System Lag Tests for Augmented and Virtual Environments

Colin Swindells, John C. Dill
School of Computing Science
Simon Fraser University
8888 University Drive
Burnaby, BC V5A 1S6
Tel: (604) 291-4371
E-mail: {swindell, dill}@cs.sfu.ca

Kellogg S. Booth
Department of Computer Science
University of British Columbia
201-2366 Main Mall
Vancouver, BC V6T 1Z4
Tel: (604) 822-8193
E-mail: ksbooth@cs.ubc.ca

ABSTRACT
We describe a simple technique for accurately calibrating the temporal lag in augmented and virtual environments within the Enhanced Virtual Hand Lab (EVHL), a collection of hardware and software to support research on goal-directed human hand motion. Lag is the sum of various delays in the data pipeline associated with sensing, processing, and displaying information from the physical world to produce an augmented or virtual world. Our main calibration technique uses a modified phonograph turntable to provide easily tracked periodic motion, reminiscent of the pendulum-based calibration technique of Liang, Shaw and Green. Measurements show a three-frame (50 ms) lag for the EVHL. A second technique, which uses a specialized analog sensor that is part of the EVHL, provides a "closed loop" calibration capable of sub-frame accuracy. Knowing the lag to sub-frame accuracy enables a predictive tracking scheme to compensate for the end-to-end lag in the data pipeline. We describe both techniques and the EVHL environment in which they are used.

KEYWORDS: Augmented Reality, Calibration, Lag, Sensor, Turntable, Virtual Reality.

1. INTRODUCTION

An earlier paper by Summers et al. [1999] describes the Virtual Hand Laboratory (VHL) and the calibration issues that we faced during its development. This paper extends that work in the context of the Enhanced Virtual Hand Laboratory (EVHL), where we support increased rendering complexity for augmented environments, additional sensing modalities (forces and torques using A/D inputs), and provisions for haptic and audio output to enrich the augmented environments in which experiments on goal-directed human hand motion are conducted.

Appropriate response to haptic data and accurate display of moving objects using predictive tracking both require knowing the system lag to sub-frame resolution. If lag is only known to within one frame time, it is difficult to compensate for the temporal difference between data obtained from the input sensors and data needed to render the output displays. Our earlier paper discussed various temporal and spatial calibrations that we performed on our VHL system to ensure a spatially accurate head-coupled stereo view of an augmented environment. This was achieved by guaranteeing an update rate of 60 Hz and an overall system lag that was on average 2.5 frames, a one-frame delay for sensing data, another one-frame delay for modeling & rendering, and up to one more frame of delay for the monitor to refresh the image from the frame buffer. This paper discusses subsequent work that more accurately measures system lag using two different techniques.

The first calibration technique is an "external" measurement that visually measures system lag. This was inspired by an earlier technique by Liang, Shaw and Green [1991] for calibrating a Polhemus FASTRAK sensor in which a pendulum's periodic motion provided accurate visual measurement of the system lag. Our technique uses photographs, rather than a video camera, so that motion blur can be used to gain additional "vernier" information about lag, although we could also employ a video camera if less accuracy were required.

The second calibration technique is an "internal" or closed loop measurement that enables the system to calibrate itself. It uses the system's analog sensing devices to "see" the total lag in the system by connecting the video output signal of the display to one of the analog input channels of the sensing system. The system records both the frame at which an external signal is received, and the "echo" of that signal in the display output. This gives a measure of the end-to-end system lag in terms of frame times. Improving the accuracy to sub-frame resolution is accomplished by manipulating the geometric pattern that is echoed on the display, to provide a fractional lag measurement.

The external and internal techniques are in rough agreement, and correspond to the system lag predicted by our end-to-end model of the system's data pipeline. This convergence of results provides assurance that the EVHL operates as expected, and that the experimental data collected during trials is accurate.

The remainder of this paper describes the types of experiments that are supported by the EVHL (for background information), presents a theoretical end-to-end model of the data pipeline, characterizes each processing component, and then discusses the two new calibration techniques and the results we obtained using them. The paper concludes with a comparison to some additional related work, and an explanation of how precise knowledge of lag could be used to further improve the visual quality of the augmented environments supported by the EVHL.

2. END-TO-END SYSTEM MODEL OF THE EVHL

The Enhanced Virtual Hand Lab (EVHL) supports augmented and virtual environments with a head-coupled stereo view of a relatively small (25x35x20 cubic cm) working volume. Experiments are conducted to investigate how people move their hands when performing various actions associated with common tasks such as grasping and manipulating objects. The primary hardware components are a Silicon Graphics Onyx2 workstation and a Northern Digital OPTOTRAK with 3-D position sensors based on active infrared emitting diode (IRED) markers, and an analog input capability provided by its OPTOTRAK data acquisition unit (ODAU) analog-to-digital (A/D) converters.

Custom software integrates these hardware components into a system that senses the location of the markers on objects of interest (including the subject), updates a computational model of the physical environment, and then displays a graphical "overlay" onto the working volume to provide an augmented environment in which experiments are conducted. The graphical overlay is achieved using a standard computer monitor suspended over the working volume. The monitor is reflected by a half-silvered mirror (with or without an optional black cardboard backing to make it fully reflecting) to provide either a simultaneous view of the physical and virtual objects (a mixed or augmented environment) or just the virtual objects (a purely virtual environment). Figure 1 shows a schematic diagram of the EVHL. A more detailed description of the apparatus is provided in an earlier paper by Summers et al. [1999].

Experiments Supported by the EVHL

Improvements in computer aided design, collaborative computing environments, video games, and telerobotic surgery are a few of the applications for the goal-directed human hand studies and analyses that are underway. Instead of using traditional human-computer interaction devices such as a mouse or keyboard, we are exploring hand pointing, grasping, and gestures in a 3-D augmented environment.

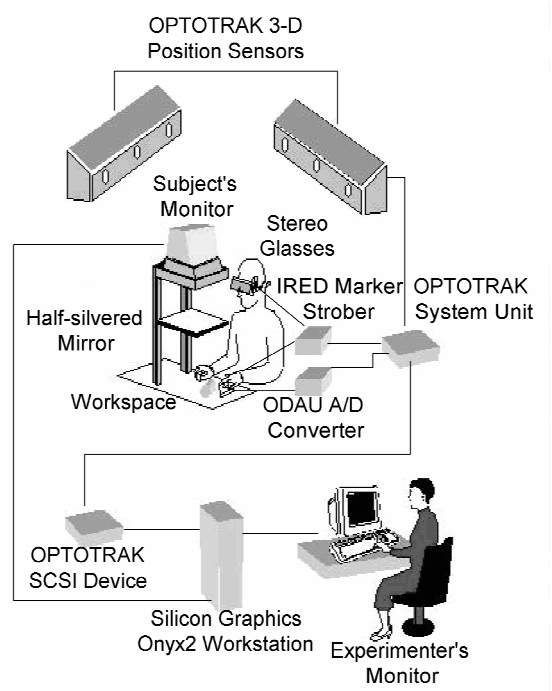

Figure 1. The Enhanced Virtual Hand Lab

The positions and orientations of subjects and objects in the working volume are tracked by one or more of the OPTOTRAK 3-D position sensors. Additionally, analog data, such as the force exerted by a subject's fingers on a grasped object or electromyography signals, are collected and time-stamped by one or more ODAU A/D channels. Both types of sensing data are available to the EVHL software, and are used to maintain a model of the physical environment in which additional virtual objects may also be present.

A typical experiment is one that analyzed how a subject positioned and oriented a small wooden cube within a workspace [Wang et al., 1998], [Wang, 1999]. Human performance was analyzed by monitoring how a subject moved the wooden cube under a variety of graphic and haptic conditions. As the subject moved the wooden cube towards a target, the OPTOTRAK recorded the position and orientation of the cube every 1/60 second. Complete trajectories of the wooden cube and the subject's hand were recorded and subsequently analyzed. In some experimental conditions the subject did not see the physical cube, but instead saw a virtual cube at the same location. Within a random subset of those trials, the virtual cube was rotated 60 degrees from the orientation of the physical cube. The goal of the experiment was to determine the effect of a mismatch between haptic and visual information on a subject's hand movements by perturbing the correspondence between the physical and virtual objects.

A more advanced set of experiments is currently underway using a wooden cube with embedded force transducers. A subject will pass the cube back and forth to the experimenter under a variety graphic and haptic conditions that do not always agree. Perturbations will include the

location, orientation, and size of the virtual cube in contrast to the actual location, orientation, and size of the physical cube. When a user grasps or releases the physical cube, the EVHL software will detect the contact forces using the ODAU's A/D channels. The position and orientation of the cube and of both the subject's and the experimenter's hands will be recorded for subsequent analysis.

System Overview of the EVHL

Figure 2 illustrates the data flow that occurs as a subject interacts with the EVHL system. Each of the system components is described in Table 1.

Table 1. System Component Descriptions

Component	*Description*
Active Marker (IRED)	Emits infrared radiation that is detected by linear charge-coupled devices within the OPTOTRAK 3-D position sensor.
OPTOTRAK 3-D Position Sensor	Determines 3-D co-ordinates in real-time by sensing IREDs. Each IRED can be sampled at frequencies up to 3200 Hz with a spatial accuracy of 0.1 mm.
Analog Signal	Represents external system inputs such as force data, electromyography data, etc.
ODAU A/D Converter	Digitizes up to 16 analog inputs in real-time. Each analog channel can be sampled at frequencies up to 90 kHz and accuracy of 5 mV.
OPTOTRAK System Unit	Collects and synchronizes data from the OPTOTRAK 3-D position sensor and ODAU A/D converter. Up to 4 MB of raw data can be stored in the System Unit's temporary buffers.
SCSI Device	Facilitates real-time communication between the OPTOTRAK System Unit and Onyx2 workstation using specialized firmware.
SGI Onyx2 Workstation	Runs specialized graphics and data collection software written with OpenGL and C++ on an IRIX 6.5 platform. Contains two 250 MHz processors, 1 GB RAM, a Reality Engine graphics card, and a single graphics pipe.
Monitor	Projects onto a half-silvered mirror to create augmented graphic objects. Contains a 350 mm x 250 mm viewable area and is refreshed at 60 Hz with 1280x1024 resolution (or two 640x1024 fields in stereo mode).

As illustrated in Figure 3, both the OPTOTRAK IREDs and the ODAU A/D channels are sampled at a minimum of 240 Hz to update the model of the user's workspace. This data set, or "frame", is time-stamped and synchronized by the OPTOTRAK System Unit and stored into a set of temporary registers within the System Unit. Additionally, the previous data frame is transferred from the temporary registers into a circular memory buffer within the System Unit. Every 1/60 second, the Onyx2 workstation reads the registers containing the most recent frame of 3-D position data and analog data via a specialized SCSI interface. Because the Onyx2 workstation only reads the OPTOTRAK System Unit at 60 Hz, many data frames within the OPTOTRAK System Unit are not read in real-time (three of every four frames in this case). During periods of inactivity, the unread (buffered) data can be transferred to the Onyx2 via block transfer mode. Since up to 4 MB of data can be buffered, the transfer may take several frame times. After the data transfer is complete, further processing, such as offline data analysis or calculations for use in real-time prediction algorithms, can be performed on the full data set.

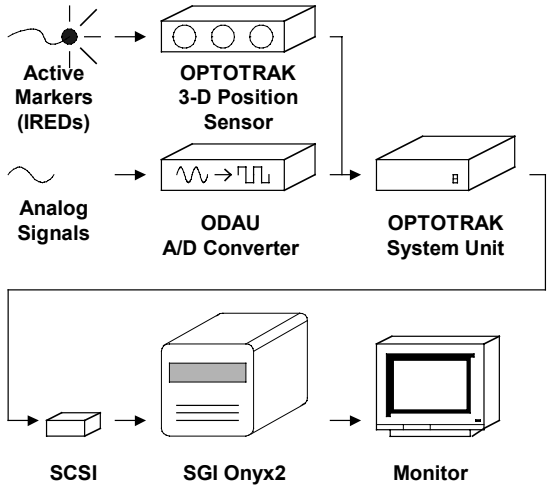

Figure 2. System Lag Data Flow

Retrieving data from the OPTOTRAK System Unit requires two steps – a prefatory "request" followed by a data "get". Alternating request/get commands for the OPTOTRAK and ODAU are shown in the top portion of Figure 3.

Once the most recent frame of OPTOTRAK data is obtained by the Onyx2 workstation, left-eye and right-eye scene graphs are updated and rendered. This is based on a model of the physical world that includes the objects on which sensors have been placed and additional virtual objects that are not present in the physical workspace. The middle portion of Figure 3 illustrates this.

These scenes are successively displayed in field-sequential stereo, as in the bottom portion of Figure 3, left eye first, right eye second, on a monitor with a 1280x1024 screen resolution and 60 Hz refresh rate (640x1024 for each field and a 120 Hz field rate). The monitor projects the scenes onto a half-silvered mirror so that the subject can see the physical and virtual workspaces superimposed as a single augmented workspace (see Figure 1). The face of the monitor, the mirror, and the surface of the table are all parallel to each other and to the floor.

Temporal System Lag Model

As shown in Figure 3, our end-to-end system lag model is approximately three 1/60-second frames (i.e. 50 ms). We

define end-to-end system lag as the time duration between the sensing of the user (i.e. 3-D position markers) and when the graphical feedback is displayed onto our half-silvered mirror. The worst-case lag is for pixels at the bottom right corner of the right-eye field, which will be displayed last; the best-case lag is pixels at the upper left corner of the left-eye field, which will be displayed first. Because the OPTOTRAK System Unit and Onyx2 clocks are not synchronized, processes on both systems will be out of phase by some amount. Additionally, the slack time at the end of a software update (see Figure 3) may overlap with the start of the monitor refresh. For our calibration, we disabled stereo viewing because we used a photographic camera to view the images. The timing diagram for *Monitor Refresh* in Figure 3 is thus for the monocular (non-stereo) case. The timing diagram for stereo is effectively the same diagram "doubled up", with half of the scan lines in each field.

Each of the following groups of system components contributes to a one frame delay because the three stages of processing shown in Figure 3 are pipelines, that is they take place sequentially as data flows from the sensing system to the modeling & rendering system, and finally to the refresh buffer. Of course each of these stages is actually operating in parallel, the essence of pipeline architecture, but each with data for a different frame.

- The OPTOTRAK System Unit and related components (3-D position sensor and A/D converter) buffer data that is retrieved by the Onyx2 every 1/60 second.
- EVHL software on the Onyx2 workstation has a basic cycle time of 1/60 second during which it retrieves the sensed data, updates its model of the workspace, and renders left-eye and right-eye images appropriate for the subject's current eye positions as calculated by the model.
- The 120 Hz field sequential monitor refresh requires 1/60 second to read both the left-eye and right-eye images from the Onyx2's frame buffer and display them. The stereo glasses require some time turn on and off between fields, but this does not add to the system lag, so we do not consider it here. (If the switching time is too long, there will be "bleeding" of one field into the other, so the left eye sees part of the right eye's image and vice versa, which will effect the quality of the stereo image, but not the lag.)

The above model is specific to our hardware and software configuration, but the general pipelined structure is similar in many systems. There is a sensing stage, a modeling & rendering stage, and a display or refresh stage that all operate in lockstep as a pipeline. Each phase is, at any given time, working on a different frame. When frame N is being refreshed from the frame buffer, frame N+1 is being rendered and data for frame N+2 is being acquired by the sensing system. Differences between systems include the nominal frame time (or update rate), which can be much less that 60 Hz if modeling & rendering are complex. Other differences may be due to multi-threading (on a single processor) or true multiprocessing (when more than one processor is available). In this case, there may be both pipelined parallelism and parallelism due to multiple CPUs in addition to the multiple processors that comprise the sensing, modeling & rendering, and display refresh subsystems.

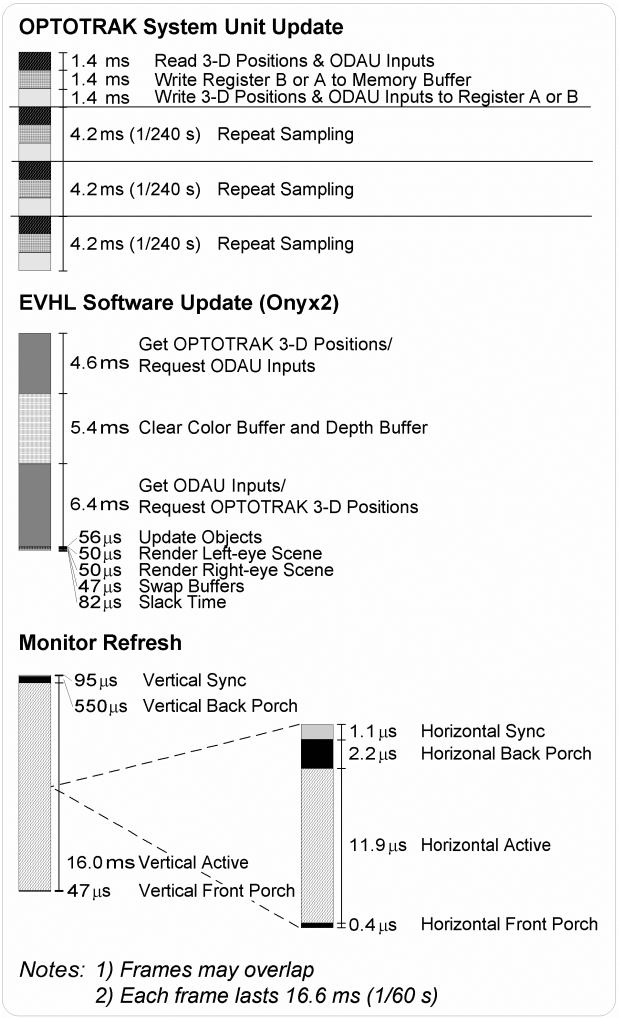

Figure 3. Temporal System Lag Models

Some of work cited later in the *Related Research* section addresses these issues through mechanisms such as "just in time" use of sensing data to make image-based "warping" corrections that partially compensate for movement that takes place between the time the sensing subsystem data is used to update the model and the time rendering is completed. These and other embellishments on the basic three-stage pipeline can be adapted to our model.

We believe that our calibration techniques are equally applicable to these more general architectures because our techniques are independent of the specific internal details of the architecture, relying mainly on measurements that compare the inputs to the system with its outputs. Our earlier paper provides a more complete description of the

system and the calibration steps required to ensure accurate estimates for the eye positions and the positions of objects in the workspace [Summers et al., 1999]. A prediction for frame rates (number of frames per second) and lag (delay between when sensed data is acquired and when it is finally seen by the user) was based on specifications provided by the OPTOTRAK and Onyx2 manufacturers (Northern Digital, Inc. and Silicon Graphics Inc., respectively). To confirm the theoretical model, tests were performed using the EVHL software to ensure that no data was lost when recording 3-D position and analog data at 240 Hz using the OPTOTRAK equipment. This confirmed that the Onyx2 and OPTOTRAK remain relatively synchronized (i.e., the frame rates were the same, even though the phases may have differed by a constant amount) during an experiment. Multiple timing tests of the EVHL software's 60 Hz display frame rate agreed within 2%. In addition, the monitor's horizontal and vertical refresh signals were verified to be within 5% of the predicted values using an oscilloscope. This confirmed that the system throughput was the predicted 60 Hz frame rate. Measurement of lag was performed visually, and estimated to be about three frames (50 ms), again in agreement with the theoretical model.

While such visual estimates of lag are adequate for the experiments we have run to date, future experiments will require more accurate estimates of lag. Although the OPTOTRAK System Unit collects 3-D position and analog data within 1/240 second, the OPTOTRAK System Unit and Onyx2 are not currently synchronized to the same clock. This means that both the frequency and phase of the two clocks might differ. Consequently, a data frame collected by the EVHL software from the OPTOTRAK System Unit buffer could be up to 1/240 second older than necessary (i.e. data could be read by the Onyx2 just before a new data frame is collected by the OPTOTRAK). A solution to this problem is currently being explored, and will need to be fully resolved before implementing additional modalities into the EVHL system – such as haptics and audio. The current work is the first step. We assume that the frequencies of the two clocks are quite close, and thus that the phases drift only slowly with respect to each other.

A second concern is that as we develop graphic scenes more complex than a small collection of geometric primitives, we must reduce the large delays of 4.6 ms and 5.4 ms associated with reading real-time data from the OPTOTRAK System Unit. Unfortunately, the 3-D position data and analog data cannot be read at the same time due to OPTOTRAK system limitations. The obvious solution is to overlap the collection of sensing data with the modeling & rendering, but this adds to overall lag because the data will be one frame "older" when it is used. This is a standard tradeoff, encountered in most systems and solved using predictive methods that will estimate current position based on the sensed data for previous positions. Similar predictive schemes might also be used to estimate more accurate force information. Because of the possible differences between the OPTOTRAK and the Onyx2 clocks, it is important to have estimates for the various lags that are accurate to sub-frame resolution.

3. CALIBRATION TECHNIQUES

We have devised two different calibration procedures for estimating lag. The first is similar to methods used in other systems, and requires little specialized hardware. The second is a novel approach that uses the specialized OPTOTRAK sensing hardware to support on-line calibration of lag during an experiment.

External Calibration

We placed three IREDs on a physical disk, and rotated it in the workspace using a modified phonograph turntable (see Figure 4). The turntable's angular velocity can be adjusted from 0 to 100 rpm using two knobs connected to potentiometers – one for coarse adjustment and one for fine adjustment. Using the OPTOTRAK 3-D position sensor and custom software, we determined the turntable's period within 1 ms and observed no discernable speed variation over a number of measurements. This was accomplished by mounting an IRED on a 16 mm diameter physical disk attached to the turntable using a mechanical linkage constructed for this purpose. The disk was black, to make it easy to see any virtual objects that might be displayed superimposed on it.

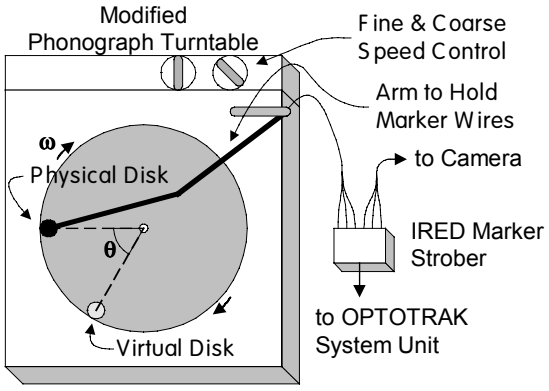

Figure 4. Turntable for External Calibration

As the physical disk spins on the turntable, the OPTOTRAK determines the IRED's position and sends the 3-D co-ordinates to the Onyx2 workstation. A white virtual disk is then rendered, ideally superimposed on the physical disk. In practice, by the time the virtual disk is rendered and displayed, the physical disk will have moved such that the physical and virtual disks are separated by θ radians. Measuring θ and knowing the turntable's constant angular velocity, ω, allows us to calculate the system lag, t_l, using a simple formula:

$$t_l = \frac{\theta}{\omega} \tag{1}$$

There is a trade-off between blur and resolution when choosing an angular velocity. Increasing the turntable speed will increase the angular difference, θ, between the

virtual and physical disks, thereby increasing the spatial resolution used for system lag calculations. However, faster moving objects will blur more when captured using still or video photography. Faster shutter speeds also let less light into the photographic equipment, so more light is needed to obtain a discernable image. Unfortunately, increased room lighting tends to 'wash out' the monitor display (see Figure 3.18 in the textbook by Ware [2000]).

We photographed the physical and virtual disks with an analog (film) camera placed where a user would normally sit (see Figure 5). The position and orientation of the camera were tracked in real-time by monitoring three IREDs on the side of the camera, much as a subject's head is tracked during experimental use of the EVHL. We used ASA 3200 black-and-white film, and a shutter speed of 1/250 second to minimize motion blur. Because light from a flash would reflect off the half-silvered mirror, the turntable was illuminated by a light placed under the mirror, and the room was kept dim, with only indirect lighting. Black-and-white film was used to avoid color shifting phenomena under the fluorescent lighting and for increased resolution.

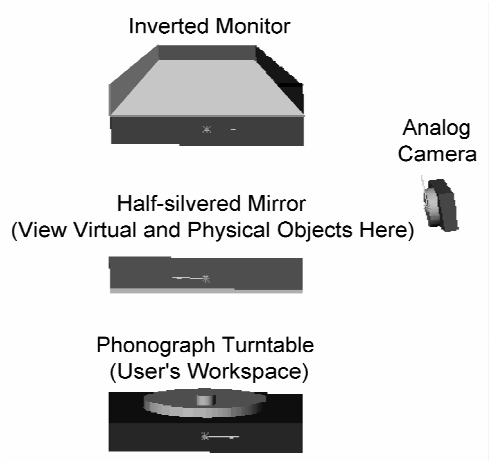

Figure 5. Workspace Setup for External Calibration

Because the physical disk rotates in a known plane on the turntable, we only need one IRED (not three) to determine the disk position at any point in time. However, physical objects with up to six degrees of freedom (DOF) are usually tracked with the EVHL system during experiments, and these require three IREDs. Tracking these 6-DOF physical objects, and rendering appropriate augmented objects into the workspace, requires more time than the corresponding operations for objects with fewer degrees of freedom.

To ensure that we obtain worst case estimates for lag, our external calibration reads the full complement of IREDs, even though some of the data is not required. It also performs a general position and orientation calculation to ensure that the computational overhead is as high as for regular usage. In this way we know that we have measured temporal lag that would be experienced by a subject in an experiment.

To verify accurate calibration of the virtual disk to the physical disk, four photographs were taken with the turntable in a stationary position. Photographs were taken with the physical disk at the left, right, front, and back sides of the workspace. An example photograph is shown in Figure 7a. At the left side of the photograph, the white virtual disk is closely superimposed on top of the black physical disk, as we would expect.

Immediately after taking the pictures of the stationary turntable, a series of 62 photographs was taken with the turntable spinning at 100 rpm. Figure 7b shows an example photograph with the virtual disk lagging the physical disk by about 30°. On average, the virtual disk was found to lag the physical disk by 29.5°, with a standard deviation of 0.6°. This corresponds to a 49 ms delay, which closely matches our theoretical system lag model of 50 ms (i.e. three 1/60-second frames). Our measurement provides a worst case lag because the virtual disk is being displayed at the very bottom of the screen (last scan line) when it is at the left side of the photograph.

Some of the photographs revealed only a faded virtual disk or no disk at all. This behavior was expected, because the fast phosphor (persistence is 10-60 μs) does not fluoresce for the entire 1/60-second period between refresh scans. A back-of-the-envelope calculation suggests that the virtual disk will be faded or invisible in one half to three quarters of the photographs for this reason.

As an additional check, we analyzed the blurry black angular lines in the photographs (see Figure 7b) to verify that the LP turntable was spinning at 100 rpm. The photographs were taken with a shutter speed of 1/250 second, so the black angular lines should blur by about 2.5° if the turntable rotates at 100 rpm. This was determined to be the case, so we are confident that the turntable was spinning at a relatively constant speed of 100 rpm. We further confirmed this using exposures of 1/125 and 1/500, for which we observed double and half the blur angle, respectively, and by examining the 3D position data obtained from the OPTOTRAK, which indicated a periodicity of 100 rpm for the IRED position.

Internal Calibration

We fed the red, green, and blue (RGB) monitor signals into the ODAU A/D channels, and then used the EVHL software to render a colored scene on an initially black monitor. Full intensity pixels (RGB) have a voltage 1.13 V higher than non-illuminated (black) pixels. Consequently, as the screen changes from black to a color, the EVHL software should detect a voltage increase on one or more ODAU A/D channels. The time from when the EVHL software renders a colored scene until the software detects a voltage change on an ODAU analog channel represents the system lag. The EVHL log files, containing the 240 Hz ODAU data, were analyzed to determine the number of input samples until the display output voltage changed as a result of change in display color from black to full intensity.

Sub-frame timing information was obtained by partially filling the screen from the bottom with a solid color. As less and less of the screen is colored, the ODAU A/D channels detect voltage changes later and later in a frame because the monitor refreshes from top to bottom. Point "V" in Figure 6 depicts a screen that is half colored. As more or less of the screen is colored, point "V" shifts left or right, respectively, thereby providing sub-frame timing information.

For calibration purposes, we modified the ODAU sampling rate for one channel to 9000 Hz, allowing us to observe subtle sub-frame voltage changes. This provides 150 samples per display frame, roughly one for every eight or nine visible scan lines.

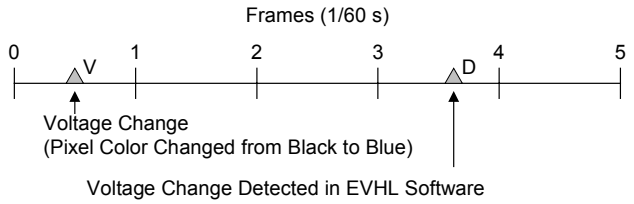

Figure 6. Timeline for Internal Calibration

Unfortunately, the RGB monitor signals are only active for 75% of the time (see Figure 3) because approximately 25% of the signal is devoted to vertical and horizontal retrace and other sub-signals. The green signal has horizontal and vertical synchronization information, whereas the red and blue signals contain only vertical synchronization information. We do not have precise control over the sample time of the ODAU A/D converter. This means that a signal may be sampled during its synchronization phase instead of its active phase. Consequently, we may detect no voltage change even though the signal is active. To compensate for this problem, we used a blue signal (not green), and conducted tests in batches of 10 or 20 trials. The lowest time lag estimate within a batch of trials was taken to represent the actual system lag.

This technique worked reasonably well, indicating an average lag of 1.7 frames from the time the software first begins updating its scene graph until the full-intensity color is recorded in the OPTOTRAK's log file. One additional frame should be added for a complete closed loop, for a total of 2.7 frame times, which is a little less than the three frame times estimated by the external measurement. This agrees with our intuition, which is that the 60 Hz sample rate used in the run-time system is rounding the lag upward to an integral number of frames. Using the internal technique provides a better estimate.

4. DISCUSSION AND CONCLUSIONS

Lag is always a concern for augmented and virtual environments. We chose the OPTOTRAK sensors for a number of reasons, but a primary consideration was a concern for low latency in sensing 3-D positions. Liang, Shaw and Green [1991] report a 90 ms lag using a Polhemus FASTRAK sensor. Newer models of the Polhemus have less latency, and are generally considered superior to other trackers. Our measurement of a 50 ms end-to-end lag for the OPTOTRAK is encouraging, especially because this includes all of the overhead present in our actual application – we conducted our tests using the same software that runs the experiments.

Modifications to our techniques

Our external calibration measured the angular difference between the physical and virtual disks to determine lag. A more visual approach would be to allow the viewer to adjust a predictive phase angle until the virtual disk exactly superimposes onto the physical disk. This would provide a simple interactive calibration technique, although it would probably not be as accurate as our film-based approach. Similarly, we could use a video camera (as did Shaw et al.), instead of film, to obtain quicker access to the images for analysis, but this would sacrifice both spatial and temporal resolution compared to film. A second advantage of using video for the external calibration is that it could be fed directly into the system, which could employ standard image processing techniques to automate the estimation of lag.

The internal calibration was analyzed off-line, by examining log files produced by the standard EVHL software. This could easily be done entirely on-line, although we have not yet figured out how to obtain the analog sensing data at faster than 60 Hz except through the batch-mode log files. This appears to be a limitation of the OPTOTRAK system architecture. We are pursing this issue with the manufacturer.

Even limited to 60 Hz, an on-line lag calibration could be useful for maintaining synchronism between the OPTOTRAK and the Onyx2. Our internal calibration technique uses the output video signal to encode timing information, so it cannot be used during an experimental trial. But between trials, it is possible to do this. The Onyx2 supports multiple video output channels, so we could sample an otherwise unused video output channel to obtain a continuous estimate of sensor lag during experiments. Our basic assumption that the OPTOTRAK and Onyx2 are almost synchronized means that lag will not change very rapidly. Our internal calibration technique works well in this situation.

Applications of our techniques

Knowledge of system lag can be used to further improve the visual quality of the augmented environments supported by the EVHL. A precise lag estimate allows us to:

- *Compare tradeoffs between various hardware and software configurations*. For example, according to the model in Figure 3, removing the read of the ODAU A/D channels would free up about 6 ms of time that could be used to increase the complexity of graphic scenes without affecting the temporal lag. For experiments that do not involve force sensing or other analog inputs, graphics could be significantly improved.

- *Utilize predictive tracking algorithms when updating an augmented scene*. Because the temporal lag is constant, we can generate an augmented scene for a future point in time using common prediction algorithms. The augmented scene should match a subject's current state – not the state 50 ms in the past. Situations where the user quickly interacts with the system (e.g. fast hand movement) require predictive tracking with an accurate estimate of the lag time or the prediction might be even worse than no prediction at all.

- *Identify system bottlenecks and suggest methods for system enhancement*. Our initial system lag tests helped isolate a communications delay in the OPTOTRAK SCSI device that was causing data frames to be lost under certain system configurations. Once the SCSI device firmware was updated, data frames were no longer lost.

- *Aid the complete integration of additional system modalities*. Feedback of all modalities must be precisely synchronized to provide a compelling experience for subjects, and meaningful data for research experiments. Precise knowledge of the system lag enables us to monitor the subject and the environment, while continually providing feedback that is appropriately synchronized.

Related Research

Earlier research has identified end-to-end lag, or latency, as a critical problem for augmented reality systems. Ellis et al. [1997] conducted experiments using tasks similar to those in the EVHL. Their results suggest a 50 ms threshold for latency, above which asymptotic performance is degraded. Shaw et al. [1993] discuss the problem of lag in distributed virtual environments; Jacobs, Livingston, and State [1997] describe techniques for adjusting data from various sensors with different latency to achieve a consistent model of the physical and virtual environments; and Olano et al. [1995] describe a number of ways in which the rendering pipeline can be reorganized to reduce end-to-end latency. Mark, McMillan and Bishop [1997] look specifically at post-rendering image warps as a mechanism for compensating for mis-registration due to lag; they adjusted the rendered images based on just-in-time sensing data that was more recent than the data used to render the images.

Many augmented reality systems use prediction-based techniques to compensate for lag. A comparison of "Grey systems" and Kalman filtering for prediction was conducted by Wu and Ouhyoung [1995]. Both techniques improved performance when lag is as high as 120 ms, but Kalman filtering provided the least "jitter".

State et al. [1996] discuss augmented reality systems for surgical biopsies and how hybrid tracking systems can be used to gain greater accuracy. Their system uses video-based augmented reality, in which virtual images are merged with video images of the physical scene. This affords the opportunity to delay the video to compensate for lag in the virtual images. Such video augmentation is not an option for our system since we merge virtual images with the user's own vision of the physical scene. In systems where video-based augmented reality is used, our calibration techniques could be used to obtain accurate static or dynamic estimates of lag.

Our work addresses the specific question of how to accurately measure end-to-end lag. Once this is known, all of the other techniques can be applied. We intend to use our results to design more accurate prediction schemes by incorporating our static measurement of sub-frame lag, and perhaps a dynamic measurement, under the assumptions noted earlier that end-to-end lag does not vary rapidly during an experiment.

We believe that our techniques can be applied to other systems as well. The equipment required for our basic calibration, a modified turntable, is easily constructed from relatively inexpensive parts and seems better suited to many situations than the pendulum used by Liang, Shaw, and Green. Sub-frame calibration requires closed loop sensing using analog signals, which may not be available in some systems, but for those that do have analog sensing (often the case when haptic devices are included) this technique too can be used.

Future work

A few problems remain to be solved. The closed loop calibration uses the video signals as an input to the A/D channels, but the signals are not exactly what we want because of the vertical and horizontal blanking intervals (i.e. non-pixel information). The signals could be conditioned with simple filter circuits to eliminate most of these problems. This would provide a more robust internal calibration procedure, especially if the input A/D channels on the ODAU were used to monitor secondary video output channels on the Onyx2 reserved for this purpose. We would then have a completely closed loop calibration procedure that could be invoked during an experimental trial, not just between trials. This would allow us to adaptively change the parameters for prediction filters.

A similar closed loop calibration technique ought to be possible for determining the lag in audio output when we add that to the system. Connecting the audio output voltages to the ODAU inputs provides a mechanism for measuring the end-to-end delay in the audio display in a manner similar to what we have done for the graphic display. Using this to measure lag in a haptic display might be more difficult because of the much higher frequencies required for proper force feedback.

ACKNOWLEDGEMENTS

The authors thank the students, staff, and faculty of the Enhanced Virtual Hand Lab for their guidance, help, and patience. In particular, we would like to thank Valerie Summers, Frank Henigman, and Dr. Evan Graham for their help with software development, and Drs. Christine MacKenzie and Shahram Payandeh for their critiques and suggestions. Special thanks to Ron Long for shooting and developing the black & white photography.

REFERENCES

1. Ellis, S. R., Bréant, F., Menges, B., Jacoby, R., and Adelstein, B. D. (1997). Factors influencing operator interaction with virtual objects viewed via head-mounted see-through displays: viewing conditions and rendering latency. In *Proceedings of the IEEE 1997 Virtual Reality Annual International Symposium* (Los Alamitos, California), IEEE Computer Society Press, pp. 138-145.

2. Jacobs, M.C., Livingston, M.A., and State, A. (1997). Managing latency in complex augmented reality systems. In *Proceedings of the Symposium on Interactive 3D Graphics* (April 27-30, Providence, Rhode Island), ACM/SIGGRAPH, pp. 49-54.

3. Liang, J., Shaw, C., and Green, M. (1991). On temporal-spatial realism in the virtual reality environment. In *Proceedings of the Symposium on User Interface Software and Technology* (November 11-13, Hilton Head, South Carolina), ACM/SIGGRAPH, pp. 19-25.

4. Mark, W.R., McMillan, L., and Bishop, G. (1997). Post-rendering 3D warping. In *Proceedings of the Symposium on Interactive 3D Graphics* (April 27-30, Providence, Rhode Island), ACM/SIGGRAPH, pp. 7-16.

5. Olano, M., Cohen, J., Mine, M., and Bishop, G. (1995) Combatting rendering latency. In *Proceedings of the Symposium on Interactive 3D Graphics* (April 9-12, Monterey, California), ACM/SIGGRAPH, pp. 19-24.

6. Shaw, C., Green, M., Liang, J., and Sun, Y. (1993). Decoupled simulation in virtual reality with the MR Toolkit. In *ACM Transactions on Information Systems*. 11(3):287-317 (July).

7. State, A., Livingston, M.A., Garrett, W.F., Hirota, G., Whitton, M.C., Pisano, E.D., and Fuchs, H. (1996) Technologies for augmented reality systems: realizing ultrasound-guided needle biopsies. In *ACM Proceedings of the Conference on Computer Graphics* (August 4-9, New Orleans, Louisiana), ACM/SIGGRAPH, pp. 439-446.

8. Summers, V.A., Booth, K.S., Calvert, T., Graham, E.D., and MacKenzie, C.L. (1999). Calibration for augmented reality experimental testbeds. In *Proceedings of the Symposium on Interactive 3-D Graphics* (April 26-28, Atlanta, Georgia), ACM/SIGGRAPH, pp. 155-162.

9. Wang, Y. (1999). *Object Transportation and Orientation in Virtual Environments*. Ph.D. Thesis. Simon Fraser University, Burnaby, British Columbia.

10. Wang, Y., MacKenzie, C.L., Summers, V.A., and Booth, K.S. (1998). The structure of object transportation and orientation in human-computer interaction. In *Proceedings of the Conference on Human Factors in Computing Systems* (April 18-21, Los Angeles, California), CHI, pp. 292-299.

11. Ware, C. (2000). *Information Visualization*. Morgan Kaufman Publishers. San Francisco, CA.

12. Wu, J-R, and Ouhyoung, M. (1995). A 3D tracking experiment on latency and its compensation methods in virtual environments. In *Proceedings of the ACM Symposium on User Interface Software and Technology* (November 15-17, Pittsburgh, Pennsylvania), ACM/CHI, pp. 41-49.

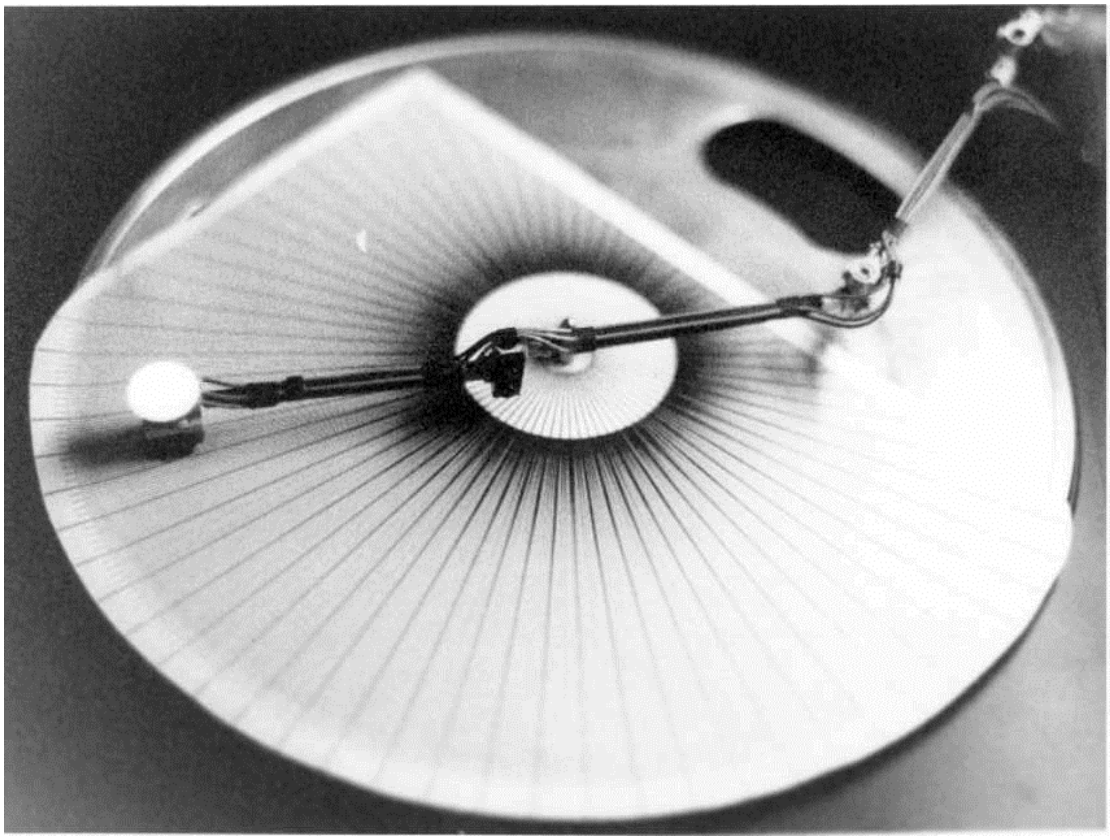

Figure 7a. Stationary Physical and Virtual Disks

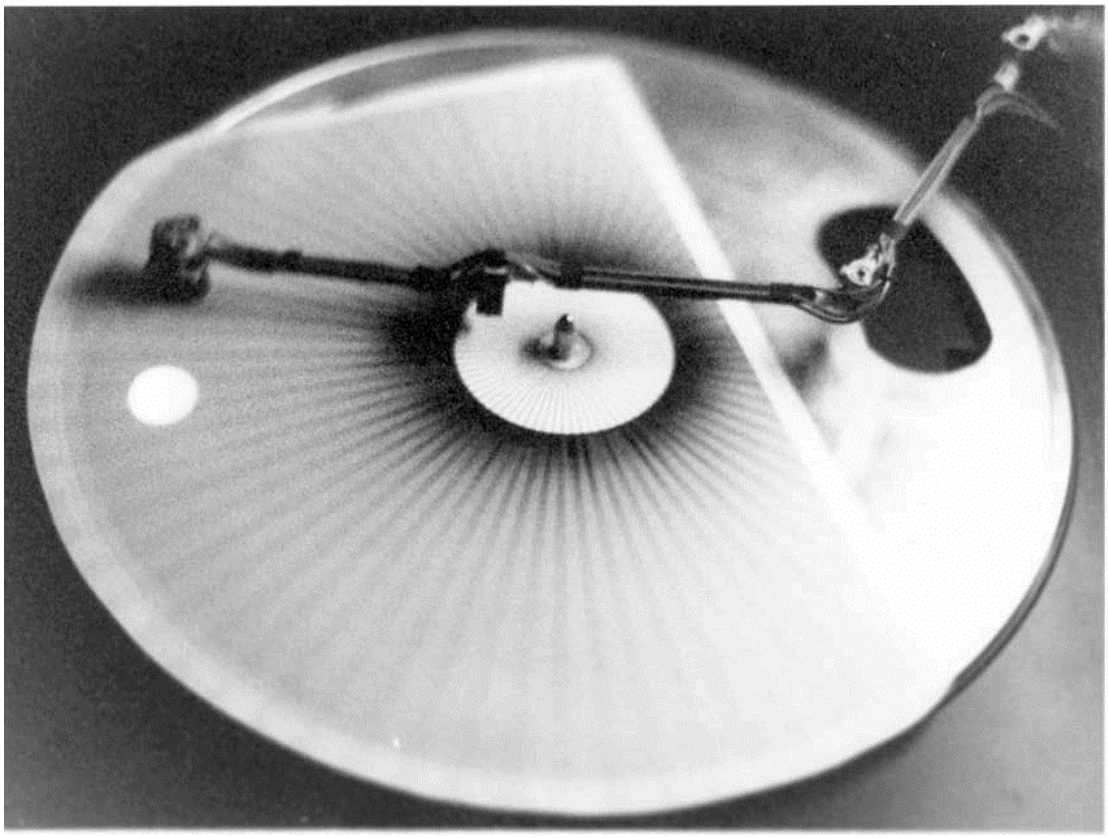

Figure 7b. Physical and Virtual Disks Rotating at 100 rpm