

Technologies for Augmented Reality Systems: Realizing Ultrasound-Guided Needle Biopsies

Andrei State, Mark A. Livingston, William F. Garrett, Gentaro Hirota,
Mary C. Whitton, Etta D. Pisano, MD*
and
Henry Fuchs

Departments of Computer Science and *Radiology
University of North Carolina at Chapel Hill

<http://www.cs.unc.edu/~us/>

ABSTRACT

We present a real-time stereoscopic video-see-through augmented reality (AR) system applied to the medical procedure known as ultrasound-guided needle biopsy of the breast. The AR system was used by a physician during procedures on breast models and during non-invasive examinations of human subjects. The system merges rendered live ultrasound data and geometric elements with stereo images of the patient acquired through head-mounted video cameras and presents these merged images to the physician in a head-mounted display. The physician sees a volume visualization of the ultrasound data directly under the ultrasound probe, properly registered within the patient and with the biopsy needle. Using this system, a physician successfully guided a needle into an artificial tumor within a training phantom of a human breast.

We discuss the construction of the AR system and the issues and decisions which led to the system architecture and the design of the video see-through head-mounted display. We designed methods to properly resolve occlusion of the real and synthetic image elements. We developed techniques for real-time volume visualization of time- and position-varying ultrasound data. We devised a hybrid tracking system which achieves improved registration of synthetic and real imagery and we improved on previous techniques for calibration of a magnetic tracker.

CR Categories and Subject Descriptors: I.3.7 [Three-Dimensional Graphics and Realism]: *Virtual Reality*, I.3.1: [Hardware Architecture]: *Three-dimensional displays*, I.3.6 [Methodology and Techniques]: *Interaction techniques*, J.3 [Life and Medical Sciences]: *Medical information systems*.

Additional Keywords and Phrases: Augmented reality, stereo video see-through head-mounted display, ultrasound echography, 3D medical imaging, BSP tree,

calibration, registration.

1 INTRODUCTION

Since the early days of computer graphics, people have wanted to merge synthetic imagery with their view of the surroundings to create an enhanced view of reality. The range of applications that can potentially benefit from augmented reality (AR) technology includes architecture, mechanical repair, circuit wiring, and health care. Due to a few key technical problems—the quality of merged display systems, occlusion conflicts between real and synthetic objects, real-time image generation, and registration of real and synthetic objects—very few AR systems have been placed in users' hands.

We made advances on the above issues while building an AR system designed to assist a physician with ultrasound-guided needle biopsy of the breast or with cyst aspiration. Figure 1 shows the stereo view displayed in the head-mounted display (HMD) worn by the physician during the AR-guided insertion procedure. With conventional methods the physician has only a non-registered two-dimensional ultrasound image (and perhaps pre-scan 2D medical imagery such as X-ray images) to assist her in the inherently three-dimensional task of guiding a needle to a biopsy target. Our system displays a synthetic opening, or *pit*, into the patient and one or more ultrasound *slices* that are emitted by the tracked hand-held ultrasound probe. We hope that presenting such imagery in the proper context of patient anatomy will make these widely-practiced needle biopsies easier to perform both in the breast and, eventually, also in other, less accessible parts of the body. Figure 2 shows the physician examining a patient in our lab and Figure 3 shows the image she sees in her HMD.

This paper describes the hardware and software of the AR system used by the physician during experiments with training models and human subjects. Section 2 briefly summarizes previous work in AR, and Section 3 discusses some of the issues in building such systems. We present our system in detail in Section 4. Section 5 focuses on issues related to the AR HMD. Section 6 describes how we achieve proper occlusion relationships. Section 7 presents our new real-time volume visualization technique for ultrasound data. Section 8 discusses techniques used to improve registration. We discuss what we have learned from the development and the operation of our system in Section 9. We conclude with a discussion of likely future work in Section 10.

2 PREVIOUS WORK

AR is not a new concept. Ivan Sutherland's original HMD allowed the user to see both the real world and virtual objects [Sutherland68]. The VCASS system [Furness86] used an

CB #3175 Sitterson Hall, Chapel Hill, NC 27599-3175. Tel: +1-919-962-1700.
E-mail: {state, livingst, garrett, hirota, whitton, fuchs}@cs.unc.edu

*Department of Radiology, 503 Old Infirmary 226, CB #7510, Chapel Hill, NC 27599-7510. Tel: +1-919-966-6957. E-mail: pisano@rad.unc.edu

Permission to make digital or hard copies of part or all of this work or personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee.

© 1996 ACM-0-89791-746-4/96/008...\$3.50

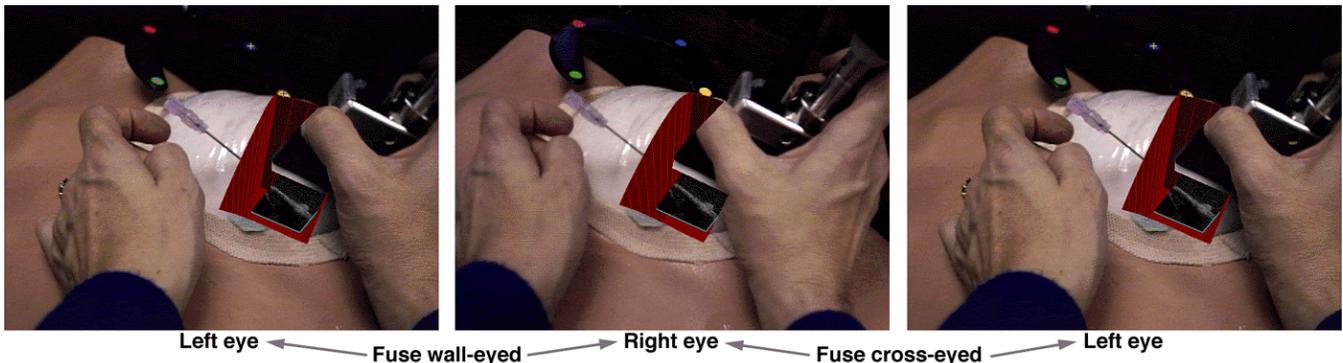


Figure 1. Real-time stereo HMD views from AR system in use during ultrasound-guided needle insertion into phantom training breast. Both the needle and the target lesion are visible in the live ultrasound slice attached to the tracked hand-held probe. Note synthetic opening into the breast and accurate registration between the needle and its image in the ultrasound slice.

optical see-through HMD to superimpose flight and target data onto a pilot's view. The Boeing Company has a group working on an AR system to guide a technician in building a wiring harness for an airplane electrical system [Sims94]. A similar system places text labels of engine parts in the user's view when he points at the real object [Rose94].

Medical applications of AR include visualization and training for surgical interventions. Such systems have been demonstrated by groups at the MIT AI Lab [Mellor95], at Brigham & Women's Hospital [Lorensen93], and elsewhere. These systems use MRI or CT data that is collected before the procedure, then registered to the patient during the procedure. Previous systems built in our lab have demonstrated AR ultrasound visualization [Bajura92, State95, State94]. These earlier systems were limited to non-invasive procedures such as exploratory visualization of a fetus in the womb.

3 AR SYSTEM DESIGN

To a computer graphics veteran, building our AR system may

Figure 2. Physician wearing AR HMD with stereo cameras examines patient in preparation for biopsy. The ultrasound probe is attached to a mechanical tracking arm.



appear straightforward or even simple. It appears that all we must do is add a real-time rendering of ultrasound data to a live video background. However, realizing a working system requires addressing the technical problems of stereo AR, real-time volume rendering of time- and position-varying ultrasound data, and precise registration of real and synthetic image elements. In this section, we describe and justify our choice of technologies for various system components.

3.1 Merging real and synthetic imagery

By definition, an AR system must allow the user to see the real world. There are two common technologies for AR HMDs: *optical see-through* and *video see-through*. In the former, beam splitters (e. g., half-silvered mirrors) optically combine light from the environment with computer-generated display elements. In the latter, video cameras mounted on the HMD acquire images of the real world which are then electronically combined with computer-generated imagery. This can be accomplished via analog technology (e.g. luminance keying or chroma keying) or via real-time digital video capture and digital compositing, as in our current system. The combined video images are then displayed in the (conventional, "opaque") HMD. More details about these technologies can be found in [Azuma95].

For our system, we chose video see-through (a decision which dates back to [Bajura92]). The four key advantages of video see-through over optical see-through which led to this decision are:

- (1) support for proper occlusion relationships between real and virtual objects (provided that depth of real objects is known or can be determined) [Wloka95].
- (2) ability to balance the brightness of synthetic and real

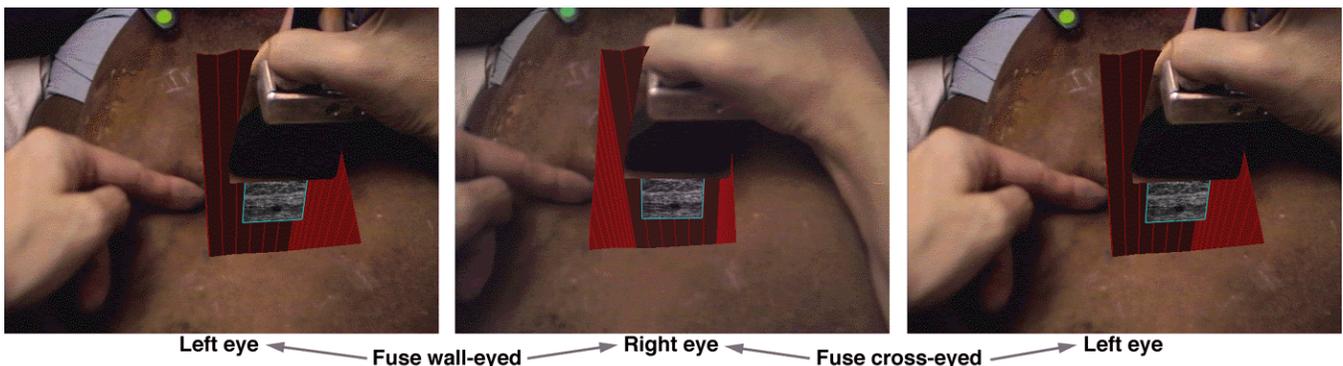


Figure 3. Real-time stereo HMD view during patient examination (Figure 2) showing a single ultrasound slice properly positioned within the patient's breast. The physician's left index finger points towards the cyst (visible as a dark spot in the ultrasound slice).

imagery on a pixel-by-pixel basis.

- (3) ability to match the latencies (delays) of synthetic and real imagery [Bajura95].
- (4) ability to use vision-based registration strategies, such as video tracking of landmarks—provided that the video image is digitized and available for image processing.

The above items are critical to the design of our system. We accept, for now, the drawbacks of video see-through technology: low resolution for the real-world imagery and the spatial offset between the user's eyes and the locations of the HMD-mounted cameras used to acquire the real world images. These problems will be solved in future HMDs [Colucci95].

In video see-through, the cameras introduce several issues not present in optical see-through. The field of view and angle of convergence (and thus the amount of stereo overlap) of the cameras should match those of the HMD in order for the synthetic imagery to have the same alignment and perspective as the real world imagery. This requires measuring the intrinsic parameters of the cameras.

3.2 Tracking

Accurate tracking is crucial for precise registration of real and synthetic imagery, especially in a medical application where surgical intervention is to be performed under AR guidance. In our system we must track the physician's head and the hand-held ultrasound probe.

In selecting a head tracking system, we wanted to avoid encumbering the physician, so we chose a magnetic tracker. Unfortunately, the metallic structures in our lab interfere with its accuracy. To overcome this, we combine magnetic tracking with vision-based landmark tracking for improved registration.

Tracking of the ultrasound probe must be extremely precise for correct registration of ultrasound slices (both slice-to-slice and slice-to-patient). The probe is usually positioned over a confined area of the patient's body and gathers data only when in contact with the patient, so it has a small working and tracking volume. Furthermore, the probe is already tethered to the ultrasound machine. We therefore elected to accurately track the probe with a 6-degree-of-freedom mechanical tracker even though it hinders probe motion to a certain extent.

3.3 Image generation platform

The success of [Cullip93, Cabral94] and the availability of fast hardware for rendering texture-mapped polygons led us to choose hardware texture mapping for ultrasound data visualization. Our system requires an image generation platform that supports real-time video acquisition for three channels: one ultrasound video stream and two camera video streams. The latter are inspected by vision-based registration algorithms. On the output side, the system must be able to generate two video output streams for the HMD as well as a user interface screen.

4 SYSTEM DESCRIPTION

4.1 System Configuration

The principal hardware platform of our system is a Silicon Graphics Onyx™ Reality Engine²™ graphics workstation (Onyx) equipped with a Sirius Video™ real-time video capture device (Sirius) and a Multi-Channel Option™ (MCO) that outputs multiple video streams that can be used for the left and right eye displays. The Sirius simultaneously acquires video signals from a PIE Medical Scanner 200 ultrasound machine (PIE) and from two Panasonic GP-KS102 head-mounted CCD video cameras equipped with Cosmimar F1.8 12.5 mm lenses (28° field of view, selected for minimal optical distortion). Through its serial ports, the Onyx acquires tracking data from

two trackers: a FARO Metrecom IND-1 mechanical arm (FARO), which tracks the ultrasound probe, and an Ascension Flock of Birds™ magnetic tracker for the user's head. The Onyx generates stereo video signals to be displayed within the Virtual Research VR-4 HMD (VR-4). PIE, VR-4 with cameras and Flock sensor, and FARO are all visible in Figure 2.

4.1.1 Video Input

The Sirius allows simultaneous acquisition and digital processing of two video streams. This constraint required us to combine the video streams from the two head-mounted cameras into a single stream for the purpose of acquisition (and to devise a software method to split them again after capture for the purpose of stereo output). A commercial analog multiplexer (QD Technology QD-1110) combines the camera video signals into a single analog signal by selecting odd video fields from one camera and even fields from the other.

The Sirius has two digital video inputs but only one analog video input. Both the ultrasound machine and the multiplexer produce analog video streams. Hence, one of these two streams had to be converted into digital format. Since we were willing, at the time, to compromise the ultrasound imaging subsystem but not the overall "AR feel," and since the PIE's video had to be time-base corrected, we opted to convert the ultrasound video to digital format. The time-base correction and the conversion to digital format both introduce lag into this video stream.

The constraints imposed by our system and by the hardware platform are summarized in the following list:

- (1) The Sirius can capture and process only two video streams simultaneously.
- (2) Video streams captured by the Sirius can be routed to main memory, texture memory, or the frame buffer. The frame buffer and texture memory, however, cannot be used as destinations at the same time. Hence main memory must be one of the destinations for our two streams, while the other can be either frame buffer or texture memory.
- (3) Separation of a field-multiplexed video stream is best done if the stream is captured into the frame buffer (fast). It can also be done if captured into main memory (albeit slow), but it is virtually impossible if captured into texture memory.
- (4) The camera video contains landmarks which must be detected; therefore it cannot be captured into texture memory. If it is captured into main memory (convenient for inspection), it will have to be copied to the frame buffer for display. If it goes into the frame buffer, it will have to be copied into main memory for landmark search (moderately slow, since in practice only the areas of the image containing landmarks need to be transferred and inspected).
- (5) The Sirius captures video into texture memory only in a 512×1024×24-bit format, severely limiting the number of frames that can be held at any one time. This is unacceptable since we want to be able to simultaneously display multiple textured ultrasound slices. However, the ultrasound video images are monochrome and could easily fit into 256×256×8 bits of texture memory each (with downsampling). Therefore ultrasound video cannot be captured directly into texture memory, although it must eventually be loaded into texture memory so that we can take advantage of the hardware texturing capability.
- (6) We had only one Sirius unit and one graphics pipeline. From this (over-determined) set of constraints it follows that the (digital) ultrasound video signal must be captured into main memory, where it is resampled (by the CPU) into the 256×256×8-bit format, and then transferred into texture

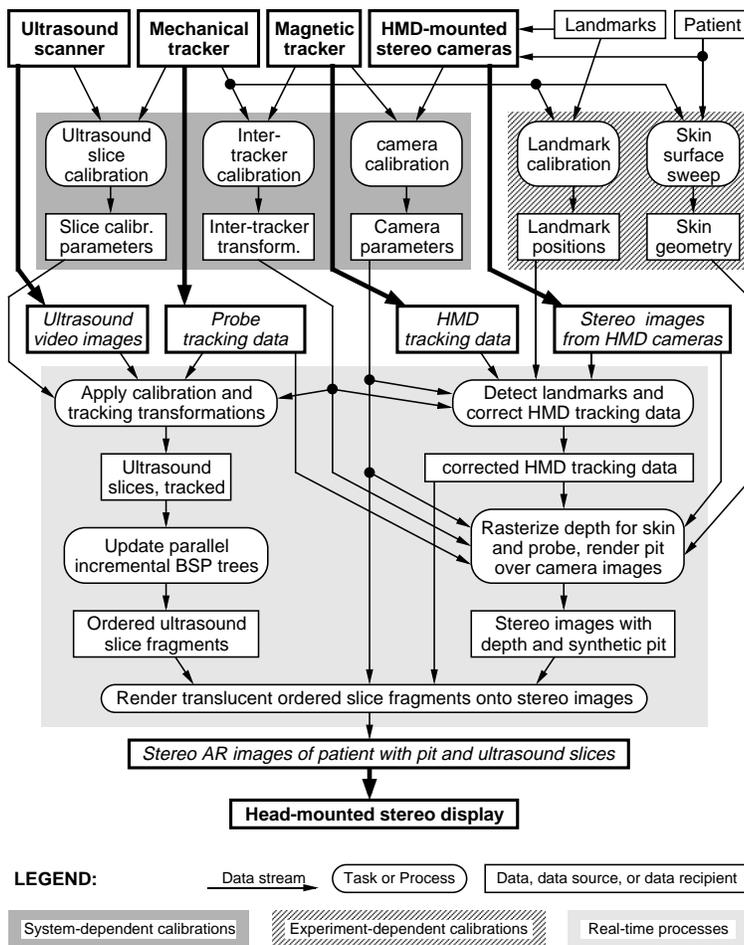


Figure 4. Data flow within stereo AR system for ultrasound visualization

memory. The combined (analog) camera video signal is captured by the Sirius into the frame buffer.

4.1.2 Tracking Input

The Flock magnetic tracker and the FARO mechanical tracker are attached to the Onyx via two dedicated 38,400 baud serial lines. The FARO operates at a maximum rate of 27 Hz, the Flock at 103 Hz (both, however, are read asynchronously within the main loop software on the Onyx). The ultrasound probe is mounted on the FARO arm with a custom-built mount. The Flock receiver is mounted on a plastic arm which is rigidly attached to the stereo camera rig on the VR-4 HMD.

4.1.3 Video Output

We selected an MCO configuration that simultaneously transmits a 1280×1024 high-resolution image and two 640×480 VGA images. The high-resolution image is used for the user interface. The two VGA images contain the viewports for camera image capture and hence for the AR imagery. The VGA signals are fed via commercial VGA-to-S-Video scan converters (Extron Super Emotia) into the left and right eye displays of the VR-4. We also carry these two signals to standard VGA monitors in the lab so that people other than the HMD wearer can observe. The monitors for all three output signals are visible in Figure 2.

4.2 System Operation

Figure 4 shows the data flow within the system. Across the top are the input sources. The system captures input data from

four different sources that “sample” the real world—two video streams (camera, ultrasound) and two tracking data streams (head, probe). These four streams are processed into the stereoscopic AR HMD display that is the system’s output (bottom of the diagram). The upper third of the diagram shows the calibration procedures that must be performed before system operation; the lower two thirds depicts the data flow required to produce a stereo image pair.

4.2.1 Calibration

The first set of calibration procedures are system-dependent calibrations of the input video streams and tracking streams. The ultrasound machine is calibrated to determine a transformation between pixels in the ultrasound video stream and the local coordinate system of the tracked ultrasound probe (a plane equation and scale factors for the ultrasound slice). The procedure also calibrates the area of the ultrasound video image that contains scanned data (a polygonal outline for the ultrasound slice) [State94]. During system operation, this area is resampled into the 256×256×8 format to be loaded into texture memory.

A calibration procedure similar to [Bajura95] is performed on the camera-sensor rig to determine the transformation between the head tracking sensor and the cameras’ local coordinate systems. The cameras’ intrinsic parameters (location of the center of projection, field of view) are also determined, albeit with limited accuracy [State96]. Finally, a transformation between the coordinate systems of the magnetic and mechanical trackers is determined by calibrating each to a reference coordinate system (a lab-mounted wooden frame).

The second set of calibrations are experiment-dependent calibrations. First, we sweep the patient’s skin with the mechanical tracker and acquire 3D points on the surface. Section 6 describes how we use this surface to generate proper occlusion cues. Second, we record the precise location of the landmarks used by the hybrid head tracking algorithm (described in Section 8).

4.2.2 Real-time processing

The bottom two thirds of Figure 4 shows the real-time processes implemented on the Onyx. For each (stereo) frame to be generated, the software captures one frame from the ultrasound video stream, one frame from the multiplexed camera image stream, and readings from each of the two trackers.

The ultrasound video defines a texture for the ultrasound slice polygon to be displayed. The slice’s position and orientation is determined by the probe tracking data together with the off-line calibration parameters. The slice polygon is processed by a dual BSP tree algorithm described in Section 7.

The multiplexed stereo camera video frame is captured into the frame buffer and split into the left and right eye images (odd fields into the left eye image, even fields into the right). Each scan line is duplicated in order to preserve aspect ratio. The left and right eye areas of the frame buffer correspond to the output regions of the MCO’s two VGA channels.

The Flock report is used to estimate the locations of the landmarks visible in the camera images. The landmarks’ actual positions in the video images are then determined by image analysis, and corrected position and orientation data is computed for each camera. Based on this data, the pit and other geometric elements are rendered on top of the video image background, as described in Section 6. Finally, the ultrasound

slices are rendered using the dual BSP tree. The rendering stages are executed twice, once for each eye.

5 STEREO VIDEO-SEE-THROUGH HMD

Without stereo depth cues, the physician user of our system cannot assess the distance to the patient or the depth of a lesion within the breast. The construction of a stereo video-see-through HMD (visible in Figure 2) was key to physician acceptance of our system and to the start of patient trials.

5.1 Head-mounted cameras

We mounted cameras on the front of the HMD on top of the housing for the LCD displays. In this arrangement, the camera's centers of projection are located approximately 5 cm above and 8 cm in front of the wearer's eyes, who must learn to compensate for this constant eye offset. Even after accommodation training, we expect the user's performance to be impaired [Rolland95].

For the mount, we used an interpupillary distance of 64 mm and chose a fixed convergence angle of 4°. The horizontal field of view of the lenses is 28°, producing a stereo overlap of roughly 80% at a working distance of 50 cm. The 4° convergence is a compromise; we could achieve 100% overlap with a larger convergence angle (about 7.4°), but then viewing the images inside the HMD would cause eye strain since the display convergence angle does not match the camera convergence angle. (The convergence angle of 4° also makes the stereograms in this paper slightly difficult to fuse.)

The limited light sensitivity of our cameras causes problems for the image analysis technique mentioned in Section 8.2. In order to get enough light for good image quality and landmark tracking, the iris of the cameras must be opened to the point that the depth of field is less than the depth extent of the working volume in our application. We therefore manually adjust the focus on the cameras as necessary.

5.2 Head-mounted display

The VR-4 weighs over 2 pounds before the 1.5 pounds of camera and fixtures are added. Most of this weight is concentrated around the user's eyes, making the device very front-heavy. A counterweight provides balance, but nearly doubles the weight of the HMD.

The horizontal field of view in the VR-4 is approximately 40° for each eye, compared to 28° in each camera. This mismatch leads to a "telephoto" viewing experience similar to that of using binoculars. The VR-4 displays can be set to a convergence angle of either zero or three degrees.

The stereo images acquired by the Sirius have a resolution of 646×243 for each eye (due to left-right field multiplexing). The complete AR views generated by the system have a final resolution of 640×480 for each eye. However, the resolution of the VR-4 is only roughly 250×230. Within those pixels we map the 256×256 ultrasound data slice (downsampled from the original 512×512) to a small fraction of the screen—about 40×40 for the image in Figure 1. In Figure 1, a 3 mm breast lesion would image on approximately 3×3 VR4 HMD pixels, and a 22-gauge needle would appear 0.7 pixels thick. The HMD resolution is hence adequate for training phantom experiments—where thicker needles can be used—but insufficient for human subject trials.

6 OCCLUSION

To present correct occlusion cues in our visualization, we must enhance the pure *RGB* color information acquired by the HMD-mounted cameras with proper depth (or *z*) values. We use a set of geometric elements that are rendered in depth (*z*) only (and not in *RGB*). The top edges of the pit must be spatially aligned

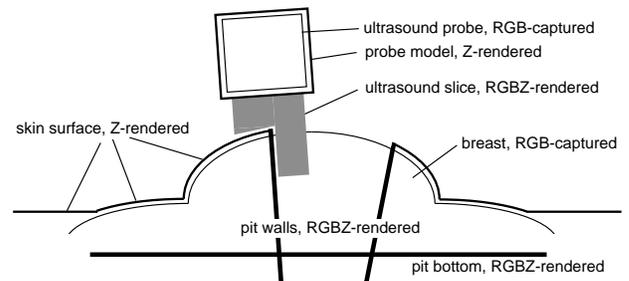


Figure 5. Geometric elements used for the synthesis of occlusion cues

with the patient's skin, otherwise the pit and the rendering of the ultrasound data will appear pasted on or swimming across the patient's skin, rather than appearing (and staying) properly positioned with respect to the patient (Figure 5).

For both correct occlusion and correct registration of the pit, we must know the location and shape of the patient's skin in 3D lab space. To acquire this information, we sweep the tip of the FARO over the patient's skin and collect (unordered) 3D points from the surface. (This assumes that the patient will not move during or after the procedure.) The collection procedure is one of the experiment-specific calibrations in Section 4.2.1.

To resolve occlusion, the unordered set of points must be converted into a polygonal surface. Techniques such as Delaunay triangulation can be applied only to 2D arrays of points; we therefore exploit the shape of the human body and convert the points into cylindrical coordinates using a horizontal cylinder axis that is roughly aligned with the patient's spinal column. Delaunay triangulation is then applied in the cylinder's height-angle domain, under the assumption that the surface to be constructed is a radius field—i. e., that it can be expressed as a function $radius = radius(height, angle)$, similar to Cyberware™ scans. We then resample the mesh output from the Delaunay triangulation into a regular grid in the height and angle dimensions.

The regular grid of the resampled triangles, together with the cylindrical coordinate system in which it is defined, is then used to create a polygonal model of the patient's skin surface, as well as a polygonal model of the pit, which is embedded within the skin surface model. *Z*-rendering of the surface model minus the pit opening results in a correct *z*-buffer for the patient surface. *RGBZ*-rendering of the pit model results in a colored and *z*-buffered pit; As a result of this process, synthetic elements such as the ultrasound slice attached to the transducer can now penetrate into or disappear below the skin of the patient, except within the pit, where they remain visible.

In addition, we *z*-render a polygonal model for the (FARO-tracked) ultrasound probe, which enables the probe to occlude synthetic image elements if it passes in front of them. For example, if the probe is positioned between the HMD cameras and the pit, the probe obscures the pit (Figures 1, 3). We do not track other real world elements such as the physician's hands, which hold the probe and may also pass between the HMD cameras and the patient's breast, so we cannot eliminate all depth conflicts from the user's field of view. Nevertheless, the implementation of depth images for certain components of the real world has significantly enhanced our visualization.

7 REAL-TIME INCREMENTAL VOLUME VISUALIZATION

The AR system must produce stereo visualizations in near real-time (at least 10 stereo frames per second) from a dynamic volumetric target (for example, a cyst within the breast plus a moving needle). While the dataset is a 3D volume, the data is acquired as a sequence of 2D slices. We have implemented a

Frame	Scene 1	Tree 1	Scene 2	Tree 2
1				
2				
3				
4				
5				
6				
7				
8				

Ultrasound slices: new active expired

BSP tree fragments:

Figure 6. Dual parallel BSP tree for $n = 3$. The shaded areas represent the tree selected for rendering. Both trees are updated.

dynamic volumetric display that maintains a set of such slices in the system. During the generation of each output frame, as a new frame is acquired from the ultrasound video stream, one new slice is added to the set and the oldest slice is removed from it. Thus the system always displays the n most recent slices. Volume reconstruction (even incremental reconstruction [Ohbuchi92]) of a set of slices into a regular grid is computationally too expensive. Instead, we use the Onyx' texturing hardware to visualize the slices as polygons with translucent textures. The shape of the polygon is determined as part of the calibrations described in Section 4.2.1. The texture for the polygon comes from the ultrasound video frame.

Rendering of the volume is accomplished by rendering the collection of (possibly intersecting) translucent textured polygons. Due to its hardware texturing capability and large texture memory, the Onyx is well-suited for this kind of volume visualization. The translucent polygons must be presented to the graphics pipeline in back-to-front order. We use a binary space partition (BSP) tree to establish the order [Fuchs80].

The set of n polygons contained in the BSP tree constantly changes. We delete a polygonal slice from the BSP tree as each new slice is added. Unfortunately deletion of a polygon from a BSP tree is more expensive than insertion,

particularly if polygon fragmentation and the associated decrease in efficiency are to be avoided. Leaving expired polygons in the tree while tagging them as expired (so that they are not rendered) is also problematic when new slices are being added at each frame: the size of the tree can grow as the square of the number of insertions [Fuchs80].

We solve the BSP tree update problem by constructing and maintaining two BSP trees, out of phase in time, in the following manner: Let n be the number of ultrasound slices the user wants displayed (Figure 6). As the first n slices arrive during frames 1 through n , they are added to only the first tree. When slice $n+1$ arrives, it is inserted into the first BSP tree, slice 1 is marked "expired" (but is not removed) from the tree, and the second BSP tree is started with this single slice, $n+1$. With slices $n+2$ through $2n$, the new slice is inserted into both trees, the appropriate old slice in the first tree is marked "expired," and the first tree continues to be rendered. After processing slice $2n$, the second tree contains (exactly) the most recent n slices and no "expired" slices and unnecessary fragments to slow down the traversal and rendering. The rendering is now switched to the second tree, the first tree is deleted (to be initialized next with slice $2n+1$). The procedure continues in this way, always rendering from the older tree until the newer one contains n slice images.

Figure 7 is an HMD view with a volume visualization of a lesion within a breast training phantom penetrated by a needle.

8 IMPROVING REGISTRATION

The breast biopsy task requires very high precision. The physician may be required to place a thin needle – for example, 22 gauge (0.7 mm diameter) for cyst aspiration, 14 gauge (2.1 mm diameter) for biopsy – into a 3 mm cyst. Of the trackers we have tried, none have the accuracy and precision required for this medical application. We therefore combine a mechanical tracker (FARO), a magnetic tracker (Flock) corrected by a lookup table, and vision-based tracking to achieve improved registration of real imagery (patient, ultrasound probe, biopsy needle) and synthetic imagery (ultrasound slices, rendered visual and occlusion cues).

8.1 Correction table for magnetic tracker

The Flock is the primary head tracker. Since our lab has metal in the floor, ceiling, and light fixtures, and since we use metal objects and electric fields within the work environment (FARO, PIE), there is a significant amount of static distortion of the magnetic field. Distortion is one of the possible sources for the Flock's tracking errors of up to 10 cm in position and up to 10° in orientation within our tracking area. Removing the sources of distortion is often not desirable (in the case of

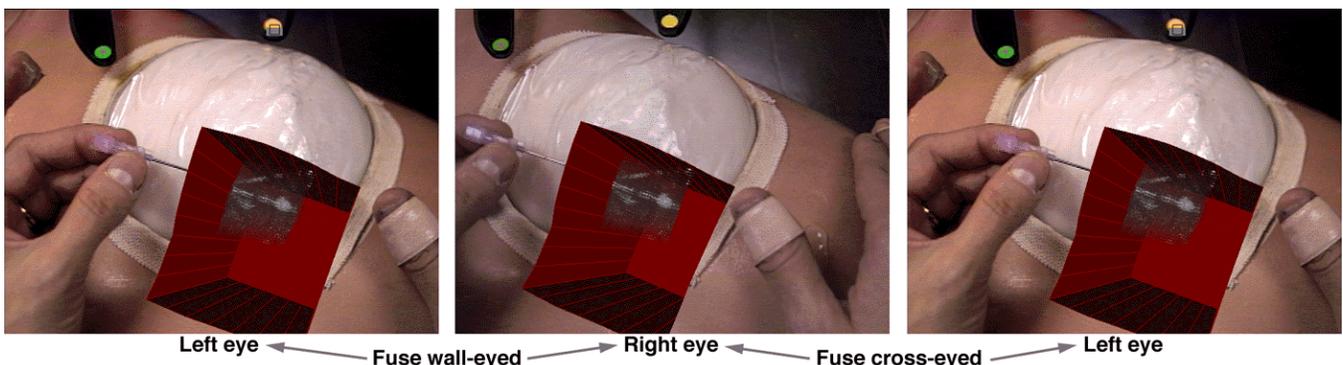


Figure 7. Real-time stereo HMD view with ultrasound volume display. A needle has been inserted into the breast phantom; the inside of the phantom has been imaged with the probe, resulting in a volume representation of the inside of the breast. The needle is registered with its image inside the ultrasound volume. The phantom also contains needle traces from previous insertion attempts.

FARO and PIE) or not possible (floor, ceiling, lights). Expanding upon the work of others [Bryson92, Ghazisaedy95], we therefore statically calibrated the Flock.

For the calibration procedure, we affix the Flock sensor to the FARO (buffered by plastic) and collect sample points at thousands of arbitrary locations in the work environment. We then determine the error in the Flock reports by comparing Flock and FARO readings. We then resample these error values into a rectilinear look-up table.

The calibrated Flock performs quite well for position, with errors by 80 percent, down to an average post-correction error of 0.5 cm. The calibration does not enjoy such success for orientation correction, however. It reduces orientation error by only 40 percent, down to 1.4 degrees on average. Further details can be found in [Livingston95].

8.2 Vision-based tracking

Even after table-based correction, the Flock is not sufficiently accurate for the application's registration requirements. We therefore use a hybrid head tracking algorithm and image landmarks to obtain higher accuracy. Our landmarks are fluorescent discs positioned in view of the HMD-mounted cameras, typically close to the sterile field (visible in Figure 1.). The positions of the landmarks in world space are known; they are calibrated with the FARO as part of the experiment-specific calibrations outlined in Section 4.2.1.

Landmark tracking is performed by the Onyx CPU, using the stereo images captured into the frame buffer by the Sirius. The software attempts to predict expected landmark positions. This minimizes the size of the pixel arrays that must be transferred from the frame buffer into main memory and searched for landmarks. If a single landmark is detected, two out of three degrees of freedom of the camera's orientation can be corrected under simplifying assumptions [Bajura95]. With two landmarks, orientation can be corrected fully, again under the same assumptions. With three landmarks, camera position and orientation can be determined completely to within a sign. With knowledge of the transformation between the camera coordinate systems of the stereo cameras—a system-dependent calibration not mentioned in Section 4.2.1—the system can even correct both cameras with landmarks detected in only one of the cameras. Furthermore, the system stores the correction and applies it to the raw Flock reading even if no landmarks are visible. However, the quality of the correction degrades rapidly in such cases and the HMD wearer is required to keep the landmarks in view. An improved version of our hybrid tracking system is described in [State96].

9 SUMMARY AND CONCLUSIONS

Using the AR system described here, a physician successfully guided a biopsy needle into an artificial tumor within a life-sized breast model. The system is sufficiently robust and accurate for the physician to report that the procedure on the breast model was easy. The key efforts that led to this milestone were the construction of a stereo input and stereo output video see-through head-mounted display, methods for properly resolving occlusion between real and synthetic objects, a new real-time volume visualization method using parallel BSP trees, a closed-loop vision-based head tracking algorithm, and judiciously applied calibration techniques for all input data streams (cameras, trackers, ultrasound probe).

An AR visualization viewed by an "over-the-shoulder" observer during two patient case studies demonstrated some of the remaining problems with the AR system. The resolution of the HMD is insufficient for good visualization of the ultrasound image. The HMD with stereo cameras is too heavy. The image

landmarks are difficult to keep in view and unoccluded for the physician. Finally, the data captured in real-time from the four input streams is not synchronized. While we have significantly improved spatial registration, temporal registration (synchronization) remains a problem.

10 FUTURE WORK

We envision the introduction of a system such as ours into the operating room, but advances in several areas are required before this goal can be realized. First, the HMD should be considerably lighter and feature higher-quality image acquisition (cameras) as well as higher-resolution displays. In future video-see-through HMDs the optical paths of camera and user's eye should be aligned, in order to eliminate the eye offset problem. This can be accomplished by folding the camera's optical path with mirrors. Second, head tracking should be more accurate and less dependent on (or even completely independent of) landmarks. We are investigating a better orientation calibration for the magnetic tracker and are considering alternatives such as optical trackers, to be used alone or as part of a hybrid tracking technique.

The input streams must be synchronized (temporally registered). Data from these is currently captured asynchronously, at discrete intervals, from the four input devices. Ideally, all four signals should sample the real world at the same moment in time, thus ensuring that the stereo AR display shows a consistent enhanced view of the real world, albeit delayed with respect to the real world by the time it took to synthesize the view. In practice, each of the streams has a certain amount of lag associated with it. Precise knowledge about the lag in each stream holds the potential for eliminating or compensating for lag differences between streams. We have begun to devise experiments and software organization strategies for this purpose.

An operative system should address a number of additional problems. The noisy quality of ultrasound images of human tissue makes targets such as cysts or tumors difficult to recognize and even more challenging to visualize volumetrically. (The image in Figure 7 was acquired in a training phantom.) It is therefore necessary to explore techniques for improved real-time identification (segmentation) and visualization of cysts and lesions in human breast tissue. Finally, methods to track the skin surface and its deformations in real time are required.

ACKNOWLEDGMENTS

It takes many people to realize a complex system like the one described here. We wish to express our gratitude to John Airey, Ronald T. Azuma, Michael Bajura, Andrew Brandt, Gary Bishop, David T. Chen, D'ardo Colucci, Darlene Freedman, Arthur Gregory, Stefan Gottschalk, David Harrison, Linda A. Houseman, Marco Jacobs, Fred Jordan, Kurtis Keller, Amy Kreiling, Shankar Krishnan, Dinesh Manocha, Mark McCarthy, Michael North, Ryutarou Ohbuchi, Stephen M. Pizer, Scott Pritchett, Russell M. Taylor II, Chris Tector, Kathy Tesh, John Thomas, Greg Turk, Peggy Wetzel, Steve Work, the anonymous patients, the Geometry Center at the University of Minnesota, PIE Medical Equipment B.V., Silicon Graphics, Inc., and the UNC Medical Image Program Project (NCI P01 CA47982).

We thank the anonymous reviewers for their comments and criticism.

This work was supported in part by ARPA DABT63-93-C-0048 ("Enabling Technologies and Application Demonstrations for Synthetic Environments"). Approved by ARPA for Public Release—Distribution Unlimited. Additional

support was provided by the National Science Foundation Science and Technology Center for Computer Graphics and Scientific Visualization (NSF prime contract 8920219).

REFERENCES

- AZUMA, R. A Survey of Augmented Reality. SIGGRAPH 1995 Course Notes #9 (Developing Advanced Virtual Reality Applications), pp. 20-1 through 20-38.
- AZUMA, R., BISHOP, G. Improving Static and Dynamic Registration in an Optical See-through HMD. Proceedings of SIGGRAPH 94 (Orlando, Florida, July 24-29, 1994). In *Computer Graphics Proceedings, Annual Conference Series, 1994*, ACM SIGGRAPH, pp. 197-204.
- BAJURA, M., FUCHS, H., OHBUCHI, R. Merging Virtual Objects with the Real World. Proceedings of SIGGRAPH '92 (Chicago, Illinois, July 26-31, 1992). In *Computer Graphics* 26, 2 (July 1992), ACM SIGGRAPH, New York, 1992, pp. 203-210.
- BAJURA, M., NEUMANN, U. Dynamic Registration Correction in Video-Based Augmented Reality Systems. *IEEE Computer Graphics and Applications* (September, 1995), pp. 52-60.
- BARBER, C.B., DOBKIN, D.P., HUHDANPAA, H. The Quickhull Algorithm for Convex Hull. *Geometry Center Technical Report GCG53*, 1993.
- BRYSON, S. Measurement and Calibration of Static Distortion of Position Data from 3D Trackers. *SPIE Stereoscopic Displays and Applications III*, 1992, pp. 244-255.
- CABRAL, B., CAM, N., FORAN, J. Accelerated Volume Rendering and Tomographic Reconstruction Using Texture Mapping Hardware. *Proceedings of the 1994 Symposium on Volume Visualization* (Washington, D. C., October 1994), pp. 91-97, 131, back cover.
- COLUCCI, D., CHI, V. Computer Glasses: A Compact, Lightweight, and Cost-effective Display for Monocular and Tiled Wide Field-of-View Systems. *Proceedings of SPIE*, Vol. 2537, pp. 61-70 (1995).
- CULLIP, T.J., NEUMANN, U. Accelerating Volume Reconstruction With 3D Texture Hardware. *University of North Carolina Department of Computer Science Technical Report TR93-027*.
- FUCHS, H., ABRAM, G., GRANT, E. Near Real-time Shaded Display of Rigid Objects. Proceedings of SIGGRAPH '83 (July 1983). In *Computer Graphics*; 17, 3 (July 1983), ACM SIGGRAPH, New York, 1983, pp. 65-72.
- FUCHS, H., KEDEM, Z., NAYLOR, B. On Visible Surface Generation by a Priori Tree Structures. Proceedings of SIGGRAPH '80 (July 1980). In *Computer Graphics*; 14, 3 (July 1980), ACM SIGGRAPH, New York, 1980, pp. 124-133.
- FURNESS, T. The super cockpit and its human factors challenges. *Proceedings of the Human Factors Society*, 30, 1968, pp. 48-52.
- GARRETT, W.F., FUCHS, H., STATE, A., WHITTON, M.C. Real-Time Incremental Visualization of Dynamic Ultrasound Volumes Using Parallel BSP Trees. *University of North Carolina Department of Computer Science Technical Report TR96-018*.
- GHAZISAEDY, M., ADAMCZYK, D., SANDIN, D., KENYON, R., DEFANTI, T. Ultrasonic Calibration of a Magnetic Tracker in a Virtual Reality Space. *Proceedings of the Virtual Reality Annual International Symposium '95*, IEEE Computer Society, Los Alamitos, CA, 1995, pp. 179-188.
- LALOUCHE, R.C., BICKMORE, D., TESSLER, F., MANKOVICH, H.K., KANGARALOO, H. Three-dimensional reconstruction of ultrasound images. *SPIE '89, Medical Imaging*, SPIE, 1989, pp. 59-66.
- LIVINGSTON, M.A., STATE, A. Magnetic Tracker Calibration for Improved Registration in Augmented Reality Systems. *University of North Carolina Department of Computer Science Technical Report TR95-037*.
- LORENSEN, W., CLINE, H., NAFIS, C., KIKINIS, R., ALTOBELLI, D., GLEASON, L. Enhancing reality in the operating room. *Proceedings of Visualization 1993* (Los Alamitos, CA, October 1993), pp. 410-415.
- MELLOR, J.P. Enhanced Reality Visualization in a Surgical Environment. MS Thesis, MIT Department of Electrical Engineering, 1995.
- NAYLOR, B., AMANATIDES, J., THIBAUT, W. Merging BSP Tree Yields Polyhedral Set Operations. Proceedings of SIGGRAPH '90 (Dallas, Texas, August 1990). In *Computer Graphics*; 24, 4 (August 1990), pp. 115-124.
- NELSON, T., ELVINS, T. Visualization of 3D Ultrasound Data. *IEEE Computer Graphics and Applications*, (November, 1993), pp. 50-57.
- OHBUCHI, R., CHEN, D.T., FUCHS, H. Incremental Volume Reconstruction and Rendering for 3D Ultrasound Imaging. *University of North Carolina Department of Computer Science Technical Report TR92-037*.
- ROBINETT, W., ROLLAND, J.P. A Computational Model for the Stereoscopic Optics of a Head-Mounted Display. *Presence*, Vol. 1, No. 1 (Winter 1992), pp. 45-62.
- ROLLAND, J.P., BIOCCA, F., BARLOW, T., KANCHERLA, A.R. Quantification of Adaptation to Virtual-Eye Location in See-Thru Head-Mounted Displays. *Proceedings of the Virtual Reality: Annual International Symposium '95*, IEEE Computer Society, Los Alamitos, CA, 1995, pp. 56-66.
- ROSE, E., BREEN, D., AHLERS, K., CRAMPTON, C., TUCERYAN, M., WHITAKER, R., GREER, D. Annotating Real-World Objects Using Augmented Reality. *Proceedings of Computer Graphics International 1995* (Leeds, UK, June, 25-30, 1995), pp. 357-370.
- SIMS, D. New Realities in Aircraft Design and Manufacture. *IEEE Computer Graphics and Applications*, (March, 1994), p. 91.
- STATE, A., CHEN, D.T., TECTOR, C., BRANDT, A., CHEN, H., OHBUCHI, R., BAJURA, M., FUCHS, H. Case Study: Observing a Volume-Rendered Fetus within a Pregnant Patient. *Proceedings of IEEE Visualization '94*, October 1994, pp. 364-368, CP-41.
- STATE, A., HIROTA, G., CHEN, D.T., GARRETT, W.F., LIVINGSTON, M.A. Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking. Proceedings of SIGGRAPH '96 (New Orleans, LA, August 4-9, 1996). In *Computer Graphics Proceedings, Annual Conference Series, 1996*, ACM SIGGRAPH.
- STATE, A., MCALLISTER, J., NEUMANN, U., CHEN, H., CULLIP, T.J., CHEN, D.T., FUCHS, H. Interactive Volume Visualization on a Heterogeneous Message-Passing Multicomputer. *Proceedings of the 1995 Symposium on Interactive 3D Graphics* (Monterey, CA, April 1995), pp. 69-74, 208, front cover.
- SUTHERLAND, I. A head-mounted three dimensional display. *Proceedings of the Fall Joint Computer Conference*, 1968.
- WLOKA, M.M., ANDERSON, B.G. Resolving Occlusion in Augmented Reality. *Proceedings of the 1995 Symposium on Interactive 3D Graphics* (Monterey, CA, April 1995), pp. 5-12.