

Psychophysical Evaluation of *In-Situ* Ultrasound Visualization

Bing Wu, Roberta L. Klatzky, Damion Shelton, and George D. Stetten

Abstract—We present a novel psychophysical method for evaluating ultrasonography based on Real-Time Tomographic Reflection (RTTR), in comparison to Conventional Ultrasound (CUS). The method measures the user's perception of the location of an ultrasound-imaged target independently from assessing the action employed to reach it. Three experiments were conducted with the Sonic Flashlight (SF), an RTTR device, and CUS. The first two experiments determined subjects' perception of target location with a triangulation-by-pointing task. Depth perception with the SF was comparable to direct vision, while CUS caused considerable underestimation of target depth. Binocular depth information in the SF was shown to significantly contribute to its superiority. The third experiment tested subjects in an ultrasound-guided needle insertion task. Because the SF provides visualization of the target at its actual location, subjects performed insertions faster and more accurately by using the SF rather than CUS. Furthermore, the trajectory analysis showed that insertions with the SF generally went directly to the target along the desired path, while CUS often led to a large deviation from the correct path consistent with the observed underestimation of target depth. These findings lend great promise to the use of RTTR-based imaging in clinical practice and provide precise means of assessing efficacy.

Index Terms—Psychology, evaluation/methodology, artificial, augmented, and virtual realities, image display, medical information systems, real time.

1 INTRODUCTION

MANY medical procedures require accurate insertion of a needle into the human body. Common procedures where needle placement is an important task include peripherally inserted central catheter (PICC) insertion and needle biopsies. Failure rates for initial insertion based on landmark guidance have been found to be relatively high (18.7 percent, 15.9 percent, and 35.0 percent, respectively, for cannulation of internal jugular vein, subclavian vein, and femoral vein, by fairly experienced or supervised physicians [1]). To improve successful treatment and minimize complications, ultrasound is often used to examine the portion of the body where the procedure will take place and provide intraoperative real-time feedback. Fixed guides may be mounted on the ultrasound transducer to aid in the accurate placing of the needle. However, visual feedback (in the form of a video image) is typically provided via a computer monitor or video display, located separately from the transducer. This separation introduces the problem of integrating the image display and workspace.

A number of techniques that could potentially address this problem have been developed [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12]. They have in common the superimposition of an image on the operator's workspace,

thereby removing the need to shift focus. These image-overlay techniques fall into the broad category of *augmented reality* (AR). One approach is to either fully or partially replace the operator's direct vision by means of a head-mounted display (HMD). State and associates at UNC [4], [5] and Sauer at Siemens Corp. [6] have developed HMD-based ultrasound AR systems. These systems track the position and orientation of the operator's head, surgical tools, ultrasound transducer, and patient's body in order to correctly combine ultrasound images with a view of the patient from video cameras and present the augmented image on the HMD. Unfortunately, present HMD systems suffer from problems such as tracking lag, low resolution of the displays, and limited field of view. To overcome these limitations, Stetten et al. [7], Chang et al. [8], Stetten [9] and Masamune et al. [10] proposed a technology which we refer to as real-time tomographic reflection (RTTR). The purpose of this paper is to describe an implementation of RTTR and a psychophysical evaluation of its efficacy for needle placement in a simulated clinical procedure.

Stetten et al. and Masamune et al. separately proposed the idea of RTTR. Although the systems are similar in design, Stetten's was originally developed for real-time visualization of ultrasound, while Masamune's was focused on static display of CT images. Stetten's RTTR system functions by fixing the relative geometry of the ultrasound transducer, the display, and a half-silvered mirror to produce a virtual counterpart of the tomographic image (a sector-scan B-mode ultrasound image) at the scanned anatomy within the body (see Fig. 1). Through the half-silvered mirror, the ultrasound image is projected as if it "shines out" from the probe and illuminates the inner tissue, without its being occluded by any more proximal surface such as skin. For that reason, this implementation of RTTR has been named the Sonic Flashlight (SF). In the SF, light rays reflected from the mirror appear as if they come from the virtual slice. This brings about binocular depth cues including convergence of eyes and disparity between the left and right retinal images. With these cues, the virtual

- B. Wu is with the Department of Psychology and the Robotics Institute, Carnegie Mellon University, 436G Baker Hall, Pittsburgh, PA 15213. E-mail: bingwu@andrew.cmu.edu.
- R.L. Klatzky is with the Department of Psychology, Carnegie Mellon University, 354K Baker Hall, Pittsburgh, PA 15213. E-mail: klatzky@cmu.edu.
- D. Shelton is with the Robotics Institute, Carnegie Mellon University, 214 Smith Hall, Pittsburgh, PA 15213. E-mail: beowulf@cs.cmu.edu.
- G.D. Stetten is with the Robotics Institute, Carnegie Mellon University and the Department of Biomedical Engineering, University of Pittsburgh, 749 Benedum Hall, University of Pittsburgh, Pittsburgh, PA 15261. E-mail: george@stetten.com.

Manuscript received 7 Dec. 2004; revised 15 Mar. 2005; accepted 12 May 2005; published online 8 Dec. 2005.

For information on obtaining reprints of this article, please send e-mail to: tocg@computer.org, and reference IEEECS Log Number TVCGSI-0223-1204.

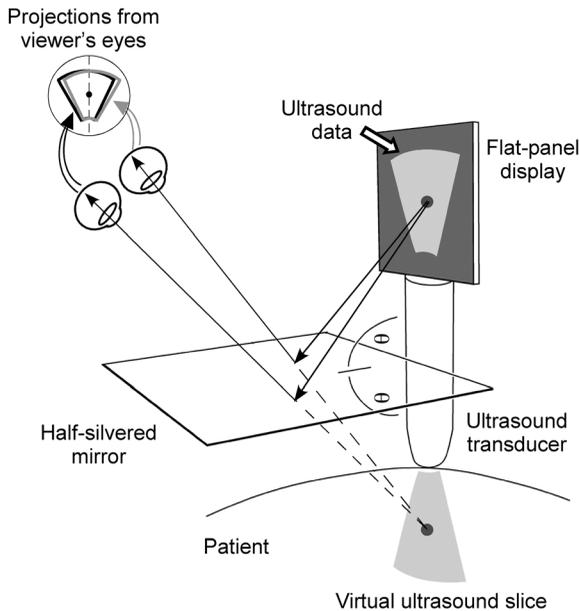


Fig. 1. Optical and perceptual characteristics of Real-Time Tomographic Reflection (RTTR). Through the half-silvered mirror, the ultrasound slice on the display is superimposed on the scanned area, creating a virtual in-situ image inside the patient. Binocular depth cues including convergence of viewer's eyes and disparity between the two retinal projections of the ultrasound slice are naturally available for localizing the target of interest in 3D space.

slice is located in 3D space by means of the viewer's natural perceptual system (see Fig. 1). There is no need for tracking the observer, the patient, or the transducer in the SF, due to the direct registration between the virtual image and the ultrasound beam. Moreover, the patient, the ultrasound image, the instrument, and the operator's hands are merged into one environment for all observers looking through the half-silvered mirror.

The present paper offers psychophysical evaluation of the Sonic Flashlight with its *in-situ* visualization in comparison to Conventional Ultrasound (CUS), in which the image is shown on a conventional display. It compares the ability of users to localize and guide a needle to targets with the two devices, and it also assesses the contribution of binocular depth cues to performance with the SF. In the first experiment, we used a triangulation procedure to assess subjects' perception of an ultrasound-imaged target in 3D space. In the second experiment, we used the same procedure to compare perception with monocular versus binocular depth cues, when using the SF. The third experiment simulated the clinical procedure of needle insertion. It required subjects to direct a needle into a phantom so as to contact a target. We then could compare the trajectory used in attempts to physically contact the target with an independent assessment of the target's perceived location.

2 PERCEPTUAL ADVANTAGES OF RTTR DISPLAY OF ULTRASOUND DATA

In this section, we briefly review three perceptual concomitants of the SF that might lead it to be superior with respect to its utility for localization of targets. They are 1) the SF's capability of providing a direct visualization of the data with binocular depth cues, 2) its temporal-spatial coupling

of visual-motor coordination, and 3) the aligning of the reference frames in which the target and the image are represented. All three of these could contribute to general superiority of the SF over CUS in the present tasks.

2.1 Direct Visualization of Ultrasound Images with Binocular Depth Cues

As shown in Fig. 1, the SF presents a visualization of an ultrasound slice residing at the location of the scanned area and, hence, enables the user to directly perceive depth information in the virtual slice. As the distance of the slice from the viewer is well within the range where normal binocular cues function (within 2m according to [14]), the human visual system provides a natural mechanism for determining the depth. The visual system can locate a target in the virtual slice from rays that bounce off the mirror, and depth can be perceived using the position of eyes (convergence cue, as shown in Fig. 1) and the proprioceptive feedback from eye muscles deforming the lens (accommodation cue). Relative depth in the slice can also be judged from the difference between the images produced in the left and right eye (binocular disparity cue, see Fig. 1).

The situation with CUS, however, is quite different (see Fig. 2a). With most commercial equipment, the CUS screen provides a metric for depth, which changes as the zoom changes under control of the user. This means that in order to have a spatial representation of target depth, the user must convert the metric into an internalized scale. This requires mental processing, which may be achieved either by translation of the metric through visual imagery and/or by verbal mediation. In either case, this is a demanding cognitive (cf. perceptual) process that is subject to error. It may also increase workload and variability. Because of the importance of this difference between the two devices, we have chosen to conduct experiments here that compare their efficacy and directly assess the role of binocular depth cues in the SF.

2.2 Temporal-Spatial Visual-Motor Coordination

The SF displays live ultrasound data in real time by superimposing ultrasound images onto the user's direct view of the site of operation. Therefore, the ultrasound data and the surgical procedure are colocated in time and space (as in [11], [12]). Typically, the operator may hold the SF in one hand while performing the procedure with the other. He or she looks through the half-silvered mirror to see the target area in the ultrasound image as well as the manipulations performed on it. This provides a completely natural perceptual-motor coupling, and fine motor behavior can be assisted by immediate, continuous, visual feedback of action results.

In contrast, because the CUS display is arbitrarily located in the greater workspace (usually above or to the side of the patient), the user must look back and forth from the ultrasound data to the site of operation. Obviously, such displaced hand-eye coordination is unnatural in terms of the perception/action coupling and, hence, the operator's performance may deteriorate [15]. Attentional shifts, lack of direct guidance from image to hand, and memory load may all contribute to error and variability with CUS.

2.3 Frames of Reference for Image and Action

There are three coordinate frames that are immediately invoked in the performance of ultrasound-guided surgeries.

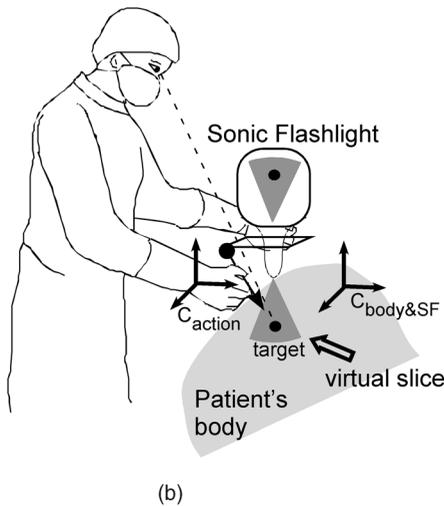
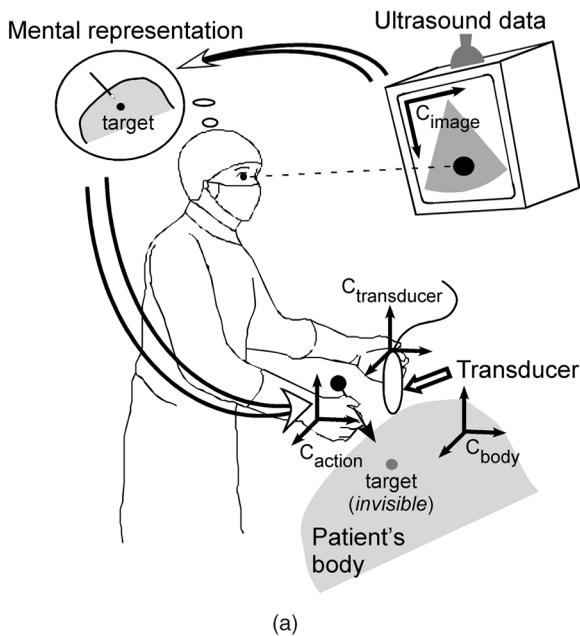


Fig. 2. Schematic of the perceptual/cognitive issues in the two devices. (a) In conventional ultrasound (CUS), there is displaced hand-eye coordination since the user has to look away from the site of operation to check information on the display. Additionally, demanding cognitive processing is required to normalize the metric of the display, align multiple frames of reference (C_{image} , C_{body} , and $C_{transducer}$ for the display, the patient's body, and the transducer, respectively), and then form a mental representation of the target for planning and guiding the action. (b) The Sonic Flashlight (SF) addresses these issues by presenting a virtual ultrasound slice at the depth of the scanned area with reference to the patient's body ($C_{body\&SF}$: the aligned reference frame for both virtual slice and the patient's body). Binocular depth information is available for localizing the target, and the user can aim the needle directly at the target, circumventing the displaced sense of hand-eye coordination. (Adapted from a figure in [13] with permission).

One is the extrinsic frame of reference provided by the patient's body, another is the egocentric frame of the operator's action, and the third is provided by the display of ultrasound data. Normal manipulation requires the fusion of the first two, and brain mechanisms for these translations are increasingly understood [16]. The addition of the ultrasound requires that the perception/action composite frame be further aligned with the ultrasound image frame. The human processing mechanisms by which this occurs

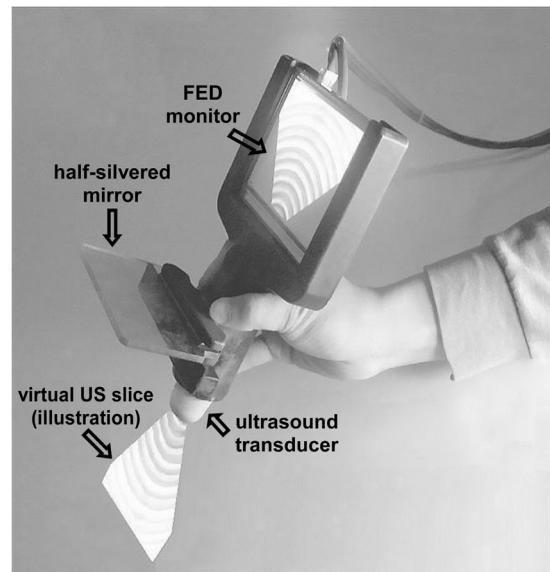


Fig. 3. The Model 4 prototype B-mode Sonic Flashlight used in the experiments described in this paper.

are less well understood, but there is little doubt that the collocation of the image with the perception/action frame facilitates this alignment. Having to align disparate frames of reference is known to impose cognitive load [17].

3 THE SONIC FLASHLIGHT

In this section, we describe details of the implementation of the SF used in the present experiments.

Fig. 3 shows the Model 4 prototype B-mode Sonic Flashlight [8]. It was constructed around a standard ultrasound machine (Pie Medical 50S Tringa, Esaote-Pie Medical Inc., Maastricht, Netherlands) producing a conventional B-mode scan along the axis of the ultrasound transducer. The real-time ultrasound video signal was sent through an analog-to-digital converter to a standard laptop computer (Dell Inspiron 8200, Dell Inc., Round Rock, TX), which provided the processing necessary to properly scale, orient, and locate the image on the display of the SF. A special flat panel monitor based on the Field Effect Display (FED) technology (5.2 inches in diagonal; Model FE524G1, Pixtech, Inc., Santa Clara, CA) was mounted along the axis of the transducer. The FED is a variation on the standard Cathode Ray Tube (CRT) permitting a flat panel configuration by generating electron beams for each pixel from individual emitter tips, instead of steering a single beam with a magnetic coil as in a standard CRT [18]. It offers good brightness and the same excellent off-angle viewing characteristics as a CRT for presenting the ultrasound image.

A half-silvered mirror was mounted perpendicular to the axis, halfway between the tip of the transducer and the bottom of the display (see Fig. 3). As a result of this configuration, the reflected ultrasound image in the mirror appears to be illuminated by the transducer, exactly occupying the space being scanned. Thanks to the semireflective characteristics of the mirror, once the device is calibrated this virtual ultrasound slice is merged at the proper scale and location into the viewer's direct vision. Moreover, the location of the virtual slice is independent of viewpoint owing to the nature of image formation by plane mirrors. Multiple people can view the virtual image simultaneously, facilitating cooperation or training.

The Model 4 of the SF used for this research weighs about 780 g, including the display and the ultrasound probe. Subjects reported no fatigue during the trials. (A newer model of the SF that weighs only 170 g is currently being tested in clinical trials.)

4 EXPERIMENT 1: PERCEIVED LOCATION OF A TARGET BY USING THE SF AND CUS

This experiment used a triangulation-by-pointing method to determine the perceived location of a target imaged through the SF and CUS.

4.1 Method

4.1.1 Subjects

Eleven naïve observers and one coauthor, three females and nine males, participated in the experiment. All subjects had normal or corrected-to-normal vision and stereo acuity better than 40" of arc. All gave informed consent.

4.1.2 Stimuli

A set of nine phantoms was constructed, as shown in Fig. 4a. Six were used in ultrasound-imaging conditions, containing water that had been colored to obscure the contents. Another three phantoms without water were used to measure subjects' baseline performances with direct vision of the target. Each phantom consisted of a plastic tank, the lid of which was cut away and covered with a screen that allowed the ultrasound image to be formed while stabilizing the probe. Inside each phantom three beads of different sizes (0.8 cm, 1.0 cm, and 1.2 cm in diameter) were mounted at three different depths: 3.5 cm, 5 cm, and 6.5 cm. Their relative locations, roughly forming an equilateral triangle around the center, were fixed within a phantom but varied between phantoms so as to create an unpredictable stimulus environment. Across phantoms, each size occurred once at each possible depth. In addition, a phantom with only one bead inside was used for practice to familiarize subjects with the experimental procedure.

4.1.3 Design and Procedure

A 3 (Depth) \times 3 (Size) \times 3 (Viewing conditions: direct vision, CUS, or SF) three-way design was implemented with three replications. All 81 trials were blocked by the viewing condition. The direct-vision condition was tested last to avoid bias; for the other two conditions, display order was counterbalanced across subjects. The testing order of phantoms was also counterbalanced.

Subjects performed the experiment binocularly. On direct-vision trials, beads inside the phantom were visible and labeled by numbers. The subject responded to them successively in an arbitrary order. On ultrasound-imaging trials, the subject was directed to find the ultrasound image of one of the three beads. If that particular bead, identified by the combination of its size and depth, had previously been used in a trial, the experimenter instructed the subject to find a novel one, until all three had been tested. Having found an appropriate bead in the ultrasound slice, the subject could localize it directly from the virtual SF image on the SF trials or, on CUS trials, infer its location by scaling the depth data read using the metric on the CUS display with reference to a standard 1 cm \times 1 cm grid printed on a sheet of paper.

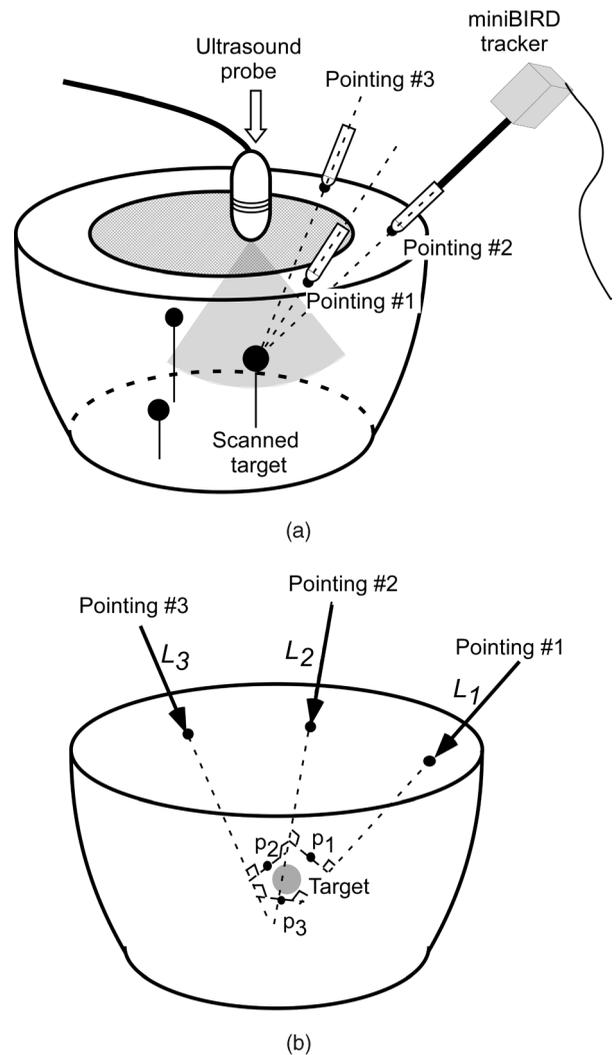


Fig. 4. (a) The water tank phantom and experimental settings for Experiments 1 and 2. (b) The algorithm to assess the perceived location of the target from pointing responses (see text).

To demonstrate the perceived location, the subject pointed a stylus mounted with a tracker (miniBIRD 500, Ascension Technology Co., Burlington, VT) at the bead. For this purpose, the stylus was placed in a response site, consisting of a rotating pen holder mounted on the side of the phantom. Three such holders were mounted around the circumference of each phantom, permitting localization by triangulation according to an algorithm described below. The subject pointed the stylus from each of the three sites in turn, to complete a trial. After all three targets in a phantom had been tested (i.e., after three trials, each with three pointing responses), the next phantom was introduced. No feedback was given to the subject regarding the accuracy of his or her performance.

Before the CUS condition, to assess errors due to subjective representation of scale, subjects were asked to draw lines they believed to be 1 cm and 1 inch in length.

4.1.4 Data Analysis

The perceived location of a target was derived from the subject's pointing behavior. Ideally, all lines along which the subject pointed during different entries would converge to where the subject judged the target to be and, hence, the

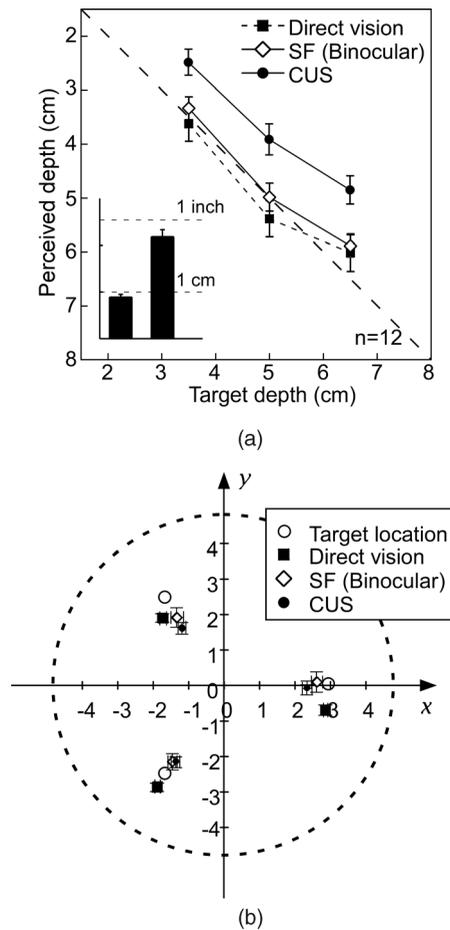


Fig. 5. (a) Mean judged depth as a function of the target depth. The inset figure shows the average length of subjects' drawing of 1 cm and 1 inch. (b) Mean judged locations in the x (subject's frontal) and y (sagittal) axes in each condition, along with target location. Error bars represent the standard error of the mean.

perceived location could be computed using any pair of pointing lines. But, in reality, few pointing lines would intersect due to variability in human performance and measurement noise. Given the lack of exact intersection, in order to assess the subject's perception, we first estimated the perceived target location from each pair of pointing lines. Midpoints were found between each pair of pointing lines, where the lines passed closest to each other (i.e., p1, p2, and p3 in Fig. 4b), using the algorithm in [19]. With large numbers of pointing lines, the intersection points would form a cluster that was normally distributed around the perceived location of the target, assuming uniformly distributed pointing errors. Here, we had three pointing lines, which yielded three such intersection points. The centroid of those points was used as an estimation of the target location perceived by the subject.

4.2 Results

Fig. 5a shows the mean perceived depth for three viewing conditions. Consider the SF and direct-vision conditions first. Clearly, the subjects' responses in the SF condition (open diamonds) were very similar to their baseline performance in the direct-vision condition (solid squares): Their judgments were relatively accurate at the target depth of 3.5 cm and 5.0 cm, but underestimated a little when the target was deeper (6.5 cm). A repeated measure ANOVA

was applied to the data to ascertain if there was any significant difference. No significant main effect or interactions involving viewing conditions were found (main effect: $F(1, 11) = 1.156, p > 0.25$; Viewing conditions \times Target depth: $F(2, 22) = 0.666, p > 0.50$; Viewing conditions \times Target size: $F(2, 22) = 3.052, p > 0.05$; Viewing conditions \times Target depth \times Target size: $F(4, 44) = 1.036, p > 0.25$).

In contrast, the subjects' judgments in the CUS condition (solid dots in Fig. 5a) deviated considerably from the actual target depth. All three target depths were underestimated. When compared with baseline data obtained from direct vision, the ANOVA revealed a significant main effect for CUS display ($F(1, 11) = 15.482, p < 0.005$). Note that such an underestimation might be attributed to the subject's inaccurate knowledge about the centimeter scale that was displayed on the screen. We assessed this contribution from the lines drawn by the subjects, one 1 cm long and one 1 inch long. The inset figure in the lower left of Fig. 5a shows the averaged length of drawn lines. Both are shorter than the specified lengths (0.89 cm and 2.18 cm), but this inaccuracy alone is too small to account for the total underestimation that was observed. Note too that subjects were shown a correct scale while responding, which should have reduced error attributable to a subjective metric. In the discussion section, other possible causes for depth underestimation will be addressed.

Fig. 5b shows the perceived x-y coordinates of targets in three viewing conditions. Compared to the two ultrasound conditions, subjects showed a small bias in sagittal distance (y) perception of the target with direct vision (main effect: $F(2, 22) = 7.356, p < 0.005$, Viewing conditions \times Sagittal distance: $F(4, 44) = 9.624, p < 0.001$). No significant difference was observed between the SF and CUS conditions (x: main effect: $F(1, 11) = 0.650, p > 0.80$, Device \times Location: $F(2, 22) = 2.613, p > 0.05$; y: main effect: $F(1, 11) = 1.553, p > 0.20$, Device \times Location: $F(2, 22) = 1.029, p > 0.35$). Note that in the experiment subjects usually placed the probe directly over the target to get the best image. Subjects thus may have been aided in x-y localization by using the probe as the reference in both ultrasound conditions.

5 EXPERIMENT 2: ROLE OF BINOCULAR DEPTH CUES IN THE SF

The purpose of this experiment was to demonstrate the role of binocular depth cues in perceptual localization of a target using the SF. Subjects were tested with one eye covered. Their judgments were compared to judgments made in the binocular SF condition of Experiment 1.

5.1 Method

The subjects and procedure were identical to the SF condition of Experiment 1, except that subjects wore an eye patch to cover their nondominant eye, constituting a monocular viewing condition that could be compared to the binocular viewing condition of Experiment 1. This condition was run at the same time as Experiment 1, in counterbalanced order with the conditions of that experiment.

5.2 Results

The solid dots in Fig. 6a show the subjects' depth estimation in the monocular viewing condition. Other symbols, represent-

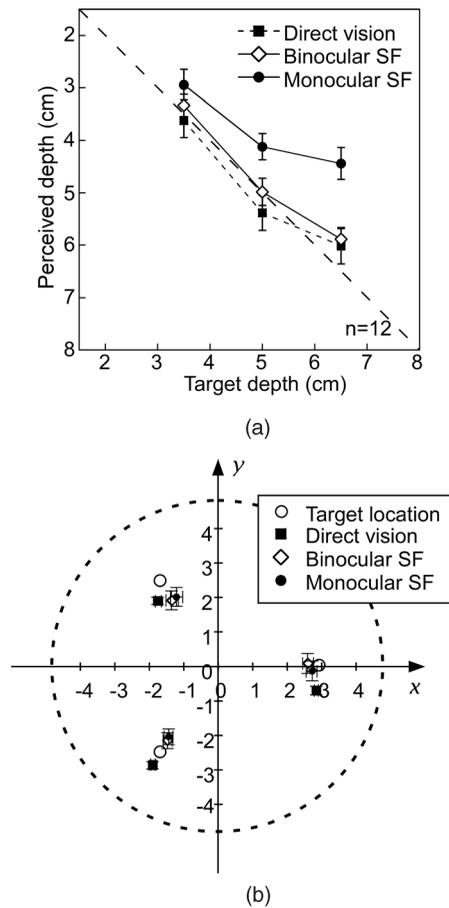


Fig. 6. (a) Mean judged depth as a function of the target depth in different viewing conditions (direct-vision, binocular SF, and monocular SF). (b) Mean judged locations in the x-y position in each condition, along with target location. Error bars represent ± 1 standard error.

ing Direct Vision versus SF, are as in Fig. 5. Without binocular cues, subjects tended to underestimate target depth. Moreover, the underestimation increased with the target depth. Subjects' perception was relatively accurate (2.94 cm) for a shallow depth of 3.5 cm, while the deepest target depth (6.5 cm) was underestimated by almost 30 percent (4.44 cm). Thus, depth in the SF image was perceptually compressed in the monocular viewing condition: the mean perceived depth increased only 1.5 cm despite the increase of target depth by 3.0 cm. A three-way repeated measure ANOVA examined the

subjects' judgments in the binocular and monocular conditions. The main effect of binocular cues was found significant ($F(1, 11) = 43.589, p < 0.001$), as well as the interaction between viewing conditions and target depth ($F(2, 22) = 10.656, p < 0.001$).

In contrast to depth, monocular perception of x-y position (Fig. 6b) showed little systematic error (x: main effect: $F(1, 11) = 0.485, p > 0.50$, Device \times Location: $F(2, 22) = 0.907, p > 0.90$; y: main effect: $F(1, 11) = 0.001, p > 0.90$, Device \times Location: $F(2, 22) = 1.527, p > 0.235$). As in the previous experiment, this might result from the subjects' utilization of the ultrasound probe as the reference in x-y localization.

Fig. 7 shows stereo pairs which the reader may use to simulate the experience of viewing a target with the SF.

6 EXPERIMENT 3: IMAGE-GUIDED NEEDLE INSERTION WITH THE SF AND CUS

In this experiment, subjects simulated clinical needle insertion under conditions similar to Experiment 1. The purpose was to compare the needle trajectory to the prediction from the perceived locations as determined in Experiment 1. In particular, as the depth of the target was underestimated with CUS, we predicted that subjects using CUS should initially aim the needle too high and only later, if at all, correct. In contrast, subjects using the SF could aim directly at the target and show a straighter trajectory.

6.1 Method

6.1.1 Subjects

Ten subjects, two females and eight males from Experiment 1, participated in this experiment.

6.1.2 Stimuli

Phantoms used in this experiment were similar to those in Fig. 4a, except that each phantom contained only one bead (1.0 cm in diameter, 5.0 cm in depth) and pen holders were replaced by needle entry points around the circumference of the phantom. The entry points were positioned with a radius of 5 cm and, thus, instituted the insertion path often used clinically with an elevation angle of 45° . Seven phantoms were created: four were used as the stimuli, in which the target depth was 5 cm; the other three were dummy phantoms with random target depths. Stimulus and dummy phantoms were tested alternately to create an unpredictable stimulus environment. A practice phantom was used to familiarize the subject with the experimental procedure.

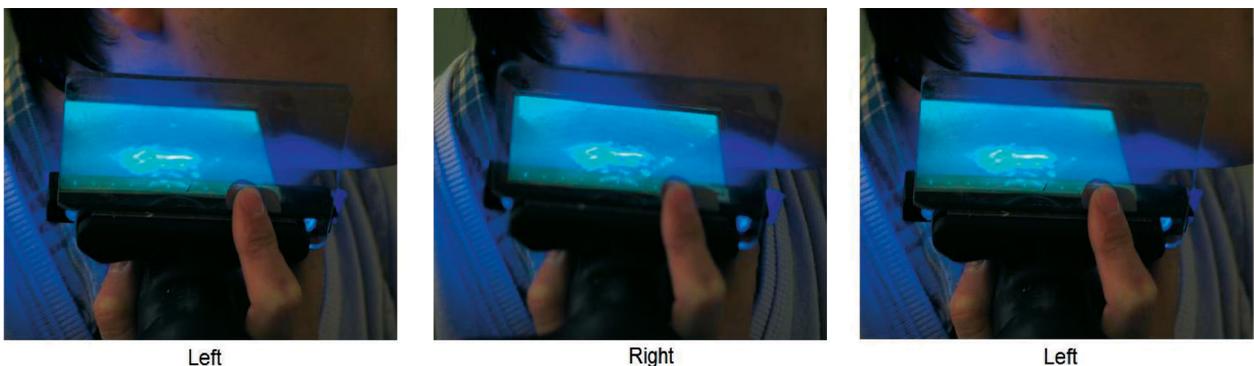


Fig. 7. Stereo image pairs to illustrate the role of binocular depth cues in perceptual localization of a target using the SF. Uncrossed fusion of the left pair or crossed fusion of the right pair shows an ultrasound slice inside the neck of a human containing the jugular vein and carotid artery.

6.1.3 Design and Procedure

Each subject participated in 24 trials, representing the crossing of three factors: visualization devices (CUS or SF), entry points (3), and replications (2). Trials were blocked by different devices; the starting device and the testing order of phantoms were counterbalanced across subjects.

Subjects performed insertions binocularly, using a dull needle mounted with a miniBIRD tracker (Ascension Technology Co.). The procedure simulated needle placements as usually done in clinical practice, keeping the needle in the plane of the ultrasound scan. At the beginning of each trial, the subject had 30 seconds to find the target, judge its location, and plan the insertion. The subject was instructed that once ready, he or she should insert the needle as quickly as possible while maintaining as high a level of accuracy as possible. So that the subject could track the tip of the needle, he or she was told to position the needle in plane with the ultrasound image during the insertion. The subject was also told not to sweep the needle to search blindly for the target, given that such operation is impractical in real clinical settings. The subject was instructed that once the target was reached, he or she should push the bead a little to prove that the needle had really touched it. If the target was not reachable along the given trajectory, the subject moved on to the next trial. The insertion trajectory was recorded by the miniBIRD with a sampling rate of 103.3Hz; the time that the needle first became visible in the ultrasound image was also recorded by the experimenter.

6.2 Results

Subjects performed insertions faster (12.5 s versus 16.8 s, $t(9) = 3.403, p < 0.005$) and with a higher success rate (88.9 percent versus 72.0 percent, $t(9) = 6.760, p < 0.001$) when using the SF rather than CUS. In a postexperiment interview, they uniformly indicated that compared to CUS, the SF seemed to make the task much easier.

For analysis, the subject's response on each trial was broken into three phases: 1) the preinsertion planning, 2) insertion before the needle entered the ultrasound slice, and 3) insertion with visual guidance from ultrasound. In the first phase, the subject decided how to insert the needle. The accuracy of planning was assessed in terms of the needle's initial orientations (azimuth and elevation) at the beginning of each insertion. As to initial azimuth, no significant errors were found in both conditions (CUS: $-1.2^\circ \pm 1.1^\circ, t(9) = 1.149, p > 0.25$; SF: $-1.1^\circ \pm 1.5^\circ, t(9) = 0.720, p > 0.45$). As in the previous experiments, this might be because subjects judged the target's x-y position by referring to the ultrasound probe over it. However, different patterns were evident in the initial elevation for SF and CUS insertions (Fig. 8). With *in-situ* visualization of the target in the SF, subjects could aim the needle at the target with reasonable accuracy ($46.1^\circ \pm 1.3^\circ$). No systematic deviation from the correct elevation (45°) was found ($t(9) = 0.891, p > 0.35$). On the contrary, underestimation in elevation was significant in the CUS condition ($39.4^\circ \pm 1.2^\circ, t(9) = 4.535, p < 0.005$). Consistent with the previous results, subjects tended to point the needle to a location shallower than the target when using CUS. Quantitatively, depth underestimation was comparable to that in Experiment 1. A projection line of the averaged elevation (the dashed arrows in Fig. 8; the gray area depicts ± 1 SE) deviated toward the perceived target locations measured in Experi-

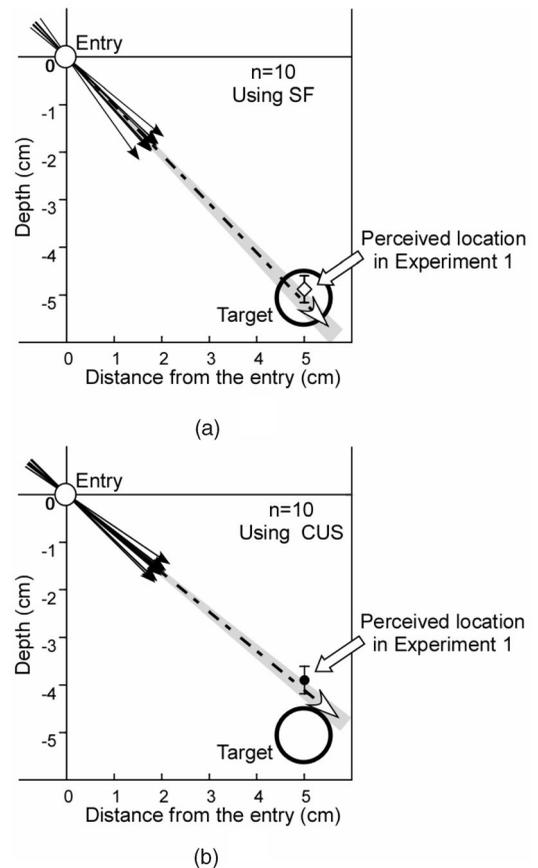
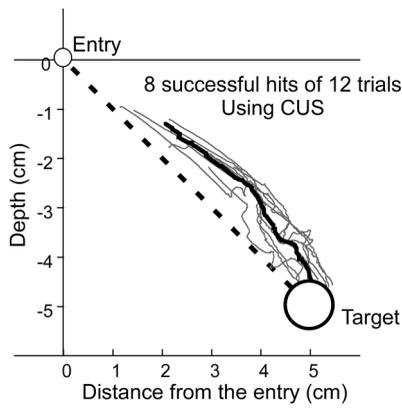


Fig. 8. The needle's initial elevation as positioned by subjects in different conditions. Solid arrows represent individual subjects' mean responses. Dashed arrows show the overall average across all subjects and gray areas represent ± 1 SE.

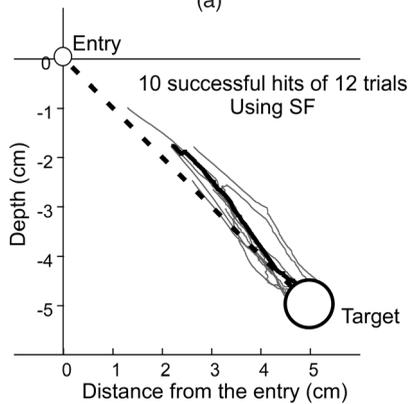
ment 1 (the open diamond in Fig. 8a and the solid dot in Fig. 8b). The results thus provide evidence that needle insertion was initially directed by perceptual localization of the target.

Fig. 9 shows trajectories of all successful insertions performed by one naïve subject. The subject inserted the needle along the initial direction until it entered the ultrasound image. On CUS trials, trajectories deviated gradually from the correct path toward a location higher than the target due to initial errors in elevation. Deviations of SF trajectories, on the other hand, were much smaller. When the maximum deviation from the correct path was used as a descriptive statistic for this trend, a significant difference between the SF and CUS conditions was found in this subject's responses ($t(16) = 2.588, p < 0.02$) as well as in the paired comparison of all subjects' responses (using the average within-subject trajectory, $t(9) = 3.944, p < 0.005$).

After the needle appeared in the ultrasound image, the subject could adjust the insertion by seeing the relative position of the needle's tip to the target. An arc-shaped pattern was usually shown with CUS by the subject correcting the underinsertion of the needle. Subjects still tended to undercompensate: the needle reached the top of the target, although the instruction was to insert the needle into the target's center. Such adjustment was generally small on SF trials, because the needle was usually inserted in the correct direction. As shown in Fig. 9 and Fig. 10, the needle went directly to the target in the SF condition and



(a)



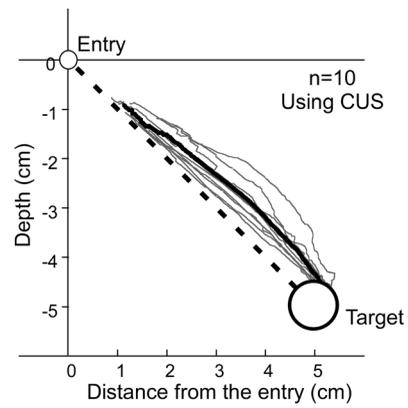
(b)

Fig. 9. Trajectories of all successful insertions performed by a naïve subject. Bold black lines depict his mean responses.

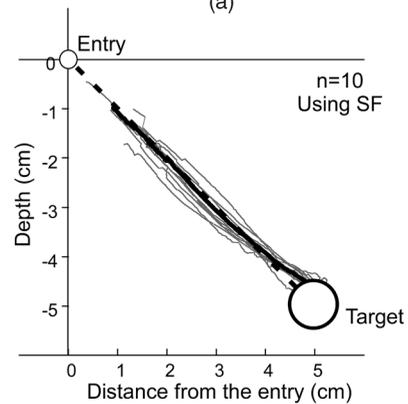
traveled a shorter distance than in the CUS condition ($t(9) = 2.511, p < 0.05$). Obviously, smaller adjustments and shorter insertion paths establish the superiority of the SF over CUS because such adjustments are difficult or impossible in real medical applications.

Several temporal measures suggest that the SF enhanced users' confidence about their performance relative to CUS. When inserting the needle, the subjects made fewer pauses (movements < 0.4 mm for at least 500 ms) in the SF condition (2.45 pauses per insertion) than the CUS condition (4.34 pauses per insertion) ($t(9) = 3.093, p < 0.02$). They also inserted the needle faster when using the SF than with CUS; however, the speed difference was significant only during the phase before the needle entered the ultrasound slice (0.80 cm/s versus 0.59 cm/s, respectively, $t(9) = 2.597, p < 0.03$). After the needle appeared in the ultrasound slice, the subjects guided it to the target with similar speeds in both conditions (0.52 cm/s for SF versus 0.44 cm/s for CUS, $t(9) = 1.639, p > 0.1$).

In addition, the intersubject variability was smaller in the SF condition than in CUS condition ($F(9,9) = 4.272, p < 0.025$). As shown in Fig. 10, considerable intersubject variation was observed in CUS insertions. Errors in the initial insertion were noticed by subjects at different points. Some subjects made no corrections until the needle had wandered far off the correct path. In contrast, the averaged individual subject's insertion with the SF was similar across all subjects, and the needle tended to go directly to the target.



(a)



(b)

Fig. 10. Averaged insertion trajectories for individual subjects using different visualization devices. Bold black lines depict the mean responses across all subjects.

7 DISCUSSION

7.1 Comparison of the SF and CUS with Respect to Localization

The present experiments provide a clear demonstration of the superiority of the SF over CUS with respect to localization of targets and perceptually guided needle insertion. Perception of depth with the binocular SF was comparable to direct vision of the target in the empty tank. With CUS, we found a systematic error in localization, such that targets were perceived closer to the surface of the phantom than their true depth.

Note that the SF and CUS presented the identical data from the same machine. Inferior performance with CUS thus results from a lack of effective visualization of the data, not the data themselves. This could stem from multiple sources. As mentioned in Section 2.1, one possibility is incorrect scaling from the centimeter scale that was displayed on the CUS screen. We assessed this contribution by asking subjects to draw lines of 1 cm and 1 inch, before the experimental trials. As can be seen from Fig. 5a, both drawings were shorter than the specified length by about 10 percent, and this could account for part of the underestimation of depth that was observed with CUS. It is likely not the only cause of underestimation, especially given that subjects could view a 1-cm standard throughout the experimental trials.

Another possible cause for underestimation of depth could be that, to the extent to which the ultrasound probe indents the deformable surface, the depth is not accurately

registered by users of CUS. Ideally, the user should have adjusted the depth reading to the target from the top of the phantom by adding the amount of deformation. Insufficient or total lack of adjustment might contribute to the underestimation of depth that was observed with CUS. In comparison, this information is automatically registered by the SF. As shown in Fig. 1, the relative geometry of the transducer, the display, and the half-silvered mirror is fixed in the SF. An indentation of the transducer thus shifts the mirror downward by the same amount and exactly counterbalances spatial changes in the ultrasound image. As a result, when seen through the mirror, the virtual image of the target remains unchanged in 3D space irrespective of indentation of the phantom (or patient) surface. This offers the SF superiority over CUS with respect to perceptual localization of targets.

7.2 Role of Binocular Cues

The monocular version of the SF also led to underestimation of depth. We conclude on this basis that the binocular cues provided by the SF contribute substantially to its superiority.

The image formation in the SF provides the observer with binocular depth information to accurately localize the virtual ultrasound slice. As shown in Fig. 1, the SF's mirror creates a virtual image from the real display: rays of light emanating from the display reflect from the half-silvered mirror to form a virtual slice behind it and, hence, appear to have originated from the location of the virtual image. Under binocular viewing, this presents two important cues for distance. The visual system can localize the virtual ultrasound slice in 3D space in terms of the apparent convergence of rays from it. Binocular disparity, the difference between the two retinal images, of the virtual image can also be used to judge relative depth within it. Giving their significance in depth perception and actions in near space (within arm's reach) [14], these cues may account for the subjects' accurate SF performance, which was comparable to the direct vision condition in Experiment 1.

The monocular viewing of the SF eliminated the aforementioned binocular cues and, hence, depth could be judged only by using less effective cues such as pictorial and motion cues and accommodation. As a result of attenuation of depth information, a systematic compression of perceived depth was found in Experiment 2, a pattern different from that in the binocular SF or CUS conditions. This phenomenon also has been found in the perception of large-scale space accessible by walking [20], [21]. An explanation might be the so-called "equidistance tendency" [22]: As information for distance is reduced, some automatic tendencies that underestimate separation-in-depth cause stimuli to tend to appear in similar depth. In addition, pictorial cues in the ultrasound image or motion cues by subjects' changing their viewpoint still provide some approximate depth information. Altogether, the subjects' perceptual space in the monocular viewing of the SF is compressed, but not totally flat in depth, as shown in Fig. 6a.

7.3 Relation of Perceived Target Location to Needle Insertion

When the subjects attempted to insert the needle into a particular location with ultrasound guidance, subjects showed little systematic error on average with the binocular SF. The averaged trajectory (over trials and subjects) went directly to the target. There was relatively little intersubject

variability as well. With CUS, in contrast, the trajectories tended to show arc-shaped deviation from the ideal path, initially starting toward a location higher than the target and then adjusting later in the trajectory.

Consistent with previous research dealing with actions in small and large space [23], [24], [25], the above results further demonstrate the perception/action linkage in very near space in the applied context of ultrasound-guided manipulation. The perceived target location has critical impact on planning and conducting needle insertion, and systematic errors in insertion could be linked to the accuracy of the subject's perceptual localization of the target. On CUS insertions, the insertion usually started in a direction that aimed at a location shallower than the target. More precisely, the extent of depth underestimation was comparable to that found in perceptual judgment of target location (Experiment 1), as shown in Fig. 8. Likewise, initial insertions with the SF also aimed at the perceived target location, which in this case was more accurate. Clearly, the simplest interpretation is that subjects consistently inserted the needle into the initially perceived target location. The perception-guided insertion continued until the needle entered the ultrasound image. Then, the need for adjustment was indicated to the subject by the deviation of the needle from the correct path perceived on the ultrasound image.

8 CONCLUSIONS

To summarize, the present paper describes a RTTR device that promises to have considerable clinical value, as it affords localization of a target from an ultrasound image with greater accuracy and reduced variability relative to conventional ultrasound. We have demonstrated that CUS and the SF lead to different percepts and accordingly to actions with different efficacy. With CUS, the arc-shaped trajectory and large adjustment resulting from the underestimation of target depth could not be achieved without tissue damage in a real application; it would be necessary to withdraw the needle and reinsert it. In contrast, insertions with the SF generally went directly to the target along the desired path. This lends great promise to the application of the SF in clinical practice for improving safety and minimizing patient discomfort. An ongoing clinical study is examining the performance of skilled operators with more than 10 years' experience with ultrasound, in using the SF to insert catheters into the deep veins of the upper arm.

ACKNOWLEDGMENTS

This work is supported by grants from NIH (#R01-EB00860) and the US National Science Foundation (0308096). The authors thank Dr. Andrei State for allowing them to use his artwork as the template for illustrations in Fig. 2. They also thank the three anonymous reviewers for their very helpful and thoughtful comments on their manuscript.

REFERENCES

- [1] S.P. Keenan, "Use of Ultrasound to Place Central Lines," *J. Critical Care*, vol. 17, no. 2, pp. 126-137, 2002.

- [2] K. Knowlton, "Computer Displays Optically Superimposed on Input Devices," *The Bell System Technical J.*, vol. 56, no. 3, pp. 367-383, Mar. 1977.
- [3] C. Schmandt, "Spatial Input/Display Correspondence in a Stereoscopic Computer Graphic Work Station," *Proc. ACM SIGGRAPH*, pp. 253-261, July 1983.
- [4] A. State, M. Livingston, G. Hirota, W. Garrett, M. Whitton, H. Fuchs, and E. Pisano, "Technologies for Augmented-Reality Systems: Realizing Ultrasound-Guided Biopsies," *Computer Graphics*, pp. 439-446, 1996.
- [5] M.R. Rosenthal, A. State, J. Lee, G. Hirota, J. Ackerman, K. Keller, E.D. Pisano, M.R. Jirutek, K.E. Muller, and H. Fuchs, "Augmented Reality Guidance for Needle Biopsies: An Initial Randomized, Controlled Trial in Phantoms," *Medical Image Analysis*, vol. 6, pp. 313-320, 2002.
- [6] F. Sauer, A. Khamene, B. Bascle, L. Schimmang, F. Wenzel, and S. Vogt, "Augmented Reality Visualization of Ultrasound Images: System Description, Calibration, and Features," *Proc. Int'l Symp. Augmented Reality, IEEE and ACM*, pp. 30-39, 2001.
- [7] G. Stetten, V. Chib, and R. Tamburo, "System for Location-Merging Ultrasound Images with Human Vision," *IEEE Proc. Applied Imagery Pattern Recognition (AIPR) Workshop*, pp. 200-205, 2000.
- [8] W. Chang, G. Stetten, L. Lobes, D. Shelton, and R. Tamburo, "Guidance of Retrobulbar Injection with Real Time Tomographic Reflection," *J. Ultrasound in Medicine*, vol. 21, pp. 1131-1135, 2002.
- [9] G. Stetten, "System and Method for Location-Merging of Real-Time Tomographic Slice Images with Human Vision," US Patent no. 6,599,247, July 2003.
- [10] K. Masamune, G. Fichtinger, A. Deguet, D. Matsuka, and R.H. Taylor, "An Image Overlay System with Enhanced Reality for Percutaneous Therapy Performed Inside CT Scanner," *Proc. Fifth Int'l Conf. Medical Image Computing and Computer-Assisted Intervention*, pp. 77-84, 2002.
- [11] J.D. Mulder and R. van Liere, "The Personal Space Station: Bringing Interaction Within Reach," *Proc. Virtual Reality Int'l Conf.*, June 2002.
- [12] "The ReachIn Display," <http://www.est-kl.com/hardware/haptic/reachin/display.html>, 2005.
- [13] H. Fuchs, M.A. Livingston, R. Raskar, D. Colucci, K. Keller, A. State, J.R. Crawford, P. Rademacher, S.H. Drake, and A.A. Meyer, "Augmented Reality Visualization for Laparoscopic Surgery," *Proc. First Int'l Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI '98)*, Oct. 1998.
- [14] J.E. Cutting and P.M. Vishton, "Perceiving Layout and Knowing Distance: The Integration, Relative Potency and Contextual Use of Different Information about Depth," *Perception of Space and Motion*, W. Epstein and S. Rogers, eds., New York: Academic Press, pp. 69-117, 1995.
- [15] F.A. Voorhorst, C.J. Overbeeke, and G.J. Smets, "Spatial Perception during Laparoscopy: Implementing Action-Perception Coupling," *Studies in Health Technology and Informatics*, vol. 39, pp. 379-386, 1997.
- [16] Y.E. Cohen and R.A. Andersen, "Multisensory Representations of Space in the Posterior Parietal Cortex," *The Handbook of Multisensory Processes*, G.A. Calvert, C. Spence, and B.E. Stein, eds., pp. 463-479, Cambridge, Mass.: MIT Press, 2004.
- [17] M.J. Sholl and T.L. Nolin, "Orientation Specificity in Representations of Place," *J. Experimental Psychology: Human Learning, Memory and Cognition*, vol. 23, pp. 1494-1507, 1997.
- [18] J. Ghayeb and R. Daniels, "Status Review of Field Emission Displays," *Proc. SPIE*, vol. 4362, pp. 319-330, 2001.
- [19] D. Sunday, "Distance between Lines and Segments with Their Closest Point of Approach," http://geometryalgorithms.com/Archive/algorithm_0106/, 2004.
- [20] T.L. Ooi, B. Wu, and Z.J. He, "Distance Determined by the Angular Declination below the Horizon," *Nature*, vol. 414, pp. 197-200, 2001.
- [21] J.W. Philbeck and J.M. Loomis, "Comparison of Two Indicators of Perceived Egocentric Distance under Full-Cue and Reduced-Cue Conditions," *J. Experimental Psychology: Human Perception and Performance*, vol. 23, pp. 72-85, 1997.
- [22] W.C. Gogel, "The Tendency to See Objects as Equidistant and Its Inverse Relation to Lateral Separation," *Psychological Monographs*, vol. 70, pp. 1-17, 1956.
- [23] J.W. Philbeck, J.M. Loomis, and A.C. Beall, "Visually Perceived Location is an Invariant in the Control of Action," *Perception and Psychophysics*, vol. 59, pp. 601-612, 1997.
- [24] R.L. Klatzky, Y. Lipka, J.M. Loomis, and R.G. Golledge, "Encoding, Learning and Spatial Updating of Multiple Object Locations Specified by 3-D Sound, Spatial Language, and Vision," *Experimental Brain Research*, vol. 149, pp. 48-61, 2003.
- [25] J. Fleming, R.L. Klatzky, and M. Behrmann, "Time Course of Planning for Object and Action Parameters in Visually Guided Manipulation," *Visual Cognition*, vol. 9, pp. 502-527, 2002.



Bing Wu, PhD, received the MS degree in neurobiology in 1997 from the Shanghai Institute of Physiology, Chinese Academy of Sciences, the MS degree in computer science in 2002, and the PhD degree in experimental psychology in 2004 from the University of Louisville. He is currently a postdoctoral fellow in the Robotics Institute and Department of Psychology at Carnegie Mellon University. His research focuses primarily on the perception of depth in the real, augmented, and virtual environments. He is the author of 15 journal and conference publications.



Roberta L. Klatzky, PhD, received the BS degree in mathematics from the University of Michigan and the PhD degree in experimental psychology from Stanford University. She is a professor of psychology at Carnegie Mellon University, where she is also on the faculty of the Center for the Neural Basis of Cognition and the Human-Computer Interaction Institute. Before coming to Carnegie Mellon, she was a member of the faculty at the University of California, Santa Barbara. Professor Klatzky is the author of more than 200 articles and chapters, and she has authored or edited four books.



Damion Shelton received the BS degree in bioengineering, the BA degree in history and philosophy of science in 2001 from the University of Pittsburgh, and the MS degree in robotics from Carnegie Mellon University in 2004. He is presently pursuing the PhD degree in robotics, also from Carnegie Mellon, with a graduate fellowship from the Whitaker Foundation. He has been the author or coauthor of nine journal or conference papers and is interested in computer assisted surgery and augmented reality applications to medicine.



George D. Stetten, MD, PhD, received the MS degree in biology from New York University, the MD degree from the State University of New York at Syracuse, and the PhD degree in biomedical engineering from the University of North Carolina at Chapel Hill. He was a research assistant professor of biomedical engineering at Duke University and is currently an associate professor of bioengineering at the University of Pittsburgh and a research scientist in robotics at Carnegie Mellon University. He is the author of 42 peer reviewed journal and conference publications, and is currently principal investigator on two NIH and one US National Science Foundation grants.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.