending approaches. Registration and sensing errors are two of the biggest problems in building effective augmented reality systems, so this paper summarizes current efforts to overcome these problems. Future directions and areas requiring further research are discussed. This survey provides a starting point for anyone interested in researching or using augmented reality.

1 Introduction

1.1 Goals

This paper surveys the current state-of-the-art in augmented reality. It describes work performed at many different sites and explains the issues and problems encountered when building augmented reality systems. It summarizes the tradeoffs and approaches taken so far to overcome these problems and speculates on future directions that deserve exploration.

A survey paper does not present new research results. The contribution comes from consolidating existing information from many sources and publishing an extensive bibliography of papers in this field. While several other introductory papers have been written on this subject (Barfield et al., 1995; Bowskill and Downie, 1995; Caudell, 1994; Drascic, 1993; Feiner, 1994a, b; Milgram et al., 1994b; Rolland et al., 1994), this survey is more comprehensive and up-to-date. This survey provides a good beginning point for anyone interested in starting research in this area.

Section 1 describes augmented reality and the motivations for developing this technology. Six classes of potential applications that have been explored are described in Section 2. Then Section 3 discusses the issues involved in building an augmented reality system. Currently, two of the biggest problems are in registration and sensing: the subjects of Sections 4 and 5. Finally, Section 6 describes some areas that require further work and research.

1.2 Definition

Augmented reality (AR) is a variation of *virtual environments* (VE), or virtual reality as it is more commonly called. VE technologies completely immerse a user inside a synthetic environment. While immersed, the user cannot

Presence, Vol. 6, No. 4, August 1997, 355–385
© 1997 by the Massachusetts Institute of Technology

Meta Platforms, Inc.
Exhibit 1112
Page 001

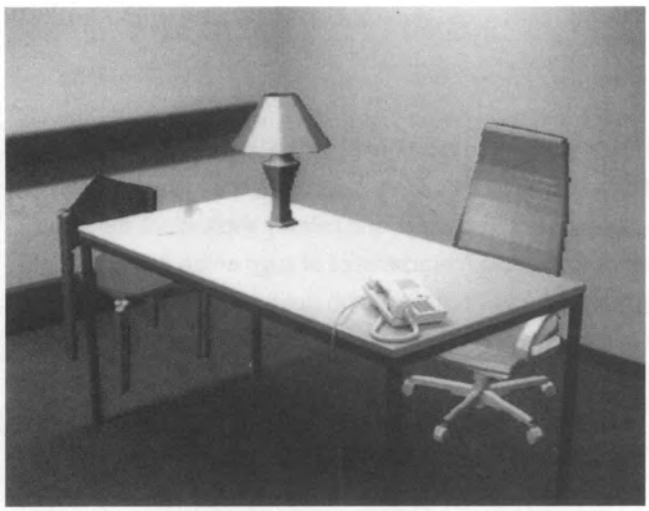

Figure 1. *Real desk with virtual lamp and two virtual chairs. (Courtesy ECRC.)*

see the real world around him. In contrast, AR allows the user to see the real world, with virtual objects superimposed upon or composited with the real world. Therefore, AR supplements reality, rather than completely replacing it. Ideally, it would appear to the user that the virtual and real objects coexisted in the same space, similar to the effects achieved in the film "Who Framed Roger Rabbit?" Figure 1 shows an example of what this coexistence might look like. It shows a real desk with a real phone. Inside this room are also a virtual lamp and two virtual chairs. Note that the objects are combined in three dimensions, so that the virtual lamp covers the real table, and the real table covers parts of the two virtual chairs. AR can be thought of as the "middle ground" between VE (completely synthetic) and telepresence (completely real) (Milgram and Kishino, 1994a; Milgram et al., 1994b).

Some researchers define AR in a way that requires the use of head-mounted displays (HMDs). To avoid limiting AR to specific technologies, this survey defines AR as any system that has the following three characteristics:

1. Combines real and virtual
2. Is interactive in real time
3. Is registered in three dimensions

This definition allows other technologies besides HMDs while retaining the essential components of AR. For example, it does not include film or 2D overlays. Films like "Jurassic Park" feature photorealistic virtual objects seamlessly blended with a real environment in 3D, but they are not interactive media. Two-dimensional virtual overlays on top of live video can be done at interactive rates, but the overlays are not combined with the real world in 3D. However, this definition does allow monitor-based interfaces, monocular systems, see-through HMDs, and various other combining technologies. Potential system configurations are discussed further in Section 3.

1.3 Motivation

Why is augmented reality an interesting topic? Why is combining real and virtual objects in 3D useful? Augmented reality enhances a user's perception of and interaction with the real world. The virtual objects display information that the user cannot directly detect with his own senses. The information conveyed by the virtual objects helps a user perform real-world tasks. AR is a specific example of what Fred Brooks calls *intelligence amplification* (IA): using the computer as a tool to make a task easier for a human to perform (Brooks, 1996).

At least six classes of potential AR applications have been explored: medical visualization, maintenance and repair, annotation, robot path planning, entertainment, and military aircraft navigation and targeting. The next section describes work that has been done in each area. While these do not cover every potential application area of this technology, they do cover the areas explored so far.

2 Applications

2.1 Medical

Doctors could use augmented reality as a visualization and training aid for surgery. It may be possible to collect 3D datasets of a patient in real time, using nonin-

Meta Platforms, Inc.
Exhibit 1112
Page 002

Figure 2. *Virtual fetus inside womb of pregnant patient. (Courtesy UNC Chapel Hill Dept. of Computer Science.)*

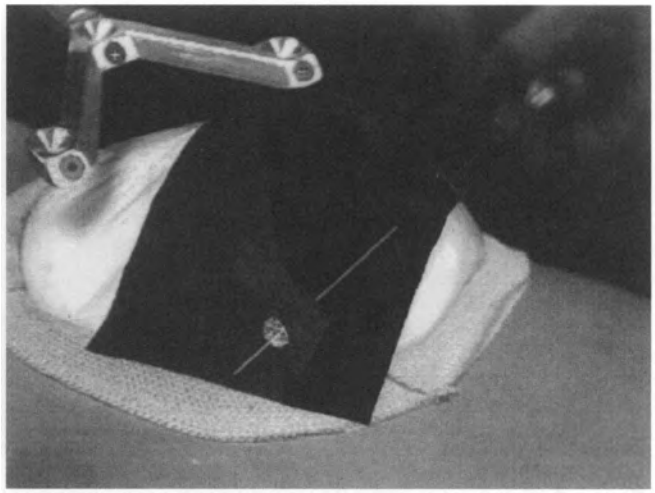

Figure 3. *Mockup of breast tumor biopsy. 3D graphics guide needle insertion. (Courtesy Andrei State, UNC Chapel Hill Dept. of Computer Science.)*

vasive sensors like magnetic resonance imaging (MRI), computed tomography scans (CT), or ultrasound imaging. These datasets could then be rendered and combined in real time with a view of the real patient. In effect, this would give a doctor "X-ray vision" inside a patient. This ability would be very useful during minimally invasive surgery, which reduces the trauma of an operation by using small incisions or no incisions at all. A problem with minimally invasive techniques is that they reduce the doctor's ability to see inside the patient, making surgery more difficult. AR technology could provide an internal view without the need for larger incisions.

AR might also be helpful for general medical visualization tasks in the surgical room. Surgeons can detect some features with the naked eye that they cannot see in MRI or CT scans, and vice versa. AR would give surgeons access to both types of data simultaneously. This information might also guide precision tasks, such as displaying where to drill a hole into the skull for brain surgery or where to perform a needle biopsy of a tiny tumor. The information from the noninvasive sensors would be directly displayed on the patient, showing exactly where to perform the operation.

AR might also be useful for training purposes (Kancherla et al., 1995). Virtual instructions could remind a novice surgeon of the required steps, without the need to look away from a patient to consult a manual. Virtual objects could also identify organs and specify locations to avoid disturbing the patient (Durlach and Mavor, 1995).

Several projects are exploring this application area. At UNC Chapel Hill, a research group has conducted trial runs of scanning the womb of a pregnant woman with an ultrasound sensor, generating a 3D representation of the fetus inside the womb and displaying that in a see-through HMD (Figure 2). The goal is to endow the doctor with the ability to see the moving, kicking fetus lying inside the womb, with the hope that this one day may become a "3D stethoscope" (Bajura et al., 1992; State et al., 1994). More recent efforts have focused on a needle biopsy of a breast tumor. Figure 3 shows a mockup of a breast biopsy operation, where the virtual objects identify the location of the tumor and guide the needle to its target (State et al., 1996b). Other groups at the MIT AI Lab (Grimson et al., 1994; Grimson et al., 1995; Mellor, 1995a, b), General Electric (Lorensen et al., 1993), and elsewhere (Betting et al., 1995; Edwards et al., 1995; Taubes, 1994) are investigating displaying MRI or CT data, directly registered onto the patient.

Meta Platforms, Inc.
Exhibit 1112
Page 003

Figure 4. *External view of Columbia printer maintenance application. Note that all objects must be tracked. (Courtesy Steve Feiner, Blair MacIntyre, and Dorée Seligmann, Columbia University.)*

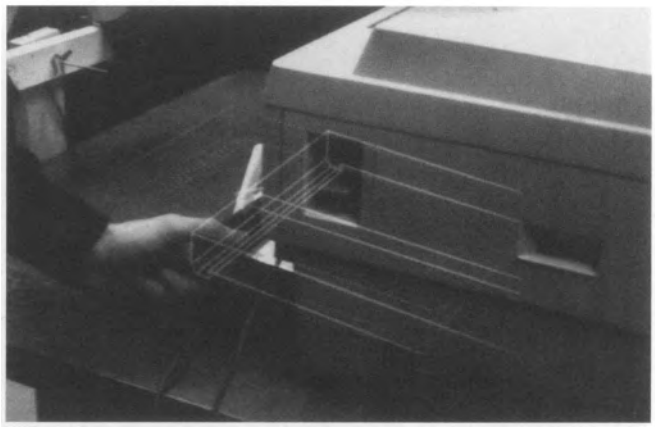

Figure 5. *Prototype laser printer maintenance application, displaying how to remove the paper tray. (Courtesy Steve Feiner, Blair MacIntyre, and Dorée Seligmann, Columbia University.)*

2.2 Manufacturing and Repair

Another category of augmented reality applications is the assembly, maintenance, and repair of complex machinery. Instructions might be easier to understand if they were available, not as manuals with text and pictures, but rather as 3D drawings superimposed upon the actual equipment, showing step-by-step the tasks that need to be done and how to do them. These superimposed 3D drawings can be animated, making the directions even more explicit.

Several research projects have demonstrated prototypes in this area. Steve Feiner's group at Columbia built a laser printer maintenance application (Feiner et al., 1993a), shown in Figures 4 and 5. Figure 4 shows an external view, and Figure 5 shows the user's view, where the computer-generated wireframe is telling the user to remove the paper tray. A group at Boeing is developing AR technology to guide a technician in building a wiring harness that forms part of an airplane's electrical system. Storing these instructions in electronic form will save space and reduce costs. Currently, technicians use large physical layout boards to construct such harnesses, and Boeing requires several warehouses to store all these boards. Such space might be emptied for other use if this application proves successful (Caudell and Mizell, 1992; Janin et al., 1993; Sims, 1994). Boeing is using a Technology Reinvestment Program (TRP) grant to investigate putting this technology onto the factory floor (Boeing TRP, 1994). Figure 6 shows an external view of Adam Janin using a prototype AR system to build a wire bundle. Eventually, AR might be used for any complicated machinery, such as automobile engines (Tuceryan et al., 1995).

Figure 6. *Adam Janin demonstrates Boeing's prototype wire bundle assembly application. (Courtesy David Mizell, Boeing.)*

Meta Platforms, Inc.
Exhibit 1112
Page 004

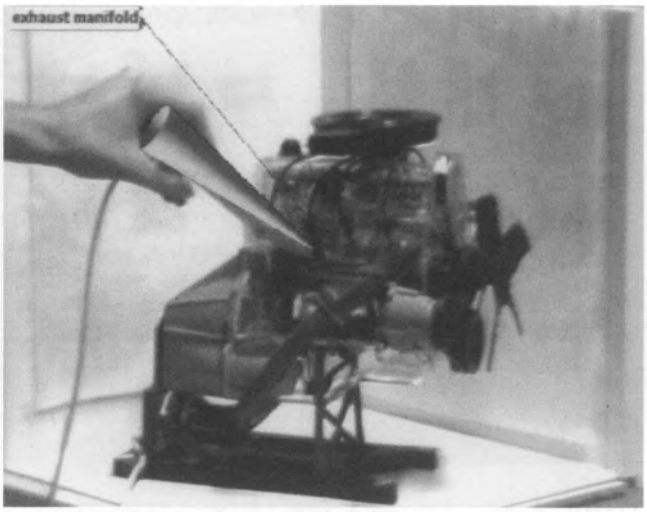

Figure 7. *Engine model part labels appear as user points at them. (Courtesy ECRC.)*

Figure 8. *Windows displayed on top of specific real-world objects. (Courtesy Steve Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon, Columbia University.)*

2.3 Annotation and Visualization

AR could be used to annotate objects and environments with public or private information. Applications using public information assume the availability of public databases to draw upon. For example, a hand-held display could provide information about the contents of library shelves as the user walks around the library (Fitzmaurice, 1993; Rekimoto, 1995; Rekimoto and Nagao, 1995). At the European Computer-Industry Research Centre (ECRC), a user can point at parts of an engine model and the AR system displays the name of the part that is being pointed at (Rose et al., 1995). Figure 7 shows this, where the user points at the exhaust manifold on an engine model and the label "exhaust manifold" appears.

Alternately, these annotations might be private notes attached to specific objects. Researchers at Columbia demonstrated this with the notion of attaching windows from a standard user interface onto specific locations in the world, or attached to specific objects as reminders (Feiner et al., 1993b). Figure 8 shows a window superimposed as a label upon a student. He wears a tracking device, so the computer knows his location. As the student moves around, the label follows his location, providing the AR user with a reminder of what he needs to talk to the student about.

AR might aid general visualization tasks as well. An architect with a see-through HMD might be able to look out a window and see how a proposed new skyscraper would change her view. If a database containing information about a building's structure was available, AR might give architects "X-ray vision" inside a building, showing where the pipes, electric lines, and structural supports are inside the walls (Feiner et al., 1995). Researchers at the University of Toronto have built a system called Augmented Reality through Graphic Overlays on Stereovideo (ARGOS) (Milgram et al., 1995), which among other things is used to make images easier to understand during difficult viewing conditions (Drascic et al., 1993). Figure 9 shows wireframe lines drawn on top of a space shuttle bay interior, while in orbit. The lines make it easier to see the geometry of the shuttle bay. Similarly, virtual lines and objects could aid navigation and scene understanding during poor visibility conditions, such as underwater or in fog.

2.4 Robot Path Planning

Teleoperation of a robot is often a difficult problem, especially when the robot is far away, with long delays in the communication link. Under this circumstance, instead of controlling the robot directly, it may be preferable to instead control a virtual version of the robot. The user plans and specifies the robot's actions by

Meta Platforms, Inc.
Exhibit 1112
Page 005

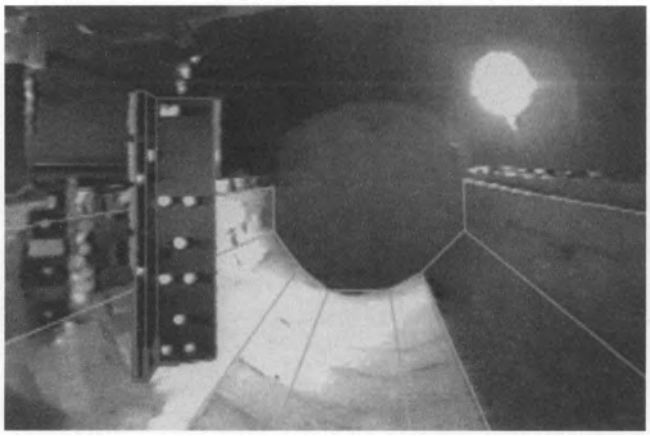

Figure 9. *Virtual lines help display geometry of shuttle bay, as seen in orbit. (Courtesy David Drascic and Paul Milgram, U. Toronto.)*

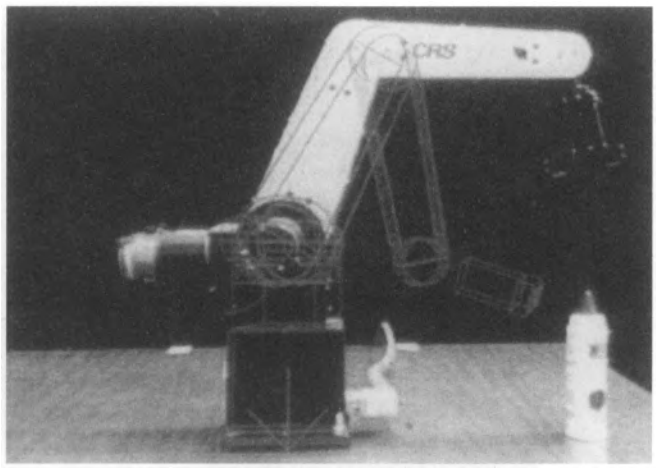

Figure 10. *Virtual lines show a planned motion of a robot arm (Courtesy David Drascic and Paul Milgram, U. Toronto.)*

manipulating the local virtual version, in real time. The results are directly displayed on the real world. Once the plan is tested and determined, then user tells the real robot to execute the specified plan. This avoids pilot-induced oscillations caused by the lengthy delays. The virtual versions can also predict the effects of manipulating the environment, thus serving as a planning and previewing tool to aid the user in performing the desired task. The ARGOS system has demonstrated that stereoscopic AR is an easier and more accurate way of doing robot path planning than traditional monoscopic interfaces (Drascic, 1993; Milgram et al., 1993). Others have also used registered overlays with telepresence systems (Kim, 1993; Kim, 1996; Oyama et al., 1993; Yoo and Olano, 1993; Tharp et al., 1994). Figure 10 shows how a virtual outline can represent a future location of a robot arm.

2.5 Entertainment

At SIGGRAPH '95, several exhibitors showed "Virtual Sets" that merge real actors with virtual backgrounds, in real time and in 3D. The actors stand in front of a large blue screen, while a computer-controlled motion camera records the scene. Since the camera's location is tracked and the actor's motions are scripted, it is possible to digitally composite the actor into a 3D virtual background. For example, the actor might appear to stand inside a large virtual spinning ring, where the front part of the ring covers the actor while the rear part of the ring is covered by the actor. The entertainment industry sees this as a way to reduce production costs because creating and storing sets virtually is potentially cheaper than constantly building new physical sets from scratch. The ALIVE project from the MIT Media Lab goes one step further by populating the environment with intelligent virtual creatures that respond to user actions (Maes, 1995).

2.6 Military Aircraft

For many years, military aircraft and helicopters have used head-up displays (HUDs) and helmet-mounted sights (HMS) to superimpose vector graphics upon the pilot's view of the real world. Besides providing basic navigation and flight information, these graphics are sometimes registered with targets in the environment, providing a way to aim the aircraft's weapons. For example, the chin turret in a helicopter gunship can be slaved to the pilot's HMS, so the pilot can aim the chin turret simply by looking at the target. Future generations of combat aircraft will be developed with an HMD built into the pilot's helmet (Wanstall, 1989).

3 Characteristics

This section discusses the characteristics of AR systems and design issues encountered when building an

Meta Platforms, Inc.
Exhibit 1112
Page 006

AR system. Section 3.1 describes the basic characteristics of augmentation. There are two ways to accomplish this augmentation: optical or video technologies. Section 3.2 discusses their characteristics and relative strengths and weaknesses. Blending the real and virtual poses problems with focus and contrast (Section 3.3), and some applications require portable AR systems to be truly effective (Section 3.4). Finally, Section 3.5 summarizes the characteristics by comparing the requirements of AR against those for virtual environments.

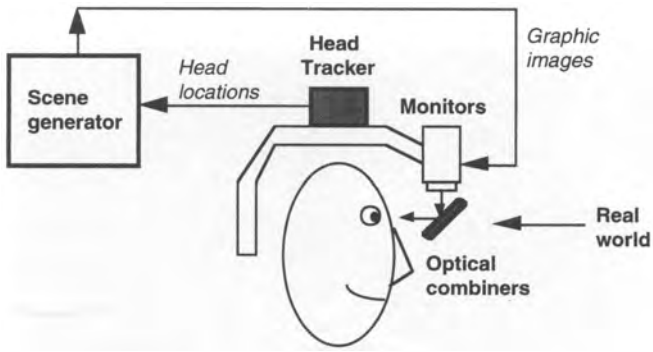

Figure 11. *Optical see-through HMD conceptual diagram.*

3.1 Augmentation

Besides *adding* objects to a real environment, augmented reality also has the potential to *remove* them. Current work has focused on adding virtual objects to a real environment. However, graphic overlays might also be used to remove or hide parts of the real environment from a user. For example, to remove a desk in the real environment, draw a representation of the real walls and floors behind the desk and "paint" that over the real desk, effectively removing it from the user's sight. This has been done in feature films. Doing this interactively in an AR system will be much harder, but this removal may not need to be photorealistic to be effective.

Augmented reality might apply to all senses, not just sight. So far, researchers have focused on blending real and virtual images and graphics. However, AR could be extended to include sound. The user would wear headphones equipped with microphones on the outside. The headphones would add synthetic, directional 3D sound, while the external microphones would detect incoming sounds from the environment. This would give the system a chance to mask or cover up selected real sounds from the environment by generating a masking signal that exactly canceled the incoming real sound (Durlach and Mavor, 1995). While this would not be easy to do, it might be possible. Another example involves haptics. Gloves with devices that provide tactile feedback might augment real forces in the environment. For example, a user might run his hand over the surface of a real desk. Simulating such a hard surface virtually is fairly difficult, but it is easy to do in reality. Then the tactile effectors in the glove can augment the feel of the desk, perhaps making it feel rough in certain spots. This capability might be useful in some applications, such as providing an additional cue that a virtual object is at a particular location on a real desk (Wellner, 1993).

3.2 Optical Versus Video

A basic design decision in building an AR system is how to accomplish the combining of real and virtual. Two basic choices are available: optical and video technologies. Each has particular advantages and disadvantages. This section compares the two and notes the tradeoffs. For additional discussion, see Rolland et al. (1994).

A see-through HMD is one device used to combine real and virtual. Standard *closed-view HMDs* do not allow any direct view of the real world. In contrast, a *see-through HMD* lets the user see the real world, with virtual objects superimposed by optical or video technologies.

Optical see-through HMDs work by placing optical combiners in front of the user's eyes. These combiners are partially transmissive, so that the user can look directly through them to see the real world. The combiners are also partially reflective, so that the user sees virtual images bounced off the combiners from head-mounted monitors. This approach is similar in nature to *head-up displays* (HUDs) commonly used in military aircraft, except that the combiners are attached to the head. Thus, optical see-through HMDs have sometimes been described as a "HUD on a head" (Wanstall, 1989). Figure 11 shows a conceptual diagram of an optical see-

Meta Platforms, Inc.
Exhibit 1112
Page 007

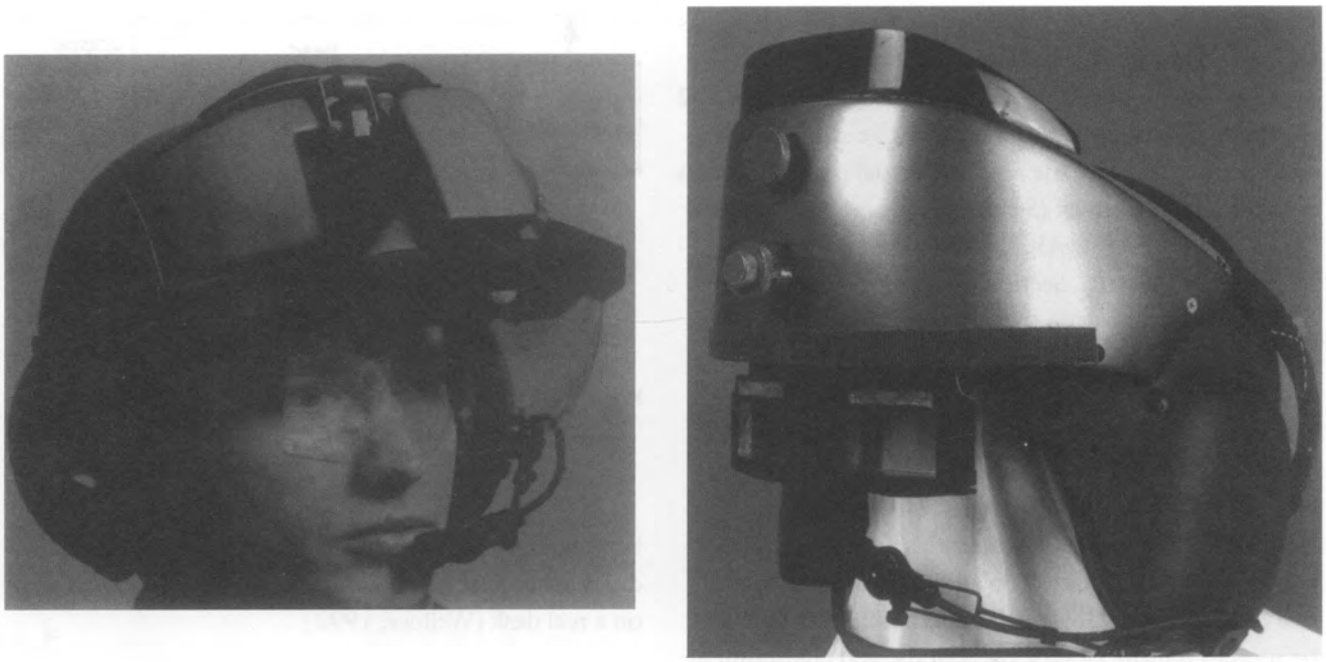

Figure 12. *Two optical see-through HMDs, made by Hughes Electronics.*

through HMD. Figure 12 shows two optical see-through HMDs made by Hughes Electronics.

The optical combiners usually reduce the amount of light that the user sees from the real world. Since the combiners act like half-silvered mirrors, they let in only some of the light from the real world, so that they can reflect some of the light from the monitors into the user's eyes. For example, the HMD described in (Holmgren, 1992) transmits about 30% of the incoming light from the real world. Choosing the level of blending is a design problem. More sophisticated combiners might vary the level of contributions based upon the wavelength of light. For example, such a combiner might be set to reflect all light of a certain wavelength and none at any other wavelengths. This approach would be ideal with a monochrome monitor. Virtually all the light from the monitor would be reflected into the user's eyes, while almost all the light from the real world (except at the particular wavelength) would reach the user's eyes. However, most existing optical see-through HMDs do reduce the amount of light from the real world, so they act like a pair of sunglasses when the power is cut off.

In contrast, video see-through HMDs work by combining a closed-view HMD with one or two head-mounted video cameras. The video cameras provide the user's view of the real world. Video from these cameras is combined with the graphic images created by the scene generator, blending the real and virtual. The result is sent to the monitors in front of the user's eyes in the closed-view HMD. Figure 13 shows a conceptual diagram of a video see-through HMD. Figure 14 shows an actual video see-through HMD, with two video cameras mounted on top of a flight helmet.

Video composition can be done in more than one way. A simple way is to use chroma-keying, a technique used in many video special effects. The background of the computer graphic images is set to a specific color, say green, which none of the virtual objects use. Then the combining step replaces all green areas with the corresponding parts from the video of the real world. This step has the effect of superimposing the virtual objects over the real world. A more sophisticated composition would use depth information. If the system had depth information at each pixel for the real-world images, it

Meta Platforms, Inc.
Exhibit 1112
Page 008

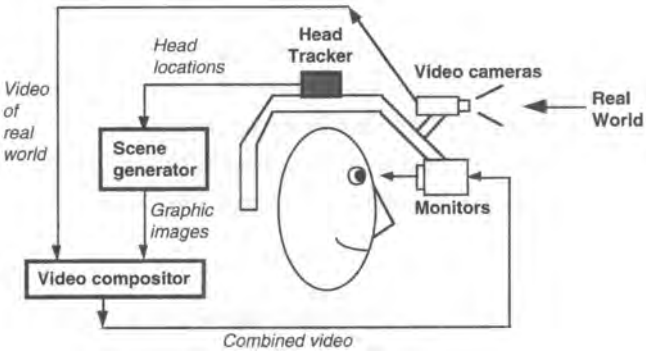

Figure 13. *Video see-through HMD conceptual diagram.*

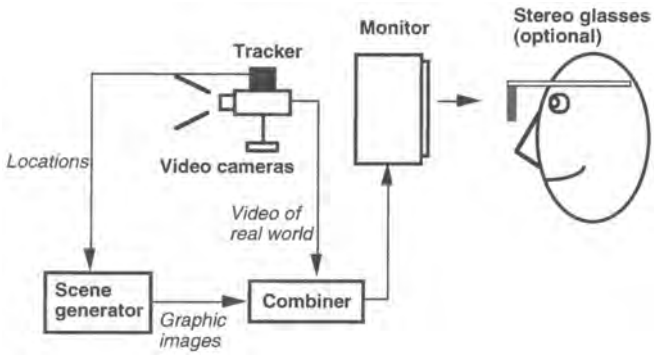

Figure 15. *Monitor-based AR conceptual diagram.*

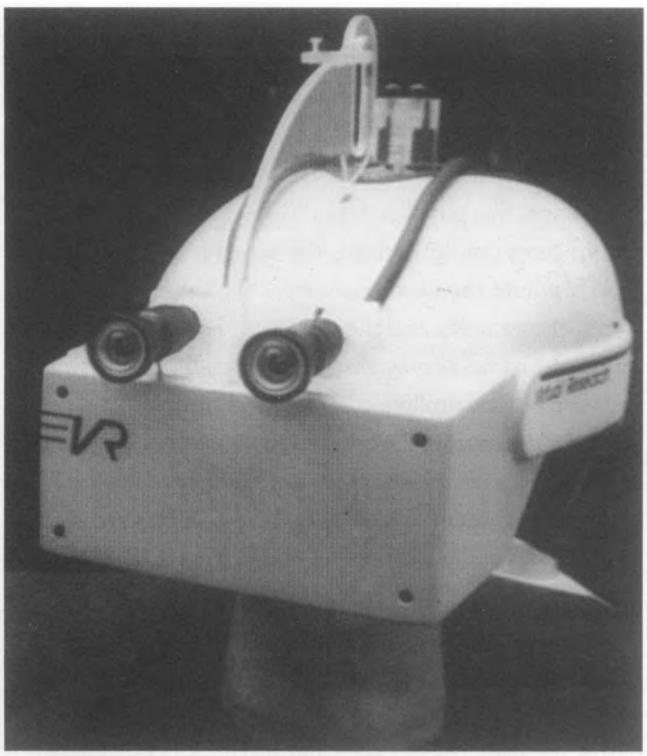

Figure 14. *An actual video see-through HMD. (Courtesy Jannick Rolland, Frank Biocca, and UNC Chapel Hill Dept. of Computer Science. Photo by Alex Treml.)*

Figure 16. *External view of the ARGOS system, an example of monitor-based AR. (Courtesy David Drascic and Paul Milgram, U. Toronto.)*

could combine the real and virtual images by a pixel-by-pixel depth comparison. This would allow real objects to cover virtual objects and vice versa.

AR systems can also be built using monitor-based configurations, instead of see-through HMDs. Figure 15 shows how a monitor-based system might be built. In this case, one or two video cameras view the environment. The cameras may be static or mobile. In the mobile case, the cameras might move around by being attached to a robot, with their locations tracked. The video of the real world and the graphic images generated by a scene generator are combined, just as in the video see-through HMD case, and displayed in a monitor in front of the user. The user does not wear the display device. Optionally, the images may be displayed in stereo on the monitor, which then requires the user to wear a pair of stereo glasses. Figure 16 shows an external view

Meta Platforms, Inc.
Exhibit 1112
Page 009

of the ARGOS system, which uses a monitor-based configuration.

Finally, a monitor-based optical configuration is also possible. This is similar to Figure 11 except that the user does not wear the monitors or combiners on her head. Instead, the monitors and combiners are fixed in space, and the user positions her head to look through the combiners. This configuration is typical of head-up displays on military aircraft, and at least one such configuration has been proposed for a medical application (Peuchot et al., 1995).

The rest of this section compares the relative advantages and disadvantages of optical and video approaches, starting with optical. An optical approach has the following advantages over a video approach:

Simplicity. Optical blending is simpler and cheaper than video blending. Optical approaches have only one "stream" of video to worry about: the graphic images. The real world is seen directly through the combiners, and that time delay is generally a few nanoseconds. Video blending, on the other hand, must deal with separate video streams for the real and virtual images. Both streams have inherent delays in the tens of milliseconds. Digitizing video images usually adds at least one *frame time* of delay to the video stream, where a frame time is the length of time it takes to completely update an image. A monitor that completely refreshes the screen at 60 Hz has a frame time of 16.67 ms. The two streams of real and virtual images must be properly synchronized or temporal distortion results. Also, optical see-through HMDs with narrow field-of-view combiners offer views of the real world that have little distortion. Video cameras almost always have some amount of distortion that must be compensated for, along with any distortion from the optics in front of the display devices. Since video requires cameras and combiners that optical approaches do not need, video will probably be more expensive and complicated to build than optically based systems.

Resolution. Video blending limits the resolution of what the user sees, both real and virtual, to the resolution of the display devices. With current displays, this resolution is far less than the resolving power of the fovea. Optical see-through also shows the graphic images at the resolution of the display device, but the user's view of the real world is not degraded. Thus, video reduces the resolution of the real world, while optical see-through does not.

Safety. Video see-through HMDs are essentially modified closed-view HMDs. If the power is cut off, the user is effectively blind. This is a safety concern in some applications. In contrast, when power is removed from an optical see-through HMD, the user still has a direct view of the real world. The HMD then becomes a pair of heavy sunglasses, but the user can still see.

No Eye Offset. With video see-through, the user's view of the real world is provided by the video cameras. In essence, this puts his "eyes" where the video cameras are. In most configurations, the cameras are not located exactly where the user's eyes are, creating an offset between the cameras and the real eyes. The distance separating the cameras may also not be exactly the same as the user's interpupillary distance (IPD). This difference between camera locations and eye locations introduces displacements from what the user sees compared to what he expects to see. For example, if the cameras are above the user's eyes, he will see the world from a vantage point slightly higher than he is used to. Video see-through can avoid the eye offset problem through the use of mirrors to create another set of optical paths that mimic the paths directly into the user's eyes. Using those paths, the cameras will see what the user's eyes would normally see without the HMD. However, this adds complexity to the HMD design. Offset is generally not a difficult design problem for optical see-through displays. While the user's eye can rotate with respect to the position of the HMD, the resulting errors are tiny. Using the eye's center of rotation as the viewpoint in the computer graphics model should eliminate any need for eye tracking in an optical see-through HMD (Holloway, 1995).

Video blending offers the following advantages over optical blending:

Meta Platforms, Inc.
Exhibit 1112
Page 010

Flexibility in Composition Strategies. A basic problem with optical see-through is that the virtual objects do not completely obscure the real world objects, because the optical combiners allow light from both virtual and real sources. Building an optical see-through HMD that can selectively shut out the light from the real world is difficult. In a normal optical system, the objects are designed to be in focus at only one point in the optical path: the user's eye. Any filter that would selectively block out light must be placed in the optical path at a point where the image is in focus, which obviously cannot be the user's eye. Therefore, the optical system must have *two* places where the image is in focus: at the user's eye and the point of the hypothetical filter. This requirement makes the optical design much more difficult and complex. No existing optical see-through HMD blocks incoming light in this fashion. Thus, the virtual objects appear ghostlike and semitransparent. This appearance damages the illusion of reality because occlusion is one of the strongest depth cues. In contrast, video see-through is far more flexible about how it merges the real and virtual images. Since both the real and virtual are available in digital form, video see-through compositors can, on a pixel-by-pixel basis, take the real, or the virtual, or some blend between the two to simulate transparency. Because of this flexibility, video see-through may ultimately produce more compelling environments than optical see-through approaches.

Wide Field of View. Distortions in optical systems are a function of the radial distance away from the optical axis. The further one looks away from the center of the view, the larger the distortions get. A digitized image taken through a distorted optical system can be undistorted by applying image processing techniques to unwarp the image, provided that the optical distortion is well characterized. This requires significant amounts of computation, but this constraint will be less important in the future as computers become faster. It is harder to build wide field-of-view displays with optical see-through techniques. Any distortions of the user's view of the real world must be corrected *optically,* rather than digitally, because the system has no digitized image of the real world to manipulate. Complex optics are expensive and add weight to the HMD. Wide field-of-view systems are an exception to the general trend of optical approaches being simpler and cheaper than video approaches.

Matching the Delay between Real and Virtual Views. Video offers an approach for reducing or avoiding problems caused by temporal mismatches between the real and virtual images. Optical see-through HMDs offer an almost instantaneous view of the real world but a delayed view of the virtual. This temporal mismatch can cause problems. With video approaches, it is possible to delay the video of the real world to match the delay from the virtual image stream. For details, see Section 4.3.

Additional Registration Strategies. In optical see-through, the only information the system has about the user's head location comes from the head tracker. Video blending provides another source of information: the digitized image of the real scene. This digitized image means that video approaches can employ additional registration strategies unavailable to optical approaches. Section 4.4 describes these in more detail.

Matching the Brightness of Real and Virtual Objects. This advantage is discussed in Section 3.3.

Both optical and video technologies have their roles, and the choice of technology depends on the application requirements. Many of the mechanical assembly and repair prototypes use optical approaches, possibly because of the cost and safety issues. If successful, the equipment would have to be replicated in large numbers to equip workers on a factory floor. In contrast, most of the prototypes for medical applications use video approaches, probably for the flexibility in blending real and virtual and for the additional registration strategies offered.

3.3 Focus and Contrast

Focus can be a problem for both optical and video approaches. Ideally, the virtual should match the real. In a video-based system, the combined virtual and real im-

Meta Platforms, Inc.
Exhibit 1112
Page 011

age will be projected at the same distance by the monitor or HMD optics. However, depending on the video camera's depth-of-field and focus settings, parts of the real world may not be in focus. In typical graphics software, everything is rendered with a pinhole model, so all the graphic objects, regardless of distance, are in focus. To overcome this, the graphics could be rendered to simulate a limited depth of field, and the video camera might have an autofocus lens.

In the optical case, the virtual image is projected at some distance away from the user. This distance may be adjustable, although it is often fixed. Therefore, while the real objects are at varying distances from the user, the virtual objects are all projected to the same distance. If the virtual and real distances are not matched for the particular objects that the user is looking at, it may not be possible to clearly view both simultaneously.

Contrast is another issue because of the large dynamic range in real environments and in what the human eye can detect. Ideally, the brightness of the real and virtual objects should be appropriately matched. Unfortunately, in the worst case scenario, this means the system must match a very large range of brightness levels. The eye is a logarithmic detector, where the brightest light that it can handle is about 11 orders of magnitude greater than the smallest, including both dark-adapted and light-adapted eyes. In any one adaptation state, the eye can cover about six orders of magnitude. Most display devices cannot come close to this level of contrast. This limitation is a particular problem with optical technologies, because the user has a direct view of the real world. If the real environment is too bright, it will wash out the virtual image. If the real environment is too dark, the virtual image will wash out the real world. Contrast problems are not as severe with video, because the video cameras themselves have limited dynamic response. The view of both the real and virtual is generated by the monitor, so everything must be clipped or compressed into the monitor's dynamic range.

3.4 Portability

In almost all virtual environment systems, the user is not encouraged to walk around much. Instead, the user navigates by "flying" through the environment, walking on a treadmill, or driving some mockup of a vehicle. Whatever the technology, the result is that the user stays in one place in the real world.

Some AR applications, however, will need to support a user who will walk around a large environment. AR requires that the user actually be at the place where the task is to take place. "Flying," as performed in a VE system, is no longer an option. If a mechanic needs to go to the other side of a jet engine, she must physically move herself and the display devices she wears. Therefore, AR systems will place a premium on portability, especially the ability to walk around outdoors, away from controlled environments. The scene generator, the HMD, and the tracking system must all be self-contained and capable of surviving exposure to the environment. If this capability is achieved, many more applications that have not been tried will become available. For example, the ability to annotate the surrounding environment could be useful to soldiers, hikers, or tourists in an unfamiliar new location.

3.5 Comparison Against Virtual Environments

The overall requirements of AR can be summarized by comparing them against those for virtual environments, for the three basic subsystems that they require.

1. Scene Generator. Rendering is not currently one of the major problems in AR. VE systems have much higher requirements for realistic images because they completely replace the real world with the virtual environment. In AR, the virtual images only supplement the real world. Therefore, fewer virtual objects need to be drawn, and they do not necessarily have to be realistically rendered in order to serve the purposes of the application. For example, in the annotation applications, text and 3D wireframe drawings might suffice. Ideally, photorealistic graphic objects would be seamlessly merged with the real environment (see Section 7), but more basic problems have to be solved first.

Meta Platforms, Inc.
Exhibit 1112
Page 012

2. Display Device. The display devices used in AR may have less stringent requirements than VE systems demand, again because AR does not replace the real world. For example, monochrome displays may be adequate for some AR applications, while virtually all VE systems today use full color. Optical see-through HMDs with a small field of view may be satisfactory because the user can still see the real world with his peripheral vision; the see-through HMD does not shut off the user's normal field of view. Furthermore, the resolution of the monitor in an optical see-through HMD might be lower than what a user would tolerate in a VE application, since the optical see-through HMD does not reduce the resolution of the real environment.
3. Tracking and Sensing. While in the previous two cases AR had lower requirements than VE, that is not the case for tracking and sensing. In this area, the requirements for AR are much stricter than those for VE systems. A major reason for this is the registration problem, which is described in the next section. The other factors that make the tracking and sensing requirements higher are described in Section 5.

4 Registration

4.1 The Registration Problem

One of the most basic problems currently limiting augmented reality applications is the registration problem. The objects in the real and virtual worlds must be properly aligned with respect to each other, or the illusion that the two worlds coexist will be compromised. More seriously, many applications *demand* accurate registration. For example, recall the needle biopsy application. If the virtual object is not where the real tumor is, the surgeon will miss the tumor and the biopsy will fail. Without accurate registration, augmented reality will not be accepted in many applications.

Registration problems also exist in virtual environments, but they are not nearly as serious because they are harder to detect than in augmented reality. Since the user only sees virtual objects in VE applications, registration errors result in visual-kinesthetic and visual-proprioceptive conflicts. Such conflicts between different human senses may be a source of motion sickness (Pausch et al., 1992). Because the kinesthetic and proprioceptive systems are much less sensitive than the visual system, visual-kinesthetic and visual-proprioceptive conflicts are less noticeable than visual-visual conflicts. For example, a user wearing a closed-view HMD might hold up her real hand and see a virtual hand. This virtual hand should be displayed exactly where she would see her real hand, if she were not wearing an HMD. But if the virtual hand is wrong by five millimeters, she may not detect that unless actively looking for such errors. The same error is much more obvious in a see-through HMD, where the conflict is visual-visual.

Furthermore, a phenomenon known as *visual capture* (Welch, 1978) makes it even more difficult to detect such registration errors. Visual capture is the tendency of the brain to believe what it sees rather than what it feels, hears, and so on. That is, visual information tends to override all other senses. When watching a television program, a viewer believes the sounds come from the mouths of the actors on the screen, even though they actually come from a speaker in the TV. Ventriloquism works because of visual capture. Similarly, a user might believe that her hand is where the virtual hand is drawn, rather than where her real hand actually is, because of visual capture. This effect increases the amount of registration error users can tolerate in virtual environment systems. If the errors are systematic, users might even be able to adapt to the new environment, given a long exposure time of several hours or days (Welch, 1978).

Augmented reality demands much more accurate registration than do virtual environments (Azuma, 1993). Imagine the same scenario of a user holding up her hand, but this time wearing a see-through HMD. Registration errors now result in visual-visual conflicts between the images of the virtual and real hands. Such conflicts are easy to detect because of the resolution of the human eye and the sensitivity of the human visual system to differences. Even tiny offsets in the images of the real and virtual hands are easy to detect.

What angular accuracy is needed for good registration in augmented reality? A simple demonstration will show

Meta Platforms, Inc.
Exhibit 1112
Page 013

the order of magnitude required. Take out a dime and hold it at arm's length, so that it looks like a circle. The diameter of the dime covers about 1.2 to 2.0 degrees of arc, depending on your arm length. In comparison, the width of a full moon is about 0.5 degrees of arc! Now imagine a virtual object superimposed on a real object, but offset by the diameter of the full moon. Such a difference would be easy to detect. Thus, the angular accuracy required is a small fraction of a degree. The lower limit is bounded by the resolving power of the human eye itself. The central part of the retina is called the *fovea;* it has the highest density of color-detecting cones, about 120 per degree of arc, corresponding to a spacing of half a minute of arc (Jain, 1989). Observers can differentiate between a dark and light bar grating when each bar subtends about one minute of arc, and under special circumstances they can detect even smaller differences (Doenges, 1985). However, existing HMD trackers and displays are not capable of providing one minute of arc in accuracy, so the present achievable accuracy is much worse than that ultimate lower bound. In practice, errors of a few pixels are detectable in modern HMDs.

Registration of real and virtual objects is not limited to AR. Special-effects artists seamlessly integrate computer-generated 3D objects with live actors in film and video. The difference lies in the amount of control available. With film, a director can carefully plan each shot, and artists can spend hours per frame, adjusting each by hand if necessary, to achieve perfect registration. As an interactive medium, AR is far more difficult to work with. The AR system cannot control the motions of the HMD wearer. The user looks where she wants, and the system must respond within tens of milliseconds.

Registration errors are difficult to adequately control because of the high accuracy requirements and the numerous sources of error. These sources of error can be divided into two types: static and dynamic. *Static* errors are the ones that cause registration errors even when the user's viewpoint and the objects in the environment remain completely still. *Dynamic* errors are the ones that have no effect until either the viewpoint or the objects begin moving.

For current HMD-based systems, dynamic errors are by far the largest contributors to registration errors, but static errors cannot be ignored either. The next two sections discuss static and dynamic errors and what has been done to reduce them. See Holloway (1995) for a thorough analysis of the sources and magnitudes of registration errors.

4.2 Static Errors

The four main sources of static errors are:

- Optical distortion
- Errors in the tracking system
- Mechanical misalignments
- Incorrect viewing parameters (e.g., field of view, tracker-to-eye position and orientation, interpupillary distance)

4.2.1 Distortion in the Optics. Optical distortions exist in most camera and lens systems, both in the cameras that record the real environment and in the optics used for the display. Because distortions are usually a function of the radial distance away from the optical axis, wide field-of-view displays can be especially vulnerable to this error. Near the center of the field of view, images are relatively undistorted, but far away from the center, image distortion can be large. For example, straight lines may appear curved. In a see-through HMD with narrow field-of-view displays, the optical combiners add virtually no distortion, so the user's view of the real world is not warped. However, the optics used to focus and magnify the graphic images from the display monitors can introduce distortion. This mapping of distorted virtual images on top of an undistorted view of the real world causes static registration errors. The cameras and displays may also have nonlinear distortions that cause errors (Deering, 1992).

Optical distortions are usually systematic errors, so they can be mapped and compensated. This mapping may not be trivial, but it is often possible. For example, (Robinett and Rolland, 1992) describes the distortion of one commonly used set of HMD optics. The distortions might be compensated by additional optics. Edwards and colleages (1993) describe such a design for a video see-through HMD. Eliminating distortion can be a diffi-

Meta Platforms, Inc.
Exhibit 1112
Page 014

cult design problem, though, and it adds weight, which is not desirable in HMDs. An alternate approach is to do the compensation digitally by image-warping techniques, both on the digitized video and the graphic images. Typically, this involves predistorting the images so that they will appear undistorted after being displayed (Watson and Hodges, 1995). Another way to perform digital compensation on the graphics is to apply the predistortion functions on the vertices of the polygons, in screen space, before rendering (Rolland and Hopkins, 1993). This requires subdividing polygons that cover large areas in screen space. Both digital compensation methods can be computationally expensive, often requiring special hardware to accomplish in real time. Holloway determined that the additional system delay required by the distortion compensation adds more registration error than the distortion compensation removes, for typical head motion (Holloway, 1995).

4.2.2 Errors in the Tracking System. Errors in the reported outputs from the tracking and sensing systems are often the most serious type of static registration errors. These distortions are not easy to measure and eliminate, because that requires another "3D ruler" that is more accurate than the tracker being tested. These errors are often nonsystematic and difficult to fully characterize. Almost all commercially available tracking systems are not accurate enough to satisfy the requirements of AR systems. Section 5 discusses this important topic further.

4.2.3 Mechanical Misalignments. Mechanical misalignments are discrepancies between the model or specification of the hardware and the actual physical properties of the real system. For example, the combiners, optics, and monitors in an optical see-through HMD may not be at the expected distances or orientations with respect to each other. If the frame is not sufficiently rigid, the various component parts may change their relative positions as the user moves around, causing errors. Mechanical misalignments can cause subtle changes in the position and orientation of the projected virtual images that are difficult to compensate. While some alignment errors can be calibrated, for many others it may be more effective to "build it right" initially.

4.2.4 Incorrect Viewing Parameters. Incorrect viewing parameters, the last major source of static registration errors, can be thought of as a special case of alignment errors for which calibration techniques can be applied. Viewing parameters specify how to convert the reported head or camera locations into viewing matrices used by the scene generator to draw the graphic images. For an HMD-based system, these parameters include:

- Center of projection and viewport dimensions
- Offset, both in translation and orientation, between the location of the head tracker and the user's eyes
- Field of view

Incorrect viewing parameters cause systematic static errors. Take the example of a head tracker located above a user's eyes. If the vertical translation offsets between the tracker and the eyes are too small, all the virtual objects will appear lower than they should.

In some systems, the viewing parameters are estimated by manual adjustments, in a nonsystematic fashion. Such approaches proceed as follows: Place a real object in the environment and attempt to register a virtual object with that real object. While wearing the HMD or positioning the cameras, move to one viewpoint or a few selected viewpoints and manually adjust the location of the virtual object and the other viewing parameters until the registration "looks right." This action may achieve satisfactory results if the environment and the viewpoint remain static. However, such approaches require a skilled user and generally do not achieve robust results for many viewpoints. Achieving good registration from a single viewpoint is much easier than achieving registration from a wide variety of viewpoints using a single set of parameters. Usually what happens is satisfactory registration at one viewpoint, but when the user walks to a significantly different viewpoint, the registration is inaccurate because of incorrect viewing parameters or tracker distortions. This means many different sets of parameters must be used—a less than satisfactory solution.

Another approach is to directly measure the parameters, using various measuring tools and sensors. For

Meta Platforms, Inc.
Exhibit 1112
Page 015

example, a commonly used optometrist's tool can measure the interpupillary distance. Rulers might measure the offsets between the tracker and eye positions. Cameras could be placed where the user's eyes would normally be in an optical see-through HMD. By recording what the camera sees of the real environment through the see-through HMD, one might be able to determine several viewing parameters. So far, direct measurement techniques have enjoyed limited success (Janin et al., 1993).

View-based tasks are another approach to calibration. These ask the user to perform various tasks that set up geometric constraints. By performing several tasks, enough information is gathered to determine the viewing parameters. For example, Azuma and Bishop (1994) asked a user wearing an optical see-through HMD to look straight through a narrow pipe mounted in the real environment. This sets up the constraint that the user's eye must be located along a line through the center of the pipe. Combining this with other tasks created enough constraints so that all the viewing parameters could be measured. Caudell and Mizell (1992) used a different set of tasks, involving lining up two circles that specified a cone in the real environment. Oishi and Tachi (1996) move virtual cursors to appear on top of beacons in the real environment. All view-based tasks rely upon the user accurately performing the specified task and assume the tracker is accurate. If the tracking and sensing equipment is not accurate, then multiple measurements must be taken and optimizers used to find the "best-fit" solution (Janin et al., 1993).

For video-based systems, an extensive body of literature exists in the robotics and photogrammetry communities on camera calibration techniques; see the references in (Lenz and Tsai, 1988) for a start. Such techniques compute a camera's viewing parameters by taking several pictures of an object of fixed and sometimes unknown geometry. These pictures must be taken from different locations. Matching points in the 2D images with corresponding 3D points on the object sets up mathematical constraints. With enough pictures, these constraints determine the viewing parameters and the 3D location of the calibration object. Alternatively, they can serve to drive an optimization routine that will search for the best set of viewing parameters that fits the collected data. Several AR systems have used camera calibration techniques, including (Drascic and Milgram, 1991; ARGOS, 1994; Bajura, 1993; Tuceryan et al., 1995; Whitaker et al., 1995).

4.3 Dynamic Errors

Dynamic errors occur because of system delays, or lags. The *end-to-end system delay* is defined as the time difference between the moment that the tracking system measures the position and orientation of the viewpoint to the moment when the generated images corresponding to that position and orientation appear in the displays. These delays exist because each component in an augmented reality system requires some time to do its job. The delays in the tracking subsystem, the communication delays, the time it takes the scene generator to draw the appropriate images in the frame buffers, and the scanout time from the frame buffer to the displays all contribute to end-to-end lag. End-to-end delays of 100 ms are fairly typical on existing systems. Simpler systems can have less delay, but other systems have more. Delays of 250 ms or more can exist on slow, heavily loaded, or networked systems.

End-to-end system delays cause registration errors only when motion occurs. Assume that the viewpoint and all objects remain still. Then the lag does not cause registration errors. No matter how long the delay is, the images generated are appropriate, since nothing has moved since the time the tracker measurement was taken. Compare this to the case with motion. For example, assume a user wears a see-through HMD and moves her head. The tracker measures the head at an initial time t. The images corresponding to time t will not appear until some future time t_2, because of the end-to-end system delays. During this delay, the user's head remains in motion, so when the images computed at time t finally appear, the user sees them at a different location than the one for which they were computed. Thus, the images are incorrect for the time they are actually viewed. To the user, the virtual objects appear to

Meta Platforms, Inc.
Exhibit 1112
Page 016

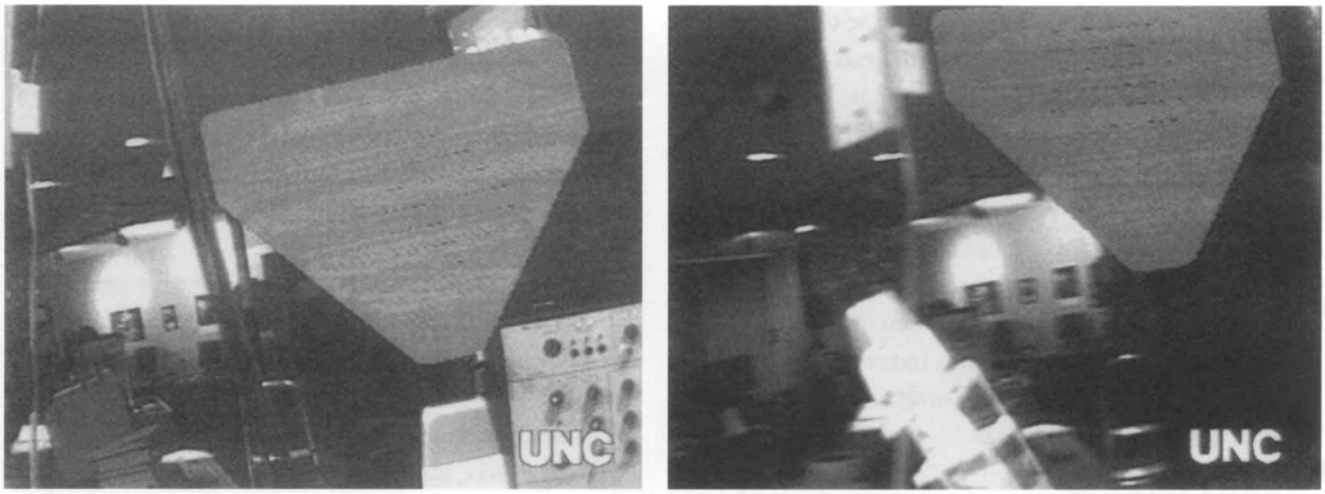

Figure 17. *Effect of motion and system delays on registration. Picture on the left is a static scene. Picture on the right shows motion. (Courtesy UNC Chapel Hill Dept. of Computer Science.)*

"swim around" and "lag behind" the real objects. This delay was graphically demonstrated in a videotape of UNC's ultrasound experiment shown at SIGGRAPH '92 (Bajura et al., 1992). In Figure 17, the picture on the left shows what the registration looks like when everything stands still. The virtual gray trapezoidal region represents what the ultrasound wand is scanning. This virtual trapezoid should be attached to the tip of the real ultrasound wand, as it is in the picture on the left, where the tip of the wand is visible at the bottom of the picture, to the left of the "UNC" letters. But when the head or the wand moves, large dynamic registration errors occur, as shown in the picture on the right. The tip of the wand is now far away from the virtual trapezoid. Also note the motion blur in the background, caused by the user's head motion.

System delays seriously hurt the illusion that the real and virtual worlds coexist because they cause large registration errors. With a typical end-to-end lag of 100 ms and a moderate head rotation rate of 50 degrees per second, the angular dynamic error is 5 degrees. At a 68 cm arm length, this results in registration errors of almost 60 mm. System delay is the largest single source of registration error in existing AR systems, outweighing all others *combined* (Holloway, 1995).

Methods used to reduce dynamic registration fall under four main categories:

- Reduce system lag
- Reduce apparent lag
- Match temporal streams (with video-based systems)
- Predict future locations

4.3.1 Reduce System Lag. The most direct approach is simply to reduce, or ideally eliminate, the system delays. If there are no delays, there are no dynamic errors. Unfortunately, modern scene generators are usually built for throughput, not minimal latency (Foley et al., 1990; Mine, 1993). It is sometimes possible to reconfigure the software to sacrifice throughput to minimize latency. For example, the SLATS system completes rendering a pair of interlaced NTSC images in one field time (16.67 ms) on Pixel-Planes 5 (Olano et al., 1995). Being careful about synchronizing pipeline tasks can also reduce the end-to-end lag (Wloka, 1995).

System delays are not likely to completely disappear anytime soon. Some believe that the current course of technological development will automatically solve this problem. Unfortunately, it is difficult to reduce system delays to the point where they are no longer an issue.

Meta Platforms, Inc.
Exhibit 1112
Page 017

Recall that registration errors must be kept to a small fraction of a degree. At the moderate head rotation rate of 50 degrees per second, system lag must be 10 ms or less to keep angular errors below 0.5 degrees. Just scanning out a frame buffer to a display at 60 Hz requires 16.67 ms. It might be possible to build an HMD system with less than 10 ms of lag, but the drastic cut in throughput and the expense required to construct the system would make alternate solutions attractive. Minimizing system delay is important, but reducing delay to the point where it is no longer a source of registration error is not currently practical.

4.3.2 Reduce Apparent Lag. Image deflection is a clever technique for reducing the amount of apparent system delay for systems that only use head orientation (Burbidge and Murray, 1989; Regan and Pose, 1994; Riner and Browder, 1992; So and Griffin, 1992). It is a way to incorporate more recent orientation measurements into the late stages of the rendering pipeline. Therefore, it is a feed-forward technique. The scene generator renders an image much larger than needed to fill the display. Then, just before scanout, the system reads the most recent orientation report. The orientation value is used to select the fraction of the frame buffer to send to the display, since small orientation changes are equivalent to shifting the frame buffer output horizontally and vertically.

Image deflection does not work on translation, but image-warping techniques might (Chen and Williams, 1993; McMillan and Bishop, 1995a, b). After the scene generator renders the image based upon the head tracker reading, small adjustments in orientation and translation could be done after rendering by warping the image. These techniques assume knowledge of the depth at every pixel, and the warp must be done much more quickly than rerendering the entire image.

4.3.3 Match Temporal Streams. In video-based AR systems, the video camera and digitization hardware impose inherent delays on the user's view of the real world. This delay is potentially a blessing when reducing dynamic errors, because it allows the temporal streams of the real and virtual images to be matched. Additional delay is added to the video from the real world to match the scene-generator delays in generating the virtual images. This additional delay to the video stream will probably not remain constant, since the scene-generator delay will vary with the complexity of the rendered scene. Therefore, the system must dynamically synchronize the two streams.

Note that while this reduces conflicts between the real and virtual, now *both* the real and virtual objects are delayed in time. While this may not be bothersome for small delays, it is a major problem in the related area of telepresence systems and will not be easy to overcome. For long delays, this can produce negative effects such as pilot-induced oscillation.

4.3.4 Predict. The last method is to predict the future viewpoint and object locations. If the future locations are known, the scene can be rendered with these future locations, rather than the measured locations. Then when the scene finally appears, the viewpoints and objects have moved to the predicted locations, and the graphic images are correct at the time they are viewed. For short system delays (under ~80 ms), prediction has been shown to reduce dynamic errors by up to an order of magnitude (Azuma and Bishop, 1994). Accurate predictions require a system built for real-time measurements and computation. Using inertial sensors makes predictions more accurate by a factor of 2–3. Predictors have been developed for a few AR systems (Emura and Tachi, 1994; Zikan et al., 1994b), but the majority were implemented and evaluated with VE systems, as shown in the reference list of (Azuma and Bishop, 1994). More work needs to be done on ways of comparing the theoretical performance of various predictors (Azuma, 1995; Azuma and Bishop, 1995) and in developing prediction models that better match actual head motion (Wu and Ouhyoung, 1995).

4.4 Vision-based Techniques

Mike Bajura and Ulrich Neumann (Bajura and Neumann, 1995) point out that registration based solely on the information from the tracking system is like building an "open-loop" controller. The system has no

Meta Platforms, Inc.
Exhibit 1112
Page 018

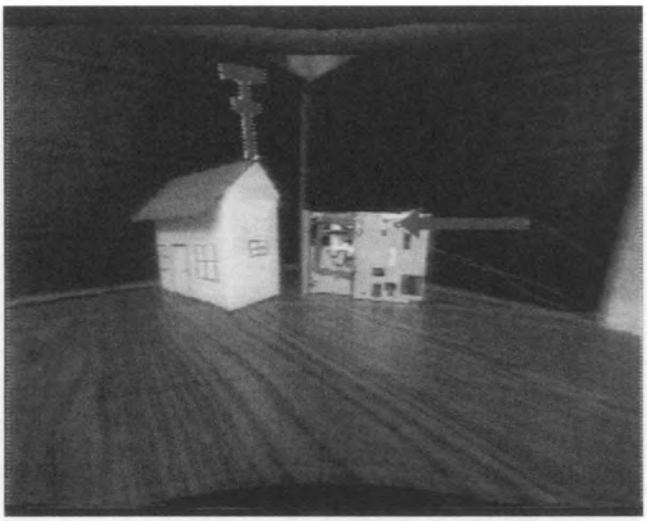

Figure 18. *A virtual arrow and virtual chimney aligned with two real objects. (Courtesy Mike Bajura, UNC Chapel Hill Dept. of Computer Science, and Ulrich Neumann, USC.)*

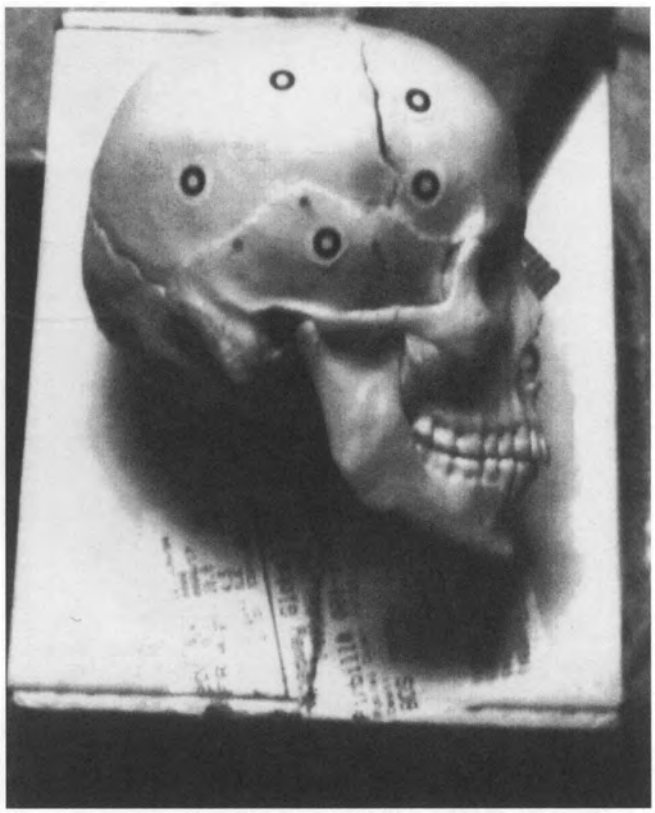

Figure 19. *Real skull with five fiducials. (Courtesy J. P. Mellor, MIT AI Lab.)*

feedback on how closely the real and virtual actually match. Without feedback, it is difficult to build a system that achieves perfect matches. However, video-based approaches can use image processing or computer vision techniques to aid registration. Since video-based AR systems have a digitized image of the real environment, it may be possible to detect features in the environment and use those to enforce registration. They call this a "closed-loop" approach, since the digitized image provides a mechanism for bringing feedback into the system.

Implementing this approach is not a trivial task. Detection and matching must run in real time and must be robust. Special hardware and sensors are often required. However, it is also not an "AI-complete" problem because this is simpler than the general computer-vision problem.

For example, in some AR applications it is acceptable to place fiducials in the environment. These fiducials may be LEDs (Bajura and Neumann, 1995) or special markers (Mellor, 1995a, b; Neumann and Cho, 1996). Recent ultrasound experiments at UNC Chapel Hill have used colored dots as fiducials (State et al., 1996a). The locations or patterns of the fiducials are assumed to be known. Image processing detects the locations of the fiducials; then those are used to make corrections that enforce proper registration.

These routines assume that one or more fiducials are visible at all times; without them, the registration can fall apart. But when the fiducials are visible, the results can be accurate to one pixel—about as close as one can get with video techniques. Figure 18, taken from (Bajura and Neumann, 1995), shows a virtual arrow and a virtual chimney exactly aligned with their desired points on two real objects. The real objects each have an LED to aid the registration. Figures 19 through 21 show registration from (Mellor, 1995a) that uses dots with a circular pattern as the fiducials. The registration is also nearly perfect. Figure 22 demonstrates merging virtual objects with the real environment, using colored dots as the fiducials in a video-based approach. In the picture on the left, the stack of cards in the center are real, but the ones on the right are virtual. Notice that these penetrate one

Meta Platforms, Inc.
Exhibit 1112
Page 019

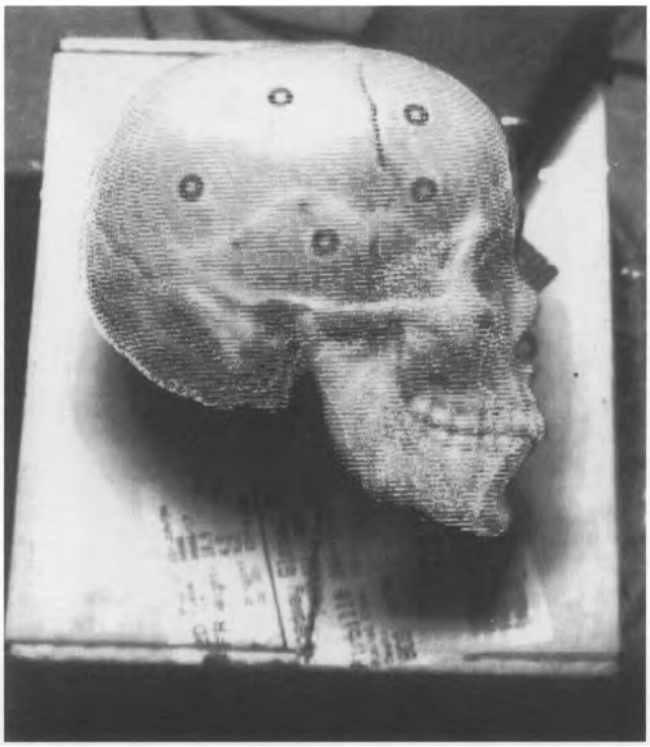

Figure 20. *Virtual wireframe skull registered with real skull. (Courtesy J. P. Mellor, MIT AI Lab.)*

Figure 21. *Virtual wireframe skull registered with real skull moved to a different position. (Courtesy J. P. Mellor, MIT AI Lab.)*

of the blocks. In the image on the right, a virtual spiral object interpenetrates the real blocks and table and also casts virtual shadows upon the real objects (State et al., 1996a).

Instead of fiducials, (Uenohara and Kanade, 1995) uses *template matching* to achieve registration. Template images of the real object are taken from a variety of viewpoints. These are used to search the digitized image for the real object. Once that is found, a virtual wireframe can be superimposed on the real object.

Recent approaches in video-based matching avoid the need for any calibration. (Kutukalos and Vallino, 1996) represents virtual objects in a non-Euclidean, affine frame of reference that allows rendering without knowledge of camera parameters. (Iu and Rogovin, 1996) extracts contours from the video of the real world, then uses an optimization technique to match the contours of the rendered 3D virtual object with the contour extracted from the video. Note that calibration-free approaches may not recover all the information required to perform all potential AR tasks. For example, these two approaches do not recover true depth information, which is useful when compositing the real and the virtual.

Techniques that use fiducials as the sole tracking source determine the *relative* projective relationship between the objects in the environment and the video camera. Although this is enough to ensure registration, it does not provide all the information one might need in some AR applications, such as the absolute (rather than relative) locations of the objects and the camera. Absolute locations are needed to include virtual and real objects that are not tracked by the video camera, such as a 3D pointer or other virtual object not directly tied to real objects in the scene.

Additional sensors besides video cameras can aid registration. Both (Mellor, 1995a, b) and (Grimson et al., 1994, 1995) use a laser rangefinder to acquire an initial depth map of the real object in the environment. Given a matching virtual model, the system can match the depth

Meta Platforms, Inc.
Exhibit 1112
Page 020

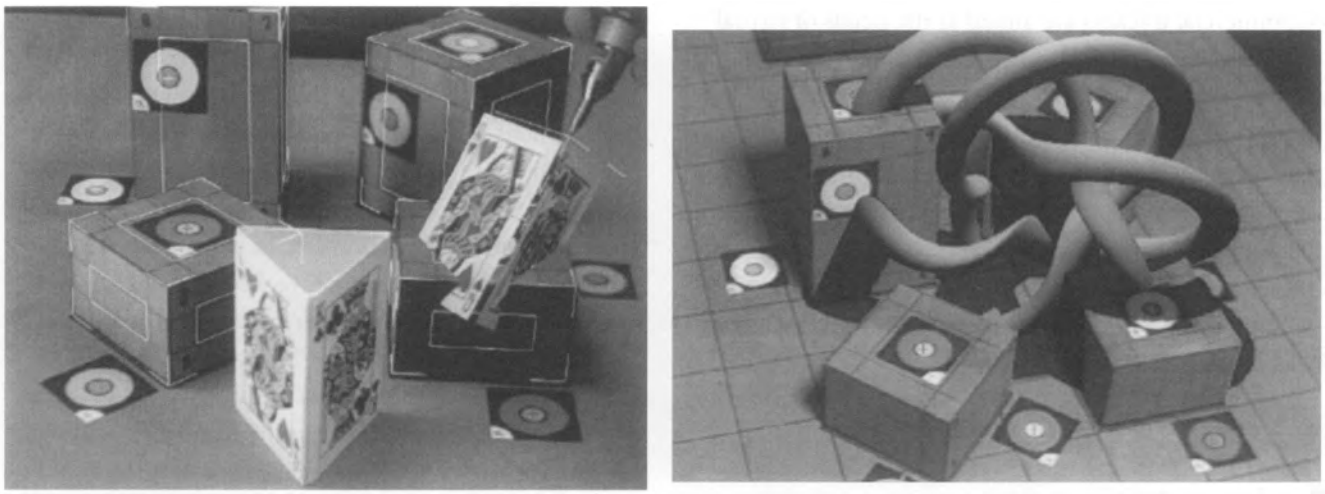

Figure 22. *Virtual cards and spiral object merged with real blocks and table. (Courtesy Andrei State, UNC Chapel Hill Dept. of Computer Science.)*

maps from the real and virtual until they are properly aligned, and that provides the information needed for registration.

Another way to reduce the difficulty of the problem is to accept the fact that the system may not be robust and may not be able to perform all tasks automatically. Then it can ask the user to perform certain tasks. The system described in Sharma and Molineros (1994) expects manual intervention when the vision algorithms fail to identify a part because the view is obscured. The calibration techniques in Tuceryan et al. (1995) are heavily based on computer vision techniques, but they ask the user to manually intervene by specifying correspondences when necessary.

4.5 Current Status

The registration requirements for AR are difficult to satisfy, but a few systems have achieved good results. (Azuma and Bishop, 1994) is an open-loop system that shows registration typically within ±5 millimeters from many viewpoints for an object at about arm's length. Closed-loop systems, however, have demonstrated nearly perfect registration, accurate to within a pixel (Bajura and Neumann, 1995; Mellor, 1995a, b; Neumann and Cho, 1996; State et al., 1996a).

The registration problem is far from solved. Many systems assume a static viewpoint, static objects, or even both. Even if the viewpoint or objects are allowed to move, they are often restricted in how far they can travel. Registration is shown under controlled circumstances, often with only a small number of real-world objects, or where the objects are already well-known to the system. For example, registration may only work on one object marked with fiducials, and not on any other objects in the scene. Much more work needs to be done to increase the domains in which registration is robust. Duplicating registration methods remains a nontrivial task, due to both the complexity of the methods and the additional hardware required. If simple yet effective solutions could be developed, that would speed the acceptance of AR systems.

5 Sensing

Accurate registration and positioning of virtual objects in the real environment requires accurate tracking of the user's head and sensing the locations of other objects in the environment. The biggest single obstacle to building effective augmented reality systems is the requirement of accurate, long-range sensors and trackers that report the locations of the user and the surrounding objects in the environment. For details of tracking technologies, see the surveys in (Ferrin, 1991; Meyer et al., 1992) and Chapter 5 of Durlach and Mavor (1995).

Meta Platforms, Inc.
Exhibit 1112
Page 021

Commercial trackers are aimed at the needs of virtual environments and motion-capture applications. Compared to those two applications, augmented reality has much stricter accuracy requirements and demands larger working volumes. No tracker currently provides high accuracy at long ranges in real time. More work needs to be done to develop sensors and trackers that can meet these stringent requirements.

Specifically, AR demands more from trackers and sensors in three areas:

- Greater input variety and bandwidth
- Higher accuracy
- Longer range

5.1 Input Variety and Bandwidth

VE systems are primarily built to handle output bandwidth, for example, the images displayed and sounds generated. The input bandwidth is tiny—for example, the locations of the user's head and hands, the outputs from the buttons and other control devices. AR systems, however, will need a greater variety of input sensors and much more input bandwidth (Buxton, personal communication, MIT Workshop on Ubiquitous Computing, Cambridge, MA, 1993). There are a greater variety of possible input sensors than output displays. Outputs are limited to the five human senses. Inputs can come from anything a sensor can detect. Robinett speculates that augmented reality may be useful in any application that requires displaying information not directly available or detectable by human senses by making that information visible (or audible, touchable, etc.) (Robinett, 1992). Recall that the proposed medical applications in Section 2.1 use CT, MRI, and ultrasound sensors as inputs. Other future applications might use sensors to extend the user's visual range into infrared or ultraviolet frequencies, and remote sensors would let users view objects hidden by walls or hills. Conceptually, anything not detectable by human senses but detectable by machines might be transduced into something that a user can sense in an AR system.

Range data is a particular input that is vital for many AR applications (Aliaga, 1997; Breen and Rose, 1996). The AR system knows the distance to the virtual objects, because that model is built into the system. But the AR system may not know where all the real objects are in the environment. The system might assume that the entire environment is measured at the beginning and remains static thereafter. However, some useful applications will require a dynamic environment, in which real objects move, so the objects must be tracked in real time. However, for some applications a depth map of the real environment would be sufficient. That would allow real objects to occlude virtual objects through a pixel-by-pixel depth value comparison. Acquiring this depth map in real time is not trivial. Sensors like laser rangefinders might be used. Many computer-vision techniques for recovering shape through various strategies (e.g., "shape from stereo," or "shape from shading") have been tried. A recent work (Wloka and Anderson, 1995) uses intensity-based matching from a pair of stereo images to do depth recovery. Recovering depth through existing vision techniques is difficult to do robustly in real time.

Finally, some annotation applications require access to a detailed database of the environment; this database is a type of input to the system. For example, the architectural application of "seeing into the walls" assumes that the system has a database describing where all the pipes, wires, and other hidden objects are within the building. Such a database may not be readily available, and even if it is, it may not be in a format that is easily usable. For example, the data may not be grouped to segregate the parts of the model that represent wires from the parts that represent pipes. Thus, a significant modeling effort may be required and should be taken into consideration when building an AR application.

5.2 High Accuracy

The accuracy requirements for the trackers and sensors are driven by the accuracies needed for visual registration, as described in Section 4. For many approaches, the registration is only as accurate as the tracker. Therefore, the AR system needs trackers that are accurate to around one millimeter and a tiny fraction of a degree, across the entire working range of the tracker.

Few trackers can meet this specification, and every

Meta Platforms, Inc.
Exhibit 1112
Page 022

technology has weaknesses. Some mechanical trackers are accurate enough, although they tether the user to a limited working volume. Magnetic trackers are vulnerable to distortion by metal that exists in many desired AR application environments. Ultrasonic trackers suffer from noise and are difficult to make accurate at long ranges because of variations in the ambient temperature. Optical technologies (Janin et al., 1994) have distortion and calibration problems. Inertial trackers drift with time. Of the individual technologies, optical technologies show the most promise because of trends toward high-resolution digital cameras, real-time photogrammetric techniques, and structured light sources that result in more signal strength at long distances. Future tracking systems that can meet the stringent requirements of AR will probably be hybrid systems (Azuma, 1993; Durlach and Mavor, 1995; Foxlin, 1996; Zikan et al., 1994b), such as a combination of inertial and optical technologies. Using multiple technologies opens the possibility of covering for each technology's weaknesses by combining their strengths.

Attempts have been made to calibrate the distortions in commonly used magnetic tracking systems (Bryson, 1992; Ghazisaedy et al., 1995). These have succeeded at removing much of the gross error from the tracker at long ranges, but not to the level required by AR systems (Holloway, 1995). For example, mean errors at long ranges can be reduced from several inches to around one inch.

The requirements for registering other sensor modes are not nearly as stringent. For example, the human auditory system is not very good at localizing deep bass sounds. For this reason subwoofer placement is not critical in a home-theater system.

5.3 Long Range

Few trackers are built for accuracy at long ranges, since most VE applications do not require long ranges. *Motion capture* applications track an actor's body parts to control a computer-animated character or for the analysis of an actor's movements. This approach is fine for position recovery, but not for orientation. Orientation recovery is based upon computed positions. Even tiny errors in those positions can cause orientation errors of a few degrees, a variation that is too large for AR systems.

Two scalable tracking systems for HMDs have been described in the literature (Ward et al., 1992; Sowizral and Barnes, 1993). A scalable system is one that can be expanded to cover any desired range, simply by adding more modular components to the system. This type of system is created by building a cellular tracking system in which only nearby sources and sensors are used to track a user. As the user walks around, the set of sources and sensors changes, thus achieving large working volumes while avoiding long distances between the current working set of sources and sensors. While scalable trackers can be effective, they are complex and by their very nature have many components, making them relatively expensive to construct.

The Global Positioning System (GPS) is used to track the locations of vehicles almost anywhere on the planet. It might be useful as one part of a long range tracker for AR systems. However, by itself it will not be sufficient. The best reported accuracy is approximately one centimeter, assuming that many measurements are integrated (so that accuracy is not generated in real time), when GPS is run in differential mode. That is not sufficiently accurate to recover orientation from a set of positions on a user.

Tracking an AR system outdoors in real time with the required accuracy has not been demonstrated and remains an open problem.

6 Future Directions

This section identifies areas and approaches that require further research to produce improved AR systems.

6.1 Hybrid Approaches

Future tracking systems may be hybrids, because combined approaches can cover weaknesses. The same may be true for other problems in AR. For example, current registration strategies generally focus on a single

Meta Platforms, Inc.
Exhibit 1112
Page 023

strategy. Future systems may be more robust if several techniques are combined. An example is combining vision-based techniques with prediction: If the fiducials are not available, the system switches to open-loop prediction to reduce the registration errors, rather than breaking down completely. The predicted viewpoints in turn produce a more accurate initial location estimate for the vision-based techniques.

6.2 Real-Time Systems and Time-Critical Computing

Many VE systems are not truly run in real time. Instead, it is common to build the system, often on UNIX, and then see how fast it runs. This approach may be sufficient for some VE applications; since everything is virtual, all the objects are automatically synchronized with each other. AR is a different story. Now the virtual and real must be synchronized, and the real world "runs" in real time. Therefore, effective AR systems must be built with real-time performance in mind. Accurate timestamps must be available. Operating systems must not arbitrarily swap out the AR software process at any time, for arbitrary durations. Systems must be built to guarantee completion within specified time budgets, rather than just "running as quickly as possible." These are characteristics of flight simulators and a few VE systems (Krueger, 1992). Constructing and debugging real-time systems is often painful and difficult, but the requirements for AR demand real-time performance.

6.3 Perceptual and Psychophysical Studies

Augmented reality is an area ripe for psychophysical studies. How much lag can a user detect? How much registration error is detectable when the head is moving? Besides questions on perception, psychological experiments that explore *performance* issues are also needed. How much does head-motion prediction improve user performance on a specific task? How much registration error is tolerable for a specific application before performance on that task degrades substantially? Is the allowable error larger while the user moves her head versus when she stands still? Furthermore, not much is known about potential optical illusions caused by errors or conflicts in the simultaneous display of real and virtual objects (Durlach and Mavor, 1995).

Few experiments in this area have been performed. Jannick Rolland, Frank Biocca, and their students conducted a study of the effect caused by eye displacements in video see-through HMDs (Rolland et al., 1995). They found that users partially adapted to the eye displacement, but they also had negative aftereffects after removing the HMD. Steve Ellis's group at NASA Ames has conducted work on perceived depth in a see-through HMD (Ellis and Bucher, 1994; Ellis and Menges, 1995). ATR (Advanced Telecommunications Research) has also conducted a study (Utsumi et al., 1994).

6.4 Portability

Section 3.4 explained why some potential AR applications require giving the user the ability to walk around large environments, even outdoors. For this reason the equipment must be self-contained and portable. Existing tracking technology is not capable of tracking a user outdoors at the required accuracy.

6.5 Multimodal Displays

Almost all work in AR has focused on the visual sense: virtual graphic objects and overlays. But, as Section 3.1 explained, that augmentation might apply to all other senses as well. In particular, adding and removing 3D sound is a capability that could be useful in some AR applications.

6.6 Social and Political Issues

Technological issues are not the only ones that need to be considered when building a real application. There are also social and political dimensions when getting new technologies into the hands of real users. Sometimes, perception is what counts, even if the technological reality is different. For example, if workers perceive lasers to be a health risk, they may refuse to use a system with lasers in the display or in the trackers, even if

Meta Platforms, Inc.
Exhibit 1112
Page 024

those lasers are eye safe. Ergonomics and ease of use are paramount considerations. Whether AR is truly a cost-effective solution in its proposed applications has yet to be determined. Another important factor is whether or not the technology is perceived as a threat to jobs and as a replacement for workers, especially now that many corporations have downsized. AR may be positively perceived in this regard, because it is intended as a tool to make the user's job easier, rather than something that completely replaces the human worker. Although technology transfer is not normally a subject of academic papers, it is a real problem. Social and political concerns should not be ignored during attempts to move AR out of the research lab and into the hands of real users.

7 Conclusion

Augmented reality is far behind virtual environments in maturity. Several commercial vendors sell complete, turnkey virtual environment systems. However, no commercial vendor currently sells an HMD-based augmented reality system. A few monitor-based "virtual set" systems are available, but today AR systems are primarily found in academic and industrial research laboratories.

The first deployed HMD-based AR systems will probably be in the application of aircraft manufacturing. Both Boeing (Boeing TRP, 1994; ARPA, 1995) and McDonnell Douglas (Neumann and Cho, 1996) are exploring this technology. The former uses optical approaches, while the latter is pursuing video approaches. Boeing has performed trial runs with workers using a prototype system but has not yet made any deployment decisions. Annotation and visualization applications in restricted, limited-range environments are deployable today, although much more work needs to be done to make them cost effective and flexible. Applications in medical visualization will take longer. Prototype visualization aids have been used on an experimental basis, but the stringent registration requirements and ramifications of mistakes will postpone common usage for many years. AR will probably be used for medical training before it is commonly used in surgery.

The next generation of combat aircraft will have helmet-mounted sights, with graphics registered to targets in the environment (Wanstall, 1989). These displays, combined with short-range steerable missiles that can shoot at targets off-boresight, give a tremendous combat advantage to pilots in dogfights. Instead of having to be directly behind his target in order to shoot at it, a pilot can now shoot at anything within a 60–90° cone of his aircraft's forward centerline. Russia and Israel currently have systems with this capability, and the United States is expected to field the AIM-9X missile with its associated Helmet-Mounted Sight in 2002 (Dornheim and Hughes, 1995; Dornheim, 1995a). Registration errors due to delays are a major problem in this application (Dornheim, 1995b).

Augmented reality is a relatively new field. Most of the research efforts have occurred in the past four years, as shown by the references listed at the end of this paper. The SIGGRAPH "Rediscovering Our Fire" report identified augmented reality as one of four areas where SIGGRAPH should encourage more submissions (Mair, 1994). Because of the numerous challenges and unexplored avenues in this area, AR will remain a vibrant area of research for at least the next several years.

One area where a breakthrough is required is tracking an HMD outdoors at the accuracy required by AR. If this is accomplished, several interesting applications will become possible. Two examples are described here: navigation maps and visualization of past and future environments.

The first application is a navigation aid to people walking outdoors. These individuals could be soldiers advancing upon their objective, hikers lost in the woods, or tourists seeking directions to their intended destination. Today, these individuals must pull out a physical map and associate what they see in the real environment around them with the markings on the 2D map. If landmarks are not easily identifiable, this association can be difficult to perform, as anyone lost in the woods can attest. An AR system makes navigation easier by performing the association step automatically. If the user's position and orientation are known, and the AR system has access to a digital map of the area, then the AR system can draw the map in 3D directly upon the user's view.

Meta Platforms, Inc.
Exhibit 1112
Page 025

The user looks at a nearby mountain and sees graphics directly overlaid on the real environment explaining the mountain's name, how tall it is, how far away it is, and where the trail is that leads to the top.

The second application is visualization of locations and events as they were in the past or as they will be after future changes are performed. Tourists that visit historical sites, such as a Civil War battlefield or the Acropolis in Athens, Greece, do not see these locations as they were in the past, due to changes over time. It is often difficult for a modern visitor to imagine what these sites really looked like in the past. To help, some historical sites stage "living history" events where volunteers wear ancient clothes and reenact historical events. A tourist equipped with an outdoors AR system could see a computer-generated version of the living history. The HMD could cover up modern buildings and monuments in the background and show, directly on the grounds at Gettysburg, where the Union and Confederate troops were at the fateful moment of Pickett's charge. The gutted interior of the modern Parthenon could be filled in by computer-generated representations of what it looked like in 430 B.C. including the long-vanished gold statue of Athena in the middle. Tourists and students walking around the grounds with such AR displays would gain a much better understanding of these historical sites and the important events that took place there. Similarly, AR displays could show what proposed architectural changes would look like before they were carried out. An urban designer could show clients and politicians what a new stadium would look like as they walked around the adjoining neighborhood, to better understand how the stadium project would affect nearby residents.

After the basic problems with AR are solved, the ultimate goal will be to generate virtual objects that are so realistic that they are virtually indistinguishable from the real environment. Photorealism has been demonstrated in feature films, but accomplishing this in an interactive application will be much harder. Lighting conditions, surface reflections, and other properties must be measured automatically, in real time. More sophisticated lighting, texturing, and shading capabilities must run at interactive rates in future scene generators. Registration must be nearly perfect, without manual intervention or adjustments. While these are difficult problems, they are probably not insurmountable. It took about 25 years to progress from drawing stick figures on a screen to the photorealistic dinosaurs in "Jurassic Park." Within another 25 years, we should be able to wear a pair of AR glasses outdoors to see and interact with photorealistic dinosaurs eating a tree in our backyard.

Acknowledgments

This paper is an updated version of my course notes for an ACM SIGGRAPH '95 class called "Developing Advanced Virtual Reality Applications," given in Los Angeles, CA, in August 1995.

I thank the anonymous reviewers for their constructive criticism and suggestions for improving this paper. I also thank the following individuals and organizations for sending pictures to include with this paper: Mike Bajura, Andrei State, and Linda Houseman, University of North Carolina at Chapel Hill Department of Computer Science; David Drascic and Paul Milgram, University of Toronto; Steve Feiner and Blair MacIntyre, Columbia University; Alessandro Giacalone, The European Computer-Industry Research Centre (ECRC) and David Breen, Caltech; J. P. Mellor, MIT AI Laboratory; David Mizell, Boeing; Ulrich Neumann, University of Southern California; Jannick Rolland, Center for Research and Engineering in Optics and Lasers (CREOL) at the University of Central Florida (rolland@creol.ucf.edu).

References

Note: Some of these references are available electronically at the following sites on the World Wide Web:
Columbia U. http://www.cs.columbia.edu/graphics/
ECRC http://www.ecrc.de/
MIT AI Lab http://www.ai.mit.edu/
UNC Chapel Hill http://www.cs.unc.edu/
U. Toronto http://vered.rose.utoronto.ca/etc-lab.html/

Aliaga, D. G. (1997). Virtual objects in the real world. *Communications of the ACM 40,* 3 (49–54).

ARGOS virtual pointer camera calibration procedure WWW

Meta Platforms, Inc.
Exhibit 1112
Page 026

page. (1994). http://vered.rose.utoronto.ca/people/david_dir/POINTER/Calibration.html.

ARPA ESTO WWW page. (1995). http://molothrus.sysplan.com/ESTO/.

Azuma, R. (1993). Tracking requirements for augmented reality. *Communications of the ACM, 36* (7), 50–51.

Azuma, R. T. (1995). *Predictive tracking for augmented reality.* Ph.D. dissertation, Dept. of Computer Science, University of North Carolina at Chapel Hill. Available as UNC-CH CS Dept. Technical Report TR95-007.

Azuma, R., & Bishop, G. (1994). Improving static and dynamic registration in a see-through HMD. *Computer Graphics (Proceedings of SIGGRAPH '94)*, Orlando, 197–204.

Azuma, R., & Bishop, G. (1995). A frequency-domain analysis of head-motion prediction. *Computer Graphics (Proceedings of SIGGRAPH '95)*, Los Angeles, 401–408.

Barfield, W., Rosenberg, C., and Lotens, W. A. (1995). Augmented-reality displays. In W. Barfield, & Furness, T. A., III (Eds.). *Virtual Environments and Advanced Interface Design* (542–575). Oxford: Oxford University Press.

Bajura, M. (1993). *Camera calibration for video see-through head-mounted display.* Technical Report TR93-048. Dept. of Computer Science, University of North Carolina at Chapel Hill.

Bajura, M., & Neumann, U. (1995). Dynamic registration correction in video-based augmented reality systems. *IEEE Computer Graphics and Applications, 15* (5), 52–60.

Bajura, M., Fuchs, H., & Ohbuchi, R. (1992). Merging virtual reality with the real world: Seeing ultrasound imagery within the patient. *Computer Graphics (Proceedings of SIGGRAPH '92), 26* (2), 203–210.

Betting, F., Feldmar, J., Ayache, N., & Devernay, F. (1995). A new framework for fusing stereo images with volumetric medical images. *Proceedings of Computer Vision, Virtual Reality, and Robotics in Medicine '95,* Nice, 30–39.

Boeing TRP WWW page. (1994). http://esto.sysplan.com/ESTO/Displays/HMD-TDS/Factsheets/Boeing.html.

Bowskill, J., & Downie, J. (1995). Extending the capabilities of the human visual system: An introduction to enhanced reality. *Computer Graphics, 29* (2), 61–65.

Breen, D. E., Whitaker, R. T., Rose, E., & Tuceryan, M. (1996). Interactive occlusion and automatic object placement for augmented reality. *Proceedings of Eurographics '96,* Futuroscope-Poitiers, 11–22.

Brooks, F. P., Jr. (1996). The computer scientist as toolsmith II. *Communications of the ACM, 39* (3), 61–68.

Bryson, S. (1992). Measurement and calibration of static distortion of position data from 3D trackers. *SPIE Proceedings Vol. 1669: Stereoscopic Displays and Applications III,* San Jose, 244–255.

Burbidge, D., & Murray, P. M. (1989). Hardware improvements to the helmet-mounted projector on the Visual Display Research Tool (VDRT) at the Naval Training Systems Center. *SPIE Proceedings Vol. 1116: Head-Mounted Displays,* 52–59.

Caudell, T. P. (1994). Introduction to Augmented Reality. *SPIE Proceedings Vol. 2351: Telemanipulator and Telepresence Technologies,* Boston, 272–281.

Caudell, T. P., & Mizell, D. W. (1992). Augmented reality: An application of heads-up display technology to manual manufacturing processes. *Proceedings of the Hawaii International Conference on System Sciences,* 659–669.

Chen, S. E., & Williams, L. (1993). View interpolation for image synthesis. *Computer Graphics (Proceedings of SIGGRAPH '93),* Anaheim, 279–288.

Deering, M. (1992). High-resolution virtual reality. *Computer Graphics (Proceedings of SIGGRAPH '92), 26* (2), Chicago, 195–202.

Doenges, P. K. (1985). Overview of computer image generation in visual simulation. *Course Notes, 14: ACM SIGGRAPH 1985,* San Francisco.

Dornheim, M. A. (1995a). U.S. fighters to get helmet displays after 2000. *Aviation Week and Space Technology, 143* (17), 46–48.

Dornheim, M. A. (1995b). Helmet-mounted sights must overcome delays. *Aviation Week and Space Technology, 143* (17), 54.

Dornheim, M. A., & Hughes, D. (1995). U.S. intensifies efforts to meet missile threat. *Aviation Week and Space Technology, 143* (16), 36–39.

Drascic, D. (1993). Stereoscopic vision and augmented reality. *Scientific Computing and Automation, 9* (7), 31–34.

Drascic, D., & Milgram, P. (1991). Positioning accuracy of a virtual stereographic pointer in a real stereoscopic video world. *SPIE Proceedings Vol. 1457: Stereoscopic Displays and Applications II,* San Jose, 302–313.

Drascic, D., Grodski, J. J., Milgram, P., Ruffo, K., Wong, P., & Zhai, S. (1993). ARGOS: A display system for augmenting reality. *Video Proceedings of the International Conference on Computer-Human Interaction '93: Human Factors in Computing Systems.* Also in *ACM SIGGRAPH Technical Video Review, 88.* Extended abstract in *Proceedings of INTERCHI '93,* 521.

Meta Platforms, Inc.
Exhibit 1112
Page 027

Durlach, N. I., & Mavor, A. S. (Eds.) (1995). *Virtual reality: Scientific and technological challenges.* Washington DC: National Academy Press.

Edwards, E., Rolland, J., & Keller, K. (1993). Video see-through design for merging of real and virtual environments. *Proceedings of the IEEE Virtual Reality Annual International Symposium '93,* Seattle, 222–233.

Edwards, P. J., Hill, D. L. G., Hawkes, D. J., Spink, R., Colchester, A. C. F., Strong, A., & Gleeson, M. (1995). Neurosurgical guidance using the stereo microscope. *Proceedings of Computer Vision, Virtual Reality, and Robotics in Medicine '95,* Nice, 555–564.

Ellis, S. R., & Bucher, U. J. (1994). Distance perception of stereoscopically presented virtual objects optically superimposed on physical objects by a head-mounted see-through display. *Proceedings of 38th Annual Meeting of the Human Factors and Ergonomics Society,* Nashville, 1300–1305.

Ellis, S. R., & Menges, B. M. (1995). Judged distance to virtual objects in the near visual field. *Proceedings of 39th Annual Meeting of the Human Factors and Ergonomics Society,* San Diego, 1400–1404.

Emura, S., & Tachi, S. (1994). Compensation of time lag between actual and virtual spaces by multi-sensor integration. *Proceedings of the 1994 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems,* Las Vegas, 463–469.

Feiner, S. (1994a). Augmented reality. *Course Notes, 2: ACM SIGGRAPH 1994, 7,* 1–11.

Feiner, S. (1994b). Redefining the user interface: Augmented reality. *Course Notes, 2: ACM SIGGRAPH 1994, 18,* 1–7.

Feiner, S., MacIntyre, B., & Seligmann, D. (1993a). Knowledge-based augmented reality. *Communications of the ACM, 36* (7), 52–62.

Feiner, S., MacIntyre, B., Haupt, M., & Solomon, E. (1993b). Windows on the world: 2D windows for 3D augmented reality. *Proceedings of the 1993 User Interface Software and Technology,* Atlanta, 145–155.

Feiner, S. K., Webster, A. C., Krueger, T. E. III, MacIntyre, B., & Keller, E. J. (1995). Architectural Anatomy. *Presence: Teleoperators and Virtual Environments, 4* (3), 318–325.

Ferrin, F. J. (1991). Survey of helmet tracking technologies. *SPIE Proceedings Vol. 1456: Large-Screen Projection, Avionic, and Helmet-Mounted Displays,* 86–94.

Fitzmaurice, G. (1993). Situated information spaces: Spatially aware palmtop computers. *Communications of the ACM, 36* (7), 38–49.

Foley, J. D., van Dam, A., Feiner, S. K., & Hughes, J. F. (1990). *Computer graphics: Principles and practice* (2nd ed.). Reading, MA: Addison-Wesley.

Foxlin, E. (1996). Inertial head-tracker sensor fusion by a complementary separate-bias Kalman Filter. *Proceedings of the Virtual Reality Annual International Symposium '96,* Santa Clara, 185–194.

Ghazisaedy, M., Adamczyk, D., Sandin, D. J., Kenyon, R. V., & DeFanti, T. A. (1995). Ultrasonic calibration of a magnetic tracker in a virtual reality space. *Proceedings of the IEEE Virtual Reality Annual International Symposium '95,* Research Triangle Park, 179–188.

Grimson, W., Lozano-Pérez, T., Wells, W., Ettinger, G., White, S., & Kikinis, R. (1994). An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* Los Alamitos, 430–436.

Grimson, W. E. L., Ettinger, G. J., White, S. J., Gleason, P. L., Lozano-Pérez, T., Wells, W. M. III, & Kikinis, R. (1995). Evaluating and validating an automated registration system for enhanced reality visualization in surgery. *Proceedings of Computer Vision, Virtual Reality, and Robotics in Medicine '95,* Nice, 3–12.

Holloway, R. (1995). *Registration errors in augmented reality.* Ph.D. dissertation. Dept. of Computer Science, University of North Carolina at Chapel Hill. Available as UNC-CH CS Dept. Technical Report TR95-016.

Holmgren, D. E. (1992). *Design and construction of a 30-degree see-through head-mounted display.* Technical Report TR92-030. Dept. of Computer Science, University of North Carolina at Chapel Hill.

Iu, S.-L., & Rogovin, K. W. (1996). Registering perspective contours with 3D objects without correspondence using orthogonal polynomials. *Proceedings of the Virtual Reality Annual International Symposium '96,* Santa Clara, 37–44.

Jain, A. K. (1989). *Fundamentals of digital image processing.* Englewood Cliffs, NJ: Prentice Hall.

Janin, A. L., Mizell, D. W., & Caudell, T. P. (1993). Calibration of head-mounted displays for augmented reality applications. *Proceedings of the Virtual Reality Annual International Symposium '93,* Seattle, 246–255.

Janin, A., Zikan, K., Mizell, D., Banner, M., & Sowizral, H. (1994). A videometric head tracker for augmented reality. *SPIE Proceedings Vol. 2351: Telemanipulator and Telepresence Technologies,* Boston, 308–315.

Meta Platforms, Inc.
Exhibit 1112
Page 028

Kancherla, A. R., Rolland, J. P., Wright, D. L., & Burdea, G. (1995). A novel virtual reality tool for teaching dynamic 3D anatomy. *Proceedings of Computer Vision, Virtual Reality, and Robotics in Medicine '95,* Nice, 163–169.

Kim, W. S. (1993). Advanced teleoperation, graphics aids, and application to time delay environments. *Proceedings of the First Industrial Virtual Reality Show and Conference,* Makuhari Meese, 202–207.

Kim, W. S. (1996). Virtual reality calibration and preview/ Predictive displays for telerobotics. *Presence: Teleoperators and Virtual Environments, 5* (2), 173–190.

Krueger, M. W. (1992). Simulation versus artificial reality. *Proceedings of IMAGE VI Conference,* Scottsdale, 147–155.

Kutulakos, K. N., & Vallino, J. (1996). Affine object representations for calibration-free augmented reality. *Proceedings of the Virtual Reality Annual International Symposium '96,* Santa Clara, CA, 25–36.

Lenz, R. K., & Tsai, R. Y. (1988). Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 10* (5), 713–720.

Lion, D., Rosenberg, C., & Barfield, W. (1993). Overlaying three-dimensional computer graphics with stereoscopic live motion video: Applications for virtual environments. *Society for Information Display International Symposium Digest of Technical Papers,* Seattle, 483–486.

Lorensen, W., Cline, H., Nafis, C., Kikinis, R., Altobelli, D., & Gleason, L. (1993). Enhancing reality in the operating room. *Proceedings of Visualization '93,* Los Alamitos, 410–415.

Maes, P. (1995). Artificial life meets entertainment: Lifelike autonomous agents. *Communications of the ACM, 38* (11), 108–114.

Mair, S. G. (1994). Preliminary Report on SIGGRAPH in the 21st Century: Rediscovering Our Fire. *Computer Graphics, 28* (4), 288–296.

McMillan, L., & Bishop, G. (1995a). Head-tracked stereoscopic display using image warping. *SPIE Proceedings 2409: Electronic Imaging Science and Technology,* San Jose, 21–30.

McMillan, L., & Bishop, G. (1995b). Plenoptic modeling. *Computer Graphics (Proceedings of SIGGRAPH '95),* Los Angeles, 39–46.

Mellor, J. P. (1995a). *Enhanced reality visualization in a surgical environment.* Master's thesis, Dept. of Electrical Engineering, MIT.

Mellor, J. P. (1995b). Realtime camera calibration for enhanced reality visualization. *Proceedings of Computer Vision, Virtual Reality, and Robotics in Medicine '95,* Nice, 471–475.

Meyer, K., Applewhite, H. L., & Biocca, F. A. (1992). A survey of position-trackers. *Presence: Teleoperators and Virtual Environments, 1* (2), 173–200.

Milgram, P., Zhai, S., Drascic, D., & Grodski, J. J. (1993). Applications of augmented reality for human-robot communication. *Proceedings of International Conference on Intelligent Robotics and Systems,* Yokohama, 1467–1472.

Milgram, P., & Kishino, F. (1994a). A taxonomy of mixed reality virtual displays. *Institute of Electronics, Information, and Communication Engineers Transactions on Information and Systems, E77-D* (9) 1321–1329.

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1994b). Augmented reality: A class of displays on the reality-virtuality continuum. *SPIE Proceedings Vol. 2351: Telemanipulator and Telepresence Technologies,* Boston, 282–292.

Milgram, P., Drascic, D., Grodski, J. J., Restogi, A., Zhai, S., & Zhou, C. (1995). Merging real and virtual worlds. *Proceedings of IMAGINA '95,* Monte Carlo, 218–230.

Mine, M. R. (1993). *Characterization of end-to-end delays in head-mounted display systems.* Technical Report TR93-001. Dept. of Computer Science, University of North Carolina at Chapel Hill.

Neumann, U., & Cho, Y. (1996). A self-tracking augmented reality system. *Proceedings of Virtual Reality Software and Technology '96,* Hong Kong, 109–115.

Oishi, T., & Tachi, S. (1996). Methods to calibrate projection transformation parameters for see-through head-mounted displays. *Presence: Teleoperators and Virtual Environments, 5* (1), 122–135.

Olano, M., Cohen, J., Mine, M., & Bishop, G. (1995). Combating graphics system latency. *Proceedings of 1995 Symposium on Interactive 3D Graphics,* Monterey, 19–24.

Oyama, E., Tsunemoto, N., Tachi, S., & Inoue, Y. (1993). Experimental study on remote manipulation using virtual reality. *Presence: Teleoperators and Virtual Environments, 2* (2), 112–124.

Pausch, R., Crea, T., & Conway, M. (1992). A literature survey for virtual environments: Military flight simulator visual systems and simulator sickness. *Presence: Teleoperators and Virtual Environments, 1* (3), 344–363.

Peuchot, B., Tanguy, A., & Eude, M. (1995). Virtual reality as an operative tool during scoliosis surgery. *Proceedings of Computer Vision, Virtual Reality, and Robotics in Medicine '95,* Nice, 549–554.

Meta Platforms, Inc.
Exhibit 1112
Page 029

Regan, M., & Pose, R. (1994). Priority rendering with a virtual reality address recalculation pipeline. *Computer Graphics (Proceedings of SIGGRAPH '94)*, Orlando, 155–162.

Rekimoto, J. (1995). The magnifying glass approach to augmented reality systems. *Proceedings of ICAT '95*, Makuhari Meese.

Rekimoto, J., & Nagao, K. (1995). The world through the computer: Computer augmented interaction with real world environments. *Proceedings of the 1995 User Interface Software Technology*, Pittsburgh, 29–36.

Riner, B., & Browder, B. (1992). Design guidelines for a carrier-based training system. *Proceedings of IMAGE VI*, Scottsdale, 65–73.

Robinett, W. (1992). Synthetic experience: A proposed taxonomy. *Presence: Teleoperators and Virtual Environments, 1* (2), 229–247.

Robinett, W., & Rolland, J. (1992). A computational model for the stereoscopic optics of a head-mounted display. *Presence: Teleoperators and Virtual Environments, 1* (1), 45–62.

Rolland, J. P., & Hopkins, T. (1993). *A method of computational correction for optical distortion in head-mounted displays.* Technical Report TR93-045. Dept. of Computer Science, University of North Carolina at Chapel Hill.

Rolland, J., Holloway, R., & Fuchs, H. (1994). A comparison of optical and video see-through head-mounted displays. *SPIE Proceedings Vol. 2351: Telemanipulator and Telepresence Technologies*, Boston, 293–307.

Rolland, J., Biocca, F., Barlow, T., & Kancherla, A. (1995). Quantification of adaptation to virtual-eye location in see-thru head-mounted displays. *Proceedings of the Virtual Reality Annual International Symposium '95*, Research Triangle Park, NC, 56–66.

Rose, E., Breen, D., Ahlers, K., Crampton, C., Tuceryan, M., Whitaker, R., & Greer, D. (1995). Annotating real-world objects using augmented reality. *Proceedings of Computer Graphics International '95*, Leeds, 357–370.

Rosen, J. M., Soltanian, H., Redett, R. J., & Laub, D. R. (1996). Evolution of virtual reality: From planning to performing surgery. *IEEE Engineering in Medicine and Biology, 15* (2), 16–22.

Sharma, R., & Molineros, J. (1994). Role of computer vision in augmented virtual reality. *SPIE Proceedings Vol. 2351: Telemanipulator and Telepresence Technologies*, Boston, 220–231.

Simon, D. A., Hebert, M., & Kanade, T. (1994). Techniques for fast and accurate intra-surgical registration. *Proceedings of the First International Symposium on Medical Robotics and Computer Assisted Surgery*, 90–97.

Sims, D. (1994). New realities in aircraft design and manufacture. *IEEE Computer Graphics and Applications, 14* (2), 91.

So, R. H. Y., & Griffin, M. J. (1992). Compensating lags in head-coupled displays using head position prediction and image deflection. *Journal of Aircraft, 29* (6), 1064–1068.

Sowizral, H., & Barnes, J. (1993). Tracking position and orientation in a large volume. *Proceedings of IEEE VRAIS '93*, Seattle, 132–139.

State, A., Chen, D. T., Tector, C., Brandt, A., Chen, H., Ohbuchi, R., Bajura, M., & Fuchs, H. (1994). Case study: Observing a volume rendered fetus within a pregnant patient. *Proceedings of IEEE Visualization '94*, Washington DC, 364–368.

State, A., Hirota, G., Chen, D. T., Garrett, B., & Livingston, M. (1996a). Superior augmented reality registration by integrating landmark tracking and magnetic tracking. *Computer Graphics (Proceedings of SIGGRAPH '96)*, New Orleans, 429–438.

State, A., Livingston, M. A., Hirota, G., Garrett, W. F., Whitton, M. C., Fuchs, H., & Pisano, E. D. (1996b). Techniques for augmented-reality systems: Realizing ultrasound-guided needle biopsies. *Computer Graphics (Proceedings of SIGGRAPH '96)*, New Orleans, 439–446.

Taubes, G. (1994). Surgery in cyberspace. *Discover, 15* (12), 84–94.

Tharp, G., Hayati, S., & Phan, L. (1994). Virtual window telepresence system for telerobotic inspection. *SPIE Proceedings Vol. 2351: Telemanipulator and Telepresence Technologies*, Boston, 366–373.

Tuceryan, M., Greer, D. S., Whitaker, R. T., Breen, D., Crampton, C., Rose, E., & Ahlers, K. H. (1995). Calibration requirements and procedures for augmented reality. *IEEE Transactions on Visualization and Computer Graphics, 1* (3), 255–273.

Uenohara, M., & Kanade, T. (1995). Vision-based object registration for real-time image overlay. *Proceedings of Computer Vision, Virtual Reality, and Robotics in Medicine '95*, Nice, 13–22.

Utsumi, A., Milgram, P., Takemura, H., & Kishino, F. (1994). Effects of fuzziness in perception of stereoscopically presented virtual object location. *SPIE Proceedings Vol. 2351: Telemanipulator and Telepresence Technologies*, Boston, 337–344.

Wanstall, B. (1989). HUD on the head for combat pilots. *Interavia, 44* (April 1989), 334–338.

Meta Platforms, Inc.
Exhibit 1112
Page 030

Ward, M., Azuma, R., Bennett, R., Gottschalk, S., & Fuchs, H. (1992). A demonstrated optical tracker with scalable ork area for head-mounted display systems. *Proceedings of 1992 Symposium on Interactive 3D Graphics,* Cambridge, 43–52.

Watson, B., & Hodges, L. (1995). Using texture maps to correct for optical distortion in head-mounted displays. *Proceedings of the Virtual Reality Annual International Symposium '95,* Research Triangle Park, 172–178.

Welch, R. B. (1978). *Perceptual modification: Adapting to altered sensory environments.* New York: Academic Press.

Wellner, P. (1993). Interacting with paper on the DigitalDesk. *Communications of the ACM, 36* (7), 86–96.

Whitaker, R. T., Crampton, C., Breen, D. E., Tuceryan, M., & Rose, E. (1995). Object calibration for augmented reality. *Proceedings of Eurographics '95,* Maastricht, 15–27.

Wloka, M. M. (1995). Lag in multiprocessor virtual reality. *Presence: Teleoperators and Virtual Environments, 4* (1), 50–63.

Wloka, M. M., & Anderson, B. G. (1995). Resolving occlusion in augmented reality. *Proceedings of 1995 Symposium on Interactive 3D Graphics,* Monterey, 5–12.

Wu, J.-R., & Ouhyoung, M. (1995). A 3D tracking experiment on latency and its compensation methods in virtual environments. *Proceedings of the 1995 User Interface Software and Technology,* Pittsburgh, 41–49.

Yoo, T. S., & Olano, T. M. (1993). *Instant Hole™ (Windows onto reality).* Technical Report TR93-027. Department of Computer Science, University of North Carolina at Chapel Hill.

Zikan, K., Curtis, W. D., Sowizral, H., & Janin, A. (1994a). Fusion of absolute and incremental position and orientation sensors. *SPIE Proceedings Vol. 2351: Telemanipulator and Telepresence Technologies,* Boston, 316–327.

Zikan, K., Curtis, W. D., Sowizral, H. A., & Janin, A. L. (1994b). A note on dynamics of human head motions and on predictive filtering of head-set orientations. *SPIE Proceedings Vol. 2351: Telemanipulator and Telepresence Technologies,* Boston, 328–336.

Meta Platforms, Inc.
Exhibit 1112
Page 031