Winyu Chinthammit
winyu@hitl.washington.edu

Eric J. Seibel
eseibel@hitl.washington.edu

Thomas A. Furness, III
tfurness@hitl.washington.edu
Human Interface Technology Lab
215 Fluke Hall, Box 352142
University of Washington
Seattle, WA 98195

A Shared-Aperture Tracking Display for Augmented Reality

Abstract

The operation and performance of a six degree-of-freedom (DOF) shared-aperture tracking system with image overlay is described. This unique tracking technology shares the same aperture or scanned optical beam with the visual display, virtual retinal display (VRD). This display technology provides high brightness in an AR helmet-mounted display, especially in the extreme environment of a military cockpit. The VRD generates an image by optically scanning visible light directly to the viewer's eye. By scanning both visible and infrared light, the head-worn display can be directly coupled to a head-tracking system. As a result, the proposed tracking system requires minimal calibration between the user's viewpoint and the tracker's viewpoint. This paper demonstrates that the proposed shared-aperture tracking system produces high accuracy and computational efficiency. The current proof-of-concept system has a precision of +/− 0.05 and +/− 0.01 deg. in the horizontal and vertical axes, respectively. The static registration error was measured to be 0.08 +/− 0.04 and 0.03 +/− 0.02 deg. for the horizontal and vertical axes, respectively. The dynamic registration error or the system latency was measured to be within 16.67 ms, equivalent to our display refresh rate of 60 Hz. In all testing, the VRD was fixed and the calibrated motion of a robot arm was tracked. By moving the robot arm within a restricted volume, this real-time shared-aperture method of tracking was extended to six-DOF measurements. Future AR applications of our shared-aperture tracking and display system will be highly accurate head tracking when the VRD is helmet mounted and worn within an enclosed space, such as an aircraft cockpit.

1 Introduction

In immersive virtual reality (VR) applications, the computer renders all objects in the visual scene. In this case, the precise registration of the virtual images relative to the real world is not important. However, in augmented reality (AR) applications, the computer renders images or graphics information that augment or overlay a real scene, both of which are observed simultaneously by the user. In this regard, precise alignment of the virtual image relative to real objects is essential. The goal of AR is to enhance human interaction with the real world using computer-generated information that is related to the real-world objects. For example, bioengineering researchers have investigated using AR for image guidance during needle biopsy (Stetten, Chib, Hildebrand, & Bursee, 2001). The doctor can look directly at the patient while the AR system projects an ultrasound image over the real one, allowing for simultaneous examination of an ultrasound image and real image during

Presence, *Vol. 12, No. 1, February 2003,* 1–18

the operation. Another augmented reality application is realized in fighter cockpit head-up displays wherein information related to the aircraft systems (such as navigational waypoints) are superimposed and registered through a see-through display over the real-world scene as viewed through the display.

To superimpose relevant information over the real world, AR applications may use a display that is mounted on an object in the real world (such as an automobile dashboard or windshield), incorporated into a handheld object (display wand), or mounted on the head/helmet worn by the user. Although these displays may present state or graphic information, which is independent of where the virtual images appear in the real world, most AR applications require a precise stabilization of these images for superimposition relative to the real world. In those applications involving head-mounted displays, the performance of the head-tracking system is a critical part of the image stabilization process. It determines the user's viewpoint so that the augmenting virtual image is always aligned relative to the real object.

The main function of the head-tracking system is to signal to the computer graphics generator the instantaneous position and orientation of the head/helmet display scene so that the computer system can generate and position the graphics within that display field of view (FOV) such that the objects appear stabilized in space. Several tracking technologies are widely implemented in the AR community (such as fiducial or video tracking, optical tracking, and inertial tracking). Each approach has its own advantages and disadvantages. For example, within a constrained environment, the placement of fiducials within the physical environment has produced single-pixel tracking accuracy in several AR applications (Azuma, 1997). However, the fiducial location in world coordinates must be known and high-accuracy tracking depends on continuously maintaining a fiducial within the instantaneous FOV of the tracking cameras. Target tracking with a fiducial system may also require significant time to process the appropriate algorithms (You & Neumann, 2001).

To simplify the computational requirements for image processing, optical beam scanning has been used to track head orientation using synchronized detection in the time domain (Rolland, Davis, & Baillot, 2001). An additional feature of optical beam scanning is the higher bandwidth of the individual optical detectors sensing target motion (Sorensen, Donath, Yang, & Starr, 1989). Thus, landmark-tracking problems associated with motion blur due to the slow 30–60 Hz video framerates can be avoided (Yokokohji, Sugawara, & Yoskikawa, 2000). But the setup time for optical beam scanning system can be extensive because such a system requires major installation within the environment where the AR system will be used. Another tracking technology is inertial tracking. Inertial tracking technologies are more self-contained than fiducial tracking and optical tracking systems, and therefore the setup time is minimal. However, a major challenge in these systems is the tendency to drift resulting in tracker errors in superimposing the virtual images relative to real-world objects (You, Neumann, & Azuma, 1999).

For AR to be effective, the objects in the real and virtual environments must be properly aligned. This alignment process is called *registration,* and it is one of the fundamental problems currently limiting AR applications. For example, with inaccuracies in a head-tracking system, users may perceive movement of a virtual cup on a real table while they walk around looking at the cup relative to the table. This greatly compromises a sense of reality that the user perceives, and it compromises the utility of AR.

The sources of the registration error can be divided into two types: static and dynamic. Intuitively, static and dynamic errors occur when the user is stationary and when the user is in motion, respectively. The sources of static errors are calibration error and tracker error. Because a typical AR system consists of tracking and display subsystems, calibration procedures are required to align the viewpoint between the two. A calibration error means that the coordinate system of the tracker is not aligned with the real world. The primary sources of dynamic errors are system latency and an invalid simultaneity assumption of measurements. The computational time and camera/display framerates are often the largest contributors to the system latency.

Creating a sense of reality requires not only an accu-

rate registration but also a high quality of the display image. The luminance of augmented images must be sufficiently bright to be seen relative to the ambient light from the real environment. In a see-through mode, current AR display devices such as transmission mode liquid crystal displays have insufficient luminance to compete with light from the real environment. As a result, a virtual image that is superimposed onto the real world often appears to be ghostlike or low contrast. Such a problem can be significant for military cockpit applications wherein the pilot can encounter an extreme range of ambient lighting conditions. Next, we describe a novel display that was invented in 1991 at the Human Interface Technology Laboratory at the University of Washington.

This novel display provides a head-mounted display of sufficiently high luminance, resolution, and good form factor (weight and size) for augmented reality applications. The display device is called the *virtual retinal display* (VRD). The VRD scans laser light directly to the retina of the eye. The VRD consists of photon generators (for red, green, and blue light), light modulator, fiber optics coupler, and a two-axis mechanical scanner, which uses two oscillating mirrors to deflect a beam of modulated laser light. The beam is scanned through a combiner/beamsplitter and other optical elements to form a real image on the retina of the viewer's eye. The light beam is intensity-modulated while being scanned to render a two-dimensional image on the retina. The VRD is then perceived as a virtual image hovering in space. Even though only one pixel from the raster scan is being illuminated on the retina, the viewer sees a stable image if the framerate exceeds the critical flicker frequency of 55–60 Hz (Kelly et al., 2001). The luminance of the VRD is capable of competing with the light from the real environment; thus, we consider the VRD as an excellent alternative display device in AR applications. For a discussion on the past and current engineering developments of the VRD technology, see Johnston and Willey (1995) and www.mvis.com.

In a conventional VRD configuration, the beamsplitter reflects some energy (approximately 40%) of visible light into the environment while the remaining light is forming a visible image on the user's retina; therefore, the beam is being scanned simultaneously to the eye and into the environment. If light-sensing detectors are placed into the real-world environment where the VRD external beam is being scanned and, given the known raster-scanning pattern of the VRD, then the orientation and position of the VRD (or the user's head) can be detected at the same time an image is being projected on the retina. This artifact of the VRD suggests that it is feasible to incorporate a head-tracking functionality into the VRD. In doing so, an accurate registration and a high quality of the display image can be accomplished as we discuss later.

We have constructed a working prototype of such a head-tracking system that is directly coupled with the retinal display using the same optical pathway. As a result, the proposed system requires minimal alignment between the display coordinate system and the tracker coordinate system. This significantly reduces calibration error between the two. Furthermore, the required computational overhead can be low due to its hardware-based configuration (rather than the software-based configuration of computer vision tracking systems). This can reduce the overall system latency. Another advantage of this approach is that tracking can be accomplished with high precision because precision is limited only by the speed and accuracy of the data acquisition system and not by the scanning resolution of the VRD. Unlike computer vision techniques, the z axis distance (that is, that distance between the display/tracker to the sensing mechanism) does not affect precision.

This paper describes the operation and performance of a six-DOF shared-aperture tracking system with image overlay. We have titled our system the *shared aperture retinal scanning display and tracking system* (SARSDT). In the ensuing sections, we detail the operation of the SARSDT and assess target-tracking performance (statically and dynamically) with image overlay over real-world objects. The SARSDT is unique because the shared-aperture permits coupling of high-quality virtual information with real images and simultaneously tracks the location of the virtual images within the real-world scene. Because the optical tracker and display share the same aperture, highly accurate and efficient tracking overlay methods can be applied in AR applica-

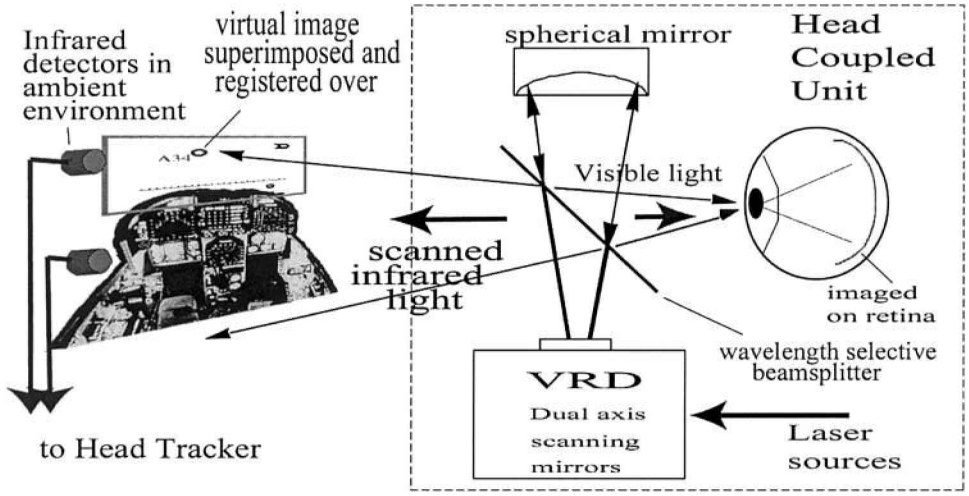

Figure 1. *Simplified operation of the shared-aperture display/tracking system.*

tions (Kutulakos & Vallino, 1998; Kato & Billinghurst, 1999). Furthermore, when a single point within the FOV is tracked across the 2D projection of the 3D scene, simplified image-plane interaction techniques for AR become feasible (Pierce et al., 1997). This shared-aperture display and tracking system can therefore alleviate shortcomings of current head-tracking systems as discussed previously.

2 System Concept

We first describe a fundamental concept of our proposed shared-aperture retinal scanning display and tracking system. It illustrates how the VRD optical path can be used simultaneously as an optical tracking system. Then, several detection techniques are described. It explains what techniques we have implemented to detect an optical scanning beam. Lastly, we explain the tracking principles of our SARSDT system.

2.1 Shared Aperture Retinal Scanning Display and Tracking System

Simply stated, the SARSDT is a VRD that has been modified to provide a tracking function by sharing coaxially the optical path of the VRD. (Figure 1 shows a simplified diagram of the system.) Visible and IR lasers generate a multispectral beam of infrared and visible light. The visible beam is modulated by the video signal (as in the normal VRD), and the IR beam is unmodulated. The IR and visual beams are combined before being manipulated by two mirrors, which scan the beams in a raster pattern. A high-speed resonant scanner scans the beams back and forth in the horizontal axis while a galvanometer mirror scans the beams in the vertical axis. (Note that, in the case of the VRD, the display is generated by active raster lines that are scanned in both directions, left to right and right to left as opposed to normal NTSC raster lines, which are only scanned in a left-to-right direction.) To draw the raster, the horizontal scanner moves the beam to draw a row of pixels at approximately 15.75 kHz, then, the vertical scanner moves the beam at VGA standards to the next line where another row of pixels is drawn with all lines refreshed at 60 Hz.

The scanned beam is then passed through a wavelength selective combiner/beamsplitter that separates the visible and IR components of the beams. The visible portion of the beam passes through the combiner and is reflected by a spherical converging mirror to form an exit pupil through which the image is transferred

through the entrance pupil of the eye to the retina. The IR portion of the beam is reflected in the opposite direction from the beamsplitter resulting in a raster of IR light being scanned within the same field of view as the display image.

Because the VRD is mounted on headgear, the outgoing tracking scanning pattern is aligned with the display scanning pattern. Given the known raster scanning pattern of both the IR and visible beams, we can determine the instantaneous orientation and position of the display scene using a set of infrared detectors installed in the environment (for example, the cockpit in figure 1). These detectors measure the instant when the scanning IR beam is coincident with each detector. The proposed tracking system is based on measuring the timing (time duration between measured events) within the raster scan of the display.

A timing measurement from the detector determines the position of the display raster relative to the position of the detector in two dimensions. If several detectors are used, the absolute position and orientation (for example, six degrees of freedom) of the head relative to the outside world can be calculated. We describe in subsection 2.2 some of the detection techniques we have developed.

2.2 Detection Techniques

The easiest way to determine the vertical orientation of the display relative to the detectors is to count the number of raster lines from the top of the display frame to the instant a line passes a detector. Given the instantaneous vertical field of view of the display, the line count accurately determines how far down the frame the display has scanned when the detector is reached. Because several scan lines may reach the detector, an analysis of successive scan lines allows the processor to pick the middle line or interpolate between lines. However, the determination of the horizontal orientation is a far more difficult issue. The following paragraphs explain how an accurate determination of the horizontal orientation is obtained.

To determine horizontal orientation of the display image, the detected IR pulse can be timed from the start of each display frame as in the vertical case. However, the exact horizontal scanning frequency in the VRD's mechanical resonant scanner (MRS) is difficult to measure to a high degree of accuracy. Therefore, the resulting error in determining a horizontal location within the overall raster scan (pixel error) dramatically increases if the detected signal is referred to only the beginning of the display frame. This error accumulates with each successive scan line. Even though the method of counting from the beginning of each frame is accurate enough for determining the vertical location as before, the horizontal location within a scan line needs to be determined separately from the vertical location.

A more straightforward method for determining the horizontal location within a scan line is to refer to the MRS driving signal at the actual beginning of every scan line. By counting the time from the beginning of that line to the detection event, the horizontal position of the display scene relative to the detector is determined. But this assumes that there is stability between the actual locations of the mirror location relative to the MRS drive signal. Unfortunately, this may not be the case. We have observed that the natural or resonant frequency of the MRS varies with environmental changes such as temperature. As a result, the phase of the driving signal no longer matches the actual phase of the scanner.

To negate the effect of such a phase drift, the actual beginning of the scan line is needed. Our initial approach to measure the actual beginning of the scan line was to put an optical sensor at the edge of the scanned field to detect the beginning of every scan line. However, the performance was susceptible to slight mechanical misalignment between the sensor strip and the actual scanned field.

To overcome this problem, we developed a dual detector approach. (Refer to figure 2 and 3.) Because multiple IR scan lines are likely to be detected in each scanned frame, the time between each IR pulse determines the general location within the scan line. However, with one detector, the direction of the horizontal scan when the beam first hits the detector in any given frame (such as going from left to right or vice versa) is arbitrary. Therefore, a second detector is placed to the

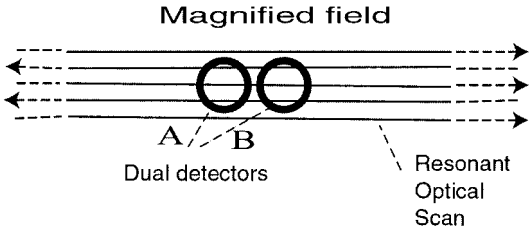

Figure 2. *The scanned IR field.*

side of the first detector along the horizontal scan axis as illustrated in figure 2, creating multiple IR pulses from each detector, as shown in figure 3. Because the scanned beam strikes the two detectors sequentially, the scan direction can be determined by the sequence of detection events.

To obtain a precise horizontal location within the scan line, one of the two detectors is chosen as the reference. As shown in figure 3, time detector A is the reference detector, and the timing regiment is when the IR beam strikes detector A to when the returning beam also strikes detector A. We assume that this time duration or measured pulse width is twice as long as the time from detector A to the scan edge or the actual beginning of the scan line. The corresponding horizontal location is then calculated by halving this timing duration. In this way, the beginning of the scan line is inferred, and the location of the detector along the horizontal line relative to that beginning is determined. Because several scan lines are likely to pass by the detector, we collect all of the events and pick the one in the middle (that is, the middle of an odd set of scan lines or interpolate if there are an even number) to determine the horizontal location of the raster relative to the detector.

To reiterate, the timing from the beginning of the frame to the time when a beam hits detector A is used to calculate the corresponding vertical location of the raster relative to the detector. Timing from the interval of successive passes of scan lines across the detectors is used to determine the horizontal location of the raster relative to the detector. By knowing the location of the detectors in the cockpit, the relative orientation of the display raster in azimuth and elevation angles can be calculated. By locating three or more detectors in the real-world environment that are within the instantaneous external scan field, the position and orientation of the display raster in six degrees of freedom can be determined. This aspect is discussed in subsection 2.3.

2.3 Tracking Principle

As stated, a detector's 2D horizontal and vertical location within the raster scan is the basis upon which we make all of our calculations of position and orientation of the display raster in space. We start by tracking in two dimensions in the time domain and then, by calculation from multiple detectors, relate that to tracking in the 3D spatial domain.

2.3.1 Tracking in the Time Domain. Tracking in the time domain is defined in two dimensions as (pixel, line) coordinates for horizontal and vertical locations in the tracking raster, respectively. We use a straightforward way to calculate the corresponding pixel and line positions from the timing measurement obtained in the detection system. For the pixel calculation, the horizontal timing is divided by a pixel clock (~40 ns/pixel) to obtain the corresponding pixel location. For the line calculation, the vertical timing is used to directly calculate the corresponding line location by dividing by the scan line time duration. It is worth noting that the error from not having an absolute scan line time duration of the MRS is unlikely to have much effect on the line accuracy. This is because the truncated result does not likely reflect the error. Consequentially, we can obtain a (pixel, line) coordinate of the target (detector) in two dimensions.

Even though 2D tracking in the time domain can be obtained via a single detector, we are only capable of superimposing an image or overlay in the direction of that single detector. For example, if the VRD is to superimpose an image slightly to the side of the detector, the registration of the augmented image is changed in location slightly as the detector moves. It is due to nonuniformity in pixel location and spacing within the display FOV. Consequently, target-tracking applications are limited. To fully use the SARSDT as an AR system, it is necessary to know where each pixel in the display

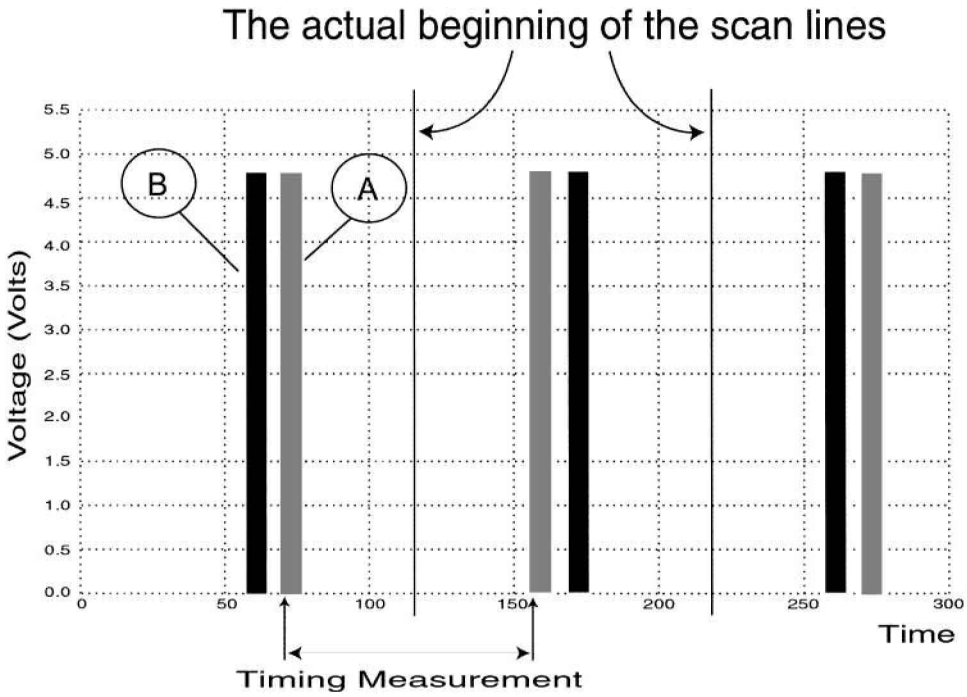

Figure 3. *The pulse train from the dual-detector system.*

actually overlays a location in the real world—the spatial information.

2.3.2 Tracking in the Spatial Domain. To fully determine the intercept of each raster pixel (beyond the single pixel detected previously), it is necessary to determine the position of that raster pixel relative to the real world. By locating several detectors whose locations in the real-world environment are known, the position and orientation of the display raster in six degrees of freedom can be determined.

A fundamental problem of converting timing information into spatial information in the world coordinate is an orientation offset in three-dimensional space between the IR scanned field axes and the world axes (Chinthammit, Seibel, & Furness, 2002). The solution to this problem is to determine a six-DOF transformation between the world coordinates and the VRD coordinates.

Generally this problem can be solved by triangulation techniques in which the detectors' positions are known relative to the real world. Typically, these techniques require that distances from emitter/source to each detector are known. For example, in the event of an earthquake, a seismograph can detect an earthquake magnitude but not a precise direction and depth where the epicenter is located. At least three seismographs are needed to triangulate a precise epicenter; see an example by Azuma and Ward (1991). Because our detection system is based on the timing incident in 2D raster scan, single-pixel detection cannot adequately determine the distance. Theoretically, an infrared sensor unit can possibly determine the distance from the signal strength of the received signal. However, the nature of our scanning field causes an infrared beam to travel quickly over a sensing area of an infrared detector. This makes it impractical to accurately determine the distance from a signal strength level.

This single-pixel detection problem in 2D coordinates can commonly be addressed as a computer vision problem such as a 3D reconstruction. Therefore, this triangulation problem can also be solved by computer vision

techniques as well. In our application, a distance from a "camera" to each "marker" is unknown whereas distances of each marker (detector) relative to the reference are known. Our known markers are located in a 2D coordinate (pixel and line) through a captured video frame. The reconstruction performs on a principle that a marker's 3D position in the real world and the corresponding 2D coordinate on the camera screen must lie along the same ray formed by the camera projection. If the accurate 3D position estimations were found, the distance between markers would be equal to the physical distances. Generally, the 3D position estimation is solved by an optimization process, which can be time consuming to reach the optimal solution (Janin, Zikan, Mizell, Banner, & Sowizral, 1994).

Another computer vision technique that requires less optimization process is to configure four markers into each vertex of the square. Using the fact that there are two sets of parallel sides and their direction vectors perpendicular to each other, an image analysis algorithm is used to estimate the position and orientation of the camera with respect to the square plane. The optimization process is applied only to compensate for the error in detected vertices coordinates that cause the direction vectors of the two parallel sides of the square not perpendicular to each other. Details on this development and algorithm are explained by Kato and Billinghurst (1999). Due to its simplicity, we implement this image analysis technique on our hardware system to determine a six-DOF of detectors' plane with respect to the VRD.

In figure 4, looking out from the VRD, we define a reference 2D plane at a fixed distance, Z_{ref}, which is an imaginary distance away from the origin of the VRD coordinate. (The choice of Z_{ref} does not affect the overall tracking accuracy.) A projection of four vertices on our reference 2D plane represents vertices on the camera screen as implemented in the original algorithm (Kato & Billinghurst, 1999). Because a 2D coordinate is defined in the VRD coordinate, converting timing information into spatial information becomes a much simpler issue. The horizontal and vertical coordinate is directly obtained from (pixel, line) coordinates in the time domain. Our 2D coordinates (or displacements) of

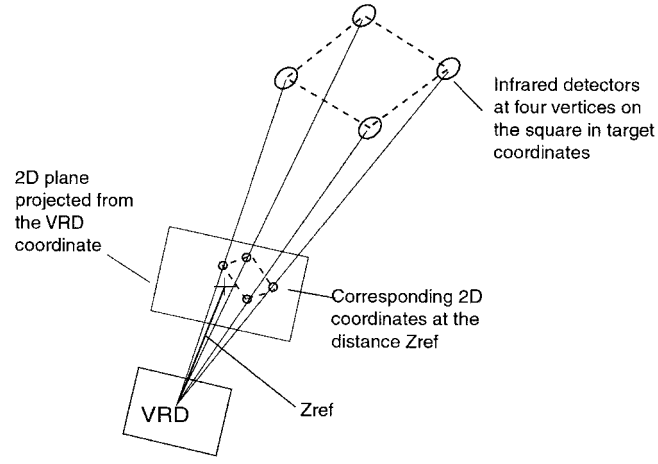

Figure 4. *The illustration of how we define a 2D plane in the VRD coordinate.*

four vertices on that plane can be calculated by the following equations.

$$Displacement_{horizontal}(t) = -\frac{K_x}{2}\cos(2\pi\, freq_x\, t_x)$$
$$Displacement_{vertical}(t) = K_y \cdot t_y \cdot freq_Y - \frac{K_y}{2} \quad (1)$$

We define subscripts (x,y) as the directions along the MRS's moving axis and the galvanometer's moving axis, respectively. Therefore (x,y) can also be referred to as the directions in the horizontal and vertical axes of the VRD scan. The t is the timing measurement in both axes or (pixel, line) coordinates from detections in time domain. *Displacement* is a calculated position in respective axes, although their calculations are derived differently for horizontal and vertical axes. The calculation of the horizontal displacement is modeled to correct for the distortion of the sinusoidal nature of the MRS, whereas a linear model is used in the calculation of the vertical displacement due to a linearity of the galvanometer motion. The displacement at the center of the scan is set to (0,0) for convenience purposes. The K is the displacement over the entire FOV of the scanner in respective axes at the distance Z_{ref}. The *freq* is the scan-

ning frequency of both scanners. In conclusion, equation (1) transforms our 2D coordinates in time domain to 2D positions in the spatial domain.

It is worth noting that this 2D position lies along the same ray (formed by the scanner) with the actual 3D position in the real world. Once the 2D position of each vertex has been determined, the image analysis algorithm can then determine the translation and orientation of the square plane with respect to the scanner.

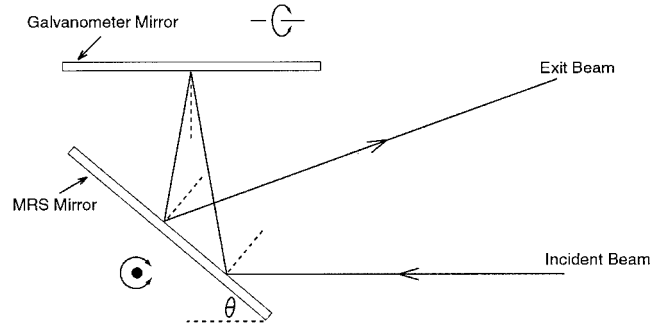

Figure 5. *Two scanning mirrors.*

3 Experimental Setting

This section first describes all of the main apparatuses in this experiment. Then, we describe our current system configuration. It is slightly different from what have been described in subsection 2.1 in terms of a moving element. Lastly, performance metrics are defined as a way to characterize our proposed system.

3.1 Experimental Apparatus

To characterize the utility and performance of the SARSDT concept, we built a breadboard optical bench as shown in figure 7. Four subsystems compose the experimental SARSDT: a modified virtual retinal display, a robot arm target mover, a detection and data acquisition subsystem, and a reticle generator. These subsystems are described in more detail in the following subsections.

3.1.1 The Virtual Retinal Display. As stated in the introduction, the VRD scans laser light directly to the retina of the eye to create augmented images. It uses two single-axis scanners to deflect a beam of modulated laser light. Currently, the VRD scans at VGA format or 640×480 pixel resolution. The vertical scan rate is equivalent to the progressive frame rate. Because the 60 Hz vertical scan rate is relatively low, an existing galvanometer scanner is used as the vertical scanner (Cambridge Instruments). Galvanometers are capable of scanning through wide angles but at frequencies that are much lower than the required horizontal frequencies. Early in our development of the VRD, we built a mechanical resonant scanner (MRS) (Johnston & Willey, 1995) to meet the horizontal scanner requirements. The only moving part is a torsional spring/mirror combination used in a flux circuit. Eliminating moving coils or magnets (contained in existing mechanical resonant scanners) greatly lowers the rotational inertia of the device; thus, a high operational frequency can be achieved. Another unique feature of the MRS is its size (the mirror is 2×6 mm). The entire scanner can be very small and is being made smaller as a micro-electromechanical system (MEMS) (Wine, Helsel, Jenkins, Urey, & Osborn, 2000).

To extend the field of view of the horizontal scanner, we arrange the scanning mirrors to cause the beam to reflect twice between the MRS and galvanometer mirrors as shown in figure 5. This approach effectively multiplies the beam deflection by 4, relative to the actual rotational angle of the horizontal scanner.

As shown in figure 6, the scanned beams pass through a beamsplitter (extended hot mirror, Edmund Scientific) and spherical optical mirror to converge and collimate the visible light at the exit pupil. When the viewer aligns her eye entrance pupil with the exit pupil of the VRD, she perceives a high-luminance display at optical infinity that is overlaid onto the natural view of the surround.

3.1.2 Light Sources. For convenience in the SARSDT breadboard system, we use only one wavelength of visible light (red) and one IR wavelength. The visible light source is a directly modulated red laser di-

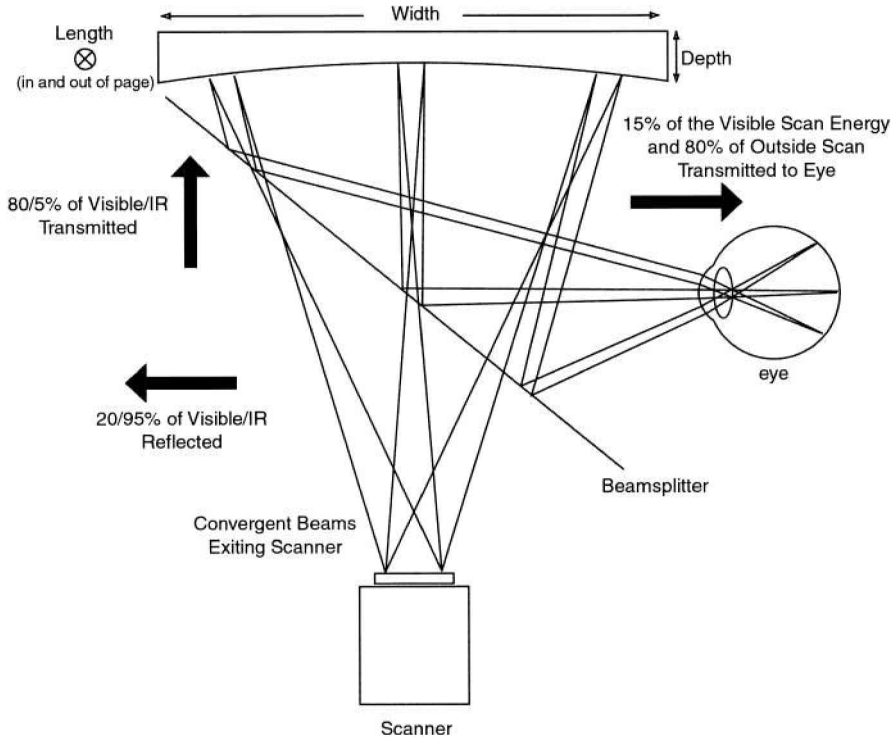

Figure 6. *Optical relay components.*

ode (635 nm, Melles Griot, Inc.) that is fiber optically coupled. The IR laser diode (1310 nm, Newport Corp.) is not modulated. Both visible and IR light are combined using an IR fiberoptic combiner (Newport Corp.) into a single multispectral beam. The result is a multiple spectra beam. The collimation of the mixed wavelength beam is individually optimized for the IR and visible wavelength before being scanned by the two-axis scanner.

3.1.3 Wavelength-Selective Beamsplitter. For the SARSDT, the normal beamsplitter in the VRD configuration is replaced by a wavelength-selective beamsplitter or hot mirror. This allows the red visible wavelength to pass through the beamsplitter with low attenuation while being highly reflective to the infrared wavelength (1310 nm). As a result, the wavelength-selective beamsplitter creates two simultaneous scans: a visible one into the eye and an IR one into the environment.

3.1.4 Robot-Arm Target Mover. A five-DOF robot arm (Eshed Robotec, Scorbot-ER 4PC) is used to test the tracking performance of the system. Because the first configuration of this concept is installed on an optical bench, it is difficult for us to move the display (as would be the case with the eventual head-coupled version of the system) relative to fixed detectors in the environment. Instead, we decided to move the detectors using the robot arm. In case of the target-tracking configuration (see subsection 3.2), we use the robot arm to move a detection circuit at the predetermined trajectory, equivalent to moving the head in the opposite trajectory. The robot arm provides a repeatable movement and reports the "truth" position. The target mover system is independently operated from the tracking system by a separate computer. Robot arm positions in five DOF (no roll axis) are recorded at 8 Hz using custom software and an interface. The reported positions can then be used for evaluating tracking performances.

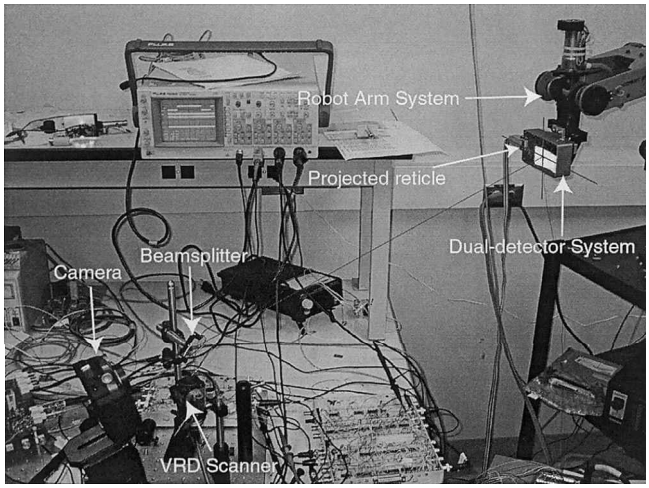

Figure 7. *SARSDT as targeting system.*

3.1.5 Detection and Data Acquisition System. The detection system consists of infrared detectors (GPD GAP 300), a low-noise amplifier, and a processing unit. Two IR detectors and a low-noise amplifier are built into one circuit contained in an aluminum box attached to the robot arm in figure 7. The signal generated from the low-noise amplifier is then sent to the processing unit. It consists of a set of flip-flops that measure specific time durations between detected signals and reference signals. The circuit architecture depends on the detection technique that is selected.

The output of the detection system is sampled by an acquisition system. It is a timer/counter board (National Instruments) with an 80 MHz clock speed. A graphical programming language, Labview, is programmed to interface with the acquisition board. The Labview program is also written to determine the detectors' locations within a display scan. A pair of line and pixel numbers define the location in two dimensions, and these values are then digitally output to a display system.

3.1.6 Reticle Generator Subsystem. To visually indicate the output of the tracking part of the SARSDT, we generate and project a reticle using the display portion of the VRD system. But, instead of projecting on the retina, we turn the visible beam around and project it on the target being moved by the robot arm. The reticle generator draws a crosshair reticle with a center at the calculated pixel and line locations. The generator consists of two identical sets of counters: one for pixel count and another for line count. Each has different sets of reference signals from the VRD electronics: the beginning of the frame signal (BOF) and the beginning of the scan line signal (BOL) for line count, and BOL and pixel clock for pixel count. The line counters are reset by the BOF and counted by the BOL. Similarly, the pixel counters are reset by the BOL and counted by the pixel clock. Both continue to count until the reported pixel and line number are reached. Turning on the same pixel position in every scan line creates the vertical portion of the crosshair, whereas turning all pixels in the reported line generates the horizontal crosshair symbol.

3.2 System Configuration

As stated earlier, it is difficult for us to move the VRD (as would be the case with the eventual head-coupled version of the system) relative to fixed detectors in the environment. Thus, we decided to move the detectors using the robot arm as in the case of the target-tracking configuration. The combination of the robot arm and the VRD projector as described above effectively simulates moving the VRD and the user's head in the environment relative to fixed locations in space. (That is, the VRD is fixed and the IR detectors are being moved by the robot arm system, as illustrated in figure 7.)

3.3 Performance Metrics

To characterize the performance of this tracking system, we consider the following performance metrics.

3.3.1 Stability. The tracker's stability refers to the repeatability of the reported positions when the detectors are stationary. It also refers to the precision of the tracking system. Ideally, the data recorded over time should have no distribution or zero standard deviation. Unfortunately, the actual results are not ideal. Inherent

in the SARSDT are two error sources: measurement error and scanner jitter. Both are random errors that limit the tracking precision, and therefore cannot be compensated by calibration procedures.

3.3.2 Static Error. Static error is defined as error that occurs when the user is stationary. The sources of static errors are calibration error and tracker error. Calibration procedures are used to calibrate a tracking system with a display system so that augmented image is aligned correctly. The calibration parameters associated with the tracking system are transformations among the scanner, the sensor unit, and the world coordinates. The parameters associated with the display system are the field of view and the nonlinearity of the MRS. Inaccurate parameters cause systematic errors, which can be compensated by calibration procedures. Furthermore, tracker error also contributes to the static error.

3.3.3 Dynamic Error. Dynamic error is defined as error that occurs when the user is in motion. The primary source of dynamic error is system latency, which is defined as the time duration from sensing a position to rendering images. It includes the computational time required in the tracking system as well as communication time required between subsystems (such as display system, sensing unit, and the computing unit). A significant latency causes the virtual image, relative to the real object in the environment, to appear to trail or lag behind during a head movement.

4 Experiments, Results, and Discussions

This section describes different experiments that are derived from the performance metrics. (See subsection 3.3.) Results and discussions are discussed in each experiment. Lastly, future works are discussed.

4.1 The Stability of the Detection System

The stability of the detection circuit was obtained by analyzing the recorded timing data when the target/detectors were stationary in the 8×6.5 deg. FOV of the scan. The standard deviations of the data were measured and the ranges are reported as $+/-$ 3 standard deviations. The horizontal axis and the vertical axis were analyzed separately. The vertical stability/precision is within 0.0126 deg. or one scan line, and the horizontal stability is $+/-$ 0.0516 deg. or 5.5 pixels. Unlike the video-based tracking system, the precision of the prototype SARSDT system is not limited by the pixelated display resolution (640×480 within the field of view), but rather by how repeatable and stable the timing measurement is.

The pixel error along the horizontal axis originates from two main sources: the VRD scanner introduces an electromechanical jitter, and our current software operational system, Windows 98, is not capable of consistently acquiring data between frames at the exact display refresh rate. Neglected are environmental noise sources, such as mirror surface degradation, scattering from dust, detector shot noise, vibration, and so forth.

Both sources of precision error contribute to the timing fluctuation in the recorded data between frames. First, multiple pixels are detected from different scan lines in one frame. Each scan line generates slightly different timing information due to the orientation of the IR sensing area with respect to the angle of the scanned beam, producing small timing fluctuations among multiple detected timing measurements or pulses in any single frame. Because the software operational system is unable to control the precise acquisition time, the order of the pulses in the processing data buffer does not correspond exactly to its proper IR detector's physical sensing area. For example, the first data in the buffer are not necessarily related to the top area of the sensing area where a beam hits first.

One method that can improve the stability of the detection circuit is a detection technique that would average multiple pulses across multiple scan lines. Thus, there would be only one averaged pulse in each frame. This would reduce the error from the operational system by selecting one pulse with respect to several pulses that have variable timing measurements. However, additional acquisition time is inevitable because the averaging techniques require the integration of detected pulses

over multiple scan lines. Offline analysis demonstrates that the horizontal stability can be reduced to +/− 0.02 deg. or +/− 2.4 pixels if the partial elimination of the scanner jitter effect and ideal acquisition time are manually and properly imposed.

The source of line error along the vertical axis is likely to be an electromechanical jitter from the VRD. Because the first pulse of multiple pixels in a buffer determines the line location, the resulting jitter causes the first pulse to move slightly vertically. Therefore, we expect a line error within one scan line.

The stability of the detection system is a fundamental error of our tracking system. It propagates through the rest of the system error.

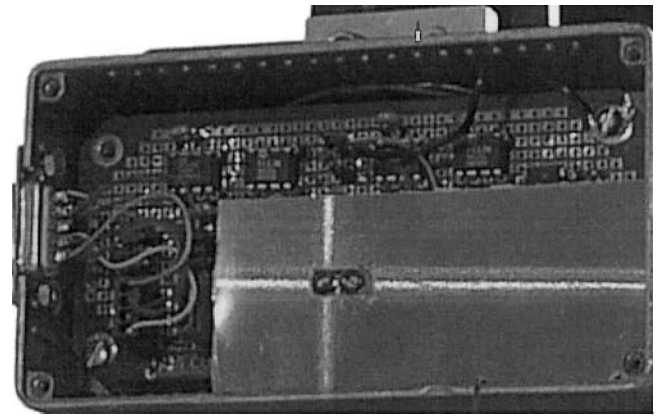

Figure 8. *Crosshair overlay on the detector.*

4.2 Static Accuracy

We start by tracking in two dimensions in the time domain and then, by calculation from multiple detectors, relate that to tracking in the 3D spatial domain.

4.2.1 Tracking in the Time Domain. To demonstrate tracking in the time domain, the dual-detector system was mounted on the robot arm system and the VRD was fixed in position as shown in figure 7. Using the IR tracking system in time domain, once the pixel location of the reference detector is calculated, a red crosshair reticle was projected onto the reference IR detector in the subsequent video frame. The red crosshair pattern from the VRD was superimposed onto the reference detector. The crosshair reticle was one pixel and one scan line wide for the vertical and horizontal lines, respectively. The robot arm was programmed to move from one corner to the next within the current 16.5 × 6.5 deg. FOV of the scan. The rectilinear motion of the target was primarily planar with a z distance of 30 in. from the VRD scanner. This tracking sequence can be viewed on our Web site (Chinthammit, Burstein, Seibel, & Furness, 2001).

A digital camera was set at the eye position to capture images of the crosshair through a custom hot mirror as described in the system configuration. Figure 8 shows the crosshair projected on the reference IR detector (left-side detector).

We then analyze the images to determine the static error. The static error at the calibration point is considered to be zero. For each image, the vertical and horizontal error was determined separately and in an angular unit or degree. Finally, an average and a standard deviation of both axes errors were calculated. The results were 0.08 +/− 0.04 and 0.03 +/− 0.02 deg. for the horizontal and vertical axes, respectively.

This static error is due to a calibration error in the display system and the sinusoidal scan motion in the horizontal axis. The display calibration requires the determination of the phase shift between the electronic drive frequency and the mechanical resonant frequency. This phase shift is used to determine when the first pixel of the scan line should be displayed to match the very edge of the horizontal mechanical scan. It can be done only visually, and the calibration error is therefore inevitable. Due to the sinusoidal scan motion, the calibration error magnitude is different along the horizontal scan. Because the calibration is taken at the middle of the scan, the registration error increases as the detector moves closer to the edge of the scan where the infrared beam travels slower.

Although the robot arm moved orthogonally from the z axis, one advantage of the current 2D tracking system is that the technique can be operated in a 3D space without having to determine the absolute positions, which is a requirement for other tracking systems. Thus, the proposed technique has a high computational

efficiency for tracking the designated object. Additionally, a more passive approach is to place retro-reflective targets in the surroundings, which concentrate all the active system components at the display. This makes this system applicable to nonmilitary and unconstrained AR applications, such as tracking the viewer's hand, wand, or mouse. For example, Foxlin and Harrington (2000) have suggested wearable gaming applications and "information cockpits" that use a wireless hand tracker worn as a small ring. The active acoustic emitter could be replaced with a passive retro-reflective film worn as a small ring. In virtual environments (VEs), an application for tracking a single outstretched finger is to select objects using line-of-sight 2D projections onto the 3D scene. A retro-reflector at the tip of a finger or wand can create the same image plane selection, manipulation, and navigation in VEs, similar to the "Sticky Finger" proposed by Pierce et al. (1997) and Zeleznik, LaViola, Feliz, and Keefe (2002). Tracking two fingers adds more degrees of freedom in the manipulation of virtual objects in VE and AR, such as the "Head Crusher" technique that can select, scale, and rotate objects (Pierce et al., 1997).

However, one limitation of our dual-detector technique is the fact that this is a line-of-sight tracking system; the working range is limited by the acceptance angle of the VRD field of view and the IR detectors. The detection system has an acceptance angle of approximately +/− 60 deg., limited by the detector's enclosure. Furthermore, the dual-detector scheme has a very limited range of tracking target movement along the roll axis because the horizontal scan lines must strike both detectors (required by the dual-detector scheme). Currently, the detectors are 5.5 mm apart, and each sensor diameter is approximately 300 microns. Consequently, the roll acceptance angle is less than +/10 deg.

To increase the acceptance angle for target roll, the detector enclosure can be removed to bring the two detectors in closer proximity. If cells are placed side by side, the roll acceptance is estimated to be +/− 45 deg. To further improve the range of roll motion, a quad-photo detector can be used. For extreme angles of roll, one pair of the dual detector system generates the pulses but not the other.

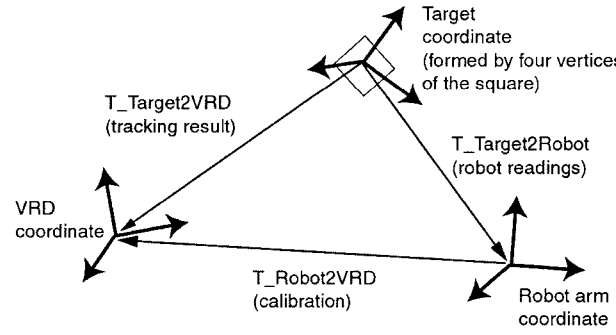

Figure 9. *Transformation matrix diagram.*

4.2.2 Tracking in the Spatial Domain. As described in the previous subsection, the superimposed image can only be accurately projected at the detector and not anywhere else in the scan field. To overlay an image on locations other than the reference detector, transformation between a time domain and a spatial domain is required.

A selected image analysis algorithm (Kato & Billinghurst, 1999), described in the tracking principal subsection 2.3.2, is used to estimate the position and orientation of the scanner with respect to the square plane. Our goal in this experiment is to determine the performance of this image-analysis tracking algorithm within our hardware system. Beside from the fact that we built only one dual-detector system, it is feasible with our repeatable target mover (robot arm) that we can efficiently conduct this experiment by moving a dual-detector system to four predetermined positions to form four vertices of the virtual square plane, 60 × 60 mm in size.

The experiment is set up to determine position and orientation (six DOF) of the square plane. The center of this square is the origin of the *Target* coordinate system. The six-DOF information is in a form of a transformation matrix, $T_{Target2VRD}$. (The diagram of this experiment is illustrated in figure 9.) We also need to take into account that there is a position offset between the reference detector and a tool center point (TCP) of the robot arm; it is defined as T_{bias}. To simplify the diagram, we do not include T_{bias} in the diagram.

The relative position and orientation of the square to the robot arm coordinate is reported by a robot arm system and defined in a transformation matrix, $T_{Target2Robot}$. Another transformation matrix is $T_{Robot2VRD}$, which related the robot arm coordinate to the VRD coordinate. This matrix has to be obtained through a calibration process because the orientation of the VRD scan is difficult to measure directly. During a calibration process, the diagram in figure 9 is still valid, with the exception that $T_{Target2VRD}$ is now reversely used to estimate $T_{Robot2VRD}$. Obviously, a $T_{Robot2VRD}$ estimate contains error from the tracking algorithm. Thus, we conducted the calibration process over multiple positions and orientations throughout the scan field to average out the error inflicted by $T_{Target2VRD}$. The final position and orientation of the calibrated $T_{Robot2VRD}$ have high correlation with our manual measurements. Then, the correction six-DOF or "truth" transformation matrix is determined from a transformation matrix, T_{truth}, which is a multiplication of two previously mentioned transformation matrices: $T_{Target2Robot}$ and $T_{Robot2VRD}$. The position and orientation obtained from $T_{Target2VRD}$ is compared to *truth* to determine the tracking accuracy.

The virtual square is maintained at 60 × 60 mm and is tracked in six DOF within the field of view of the VRD (18.5 × 6 deg.). The experiment was conducted in the z axis range from 30–35 in. away from the VRD scanner. The error is defined as the difference between the truth and the estimation pose and reported in degrees and millimeters for the orientation and translation, respectively. In translation, the error in horizontal and vertical axes of the scanner is approximately 1 mm. A more significant error is found along the z axis, producing an error of less than 12 mm within the FOV at 30–35 in. from the scanner. In orientation, the error in each of (pitch, yaw, and roll) is within 1 deg.

Overall results demonstrate that this image-analysis tracking technique performs well. A significant problem of this technique is that the farther away from the scanner the virtual square is, the projected vertices on the 2D plane become increasingly susceptible to the error introduced during the detection process. This error is often referred to as a *tracking resolution*. If a captured video image is used for the tracking purposes, its tracking resolution is limited by the image resolution. It is important to realize that our tracking resolution is not limited by the VRD display resolution but instead by the stability of our detection system discussed in subsection 4.1. To further illustrate the point, the reported error in the original work (Kato & Billinghurst, 1999) is greater than the error reported in this experiment. However, this image-analysis technique still provides our current system an alternative solution to our initial triangulation problem. A noticeable translation error in the z axis is mainly because our current experimental volume is relatively farther away in that direction than are other axes in the VRD coordinate. In other words, if the square is to be as far away in the horizontal axis as it is now in the z direction, the same error quantity can be expected.

4.3 Dynamic Accuracy

Figure 10 demonstrates that lag can be measured during dynamic tracking of the robot arm system. The dual-detector system on the robot arm was moved from one location to another horizontally and vertically as shown in the first and second columns, respectively. The robot arm moved from detection circuit position (a) to the position (c), generating a horizontal shift due to the system lag, as shown in (b). Moving the robot arm up from (d) to (f), vertically, generates a vertical shift of the crosshair below the detectors, as shown in (e). The velocities of the detectors at the photographed target motions in (b) and (e) were both approximately 100 mm/sec. (as determined from the robot arm readouts).

To measure the overall system lag, the images at the positions (b) and (e) were magnified to quantify the spatial or dynamic error. The 1.5 mm error along both axes and the known velocities of the robot arm indicate that the overall system lag is approximately within one frame or 16.67 ms, which is from the display system.

Because the detected pixel and line numbers in the current frame are displayed in the following display frame, an inevitable display lag of up to 16.67 ms occurs. This is confirmed by the experimental result. Furthermore, as expected, because the computational ex-

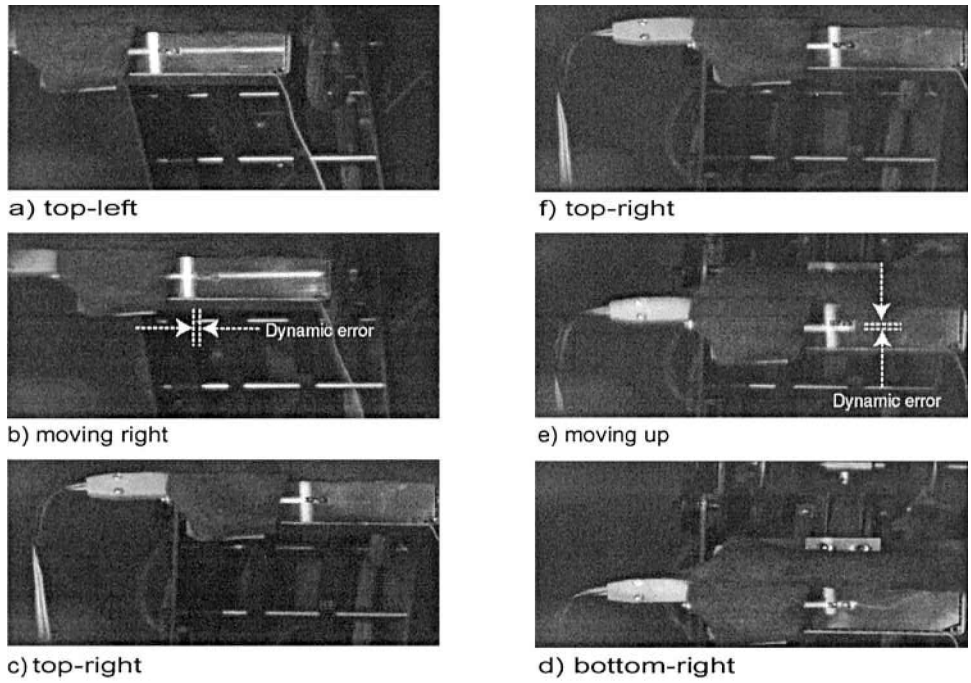

Figure 10. *Dynamic error.*

pense for the detection is minimal, its contribution to overall system lag is negligible. In the future, the computational time can be further reduced using a real time system and a faster computer, when multiple dual-detector systems are used. Furthermore, a system lag of 16.67 ms suggests that prediction algorithms and/or higher display refresh rates are to reduce latency.

4.4 Future Work

We continue to develop the proposed shared-aperture tracking display into a complete head-tracking system with six DOF. In subsection 4.2.2, we demonstrate that six DOF of the VRD can be estimated by an image-analysis technique. A noticeable error is also reported in the experiment. Therefore, we continue to develop the necessary algorithms to achieve a higher tracking accuracy.

The proof-of-concept system does not have a real-time capability and therefore results in tracking error. Currently, we are considering alternative hardware solutions to upgrade the current system into the real-time system. Furthermore, the effect of the electromechanical jitter is not well studied. A mathematical error model of the jitter can be used to improve our tracking performance (Holloway, 1997), but we first need to quantify the error and then model it.

In this paper, the reticle generator subsystem is limited to two DOF. Therefore, a user cannot fully experience a sense of reality with augmented images. Six DOF is required. Currently, we have been developing a six-DOF display subsystem that can communicate with the tracking system to display images with six DOF.

Lastly, the update rate of the shared-aperture system cannot exceed the VRD refresh rate of 60 Hz. In military cockpit applications, rapid head movements often occur. So an optical tracking system alone is not ideal for these applications. To solve this problem, we are concurrently developing a hybrid tracking system based on the shared-aperture optical tracking system presented in this paper and inertial sensors. The goal is to increase the measurement of head position to rates of 1,000 Hz while keeping the VRD update rate at only 60 Hz.

5 Conclusions

The SARSDT is a VRD that has been modified to provide a tracking function using coaxially the optical path of the VRD. In this way, the proposed tracking technology shares the same aperture or scanned optical beam with the visual display.

The tracking precision is limited only by the stability of the detection circuit. Currently, the stabilities are +/− 0.05 and +/− 0.01 deg. in the horizontal and vertical axis, respectively. Offline analysis demonstrates that the horizontal stability can be reduced to +/− 0.02 deg. if the partial elimination of the scanner jitter effect and ideal acquisition time is manually and properly imposed. Implementation of peak-detection techniques can also improve the tracking precision. The red crosshair pattern from the VRD was superimposed onto the reference detector to demonstrate tracking in the time domain. The static error was measured to be 0.08 +/−0.04 and 0.03 +/−0.02 deg. for the horizontal and vertical axes, respectively. The error was determined to be due to a calibration error in the display system and the sinusoidal scan motion in the horizontal axis.

One advantage of the current tracking system in the time domain is that the technique can be operated in a 3D space without having to determine the absolute positions. Therefore, the computational efficiency is extremely high. Consequently, this technique is effective for some AR applications such as "Sticky Finger" (Pierce et al., 1997) and "information cockpits" (Foxlin & Harrington, 2000) in which optical detectors or reflectors can be attached to the object or environment. The dynamic error is mainly due to the display lag, currently set at a 60 Hz refresh rate. The experiment supportively indicates the system lag to be within one display frame (16.67 ms). (See figure 10.)

Tracking in the time domain allows the AR system to superimpose an image only at the detector location. Tracking in the spatial domain is a solution to this problem and is therefore required. In Chinthammit et al. (2002), the result demonstrates that difficulties are inherent in transforming coordinates in the time domain to coordinates in the spatial domain. Problems with transformation are primarily due to the orientation offset between the IR scanned field and the world space in 3D space. The solution to this problem is to determine a six-DOF transformation between the robot arm coordinates and the VRD coordinates. Therefore, we have implemented an image-analysis tracking technique on our current optical tracking system to determine the six-DOF target movement. The results indicate a feasibility of such technique for other applications such as head tracking; however, more extensive development is required to achieve a very high tracking accuracy.

Acknowledgments

We would like to thank Reinhold Behringer (Rockwell Scientific Company) for the robot arm software interface and Robert A. Burstein for the 2D reticle generator. I thank Chii-Chang Gau, Michal Lahav, Quinn Smithwick, Chris Brown, and Dr. Don E. Parker for valuable contributions in the preparation of this paper. The Office of Naval Research funded this research with awards N00014-00-1-0618 and N00014-96-2-0008.

References

Azuma, R. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments, 6*(4), 355–385.

Azuma, R., & Ward, M. (1991). *Space-resection by collinearity: Mathematics behind the optical ceiling head-tracker. UNC Chapel Hill Department of Computer Science Tech. Rep.* TR 91–048.

Chinthammit, W., Burstein, R., Seibel, E. J., & Furness, T. A. (2001). Head tracking using the virtual retinal display. *The Second IEEE and ACM International Symposium on Augmented Reality.* Unpublished demonstration, available on line at (http://www.hitl.washington.edu/research/ivrd/movie/ISAR01_demo.mov)

Chinthammit, W., Seibel, E. J., & Furness, T. A. (2002). Unique shared-aperture display with head or target tracking. *Proc. of IEEE VR2002,* 235–242.

Foxlin, E., & Harrington, M. (2000). WearTrack: A self-referenced head and hand tracker for wearable computers and portable VR. *The Fourth Intl. Symp. on Wearable Computers,* 155–162.

Holloway, R. (1997). Registration error analysis for aug-

mented reality. *Presence: Teleoperators and Virtual Environments, 6*(4), 413–432.

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

Johnston, S., & Willey, R. (1995). Development of a commercial retinal scanning display. *Proceedings of the SPIE, Helmet- and Head-Mounted Display and Symbology Design Requirements II, 2465,* 2–13.

Kato, H., & Billinghurst, M. (1999). Marker tracking and HMD calibration for a video-based augmented reality conferencing system. *Proc. of the 2nd IEEE and ACM International Workshop on Augmented Reality,* 85–94.

Kelly J., Turner, S., Pryor, H., Viirre, E., Seibel, E. J., & Furness, T. A. (2001). Vision with a scanning laser display: Comparison of flicker sensitivity to a CRT. *Displays, 22*(5), 169–175.

Kutulakos, K., & Vallino, J. (1998). Calibration-free augmented reality. *IEEE Trans. Visualization and Computer Graphics, 4*(1), 1–20.

Pierce, J., Forsberg, A., Conway, M., Hong, S., Zeleznik, R., & Mine, M. (1997). Image plane interaction techniques in 3D immersive environments. *Symposium on Interactive 3D Graphics,* 39–44.

Rolland, J., Davis, L., & Baillot, Y. (2001). A survey of tracking technology for virtual environments. In W. Barfield & T. Caudell (Eds.), *Fundamentals of Wearable Computers and Augmented Reality* (pp. 67–112). Mahwah, NJ: Lawrence Erlbaum.

Sorensen, B., Donath, M., Yang, G., & Starr, R. (1989). The Minnesota scanner: A prototype sensor for three-dimensional tracking of moving body segments. *IEEE Transactions on Robotics and Automation, 5*(4), 499–509.

Stetten, G., Chib, V., Hildebrand, D., & Bursee, J. (2001). Real time tomographic reflection: Phantoms for calibration and biopsy. *Proceedings of the IEEE and ACM International Symposium on Augmented Reality,* 11–19.

Yokokohji, Y., Sugawara, Y., & Yoskikawa, T. (2000). Accurate image overlay on video see-through HMDs using vision and accelerometers. *Proc. of IEEE VR2000,* 247–254.

You, S., & Neumann, U. (2001). Fusion of vision and gyro tracking for robust augmented reality registration. *Proc. IEEE VR2001,* 71–78.

You, S., Neumann, U., & Azuma, R. (1999). Hybrid inertial and vision tracking for augmented reality registration. *Proceedings of IEEE VR '99,* 260–267.

Wine, D., Helsel, M., Jenkins, L., Urey, H., & Osborn, T. (2000). Performance of a biaxial MEMS-based scanner for microdisplay applications. *Proceedings of SPIE, 4178,* 186–196.

Zeleznik, R., LaViola, Jr., J., Feliz, D., & Keefe, D. (2002). Pop through button devices for VE navigation and interaction. *Proceedings of the IEEE Virtual Reality 2002,* 127–134.