Dynamic Virtual Convergence for Video See-through Head-mounted Displays: Maintaining Maximum Stereo Overlap throughout a Close-range Work Space

Andrei State, Jeremy Ackerman, Gentaro Hirota, Joohi Lee and Henry Fuchs
University of North Carolina at Chapel Hill
{andrei|ackerman|hirota|lee|fuchs@cs.unc.edu}

Abstract

We present a technique that allows users of video see-through head-mounted displays to work at close range without the typical loss of stereo perception due to reduced nasal side stereo overlap in most of today's commercial HMDs.

Our technique dynamically selects parts of the imaging frustums acquired by wide-angle head-mounted cameras and re-projects them for the narrower-field-of-view displays. In addition to dynamically maintaining maximum stereo overlap for objects at a heuristically estimated working distance, it also reduces the accommodation-vergence conflict, at the expense of a newly introduced disparity-vergence conflict. We describe the hardware (assembled from commercial components) and software implementation of our system and report on our experience while using this technique within two different AR applications.

1. Introduction and motivation

A video see-through head mounted display (VST-HMD) gives a user a view of the real world through one or more video cameras mounted on the display. Synthetic imagery is combined with the images captured through the cameras. The combined images are sent to the HMD. This yields a somewhat degraded view of the real world due to artifacts introduced by cameras, processing, and redisplay, but also provides significant advantages for implementers and users alike [Azuma 1997].

Most commercially available head-mounted displays have been manufactured for virtual reality applications, or, increasingly, as personal movie viewing systems. Using these off-the-shelf displays is very appealing because of the relative ease with which they can be modified for video see-through use. However, depending on the intended application, the characteristics of the displays frequently are at odds with the requirements for an AR display.

Our ongoing research focus has been on medical applications of AR. In one of our applications, ultrasound-guided needle breast biopsy (Fig. 1), a physician stands at an operating table. The physician uses a scaled, tracked, patient-registered ultrasound image delivered through our AR system to select the optimal approach to a tumor, insert the biopsy needle into the tumor, verify the needle's position, and capture a sample of the tumor. The physician wears a VST-HMD throughout the procedure. During a typical procedure the physician looks at an assistant a few meters away, medical supplies nearby, perhaps one meter away, the patient half a meter away or closer, and the collected specimen in a jar twenty centimeters from the eyes.

Most commercially available HMDs are designed to look straight ahead. However, as the object of interest (either real or virtual) is brought closer to the viewer's eyes, there is a decreasing region of stereo overlap dedicated to this object (on the nasal side). Since the image content being presented to each eye is very different, the user is presumably unable to get any depth cues from the stereo display in such situations. Users of our system have been observed to move either the object of interest or their head so that the object of interest becomes visible primarily in their dominant eye – from this configuration they can apparently resolve the stereo conflict by ignoring their non-dominant eye.

In typical implementations of video see-through displays, cameras and displays are preset at a fixed angle. Researchers have previously designed VST-HMDs while making assumptions about the normal working distance. In one design, discussed in the following section, the video cameras were preset to converge slightly in order to allow the wearer sufficient stereo overlap when viewing close objects. In another design, the convergence of cameras and displays could be selected before using the system to an angle most appropriate for the expected working distance. Converging the cameras, or both the cameras and the displays, is only practical if the user need not view distant objects as there is often not enough stereo overlap or too much disparity to fuse distant objects.

This paper discusses an alternative to physically modifying convergence of either the cameras or the displays. Our technique does not require moving parts within the HMD and is implemented fully in software.

2. Related work

In the pioneering days of VST AR work, researchers used to improvise (successfully) by mounting a single lipstick camera onto a commercial VR HMD. Even then careful consideration was given to issues such as calibration between tracker and camera [Bajura1992].

In 1995, our team assembled a stereo AR HMD [State1996]. The device consisted of a commercial VR-4 unit and a special plastic mount (attached to the VR-4 with Velcro™!), which held 2 Panasonic lipstick cameras equipped with oversized C-mount lenses. The lenses had been chosen for their extremely low distortion characteristics, since their images were digitally composited with perfect perspective CG imagery. Two important flaws of the device emerged: (1) mismatch between the fields of view of camera (28° horizontal) and display (ca. 40° horizontal) and (2) eye-camera offset or parallax (see [Azuma 1997] for an explanation), which gave the wearer the impression of being taller and closer to the surroundings than she actually was. To facilitate close-up work, the cameras were not mounted parallel to each other, but at a fixed 4° convergence angle, which was calculated to also provide sufficient stereo overlap when looking at a collaborator across the room while wearing the device.

Today many video-see-through AR systems in labs around the world are built with stereo lipstick cameras mounted on top of typical VR (opaque) or optical see-through HMDs operated in opaque mode (for example, [Kanbara2000]). Such designs will invariably suffer from the eye-camera offset problem mentioned above. (The design described in this paper is no exception, even though our new technique is not limited to such designs.)

[Fuchs1998] describes a device that was designed and built from individual LCD display units and custom-designed optics. It had two identical "eye pods." Each pod consisted of an ultra-compact display unit and a lipstick camera. The camera's optical path was folded with mirrors, similar to a periscope, making the device "parallax-free" [Takagi2000]. In addition, the fields of view of camera and display in each pod were matched. Hence, by carefully aligning the device on the wearer's head, one could achieve near perfect registration between the imagery seen in the display and the peripheral imagery visible to the naked eye around each of the compact pods. Thus this VST-HMD can be considered *orthoscopic* [Drascic1996]. Since each pod could be moved separately, the device (characterized by small field of view and high angular resolution) could be adjusted to various degrees of convergence (for close-up work or room-sized tasks), albeit not dynamically but on a per-session basis. The reason for this was that moving the pods in any way required inter-ocular recalibration. (The "head tracker" was rigidly mounted on one of the pods so there was no need to recalibrate between head tracker and eye pods.) The movable pods also allowed exact matching of the wearer's IPD.

Other researchers have also attacked the parallax problem by building devices in which mirrors or optical prisms bring the cameras "virtually" closer to the wearer's eyes. Such a design is described in detail in [Takagi2000], together with a geometrical analysis of the stereoscopic distortion of space—and thus deviation from ortho-stereoscopy—that results when specific parameters in a design are mismatched. For example, there can be a mismatch between the convergence of the cameras and the display units (such as in the device from [State1996]), or a mismatch between inter-camera distance and user IPD.

While [Takagi2000] advocates rigorous ortho-stereoscopy, other researchers have investigated how quickly users adapt to dynamic changes in stereo parameters. [Milgram1992] investigated users' judgment errors when subjected to unannounced variations in inter-camera distance. The authors determined that users adapted surprisingly quickly to the distorted space when presented with additional visual cues (virtual or real) to aid with depth scaling. Consequently, they advocate dynamic changes of parameters such as inter-camera distance or convergence distance for specific applications.

[Ware1998] describes experiments with dynamic changes in virtual camera separation within a fish tank VR system. They used a z-buffer sampling method to heuristically determine an appropriate inter-camera distance for each frame and a dampening technique to avoid abrupt changes. Their results indicate that users do not experience "large perceptual distortions," allowing them to conclude that such manipulations can be beneficial in certain VR systems.

Finally, [Matsunaga2000] describes a teleoperation system using live stereoscopic imagery (displayed on a monitor to users wearing active polarizers) acquired by motion-controlled cameras. The results indicate that users' performance was significantly improved when the cameras dynamically converged onto the target object (peg to be inserted into a hole) compared to when the cameras' convergence was fixed onto a point in the center of the working area.

3. The dynamic virtual convergence system

The [Fuchs1998] device described above had two eye pods that could be converged *physically*. As each pod was toed in for better stereo overlap at close range, the pod's video camera and display were "yawed" together (since they were co-located within the pod), guaranteeing continuous alignment between display and peripheral

imagery. Our new technique deliberately violates that constraint but uses "no moving parts," since it is implemented fully in software. Hence there is no need for recalibration as convergence is changed. It is important to note that sometimes VR or AR implementations mistakenly mismatch camera and display convergence, whereas our method intentionally decouples camera and display convergence in order to allow AR work in situations where an ortho-stereoscopic VST-HMD doesn't reach (because there are usually no display pixels close to the user's nose).

The implementation requires a VST HMD whose video cameras have a much larger field of view than the display unit (Fig. 2). Only a fraction of a camera's image (proportional to the display's field of view) is actually shown in the corresponding display via re-projection (Fig. 3). The cameras acquire enough imagery to allow full stereo overlap from close range to infinity (parallel viewing).

Thus, by enlarging the cameras' field of view, we remove the need to physically toe in the camera to change convergence. But what about the display? To preserve the above mentioned alignment between display content and peripheral vision, the display would have to physically toe in for close-up work, together with the cameras, as with the device described in [Fuchs1998]. While this is doubtlessly desirable, we have determined that it is in fact possible to operate a device with fixed, parallel-mounted displays in this way, at least for a majority of our users. This surprising finding might be easier to understand by considering that if the displays converged physically while performing a near-field task, the user's eyes would also verge inward to view the task-related objects (presumably located just in front of the user's nose). With fixed displays however, the user's eyes are viewing the very same retinal image pair, but in a configuration which requires the eyes to not verge in order for stereoscopic fusion to be achieved.

Thus virtual convergence always provides images that are aligned for parallel viewing. By preventing (relieving?) the user from converging her eyes, it allows stereoscopic fusion of extremely close objects even in display units that have little or no stereo overlap at close range. This is akin to wall-eyed fusion of certain stereo pairs in printed matter (including the images in this paper, Figs. 3-bottom, 9, and 10-bottom, HMD image only) or to the horizontal shifting of stereo image pairs on projection screens in order to reduce ghosting when using polarized glasses. It creates a disparity-vergence conflict (not to be confused with the well-known accommodation-vergence conflict present in most stereoscopic displays [Drascic1996]). For example, if we point converging cameras at an object located 1m in front of the cameras, then present the image pair to a user in a HMD with parallel displays, the user will not converge his eyes to fuse the object but will nevertheless perceive it as being much closer than infinitely far away due to the disparity present in the image pair. This indicates that the disparity depth cue dominates vergence in such situations; our system takes advantage of this fact. Also, by centering the object of interest in the camera images and presenting it on parallel displays, we all but eliminate the accommodation-vergence conflict for the object of interest, assuming that the display is collimated. In reality, HMD displays are built so that their images appear at finite but rather large (compared to the close range we are targeting) distances to the user, for example, two meters in the Sony Glasstron device we use (described below). Even so, users of a virtual convergence system will experience a significant reduction of the accommodation-vergence conflict, since virtual convergence reduces screen disparities (in our case, the screen is the virtual screen visible within the HMD). Reducing screen disparities is often recommended [Akka1992] if one wishes to reduce potential eye strain caused by the accommodation-vergence conflict. Table 1 below shows the relationships between the three depth cues accommodation, disparity and vergence for our VST-HMD with and without virtual convergence, assuming the user is attempting to perform a close-range task.

Table 1. Depth cues and depth cue conflicts for close-range work: Enabling virtual convergence maximizes stereo overlap for close-range work, but "moves" the vergence cue to infinity.

Virtual convergence setting	Available close-range stereo overlap	Where are depth cues accommodation (A), disparity (D), and vergence (V)?		Conflicts between depth cues
		Close-range	2m through ∞	
OFF	partial	D, V	A	A-D, A-V
ON	full	D	A, V	A-D, D-V

By eliminating the moving parts, we are now in a position to dynamically change the virtual convergence. Our implementation allows the computer system to make an educated guess as to what the convergence distance should be at any given time and then set the display (re)projection transformations accordingly. The following sections describe our hardware and software implementation and present some application results as well as informal user reactions to this technology.

3.1. Hardware implementation

We use a Sony Glasstron LDI-D100B stereo HMD with full-color SVGA (800x600) stereo displays, a device we have found to be very reliable, characterized by excellent image quality even when compared to considerably more expensive commercial units. (Unfortunately, it is no longer on the market.) It has a horizontal field of view of α=26°. The display-lens elements are built d=62 mm apart and cannot be moved to match a user's inter-pupillary distance (IPD). However, the displays' exit pupils are large enough [Robinett1992] for users with IPDs between roughly 50 and 75 mm. Nevertheless users with extremely small or extremely large IPDs will perceive a prismatic depth plane distortion (curvature) since they view images through off-center portions of the lenses; we do not address this issue here any further. We have mounted two Toshiba IK-M43S miniature lipstick cameras on top of this device (Fig. 4). The cameras are mounted parallel to each other. The distance between them is also 62 mm. There are no mirrors or prisms, hence there is a significant eye-camera offset (about 60-80 mm horizontally and about 20-30 mm vertically, depending on the wearer). In addition, there is an IPD mismatch for any user whose IPD is significantly larger or smaller than 62 mm.

The head-mounted cameras are fitted with 4-mm-focal-length lenses providing a field of view of approximately β=50° horizontal, nearly twice the displays' field of view. It is typical for small wide-angle lenses to exhibit barrel distortion, and in our case the barrel distortion is non-negligible and must be eliminated (per software) before attempting to register any synthetic imagery to it.

The entire head-mounted device, consisting of Glasstron, lenses, and an aluminum frame on which cameras and infrared LEDs for tracking are mounted, weighs well under 250 grams. (Weight was an important issue in this design since the device is used in extended medical experiments and is often worn by our medical collaborator for an hour or longer without interruption.)

Our AR software runs on an SGI Reality Monster equipped with InfiniteReality2 (IR2) graphics pipes and DIVO digital video I/O boards. The HMD cameras' video streams are converted from S-video to a 4:2:2 serial digital format via Miranda picoLink ASD-272p decoders and then fed to two DIVO boards. HMD tracking information is provided by an Image-Guided Technologies FlashPoint 5000 opto-electronic tracker. A graphics pipe in the SGI delivers the stereo left-right augmented images in two SVGA 60 Hz channels. These are combined into the single-channel left-right alternating 30Hz SVGA format required by the Glasstron with the help of a Sony CVI-D10 multiplexer.

3.2. Software implementation

Our AR applications are largely single-threaded, using a single IR2 pipe and a single processor. For each synthetic frame, we capture a frame from each camera via the DIVO boards. When it is important to ensure maximum image quality (considering that we will end up looking in close-up at a (re-projected) portion of an NTSC-resolution image) we capture two successive NTSC fields, even though that may lead to the well-known visible horizontal tearing effect during rapid user head motion.

DIVO-captured frames are initially deposited in main memory, from where they are transferred to texture memory. Before any graphics can be superimposed onto the camera imagery, it must be rendered on textured polygons. We use a 2D polygonal grid which is radially stretched (its corners are pulled outward) to compensate for the above mentioned lens distortion (Fig. 5), analogous to the pre-distortion technique described in [Watson1995]. The distortion compensation parameters are determined in a separate calibration procedure; we have found that both a third-degree and a fifth-degree coefficient are needed in the polynomial approximation [Robinett1992]. The stretched, video-texture-mapped polygon grids are rendered from the cameras' points of view (using tracking information from the FlashPoint unit and inter-camera calibration data acquired during yet another separate calibration procedure).

In a conventional video-see-through application one would use parallel display frustums to render the video textures—since the cameras are parallel (as recommended by [Takagi2000]). Also, the display frustums should have the same field of view as the cameras. However, for virtual convergence, we use display frustums that are verged in. Their field of view is equal to the display's field of view α. As a result of that, the user ends up seeing a reprojected (and distortion-corrected) sub-image (Fig. 6) in each eye.

The maximum convergence angle is $\delta=\beta-\alpha$, in our case approximately 24°. At that convergence angle, the stereo overlap region of space begins at a distance $z_{over,min}=0.5d\tan(90°-\beta/2)$, in our case approximately 66 mm, and full stereo overlap is achieved at a distance $z_{over,full}=d/(\tan(\beta/2)-\tan(\alpha-\beta/2))$, in our case approximately 138 mm. At that latter distance the field of view subtends an area that is $d+2z_{over,full}\tan(\alpha-\beta/2)$ wide, or approximately 67 mm in our case.

After setting the display frustum convergence, application-dependent synthetic elements are rasterized using the very same verged, narrow display frustums. For some parts of the real world we have registered geometric models (Fig. 7), and so we can rasterize those in Z only,

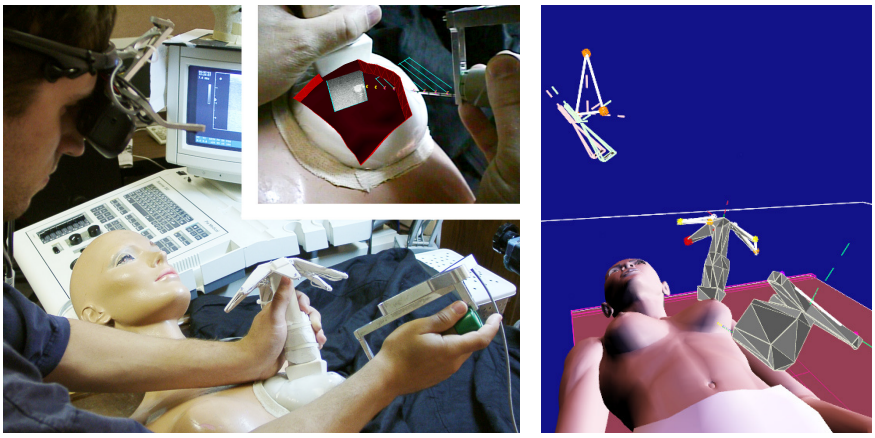

Fig. 1. Far left: AR guidance system in use on a breast biopsy training phantom. The user holds an ultrasound probe (left hand) and a biopsy needle (right hand). Inset: typical HMD view shows synthetic opening into the patient and registered ultrasound image scanning a suspicious lesion. Left: The system displays a control view with dynamic avatars for the optically tracked VST-HMD, probe, and needle

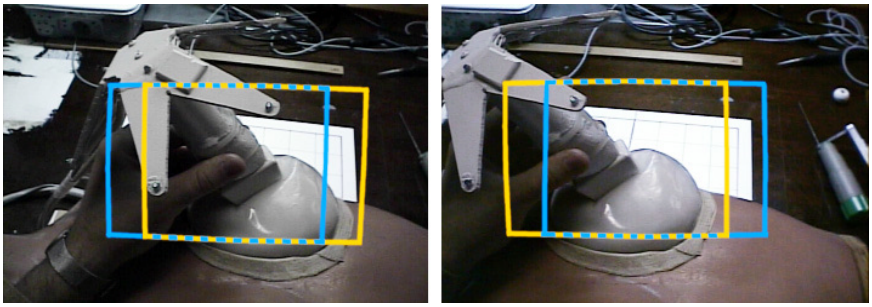

Fig. 2. Wide-angle stereo views (with barrel distortion) as acquired by the HMD cameras. The blue (virtual convergence off) and yellow (virtual convergence on) outlines show the re-projected parts of the video images corresponding to the HMD images in Fig. 3—curved because of distortion. (fuse all stereo pairs wall-eyed)

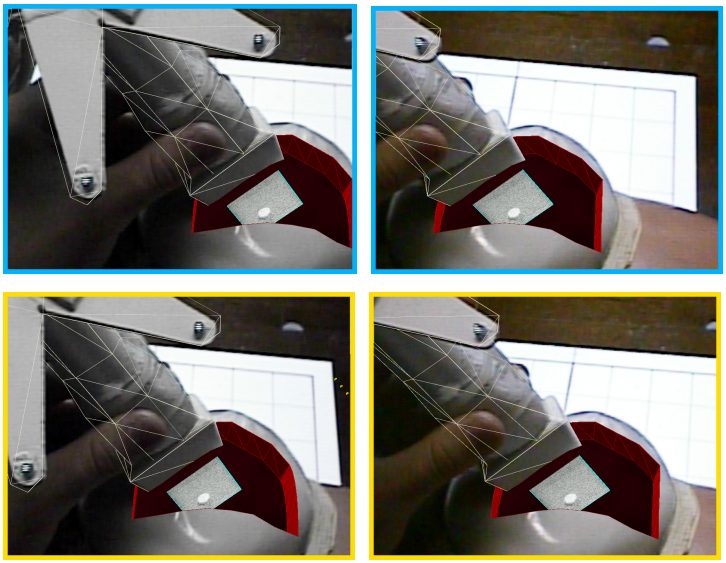

Fig. 3. Stereo images displayed in the HMD without (top) and with virtual convergence (bottom), all distortion-corrected (cf. Figs. 2, 5 and 6)

Fig. 4. VST-HMD built from Sony unit. The frame on top holds cameras and IR tracking LEDs

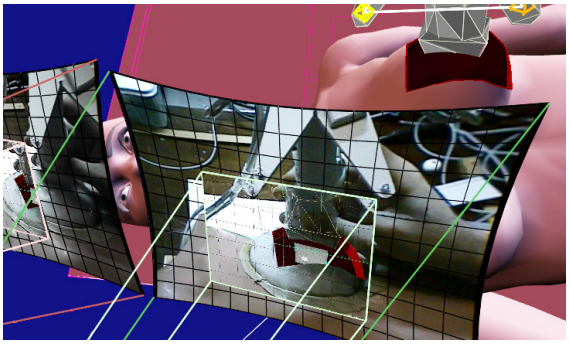

Fig. 5. A deformed polygonal grid removes the video texture distortion (exaggerated). Smaller display frustum has re-projected, distortion-corrected image shown in HMD

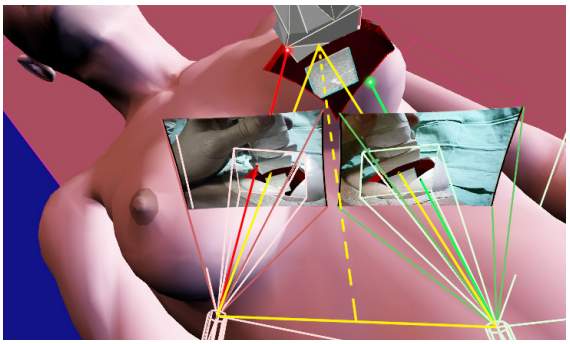

Fig. 6. Wide-FOV (camera) frustums and narrow, converged display frustums. The yellow isosceles triangle indicates display frustum convergence

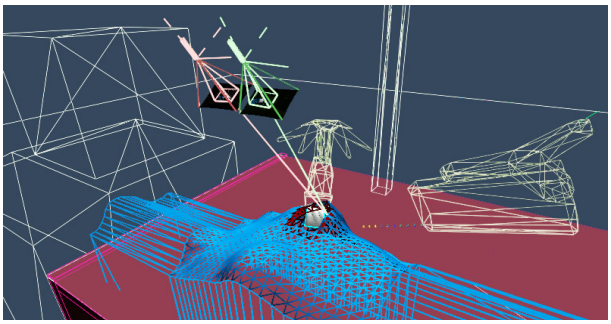

Fig. 7. Scene geometry known to the AR system

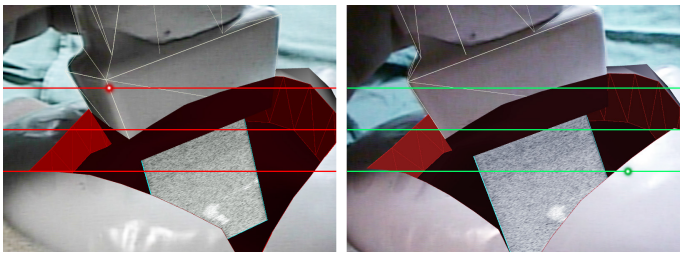

Fig. 9. Z-buffer inspection on 3 selected scan lines. The highlighted points mark the closest depth values found, corresponding to the red/green lines in Fig. 6

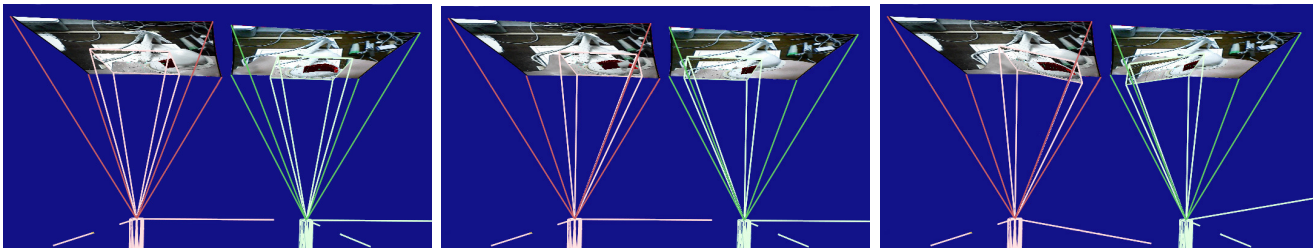

Fig. 8. Unconverged (left), sheared (center) and rotated (right) display frustums

Fig. 10. Simulated wide-angle stereo views through and “around” HMD (left and center). Virtual convergence is off in the top images and on in the bottom ones. Note how alignment between features in the display and below the display (for example, between the nipples, which are vertically aligned in the top stereo pair) is lost with virtual convergence, illustrating the new disparity-vergence conflict

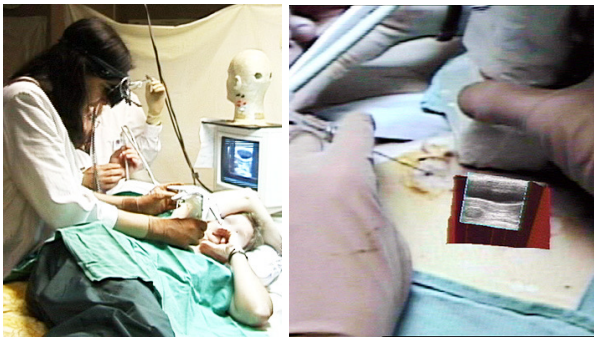

Fig. 11. Left: AR ultrasound human subject experiment and typical HMD view (right)

Fig. 12. Left: AR system used to model tracked real world objects with textures; HMD view (right)

thereby priming the Z-buffer for correct mutual occlusion between real and synthetic elements [State1996]. As shown in Fig. 7, only part of the patient surface is known. The rest is extrapolated with straight lines to approximately the size of a human. There are static models of the table and of the ultrasound machine (cf. Fig. 1), as well as of the tracked handheld objects [Lee2001]. Floor and lab walls are modeled coarsely with only a few polygons.

3.4. Sheared vs. rotated display frustums

One issue that we considered early on during the implementation phase of this technique was the question of whether the verged display frustums should be sheared or rotated (Fig. 8). Shearing the frustums keeps the image planes for the left and right eyes coplanar, thus eliminating vertical disparity or *dipvergence* [Rolland1995] between the two images. At high convergence angles (i. e., for extreme close-up work), viewing such a stereo pair in our system would be akin to wall-eyed fusion of images specifically prepared for cross-eyed fusion.

On the other hand, rotating the display frustums with respect to the camera frustums, while introducing dipvergence between corresponding features in stereo images, presents to each eye the very same retinal image it would see if the display were capable of physically toeing in (as discussed above), thereby also stimulating the user's eyes to toe in.

To compare these two methods for display frustum geometry, we implemented an interactive control (slider) in the system's user interface. For a given virtual convergence setting, we can smoothly blend between sheared and rotated frustums by moving the slider. When that happens, the HMD user perceives a curious distortion of space, maybe similar to a dynamic prismatic distortion. We did not conduct a controlled user study to determine whether sheared or rotated frustums are preferable; we merely assembled an informal group of testers (members of our group and other researchers) and there was a definite preference towards the rotated frustums method overall. However, none of the testers found the sheared frustum images more difficult to fuse than the rotated frustum ones, which is understandable given that sheared frustum stereo imagery has no dipvergence (as opposed to rotated frustum imagery). It is of course difficult to quantify the stereo perception experience without a carefully controlled study; for the time being we relied solely on users' preferences as guidance for further development.

3.5. Automating virtual convergence

Our goal was to achieve on-the-fly convergence changes under algorithmic control to allow users to work comfortably at different depths. We initially tested whether a human user could in fact tolerate dynamic virtual convergence changes at all. To this end, we implemented a user interface slider controlling convergence and assigned a human operator to continually adjust the slider while a user was viewing AR imagery in the VST-HMD. The convergence slider operator was permanently watching the combined left-right (alternating at 60Hz) SVGA signal fed to the Glasstron HMD on a separate monitor. This signal appears like a blend between the left and right eye images, and any disparity between the images is immediately apparent. The operator was continually adjusting the convergence slider, attempting to minimize the visual disparity between the images (thereby maximizing stereo overlap). This means that if most of the image consists of objects located close to the HMD user's head, the convergence slider operator tended to verge the display frustums inward. With practice our operator became quite skilled; most test users had positive reactions, with one notable exception (a member of our team) reporting extreme discomfort.

Our next goal was to create a real-time algorithmic implementation capable of producing a numeric value for display frustum convergence for each frame in the AR system. We considered three distinct approaches for this:

(1) Image content based: this is the algorithmic version of the "manual" method described above. An attractive possibility would be to use a maximization of mutual information algorithm [Viola1995]. An image-based method could run as a separate process and could be expected to perform relatively quickly since it need only optimize a single parameter. This method should be applied to the mixed reality output rather than the real world imagery to ensure that the user can see virtual objects that are likely to be of interest. Under some conditions, such as repeating patterns in the images, a mutual information method would fail by finding an "optimal" depth value with no rational basis in the mixed reality. Under most conditions however, including color and intensity mismatches between the cameras, a mutual information algorithm would appropriately maximize the stereo overlap in the left and right eye images.

(2) Z-buffer based: this approach inspects values in the Z-buffer of each stereo image pair and (heuristically) determines a likely depth value to set the convergence to. [Ware1998] gives an example for such a technique.

(3) Geometry based: this approach is similar to (2) but uses geometry data (models as opposed to pixel depths) to (again heuristically) compute a likely depth value to set the convergence to. In other words, it works

on pre-rasterization geometry, whereas (2) uses post-rasterization geometry.

Approaches (1) and (2) both operate on finished images. Thus they cannot be used to set the convergence for the current frame but only to predict a convergence value for the next frame. Conversely, approach (3) can be used to immediately compute a convergence value (and thus the final viewing transformations for the left and right display frustums) for the current frame, before any geometry is rasterized. However, as we shall see, this does not automatically exclude (1) and (2) from consideration. Rather, approach (1) was eliminated on the grounds that it would require significant computational resources. We developed a hybrid of (2) and (3), characterized by inspection of only a small subset of all Z-buffer values, and aided by geometric models and tracking information for the user's head as well as for handheld objects. The following steps describe our hybrid algorithm:

1. For each eye, the full augmented view (as described in Section 3.2) is rendered into the frame buffer (after capturing video, reading trackers, etc.).
2. For each eye, inspect the z-buffer of the finished view along 3 horizontal scan lines, located at heights $h/3$, $h/2$, and $2h/3$ respectively, where h is the height of the image (Fig. 9). Find the average of the closest depths $z_{min}=(z_{min,l}+z_{min,r})/2$. Set the convergence distance z to z_{min} for now. This step is only performed if in the previous frame the convergence distance was virtually unchanged (we use a threshold of 0.01°). Otherwise z is left unchanged from the previous frame.
3. Using tracker information, determine if application-specific geometry (for example, the all-important ultrasound image in our medical application) is within the viewing frustum of either display. If so, set z to the distance of the ultrasound slice from the HMD.
4. Calculate the average value z_{avg} during the most recent n frames, not including the current frame since the above steps can only execute on a finished frame (steps 1-2) or at least on an already calculated display frustum (step 3).
5. Set the display frustums to point to a location at distance z_{avg} in front of the HMD. Calculate the appropriate transformations, taking into account the blending factor between sheared and rotated frustums (see Section 3.4). Go to step 1.

The simple temporal filtering in step 4 is used to avoid sudden, rapid changes. It also adds a delay in virtual convergence update, which for n=10 amounts to approximately 0.5 seconds at our current frame rate of about 20 Hz (a better implementation would vary n as a function of frame rate in order to keep the delay constant). Even though this update seems slower than the human visual system's rather quick vergence response to the *diplopia* (double vision) stimulus, we have not found it jarring or unpleasant.

The conditional update of z in Step 2 prevents most self-induced oscillations in convergence distance. Such oscillations can occur if the system continually switches between two (rarely more) different convergence settings, with the z-buffer calculated for one setting resulting in the other convergence setting being calculated for the next frame. Such a configuration may be encountered even when the user's head is perfectly still and none of the other tracked objects (such as handheld probe, pointers, needle, etc.) are moved.

4. Results

Fig. 10 contains simulated wide-angle stereo views from the point of view of an HMD wearer, illustrating the difference between converged and parallel operation.

The dynamic virtual convergence subsystem has been deployed within two different AR applications. Both use the same modified Sony Glasstron HMD and the hardware and software described in Section 3. The first is an experimental AR system designed to aid physicians in performing minimally invasive procedures such as ultrasound-guided needle biopsies of the breast. This system and a number of recent experiments (Fig. 11) conducted with it are described in detail in [Rosenthal2001]. Our principal medical collaborator used the system on numerous occasions, often for one hour or longer without interruption, while the dynamic virtual convergence algorithm was active. She did not report any discomfort while or after using the system. With her help, we successfully conducted a series of experiments yielding quantitative evidence that AR-based guidance for the breast biopsy procedure is superior to the conventional guidance method in artificial phantoms [Rosenthal2001]. Other physicians, the authors of this paper and other members of our team, as well as several lab visitors have all used this system, albeit for shorter periods of time, without discomfort (except for one individual previously mentioned, who experiences discomfort whenever the virtual convergence is changed dynamically).

The second AR application to use dynamic virtual convergence is a system for modeling real objects using AR (Fig. 12). The system and the results obtained with it were described in detail [Lee2001]. Two of the authors of [Lee2001] have used that system for sessions of one hour or longer, again without noticeable discomfort (immediate or delayed).

5. Conclusions

Other authors have previously noted the conflict introduced in VST-HMDs when the camera axes are not properly aligned with the displays. While we continue to believe that this is significant, our experience with this technique suggests that violating this constraint may be advantageous in systems requiring the operator to use stereoscopic vision at several distances.

Mathematical models such as those developed by [Takagi2000] demonstrate the distortion of the visual world. These models do not demonstrate the volume of the visual world that is actually stereo-visible (i.e., visible to both eyes and within 1-2 degrees of center of stereo-fused content). Dynamically converging the cameras—whether they are real cameras as in [Matsunaga2000] or virtual cameras (i.e., display frustums) pointed at video-textured polygons as in our case—makes a greater portion of the near field around the point of convergence stereoscopically visible at all times. Most users have successfully used our AR system with dynamic virtual convergence to place biopsy and aspiration needles with high precision or to model objects with complex shapes.

Our experience suggests that the distortion of the perceived visual world is not as severe as predicted by the mathematical models if the user's eyes converge at the distance selected by the system. (If they converge at a different distance, stereo overlap is reduced and increased spatial distortion and/or eye strain may be the result. We therefore believe that our largely positive experience with this technique is due to a well-functioning convergence depth estimation algorithm.) Indeed, a substantial degree of perceived distortion is eliminated if one assumes that the operator has approximate knowledge of the distance to the point being converged on (experimental results in [Milgram1992] support this statement). Given the intensive hand-eye coordination required for our applications, it seems reasonable to conjecture that our users' perception of their visual world may be rectified by other sources of information such as seeing their own hand. Indeed, the hand may act as a "visual aid" as defined by [Milgram1992].

This type of adaptation is apparently well within the abilities of the human visual system as evidenced by the ease with which individuals adapt to new eyeglasses and to using binocular magnifying systems. On the other hand, while our approach has proved surprisingly unproblematic, we do not consider it superior to rigorous ortho-stereoscopy. We would therefore like to encourage HMD manufacturers to put more display pixels towards the wearer's nose in future designs.

6. Future Work

Our new technique reduces the accommodation-vergence conflict while introducing a disparity-vergence conflict. It may be useful to investigate whether smoothly blending between zero and full virtual convergence is useful. Also, should that a parameter to be set on a per user basis, per session basis, or dynamically?

Second, a thorough investigation of sheared vs. rotated frustums (should that be changed dynamically as well?), as well as a controlled user study for the entire system, with the goal of obtaining quantitative results, seem desirable.

Finally, we plan to use our technique on a parallax-free device. To this end, we have mounted a mirror-camera device on a Sony Glasstron HMD. This new orthoscopic device has recently been incorporated into our system and we plan to report on our experience with it in a future publication. (Of course, the term "orthoscopic" does not apply when virtual convergence is used.)

7. Acknowledgments

We thank Kurtis Keller, Etta Pisano, MD, Warren Robinett, Jannick P. Rolland, Michael Rosenthal, as well as the numerous collaborators and lab visitors who have tested our system and related their experience with it.

The geometric models for patient and physician used in the simulated images and in our system's control view were created with Curious Labs Poser 4 (formerly Metacreations Poser 4).

Funding for this work was provided by: NIH (CA 47982-10)and the STC for Computer Graphics and Scientific Visualization (NSF Cooperative Agreement ASC-8920219).

8. References

[Akka1992] Akka, Robert. "Automatic software control of display parameters for stereoscopic graphics images." SPIE Volume 1669, Stereoscopic Displays and Applications III (1992), 31-37.

[Azuma1997] Azuma, Ronald T. "A Survey of Augmented Reality." Presence: Teleoperators and Virtual Environments 6, 4 (August 1997), MIT Press, 355-385.

[Bajura1992] Bajura, Michael, Henry Fuchs, and Ryutarou Ohbuchi. "Merging Virtual Objects with the Real World: Seeing Ultrasound Imagery within the Patient." Proceedings of SIGGRAPH '92 (Chicago, IL, July 26-31, 1992). In Computer Graphics 26, #2 (July 1992), 203-210.

[Drascic1996] Drascic, David, and Paul Milgram. "Perceptual Issues in Augmented Reality."SPIE Volume 2653; Stereoscopic Displays and Virtual Reality Systems III (1996), 123-124.

[Fuchs1998] Fuchs, Henry, Mark A. Livingston, Ramesh Raskar, D'nardo Colucci, Kurtis Keller, Andrei State, Jessica R. Crawford, Paul Rademacher, Samuel H. Drake, and Anthony A. Meyer, MD. "Augmented Reality Visualization for Laparoscopic Surgery." Proceedings of Medical Image Computing and Computer-Assisted Intervention—MICCAI '98 (Cambridge, MA, USA, October 11-13, 1998), 934-943.

[Kanbara2000] Kanbara, M., T. Okuma, H. Takemura, N. Yokoya, "A Stereoscopic Video See-through Augmented Reality System Based on Real-time Vision-Based Registration." Proceedings of Virtual Reality 2000, March 2000, 255-262.

[Lee2001] Lee, Joohi, Gentaro Hirota, and Andrei State. "Modeling Real Objects Using Video See-Through Augmented Reality." Proceedings of the Second International Symposium on Mixed Reality (ISMR 2001), March 14-15, 2001, Yokohama, Japan, 19-26.

[Matsunaga2000] Matsunaga, Katsuya, Tomohide Yamamoto, Kazunori Shidoji, and Yuji Matsuki. "The effect of the ratio difference of overlapped areas of stereoscopic images on each eye in a teleoperation." SPIE Vol. 3957, Stereoscopic Displays and Virtual Reality Systems VII (2000), 236-243.

[Milgram1992] Milgram, P., and Martin Krüger. "Adaptation Effects in Stereo Due To Online Changes in Camera Configuration." SPIE Vol. 1669-13, Stereoscopic Displays and Applications III (1992), 122-134.

[Robinett1992] Robinett, Warren, and Jannick P. Rolland. "A Computational Model for the Stereoscopic Optics of a Head-Mounted Display." Presence: Teleoperators and Virtual Environments 1, 1 (Winter 1992), MIT Press, 45-62.

[Rolland1995] Rolland, Jannick, and William Gibson. "Towards Quantifying Depth and Size Perception in Virtual Environments." Presence: Teleoperators and Virtual Environments 4, 1 (Winter 1995), MIT Press, 24-49.

[Rosenthal2001] Rosenthal, Michael, Andrei State, Joohi Lee, Gentaro Hirota, Jeremy Ackerman, Kurtis Keller, Etta D. Pisano, Michael Jiroutek, Keith Muller, and Henry Fuchs. "Augmented Reality Guidance for Needle Biopsies: A Randomized, Controlled Trial in Phantoms." To appear in the Proceedings of Medical Image Computing and Computer-Assisted Intervention—MICCAI 2001 (Utrecht, The Netherlands, 14-17 October 2001).

[State1996] State, Andrei, Mark A. Livingston, Gentaro Hirota, William F. Garrett, Mary C. Whitton, Henry Fuchs, and Etta D. Pisano (MD). "Technologies for Augmented-Reality Systems: Realizing Ultrasound-Guided Needle Biopsies." Proceedings of SIGGRAPH '96 (New Orleans, LA, August 4-9, 1996). In Computer Graphics Proceedings, Annual Conference Series 1996, ACM SIGGRAPH, 439-446.

[Takagi2000] Takagi, A., S. Yamazaki, Y. Saito, and N. Taniguchi. "Development of a stereo video see-through HMD for AR systems." Proceedings of International Symposium on Augmented Reality (ISAR) 2000, 68-77.

[Viola1995] Viola, P. and W. Wells. "Alignment by Maxmization of Mutual Information." International Conference on Computer Vision, Boston, MA, 1995.

[Ware1998] Ware, Colin, Cyril Gobrect, and Mark Paton. "Dynamic adjustment of stereo display parameters." IEEE Transactions on Systems, Man and Cybernetics, 28(1), 56-65.

[Watson1995] Watson, Benjamin A., Larry F. Hodges. "Using Texture maps to Correct for Optical Distortion in Head-Mounted Displays." Proceedings of the Virtual Reality Annual Symposium '95, IEEE Computer Society Press, 1995, 172-178.