STINFO COPY

AFRL-HE-WP-TR-2007-0018

Advanced Visual and Instruction Systems for Maintenance Support (AVIS-MS)

Norman I. Badler
Jan M. Allbeck

Center for Human Modeling and Simulation
University of Pennsylvania
Philadelphia PA 19104-6389

December 2006

Final Report for June 2005 to December 2006

20070509258

Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Readiness Research Division
Logistics Readiness Branch
Wright-Patterson AFB OH 45433-7604

NOTICE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the Air Force Research Laboratory, Human Effectiveness Directorate, Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-HE-WP-TR-2007-0018 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE COMMANDER

//SIGNED//

DANIEL R. WALKER, Colonel, USAF
Chief, Warfighter Readiness Research Division
Air Force Research Laboratory

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE (DD-MM-YYYY) December 2006	2. REPORT TYPE Final	3. DATES COVERED (From - To) June 2005 – December 2006
4. TITLE AND SUBTITLE Advanced Visual and Instruction Systems for Maintenance Support (AVIS-MS)		5a. CONTRACT NUMBER FA8650-05-2-6649
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER 62202F
6. AUTHOR(S) Norman I. Badler, Jan M. Allbeck		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER 1710D226
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Center for Human Modeling and Simulation University of Pennsylvania Philadelphia PA 19104-6389		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Materiel Command Air Force Research Laboratory Human Effectiveness Directorate Warfighter Readiness Research Division Logistics Readiness Branch Wright-Patterson AFB OH 45433-7604		10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/HEAL
		11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-HE-WP-TR-2007-0018

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution in unlimited.

13. SUPPLEMENTARY NOTES
AFRL/PA cleared on 29 March 2007, AFRL/WS-07-0738.

14. ABSTRACT
Complexity, customization, and packaging of military platforms and systems increase maintenance difficulty at the same time as the available pool of skilled technical personnel may be shrinking. In this environment maintenance training, technical order presentation, and flight-line operational practice may need to adopt "just-in-time" procedural aids. Moreover, the realities of real-world maintenance may not permit the hardware indulgences and rigid controls of laboratory settings for visualization and training systems, and at the same time the actual activities of maintainers will challenge requirements for portable or wearable devices. This project has investigated technologies that may be used by Air Force maintainers for training or job aids.

15. SUBJECT TERMS Maintenance Aides, Augmented Reality, Virtual Reality, Instruction Authoring

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Alexander D. Nelson
a. REPORT UNCLASSIFIED	b. ABSTRACT UNCLASSIFIED	c. THIS PAGE UNCLASSIFIED	SAR	38	19b. TELEPHONE NUMBER (include area code)

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18

THIS PAGE LEFT INTENTIONALLY BLANK

Table of Contents

List of Figures

THIS PAGE LEFT INTENTIONALLY BLANK

Abstract

Complexity, customization, and packaging of military platforms and systems increase maintenance difficulty at the same time as the available pool of skilled technical personnel may be shrinking. In this environment maintenance training, technical order presentation, and flight-line operational practice may need to adopt "just-in-time" procedural aids. Moreover, the realities of real-world maintenance may not permit the hardware indulgences and rigid controls of laboratory settings for visualization and training systems, and at the same time the actual activities of maintainers will challenge requirements for portable or wearable devices. This project has investigated technologies that may be used by Air Force maintainers for training or job aids.

There are several modalities available for the conveyance of maintenance information, including, text, diagrams, images, speech, video, and 3 dimensional models and environments as well as live demonstrations. Currently most stored maintenance information is conveyed through text and diagrams. For this project we investigated the feasibility of using more advanced technology such as head mounted displays (HMD), fusion trackers, wearable computers, unique input devices, and AR software. We experimented with merging many of the available modalities while concentrating on the feasibility of using state of the art AR hardware. We also considered authoring systems for instructions and graphical aides that could address a plethora of possible output devices. We deemed speech input or output systems inappropriate for flight-line maintenance application due to the high noise level in the environment and the relatively poor performance of such software. Our main focus was therefore on AR and instruction authoring. Locating the user's pose and view relative to the maintained part and determining the user's hand grasps was a necessary component for understanding what task the user was engaged in or trying to execute. From demonstration systems we formulated a roadmap for re-usable instruction authoring systems and derived a set of novel requirements that should be met to make AR systems viable for training or job aids.

1

1 Overview

This project involved the investigation of Augmented Reality (AR) devices for maintenance instruction presentation for training and job aids. We developed some prototype systems to demonstrate AR potentials and examine the process of authoring instructions suitable for AR or other presentation modalities. Our ultimate purpose was therefore *to propose, design, and evaluate novel approaches to maintenance task acquisition, presentation, and validation.* Rather than take a formal human factors approach to analyzing demonstration systems, we determined that requirements analysis was sufficient to provide a viable but conservative roadmap for future AR and instruction authoring systems for Air Force maintenance applications.

The first component of our study was to integrate AR technologies with electronic instructions. While we understand the Air Force's use of Technical Orders and Interactive Electronic Technical Manuals (IETMs), we did not attempt any integration with those media so far. Likewise, the exigencies of real vehicle and system shape complexity were considered in formulating the final recommendations and future roadmap, rather than being directly incorporated in the demonstration system. We did, however, allow for the incorporation and use of 3D digital models in future AR training and job aid systems.

The second component in our study had two subparts: (1) to develop concepts for and a prototype of an AR or Virtual Reality (VR) system that enables virtual training, embedded operations (job aids), and task animations with a consistent user interface; and (2) to sketch the design of authoring tools that would allow task instructions and animations to be readily constructed without adding a requirement that the instruction authors be talented animation artists. This component led us to design and prototype a system architecture called DELITED *("Describing, Envisioning and Learning Instructions Through Expert Demonstrations")* that supports these requirements.

The third component in our study was to informally evaluate the task effectiveness of AR configurations for maintenance activities and human factors surrounding the suitability and wearability of AR/VR devices. We focused less on this component because we gathered enough experience with the devices and our prototype systems that we felt it was better to recommend a roadmap for the future than to spend time analyzing substandard solutions.

We will follow this outline somewhat in the presentation. In Section 2 we describe AR and VR systems and issues, including the hardware components utilized for this project. Section 3 describes the DELITED architecture. Section 4 evaluates our experiences and provides a roadmap for future directions in AR for maintenance applications.

2 Augmented and Virtual Reality

An *Augmented Reality* (AR) system generates for the user a composite view by superimposing virtual information onto a real scene with the goal of helping the viewer to better understand the environment. The superimposed information can be text, such as instructions, or a virtual scene. Regardless of the information type, it needs to be

displayed correctly, at the right time and in the right place, to present the user with a unified, single real visual scene.

There are two main methods for displaying the AR scene for a user:

(1) Video based see-through Augmented Reality

In this method, the real scene is captured by video cameras, and then merged with the virtual scene generated by a computer. The real and virtual scenes are integrated before the user sees them. This means that the user will be viewing the real scene indirectly through the video camera recording. A diagram of a video see-through system is shown in Figure 1.

Figure 1 Video See-through AR (Vallino 2006).
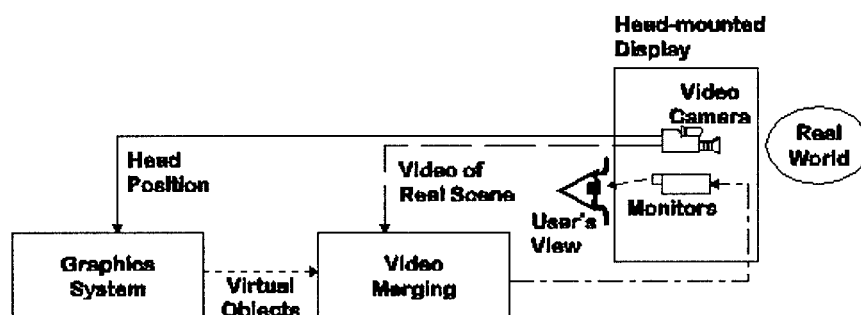
(2) Optical based see-through Augmented Reality

In an optical based see-through AR system, the user views the real scene directly, and the virtual scene is optically merged directly in the user's view, as shown in Figure 2. This optical merging can be done through the use of head-mounted displays or other projection devices.
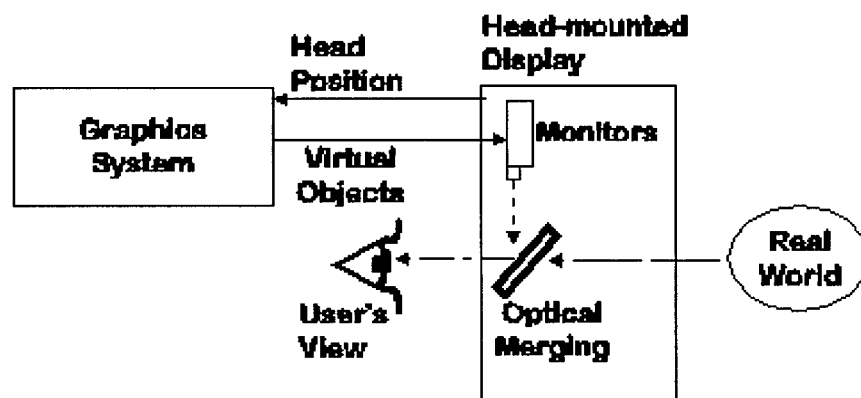

Figure 2 Optical See-through AR (Vallino 2006).

3

The greatest difference between AR and Virtual Reality (VR) is that the user in an AR system can simultaneously observe a superimposed virtual and real scene. The user in VR only views a virtual scene. VR strives for a totally immersive environment, while AR tries to merge the real world scene with a virtual scene while maintaining the user's sense of presence in the real world.

2.1 Main Challenges

The first challenge to building a successful AR system is to find a mechanism to display the real and virtual scenes at the same time, so that the virtual scene and real scene are seamlessly blended together. Video-based see-through and optical-based see-through methods are two basic solutions to solve this problem, shown as in Figure 1 and Figure 2, but there are still many open issues: *e.g.*, how to let them have the same perceptual brightness, and how to manage relative depth issues (display real objects over virtual ones and virtual objects over real objects).

The second challenge is registration (tracking). In fact, the registration problem also exists in VR and film special effects, so it is not unique to AR systems, but the requirements of accuracy and real-time performance of AR make it more difficult.

For an immersive VR system, registration is also required so that changes in the rendered scene match with the perceptions of the user, because errors here will cause conflicts between the visual system and the kinesthetic or proprioceptive (orientation) systems. Because visual perception always dominates our other sensory perceptions, a user in a VR system can accept or adjust to a visual stimulus that overrides the discrepancies with input from sensory systems. In fact, the lack of coordination between the visual and vestibular system can be exploited to make people in VR feel they are exploring a large space when in fact they are making continuous walking turns in a small area (Razzaque, Kohn et al. 2001). In contrast, errors of mis-registration in an AR system are between two visual stimuli that we are trying to fuse to be seen as one scene. AR systems are thus more sensitive to these errors.

Another challenge comes from the real time performance requirement of AR. Because the real environment is a true real-time environment, any delay or lag time in computing and displaying virtual objects will be more visible in AR when they are presented with the real scene at the same time. So a successful AR system should run as fast as the real environment, and have some mechanism to make these two scenes run synchronously.

The main challenges for AR are summarized below:

(1) Displays
    a. **See through**: AR needs see-through displays to show the real and virtual scene at the same time. But current see-through displays do not have sufficient brightness, resolution, field of view, and can not seamlessly blend a wide range of real and virtual imagery.

    b. **Delay**: Some display delay in VR may be tolerable, but any mismatch between a virtual and real scene and will make an AR system fail.

4

c. **Occlusion (Kiyokawa, Kurata et al. 2000):** Augmenting a scene need not only *add* objects to a real environment but also has the potential to *remove* them(Azuma 1997). To maintain the correct visual relationship between virtual and real objects, some real objects may be blocked.

d. **Parallax error:** Most video see-through displays have a parallax error, caused by the cameras being mounted away from the true eye optical axis.

e. **Fixed eye accommodation:** Most displays have fixed eye accommodation (focusing the eyes at a particular distance).

f. **Multimodal display:** Sometimes AR requires mixed real and virtual modalities other than just the visual modality, such as sound or haptics. There has been little work in this area.

(2) Tracking

Tracking and sensing are used to report the locations of the user and the surrounding objects in the environment, which is also the basis of registration. AR places stringent real-time demands on trackers and sensors in three areas:

a. Greater input variety and bandwidth;

b. Higher accuracy;

c. Longer range.

(3) Registration

One of the most basic problems for AR systems is the registration problem. An AR system should align its virtual and real scenes correctly, to make them appear to be in the same space. AR again presents stringent real-time and positional accuracy requirements.

(4) Interaction modality

In an AR system, objects are either real or virtual, but virtual objects cannot present haptic (physical solidity and weigh) cues. This discrepancy is a challenge for interactive AR systems.

(5) Authoring and tools

Creating the content for AR environment including 3D models, text, overlays, and interactions is also a challenge. Creating and storing semantic information with the geometric models would ease this task.

During the course of this project we examined both optical see-through AR and VR.

2.2 Head Mounted Displays

For this project we worked with three different types of head mounted displays, a true AR display (nVision Datavisor), an opaque display (eMagin Z800 3DVisor), and a sliver display (MicroOptical SV-9 PC Viewer).

n-vision datavisor

nVision Datavisor

We were able to borrow the Datavisor from another laboratory here at Penn, but normally this HMD can be purchased for approximately $25K (including the see-through option). With the see-through option, this device is capable of true AR. It allows the viewer to see virtual imagery on top of a real scene.

We were using an older version of this device, but the state-of-the-art version allows 1280 x 1024 resolution with 80° monocular field of view (FOV) and 120° maximum horizontal FOV at 180Hz and 24 bit color. The major drawback of this device is its size and weight. We feel that it would be much too cumbersome for the flight-line maintenance application. The version of the hardware that we used for testing was also not wireless. In addition to the cumbersome HMD, it required at very large control box.

eMagin Z800 3DVisor

eMagin Z800 3DVisor

We purchased this display for approximately $900. It is a much less cumbersome device and therefore more wearable device than the nVision Datavisor. It also includes a control box, but it is considerably smaller and much more portable.

This eMagin display does not afford AR. It is a completely opaque display that does support VR applications. This device would be feasible for VR training applications, but for operational environments, it does not seem feasible. The maintainer would be required to remove and replace the device throughout the maintenance task. This comfortable device has 360° horizontal FOV (with tracking device), 24 bit color at a resolution of 800 x 600, and includes stereovision. It also weighs less than 8 oz. and is USB-powered.

6

MicroOptical SV–9 PC Viewer
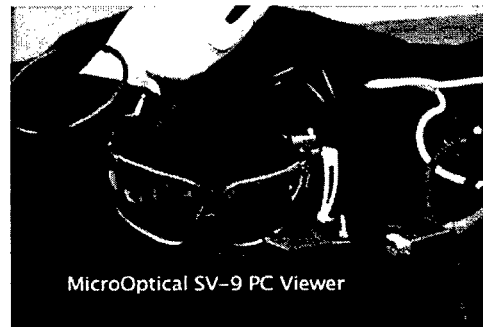
MicroOptical SV-9 PC Viewer

This display was purchased for approximately $990. This sliver display can be mounted on most glasses including safety glasses, as shown here. Unlike the eMagin visor, it obstructs only a small portion of the wearer's FOV and can be easily flipped out of the way when not in use.

Like the eMagin display, it does not afford true AR. It does, however, permit simultaneous viewing of both real and virtual scenes, though they are not superimposed. It also does not support stereovision, but out of all of the display devices this is the most wearable and practical for the flight-line maintenance application. It displays 24 bit color at a resolution of 640 x 480 and 60 Hz. It can be configured for the left or right eye and has 14° horizontal FOV. It is battery powered with a fully charged battery lasting approximately 3 hours.

2.3 Wearable Computer

We looked at a few different wearable computers, but settled on the Sony Viao U50, because it was recommended by colleagues, was lightweight, has good battery life, is affordable, is relatively powerful, and runs a standard operating system.


Sony Viao U50
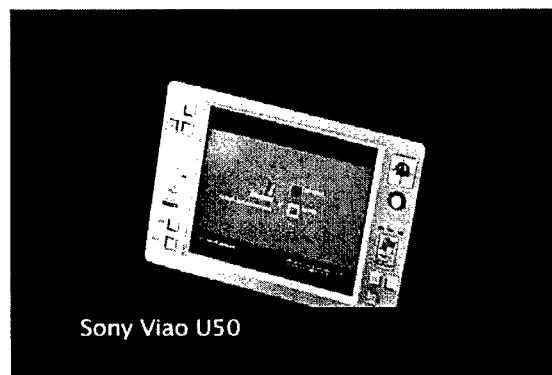
We purchased the Viao for approximately $2800. The basic specifications include Intel Celeron M 900MHz processor, 512MB RAM, 20GB hard-drive, 64BM VRAM, 5" display, 800 x 600 on screen resolution, and enhanced battery life of 5.5 hours (2.5 standard). It weights 1.21 lbs. and runs Windows XP. Though we have not yet made use

of it for this project, it also includes a touch screen. One notably missing feature of the U50 is a microphone port. This missing feature would make voice activated applications more difficult. It is possible to connect microphones through the USB ports.

2.4 Input Devices

When considering how a maintainer might interact with an instruction delivery tool, we considered traditional input mechanisms (*i.e.*, keyboards and mice). We feel that these tools are not optimal in the maintenance environment. Focusing on a computer screen and mouse and keyboard distracts from the maintenance task. Additionally, the grimy nature of maintenance is not conducive to these devices. Hence, we decided to experiment with the use of CyberGloves and hand gestures. They may be worn under traditional work gloves that would help to protect them.

The other input device that we tested is the Intersense Fusion Tracker. An important interaction with a virtual environment is synchronized movement of the eye and the virtual camera. For a maintenance task this includes positioning the camera in the virtual environment to the same point of view as the maintainer has in the real world. For this purpose, we included the fusion tracker in our experimental demonstration.

Immersion Corp. Wireless CyberGlove II
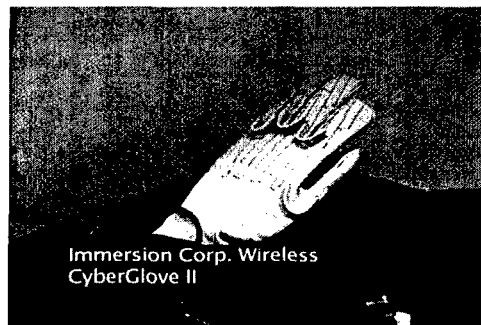
Immersion Wireless CyberGlove

This new wireless CyberGlove II system provides 22 high-accuracy joint-angle measurements. It uses resistive bend-sensing technology to transform hand and finger motions into real-time digital joint-angle data. Each sensor is extremely thin and flexible being virtually undetectable in the lightweight elastic glove. The basic CyberGlove II system includes one data glove, two batteries, a battery charger, and a USB/Bluetooth technology adapter with drivers. The CyberGlove has 0.5° resolution and repeatability to 1° . The typical data rate is 100 records per second. Its operating range is within a 30 foot radius of the USB Bluetooth adapter.

Intersense IS-1200 Fusion Tracker
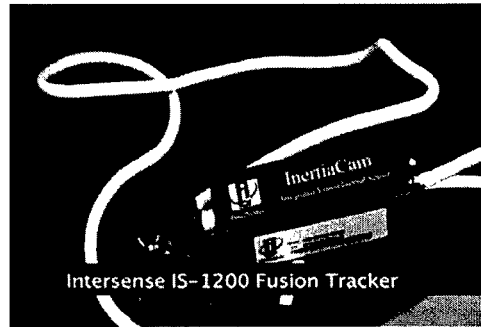
Intersense Fusion Tracker IS-1200

This is a wide-area, wearable, 6-DOF hybrid tracking and navigation system designed for AR and mobile computing applications. It uses an inertial tracker for orientation and an optical sensor for position. Its accuracy is 0.1° in orientation and 3.0 mm. in position. Circular data matrix fiducials provide up to 32,000 unique position references. The update rate is 180 Hz and it can be interfaced via Ethernet, shared memory, USB, or RS-232.

2.5 Demonstration Application

We designed a demonstration to test the individual devices as well as their interactions and applications to instruction delivery. We chose to center our demonstration on a piece of hardware that was readily available to us and has some degree of complexity, our (old) video editing rack.

Video Rack

The demonstration involves a user wearing a display device, CyberGlove, Sony Viao, and Fusion Tracker. Instructions are displayed on wearable viewer and hand signals from the CyberGlove allow the user to cycle through the instructions and activate and deactivate the devices. The Viao is the central controller for the system, running all of the necessary software and permitting the user to be entirely untethered. The Fusion Tracker can be used to track the position and orientation of the user and thereby customize the view of the virtual rack being displayed. The accompanying video shows a user wearing the MicroOptical display and CyberGlove to properly setup the video rack and copy a tape.

9

While the overall application was straightforward, the knowledge gained about the devices and issues present in instruction delivery applications were invaluable. Our first consideration was the display devices and their feasibility in this application. For the most part, all three of the display devices were easy to get working. All that was required was to properly set the display resolution and refresh rate. The DataVisor and 3DVisor had additional possible settings. The DataVisor is capable of true AR allowing the virtual and real world to merge. This would be ideal for this application facilitating the highlighting of real objects with virtual designations and information. However, it quickly became apparent that the DataVisor was not well suited for practical application. The device is rather heavy and awkward; performing maintenance instructions while wearing it would be quite difficult. The 3DVisor is much less cumbersome, but it is not a see-through display, completely blocking the user's view while it is being worn. This means that the user would have to remove it before doing the maintenance instruction and replace it again to get more information about the task.

The compact design of the MicroOptical display makes it the most feasible display for this application. It addition to its unobtrusive design that can be mounted on many different types of glasses, this display can easily be flipped out of view entirely. As with all such display devices, the resolution is small (640 x 480) and font, font sizes, and color need to be carefully chosen to ensure that the user can easily read the information being displayed. Certain things that are taken for granted when designing large size (workstation display screen) interfaces become a challenge. Font choice is important, because text needs to be legible at a small resolution. Fonts "sans serifs" are easier to read when they are smaller and bold-facing them is helpful. Color is also an important thing to consider. On small displays, light color text on a dark background is easier to read than dark text on light backgrounds.

For our demo application we constructed a simple GUI (graphical user interface) in FLTK (fast light tool kit 2006) that displayed an image of the video rack and allowed the user to cycle through an instruction set.

When considering controllers for our application, we were looking for another unobtrusive device that is also easy to use. We purchased a wireless CyberGlove which is thin enough to be worn under work gloves. Being wireless allows unencumbered movement. We wrote a hand shape recognizer in C++ using the provided SDK (software developer's kit). The code recognizes three hand shapes/gestures; an open hand to toggle activation of the recognition system, ensuring that interaction with the system is intentional, a fist gesture to move to the next instruction, and a pointing gesture to move to the previous gesture. All hand shapes must be held for a second for recognition. The software system is actually set up such that any gestures can be used. A GUI was written, again in FLTK, to allow the user to record individualized gestures for each interactive command. This allows the user to customize the interface to any comfortable and memorable gesture set.

At this stage we can visualize instructions and images on a sliver display interacting with the application through gestures recognized from CyberGlove input. A true AR system can additionally take into account the user's point of view of the scene. This enables the system to aid the user in identifying parts and states. We used the Fusion Tracker to track the position and orientation of the user's head. A mockup of the

scene was then displayed in an OpenGL window. The Fusion Tracker is small and lightweight. It is easily mounted on a helmet or cap.

2.6 Development Issues

During the implementation of our demonstration application we encountered a few issues that needed to be addressed. The viability of the displays is stated above. In the end we feel that the MicroOptical display is a quite viable choice for maintenance instruction delivery.

Overall the CyberGlove is a well-designed and reliable device. The licensing of the SDK, however, caused a few problems. When installing the SDK a code is generated. This code is then emailed to Immersion who returns another code to be entered in the authorization software to permanently unlock or authorize the software. In itself, this is not a bad procedure; however, this procedure only authorizes the software for one user on the computer where it is installed. Installing the software on another computer or reinstalling the software on the same computer or allowing another user on the computer to use the software, requires sending and receiving a new code from Immersion. While Immersion was very prompt in sending the codes, in our lab setting and particularly on a team project this authorization procedure was less than ideal.

The SDK for the CyberGlove seems well-developed, at least for this application. We extended a few of the methods easily. Our hand shape recognition code for this demonstration was not sophisticated or robust. The code includes a tolerance in the hand shape comparisons (between the stored sample and the real-time hand shapes). This tolerance is specified in degrees for each joint angle. Some preliminary experimentation has shown that, optimally, different tolerances are needed for different people. A more robust technique, perhaps a machine learning algorithm, would correct this small problem.

The Fusion Tracker was much more difficult to get working. It is a relatively new product that was not well documented. Getting the tracker fully working required several lengthy calls to technical support and returning the tracker for repair after a firmware update. The next challenge was to get the device working through a USB port instead of a serial port. Using the USB port provides the necessary power to the tracker, whereas using a serial port requires an AC power source which would restrict movement. The tracker also requires a fair amount of set up for an environment, including calculating the size and positions of the visual fiducials and attaching them. Once this setup is done for an environment it is not required again. The hardest part of dealing with the tracker was figuring out how the data that the tracker was giving us corresponded to our coordinate system. The easy part was dealing with the API. Although not every aspect has been documented yet, it was mostly intuitive. Once the server code is running on a computer, there are only a few function calls needed to receive the streaming data.

When we started this project, the initial idea was to use the tracker to identify where the user was looking at and overlay images corresponding to certain instructions onto what the user was seeing. However, the see-through display that we had was far too bulky for that idea to be practical. An alternative to overlaying images would be to use

the tracker to find out where the user is looking, and use a static image appropriate for that viewpoint. For example, if the user is supposed to press a button on the video rack, but standing farther away from the device, then put an image of the entire device on the screen while highlighting the general area that the user should be focusing on. As the user steps closer to the device, display a closer view of the video rack.

One of the major constraints of the tracker is the need for fiducial targets. In order to get correct translation and orientation information, there must me at least four fiducials in the field of view of the tracker's camera, and the tracker must be within some distance of the targets. If this is not the case, this tracker will most likely start to drift, meaning that data being returned from the tracker states that the tracker is slowly moving or rotating in some random direction. The size of the targets dictates how far away the tracker can be located while still returning reliable data. Our setup consisted of a grid of targets four inches wide spaced roughly two feet apart. This enabled us to get consistent readings up to seven feet away. In the case of an aircraft hangar this might not be a viable solution. The targets would most likely have to be situated on the device being operated on. However, this may give rise to problems while trying to maintain a view of at least four fiducials.

The Sony Viao that we used was adequate for our demonstration application. Because it is a small, wearable device it is not very powerful. We question its feasibility as the program size and complexity increase. Since our purchase of this Viao, they have discontinued this model, but they are producing newer and slightly more powerful models.

3 The DELITED Architecture

Humans excel at learning physical tasks quickly when shown example actions and given minimal verbal instruction. Typical instruction presentations may include video of a specific task performance or written text and images. There are cost, effort, and validation issues with these traditional media: they require expert video or textual instruction authoring, the time to produce useful instructional materials may exceed given time constraints, the visual media may not include crucial views or steps, and written instructions may have semantic flaws or ambiguities.

We are pursuing a new direction in multimedia instruction authoring to address and attempt to ameliorate all of these problems. Using a subject matter expert (SME) as task performer, we directly motion capture the SME's actions in the context of the actual space. While executing the task, the SME is also videotaped and audio recorded to obtain a narrative of relevant verbal instructions, annotations, and comments. The motion capture data is used to create novel views of the pre-built 3D objects being manipulated. The audio stream is used to help segment the visual and motion capture data into more atomic actions. These actions are stored as parameterized actions so that they can be used flexibly to issue instructions in video, virtual reality 3D, or illustrated text. In addition, the parameterized actions are the basic representation for re-animation of the task with a virtual human maintainer, thus providing uniform semantic execution, visualization, and verification of the instructions.

12

In the last decade or so, Badler and colleagues Bonnie Webber, Mark Steedman, Martha Palmer, and Aravind Joshi made deep inroads into understanding the nature and semantics of natural language (NL) instructions for virtual human agents (Badler, Webber et al. 1990) (Webber, Badler et al. 1995) (Badler, Palmer et al. 1999). The major outcome of these studies was the Parameterized Action Representation (PAR) for turning textual instructions in animated behaviors for multiple individual objects and agents (Badler, Bindiganavale et al. 2000). Much of this work was architectural in nature, and prototypes were constructed for domains such as vehicle maintenance, checkpoint monitoring, and even crowd behaviors (Allbeck, Kipper et al. 2002; Badler, Erignac et al. 2002). The PAR included an action database (Actionary), a NL parser, and a simple NL generator. The underlying virtual human actions included a wide range of head, eye, body, arm (reach), and locomotion behaviors and was extensible via language commands and "standing orders" through the parameterization inherent to the PAR (Bindiganavale, Schuler et al. 2000).

While working to fill the Actionary with PAR instances, Rama Bindiganavale attacked the PAR authoring problem (Bindiganavale and Badler 1998). Populating the Actionary by hand-coding PARs was possible but tedious. We began to develop tools for learning PAR parameters by observation. Using 3D motion capture data, we obtained movement exemplars for tasks such as lifting a box with two hands, drinking liquid from a mug, or touching one's nose. As anyone who works with motion capture data knows, it requires some significant clean-up and retargeting (Gleicher 2001)in order to replay a motion on a different character, since the target likely has a different body segment sizes and lengths than the original subject. Instead of mapping the source motion directly to another character, Bindiganavale generated a PAR with enough information to characterize the salient features of the captured action. Once stored as a PAR, it could be readily re-executed (retargeted) to a different human figure. The interesting part of this is the nature of the "salient features". Harking back to the early methods Badler developed, we used motion zero-crossings and other motion change features to segment the motion capture into "chunks" that became path and reach goals for the PAR. Essentially, each motion capture performance fixed one or more constraints that were stored in a PAR. For many simple tasks, *one* performance was sufficient to establish all the semantically important constraints: that is, the action description could be generalized from only one or two examples. This worked by fixing, in parallel, one or more constraints for the salient parts of the movement: thus one performance might fix end effector reach, grasp and release targets simultaneously. By determining the constraints from the actual performance we avoided both long training sequences and explicit formulation of the training objective function. Note that not all human movements may be learned so quickly; in particular, "expert" skills may require individual practice and refinement to achieve targeting accuracy, speed, coordination, and so on. Also, other expert actions with physically complex systems will require significantly more complex manual interactions.

In parallel to our physical movement learning work we were interested in understanding what features in human movement were communicatively meaningful. Classic psychological studies of movement led to gesture types such as emblems, beats, deictics, and metaphorics (Cassell, Pelachaud et al. 1994). But an alternative view was presented by human movement observers, particularly as expressed by Laban Movement

13

Analysis (LMA) (Bartenieff and Lewis 1980). The part that interested us was the notion of movement qualities. Roughly speaking, movement qualities are the *adverbs* relative to a particular movement *verb*. Significantly, while information may be conveyed by the movement "verb", the performer's attitude toward the matter is conveyed in the motion qualities. Consider the phrase "a threatening gesture": we don't have any clue what the gesture motion (verb) actually was, but we do know its performance was perceived as *threatening*. LMA was our inspiration for a motion quality representation we called EMOTE (Chi, Costa et al. 2000).

Though first designed as an animation tool (for adding motion qualities to an existing gesture form), EMOTE's emergent utility now appears to be as an intermediate representation between movement data and communicatively meaningful (linguistic) terms. For example, in recent psychological research, Ambady has shown that students who observe "thin slices" in time of a teacher in a non-verbal presentation, produce evaluations that correlate highly with long-term evaluations of the same teachers (Ambady, Bernieri et al. 2000). Surprisingly, the thin slices can be as short as 2 seconds, and the evaluations do not correlate with physical attractiveness. The study authors have subjects score teachers on so called "molar behaviors": English words associated with personality characteristics. But these molar behaviors have not obvious behavioral (movement) definitions. We postulate that the EMOTE parameters can intermediate between motion (numerical data) and such molar behaviors. Liwei Zhao and Badler showed that EMOTE parameters can be rather reliably measured in both motion capture and video stereo vision data performed by professional LMA notators (Zhao and Badler 2005) (Zhao, Lu et al. 2001). The measurements and quality recognition were handled by trained neural nets. We are working on re-engineering this system to recognize in real-time EMOTE qualities in 3D human motion capture.

A primary objective must be to reproduce the salient features (constraints) of the task and modify motion and action qualities to suit context. Context will include knowledge of objects as gauged by the human instructor's own approach to the task. Parameterized actions are the ultimate container for generalized but contextual movement information. Others have done movement-by-example (Atkeson and Schaal 1997) (Buchsbaum and Blumberg 2005) (Siskind 2001) but none with a parameterization suitable for linguistic (instruction) connections. Recent developments in "apprentice learning" (Abbeel and Ng 2004) are relevant but do not intrinsically address the communication of non-linear features across physical and linguistic channels. Thus DELITED must manage the verbalized instructions that almost always accompany the demonstration of physical actions. The PARs created from the physical demonstration will have their constraints and parameters learned from motion data as well as the verbalizations. The two communication channels will complement one another: movements will indicate locations and trajectories while language may indicate action type and movement qualities. PARs also help insure that actions are not trainer-specific. We may also learn what makes a better expert, depending on how quickly (say, by counting back-end instruction adjustment time) DELITED gets the resulting action represented and described correctly.

To pull all these threads together requires the integrated design and prototyping of the DELITED system to generate textually useful and visually validatable instructions. The DELITED environment would capture 3D motion, digital video, and audio of an expert performing some maintenance task on a physical device. The device will be previously modeled as a 3D object with separable components and manipulable parts. The initial position of the device in the motion capture space will be known or computed, but thereafter the parts will not be separately tracked by computer vision or other direct sensing means (e.g., augmented reality assembly using Bar Coded parts (Seligmann, Feiner et al. 1996)). The idea is to understand the manipulation sequence completely from the expert's own captured actions and speech utterances. It is a major hypothesis that this can be done, though we will obviously leave the door open to using computer vision cues from the live digital video feed in the future.

3.1 Outline of DELITED Components

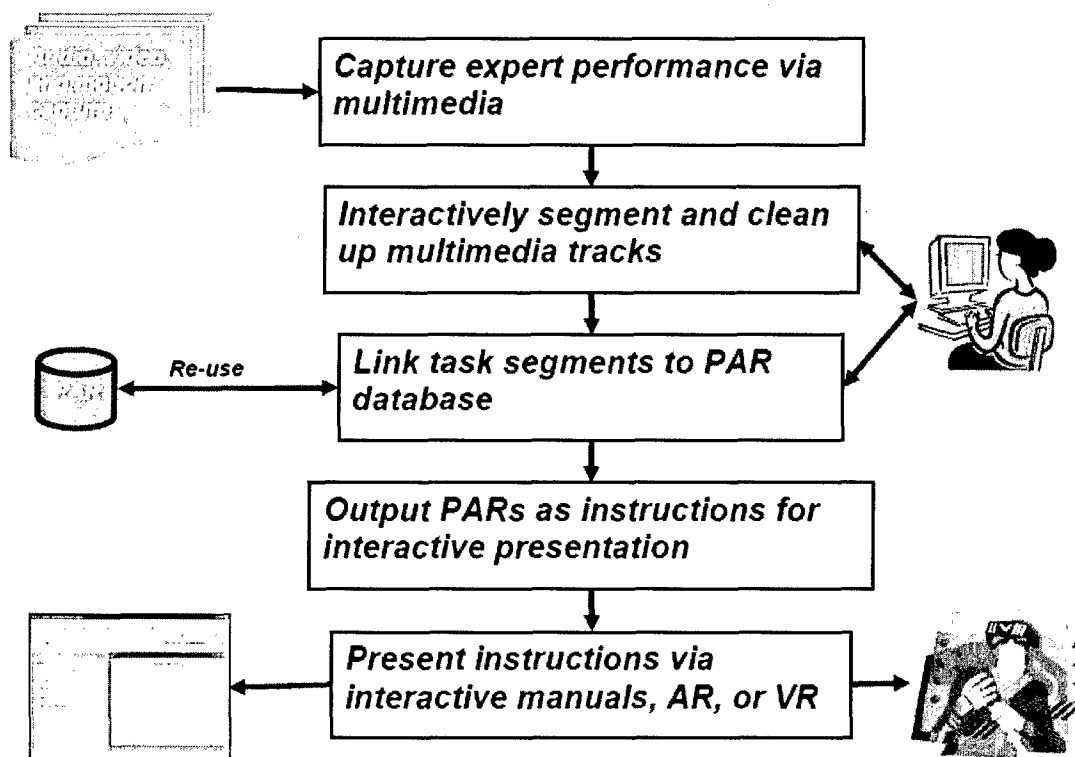An outline of the DELITED methodology follows. The general flow of information is illustrated in Figure 3.

Figure 3. DELITED data flow and architecture.

15

1. We can determine where the expert's hands are from motion capture and what handshapes they are in from the Cybergloves. A critical component of the AR system is translating the CyberGlove hand and finger joint angles into grasp types. We can determine grasp and release actions from handshape changes. Knowing what the user's hands are doing is necessary if we are to understand whether or not the user is performing the correct task action on some objects in the maintained environment.

2. From grasp and release actions and spatial proximity of object parts, we can infer the objects of the grasp or release. For grasped objects, state depends on whether another hand remains in contact, the nature of that contact (holding vs. grasping, e.g.) and physics. Having a physical simulation monitoring the situation seems prudent if not crucially essential.

3. From the motion of grasped parts we can update the 3D model to reflect state change. Note that the model itself may be under the influence of the physical simulation. Thus, e.g., we were able to model part integrity failure in the Up-Lock Hook example we did for the Air Force a number of years ago: removing one of the attachment bolts completely resulted in a spring being deprived of its pivot and retainer, so it fell out of the Up-Lock Hook assembly.

4. The 3D model of the object and its assembly must be created and marked-up with suitable semantic information such as part degrees-of-freedom, attachment types, manipulable sites, etc. Ultimately this is a considerable amount of pre-DELITED processing, but we need the mark-up until we determine how to work without it. It is possible that having the expert annotate the object interactively (verbally with manual gestures) is a good compromise that maintains the DELITED methodology: i.e., the expert can point to each part or feature and say what its attributes are. These utterances and pointing gestures can be used to tag the referenced parts with the feature tags. Thus, e.g., *"here is the power connection; it is a bayonet connector"*; *"here is the hydraulic connector; to remove it turn counterclockwise, but be sure the pressure is off"*; or *"this is the handgrip for removing the unit from the equipment bay."* We did not implement this, but it appears feasible.

5. From the motion capture stream, we can segment motions into actions. We did this in the DELITED prototype by building an interface on top of Apple Quicktime. The user sees linear editing tracks for each input modality: motion capture, audio, video stream(s). The user can manually eliminate irrelevant time segments, align motions and verbalizations, and save video or motion capture sequences as PARs. A streaming video demonstration of this system is available as "RIVET-1" at http://hms.upenn.edu/RIVET/.[1] From the audio stream we can transcribe speech to text. We can also find intonation contours and silences that may be useful in establishing action segmentation. Also, part noises may also be used to establish contacts or state changes (e.g., a component being set down on the ground might be audibly detectable). Pick-up and release of tools or other

---

[1] Note that this prototype was created as part of a complementary project with NASA.

object might be noted in speech making key connections in parameterized actions to objects that need not be detected visually.

6. Motion segmentation of action streams into PARs remains a crucial step. Durell Bouchard is studying automated motion segmentation schemes. There are several possibilities to pursue: The use of single camera video view to isolate segmentation events, find robust numerical segmentation rules, or use the emergence of specific EMOTE parameters to segment actions. We believe this is eminently realistic as some preliminary experiments by Liwei Zhao on projecting the 3D movements into a 2D plane still resulted in useful segmentations and EMOTE parameter recognition {Zhao, 2005 #118}. In addition to the motion capture data we will have the expert's audio stream. Knowing that speech instructions precede gestures, we can use speech breaks as cues to action segmentation. Our prototype relies on manual segmentations into actions (PARs).

7. While the EMOTE parameter set may be overkill for this application, it might give insight into local motion changes that are significant precursors and triggers for segmentable actions. For example, a change in the rotation speed of the thumb-index finger axis or wrist spatial path may differentiate *turning* from *loosen* or *tighten*. Posture changes may signal weight shifts that accompany movement of heavy parts or required torque application.

8. From head motion we can establish a visual line of regard and use that to augment information on the reach target.

9. Statements about cautions, warnings, and preferred practice might be included in the audio stream. The expert can be given a written check-list of items to establish prior to, during, or after the procedure is demonstrated. Some of this can even be done on a web-based fill-in (typed or menu) form: e.g., check that power is off, manage hazardous materials properly, or describe typical failure modes. There might be instances where the expert wants specific camera views. There might be a preliminary list of tools and extra parts needed and what to do with removed parts (save for re-use, stow, or dispose). Some of this information can be transcribed directly or even left in the multi-media (video and audio) part of the presentation.

10. The text processing might be facilitated by a pre-built data dictionary of tools, part, and assembly names. Also, there will be an evolving Actionary of PARs that describes actions. The names (actions) of these PARs can be used to bias the speech to text transcription to more probable utterances. Having such a limited vocabulary and allowing it to be used by the transcribing software is a feature to look for in commercial speech recognizer software.

11. From the motion segmentation, the grasp state and the parts or tools involved, PARs describing the action will be hypothesized and relevant parameters saved. Nonlinear and parallel action requirements will come from the motion capture and may also be cued by the text: e.g. *while holding the nut, insert the bolt into the hole and hand-tighten.* If one does this carefully, it might be possible to work with two individuals cooperatively (though the information capture becomes more complex).

12. Given the motion capture, video, and audio streams, plus a possible text transcription and preliminary (automatic) segmentation, the entire dataset can be presented in an interactive tool. Here the user can interactively view and fix the text, see that any PAR actions are sensible and sensibly labeled. A preliminary textual version of the instructions based on the PARs and necessary segments of the uttered narrative can be created immediately and examined. Segments of the video will be directly attached to the PARs, but the 3D models and their predicted manipulations can be created and associated with each PAR as well. A useful feature of this interface is the capability of directly authoring instructions from PARs without using captured data: by using a drag and drop menu interface to author instructions directly from the PAR Actionary and (possibly) other data sources. A prototype of this component of DELITED has also been implemented as "RIVET-2" at http://hms.upenn.edu/RIVET/.[2]

13. An animation based on the PARs can be launched as a visual instruction validation. On execution of a PAR, the 3D model can be separately manipulated for different views by the end user. Requirements here include:
    a. Need good dexterous hand model.
    b. Need robust grip and manipulation of object parts.
    c. Need motion controllers for each uninstantiated PAR type.
    d. Need to delineate object forms being manipulated, perhaps with complex manual procedures with deformable object parts. (E.g., stretching bands over pulleys, winding cable around spindle, etc.)
    e. The synthesized animation can itself be flexibly viewed as training or as an operational aid.
    f. Re-animate from a variety of body positions or orientations for individual variations or microgravity application.

14. Other issues that can be investigated and possibly handled.
    a. Semantic representation for 3D parts and assemblies.
    b. Representing part failure modes.
    c. Representing broken parts.
    d. Representing deformable, stretchable, and/or flexible parts.
    e. Animating part motion in microgravity. (Physics engine)
    f. Tool representation, database, and management.
    g. Repair components and substitutes, e.g., tape, wire, solder, etc., in an availability database (analogous to toolset).
    h. How does one move this whole information acquisition (hardware) set-up to an *in situ* site and manage real situations with complex collision avoidance, accessibility, and reach.
    i. Disassembly planner based on 3D model and semantic markup?
    j. Determining action conditionals for PAR, such as preparatory specifications, terminations, etc. Come from text and geometry?
    k. Automatic generation of re-assembly procedure as reverse of disassembly.

---

[2] Note that this prototype was also created as part of a complementary project with NASA.

l. Enumeration of necessary tools and repair items, including consumables.
m. Capability to (audio) record, save, and playback values, positions, state, etc. for proper re-assembly and operation.
n. Physical and/or kinematic simulation of modeled assembly for operation, manipulation, testing, calibration, etc.
o. Physical presentation to end user as VR or AR.
p. XML formats (DTD) for presentation interface.

People learn how to assemble, disassemble, and repair things by reading, looking, and doing – often all interleaved. Maintenance presents useful and complex procedural issues with timing, manual skills, and sensory feedback. Many of these characteristics are multi-channel: e.g., "turn nut to loosen" expresses a basic action (turn), a temporal condition (constantly), and a termination condition based on a change in resistance to movement (a motion quality). The DELITED architecture should be able to generate instructions for and simulate a significant maintenance task after being shown the basic manual skills and terminology.

4 Evaluation, Summary, and Recommendations

Throughout this project we have been concerned with testing the ease of use of these devices, their reliability, and their feasibility in the maintenance domain. We are concerned that the devices maybe too inhibiting for the maintainers, even though they are relatively compact. Ultimately, it must be determined if the benefit is worth the cost. In reviewing the benefits of using VR/AR equipment, we must consider the ease of use and type and amount of information that can be conveyed when compared to existing methods (computer screens and paper). This information must originate somewhere and somehow during the instruction authoring process.

We are interested in and would recommend investigating novel instruction authoring systems that would provide additional data to be advantaged in a VR/AR instruction delivery system. We propose expert authoring of instructions by demonstration. An expert can be captured through audio, video, and motion capture performance of a maintenance task. These modalities can then be used both in the authoring of instructions and as additional data during the delivery of the instructions. Over the past several years we have been developing a Parameterized Action Representation (PAR) (Bindiganavale, Schuler et al. 2000). We believe that PARs can be used to both recognize actions from motion capture data (Bindiganavale and Badler 1998) and fill in necessary semantics that may not be found directly in any of the audio, video, or motion capture streams. This expert authoring application is our next challenge.

5 Selected Bibliography

In-text references are listed in this section.

5.1 Surveys of Augmented Reality

[1]   Augmented Reality Page. By Jim Vallino, U. Rochester.
http://www.se.rit.edu/~jrv/research/ar/
        introduction to AR; AR Bibliography

[2]   R.T. Azuma, "A Survey of Augmented Reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355-385, 1997. [PDF, WEB]

[3]   R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent Advances in Augmented Reality," *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 34-47, Nov./Dec. 2001. [PDF, WEB]

[4]   Holger Regenbrecht, Gregory Baratoff, Wilhelm Wilke, "Augmented Reality Projects in Automotive and Aerospace Industry," *IEEE Computer Graphics and Applications*, Nov/Dec 2005. [PDF]

[5]   State of the Art in enabling technologies (AR). [PPT, WEB]

5.2 Applications in Maintenance, Industrial Design and Repair

[6]   Heather Baldwin, "Wearable Wonders," *Aviation Today*, Aug. 21, 2005. [PDF, WEB]

[7]   S. Feiner, B. MacIntyre, and D. Seligmann, "Knowledge-Based Augmented Reality," *Communications of the ACM*, vol. 36, no. 7, pp. 52-62, 1993. [PDF, WEB] (related links: http://www1.cs.columbia.edu/graphics/top.html)

[8]   Benjamin C. Lok, "Toward the Merging of Real and Virtual Spaces," *Communications of the ACM*, vol. 47, No. 8, pp.48-53, 2004. [PDF]

[9]   Thomas Reicher, "A Framework for Dynamically Adaptable Augmented Reality Systems." Thesis, Institut fur Informatik Technische Universitat Munchen, 2004. [PDF, WEB]

[10]   Dirk Reiners, Didier Stricker, Gudrun Klinker, and Stefan Müller, "Augmented Reality for Construction Tasks: Doorlock Assembly," Proceedings of IWAR'98.

[11]   Dan Curtis, David Mizell, Peter Gruenbaum, and Adam Janin, "Several Devils in the Details: Making an AR App Work in the Airplane Factory," Proceedings of IWAR'98.

[12]   Jose Molineros, Vijaimukund Raghavan, and Rajeev Sharma, "AREAS: Augmented Reality for Evaluating Assembly Sequences," Proceedings of IWAR'98.

[13] David Mizell, "Augmented Reality Applications in Aerospace," ISAR 00. [PDF, WEB]

[14] Demonstration of Augmented Reality, Rockwell Automation Fair 2003, Milwaukee, November 17-20, 2003. [WEB]

[15] F. Discenzo, D. Chung, D. Carnahan, S. L. Chen, R. Behringer, V. Sundareswaran, K. K. Wang, J. Molineros, and J. H. McGee, "Augmented Reality Supports Real-time Diagnosis and Repair of Complex Shipboard Systems," IFAC Conference on Control Applications in Marine Systems, Ancona, July 7-9, 2004.

[16] Francisco González Áñez, "Two Innovative Steps for Training on Maintenance: VIRMAN & STARMATE," American Nuclear Society Annual Meeting, June 9-13 2002. [PDF, WEB]

[17] Bernd Schwald, et al, "STARMATE: Using Augmented Reality technology for computer guided maintenance of complex mechanical elements," E2001 Conference 17-19 October 2001 - Venice – Italy. [PDF, WEB]

[18] STARMATE Public Report v2.0 (2002). [PDF, WEB]

[19] Samory Kakez, Jean Figue, Vania Conan, "A Generic Model for Immersive Documentation Environment and Applications," *British VR Journal*. [PDF, WEB]

[20] Bernd Schwald, Blandine de Laval, "An Augmented Reality System for Training and Assistance to Maintenance in the Industrial Context," Journal of WSCG, Vol. 11 No. 1, ISSN 1213-6972, February 3-7, 2003, Plzen, Czech Republic. [PDF, WEB]

[21] Bernd Schwald, Blandine de Laval, Training and Assistance to Maintenance in an Augmented Reality Environment." [PDF, WEB]

5.3 Other Applications

[22] J. Rekimoto, "NaviCam: A Magnifying Glass Approach to Augmented Reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, Aug. 1997, pp. 399-412.

[23] T. Hollerer et al., "Exploring MARS: Developing Indoor and Outdoor User Interfaces to a Mobile Augmented Reality System," *Computers and Graphics*, vol. 23, no. 6, Dec. 1999, pp. 779-785.

[24] T. Starner et al., "Augmented Reality through Wearable Computing," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, Aug. 1997, pp. 386-398.

[25]   Y. Baillot, D. Brown, and S. Julier, "Authoring of Physical Models Using Mobile Computers," Proc. of the 5th Int'l Symp. Wearable Computers, IEEE CS Press, Los Alamitos, Calif., 2001. [PDF, WEB]

[26]   W. Piekarski, B. Gunther, and B. Thomas, "Integrating Virtual and Augmented Realities in an Outdoor Application," Proc. 2nd Int'l Workshop Augmented Reality (IWAR 99), IEEE CS Press, Los Alamitos, Calif., 1999, pp. 45-54.

[27]   B. Thomas et al., "ARQuake: An Outdoor/Indoor Augmented Reality First Person Application," Proc. 4th Int'l Symp. Wearable Computers (ISWC 2000), 2000, pp. 139-146.

[28]   D. Stricker et al., "Design and Development Issues for Archeoguide: An Augmented Reality based Cultural Heritage On-Site Guide," Proc. Int'l Conf. Augmented Virtual Environments and 3D Imaging (ICAV3D 2001), Publishing ZITI, Greece, 2001, pp. 1-5. (related link: http://www.ziti.gr)

[29]   J. Rekimoto, Y. Ayatsuka, and K. Hayashi, "Augmentable Reality: Situated Communication through Physical and Digital Spaces," Proc. 2nd Int'l Symp. Wearable Computers, IEEE CS Press, Los Alamitos, Calif., 1998, pp. 68-75.

[30]   M. Billinghurst and H. Kato, "Collaborative Mixed Reality," Proc. Int'l Symp. Mixed Reality (ISMR 99), Springer-Verlag, Secaucus, N.J., 1999, pp. 261-284.

[31]   J. Rekimoto, "Transvision: A Hand-Held Augmented Reality System for Collaborative Design," Proc. Virtual Systems and Multimedia (VSMM 96), Int'l Soc. Virtual Systems and Multimedia, Gifu, Japan, 1996, pp. 85-90.

[32]   Zs. Szalavári et al., "Studierstube: An Environment for Collaboration in Augmented Reality," *Virtual Reality — Systems, Development and Application*, vol. 3, no. 1, 1998, pp. 37-49.

[33]   S. Julier et al., "The Software Architecture of a Real-Time Battlefield Visualization Virtual Environment," Proc. IEEE VR 99, IEEE CS Press, Los Alamitos, Calif., 1999, pp. 29-36.

[34]   T. Jebara et al., "Stochasticks: Augmenting the Billiards Experience with Probabilistic Vision and Wearable Computers," Proc. 1st Int'l Symp. Wearable Computers (ISWC 97), 1997, IEEE CS Press, Los Alamitos, Calif., pp. 138-145.

[35]   Steven Feiner, Tobias Höllerer, Elias Gagas, Drexel Hallaway, Tachio Terauchi, Sinem Güven, and Blair MacIntyre, "MARS - Mobile Augmented Reality Systems," http://www1.cs.columbia.edu/graphics/projects/mars/mars.html

[36] Nassir Navab, "Industrial Augmented Reality (IAR): Challenges in Design and Commercialization of Killer Apps," Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2003), p. 2, 2003.

[37] Stefan Muller, Didier Stricker, Jens Weidenhausen, "An Augmented-Reality System to support product development, production and maintenance." [PDF, WEB]

5.4 Registration

[38] J.P. Rolland, L.D. Davis, and Y. Baillot, "A Survey of Tracking Technologies for Virtual Environments," *Fundamentals of Wearable Computers and Augmented Reality*, W. Barfield and T. Caudell, eds., Lawrence Erlbaum, Mahwah, N.J., 2001, pp. 67-112. [PDF, WEB]

[39] Eric Foxlin, "Motion Tracking Requirements and Technologies," Chapter 8 in *Handbook of Virtual Environment Technology*, Kay Stanney, Editor, Lawrence Erlbaum Associates, 2002. [PDF, WEB] (related links(more papers about tracking by Eric Foxlin): http://www.isense.com/company/papers/)

[40] Ferrin, Frank J. "Survey of Helmet Tracking Technologies," SPIE Proceedings vol. 1456 Large-Screen Projection, Avionic, and Helmet-Mounted Displays (1991), 86-94.

[41] Meyer, Kenneth, Hugh L. Applewhite, and Frank A. Biocca. "A survey of Position-Trackers," *Presence: Teleoperators and Virtual Environments* 1, 2 (Spring 1992), 173-200.

[42] Y. Yokokohji, Y. Sugawara, and T. Yoshikawa, "Accurate Image Overlay on Video See-through HMDs Using Vision and Accelerometers," Proc. IEEE VR 2000, IEEE CS Press, Los Alamitos, Calif., 2000, pp. 247-254.

[43] G. Welch and G. Bishop, "Scaat: Incremental Tracking with Incomplete Information," Proc. ACM SIGGRAPH 97, ACM Press, New York, 1997, pp. 333-344.

[44] E. Foxlin, M. Harrington, and G. Pfeifer, "Constellation: A Wide-Range Wireless Motion-Tracking System for Augmented Reality and Virtual Set Applications," Proc. Siggraph 98, ACM Press, New York, 1998, pp. 371-378.

[45] G. Welch et al., "High-Performance Wide-Area Optical Tracking—The HiBall Tracking System," *Presence: Teleoperators and Virtual Environments*. vol. 10, no. 1, Feb. 2001, pp. 1-21.

[46] Y. Cho, J. Lee, and U. Neumann, "A Multi-Ring Fiducial System and an Intensity-Invariant Detection Method for Scalable Augmented Reality," Proc. Int'l

Workshop Augmented Reality (IWAR 98), A.K. Peters, Natick, Mass., 1998, pp. 147-165.

[47]   Newman, J., Wagner, M., Bauer, M., MacWilliams, A., Pintaric, T., Beyer, D., Pustka, D., Strasser, F., Schmalstieg, D., and Klinker, G.,   "Ubiquitous tracking for Augmented Reality," Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2004), 2-5 Nov. 2004, pp.192-201. [PDF, WEB]

[48]   Kiyohide Satoh, Shinji Uchiyama, Hiroyuki Yamamoto. "A Head Tracking Method Using Bird's-Eye View Camera and Gyroscope," Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2004), 2-5 Nov. 2004, pp. 202-211. [PDF, WEB]

[49]   F. Sauer et al., "Augmented Workspace: Designing an AR Testbed," Proc. Int'l Symp. Augmented Reality 2000 (ISAR 00), IEEE CS Press, Los Alamitos, Calif., 2000, pp. 47-53.

[50]   R. Azuma et al., "A Motion-Stabilized Outdoor Augmented Reality System," Proc. IEEE VR, IEEE CS Press, Los Alamitos, Calif., 1999, pp. 252-259.

[51]   R. Azuma et al., "Tracking in Unprepared Environments for Augmented Reality Systems," *Computers and Graphics*, vol. 23, no. 6, Dec. 1999, pp. 787-793.

[52]   S. You, U. Neumann, and R. Azuma, "Hybrid Inertial and Vision Tracking for Augmented Reality Registration," Proc. IEEE VR, IEEE CS Press, Los Alamitos, Calif., 1999, pp. 260-267.

[53]   K. Satoh et al., "Townwear: An Outdoor Wearable MR System with High-Precision Registration," Proc. 2nd Int'l Symp. Mixed Reality (ISMR 2001), MR Systems Lab, Yokohama, Japan, 2001, pp. 210-211.

[54]   K. Sawada, M. Okihara, and S. Nakamura, "A Wearable Attitude Measurement System Using a Fiber Optic Gyroscope," Proc. 2nd Int'l Symp. Mixed Reality (ISMR 2001), MR Systems Lab, Yokohama, Japan, 2001, pp. 35-39.

[55]   G. Simon, A.W. Fitzgibbon, and A. Zisserman, "Markerless Tracking Using Planar Structures in the Scene," Proc. Int'l Symp. Augmented Reality 2000 (ISAR 00), IEEE CS Press, Los Alamitos, Calif., 2000, pp. 120-128.

[56]   K.N. Kukulakos and J.R. Vallino, "Calibration-Free Augmented Reality," *IEEE Trans. Visualization and Computer Graphics*, vol. 4, no. 1, Jan.-Mar. 1998, pp. 1-20.

[57]   Y. Seo and K-S. Hong, "Weakly Calibrated Video-Based Augmented Reality: Embedding and Rendering through Virtual Camera," Proc. Int'l Symp. Augmented Reality 2000 (ISAR 00), IEEE CS Press, Los Alamitos, Calif., 2000, pp. 129-136.

[58]   B. Hoff and R. Azuma, "Autocalibration of an Electronic Compass in an Outdoor Augmented Reality System," Proc. Int'l Symp. Augmented Reality 2000 (ISAR 00), IEEE CS Press, Los Alamitos, Calif., 2000, pp. 159-164.

[59]   Najafi, H.; Navab, N.; Klinker, G., "Automated initialization for marker-less tracking: a sensor fusion approach," Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2004), 2-5 Nov. 2004 Page(s):79 -88. [PDF, web]

[60]   U. Neumann and S. You, "Natural Feature Tracking for Augmented Reality," *IEEE Trans. Multimedia*, vol. 1, no. 1, Mar. 1999, pp. 53-64. [PDF, WEB]

[61]   R. Behringer, "Registration for Outdoor Augmented Reality Applications Using Computer Vision Techniques and Hybrid Sensors," Proc. IEEE VR, IEEE CS Press, Los Alamitos, Calif., 1999, pp. 244-251.

[62]   V. Coors, T. Huch, and U. Kretschmer, "Matching Buildings: Pose Estimation in an Urban Environment," Proc. Int'l Symp. Augmented Reality 2000 (ISAR 00), IEEE CS Press, Los Alamitos, Calif., 2000, pp. 89-92.

[63]   B. Jiang, S. You, and U. Neumann, "Camera Tracking for Augmented Reality Media," Proc. IEEE Int'l Conf. Multimedia Expo 2000, IEEE CS Press, Los Alamitos, Calif., 2000, pp. 1637-1640.

[64]   G. Simon, A.W. Fitzgibbon, and A. Zisserman, "Markerless Tracking Using Planar Structures in the Scene," Proc. Int'l Symp. Augmented Reality 2000 (ISAR 00), IEEE CS Press, Los Alamitos, Calif., 2000, pp. 120-128.

[65]   Azuma, R. (1993). "Tracking Requirements for Augmented Reality." *Communications of the ACM* 36 (7): 50-51. [PDF, WEB]

[66]   Zikan, K., W. D. Curtis, et al. (1994). "A note on dynamics of human head motions and on predictive filtering of head-set orientations." Proceedings of SPIE Vol. 2351: Telemanipulator and Telepresence Technologies 2351: 328-336.

[67]   Andrew I. Comport, Marchand Marchand, Francois Chaumette. "A real-time tracker for markerless Augmented Reality," Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2003), p. 36, 2003.

[68]   Andrei State, Gentaro Hirota, David T. Chen, William F. Garrett, Mark A. Livingston, "Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking," Proc. SIGGRAPH '96, New Orleans, LA, 429-438, 1996. [PDF, WEB]

[69]   Gabriel Zachmann, "Distortion Correction of Magnetic Fields for Position Tracking." Proc. Computer Graphics International (CGI'97), June 23-27, 1997, Hasselt and Diepenbeek, Belgium. [PDF, WEB]

[70]   Michael Bajura and Ulrich Neumann, Dynamic Registration Correction in Video-Based Augmented Reality Systems," IEEE *Computer Graphics and Applications* 15, 5, Sept. 1995, 52-60. [PDF, WEB]

[71]   M.L. Yuan, S.K. Ong, and A.Y.C. Nee, "Registration Based on Projective Reconstruction Technique for Augmented Reality Systems," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 11, No.3, May/June 2005. [PDF, WEB

[72]   M.L. Yuan, S.K. Ong, and A.Y.C. Nee, "The virtual interaction panel: an easy control tool in Augmented Reality systems," Computer Animation and Virtual Worlds 2004; 15(3-4), pp. 425-432. [PDF, WEB]

[73]   Louise Clarke, Peter Harper, and Richard B. Reilly, "Video Based Gesture Recognition for Augmentative Communication," http://www.dinf.ne.jp/doc/english/Us_Eu/conf/tide98/129/clarke_louise.html

[74]   P. de la Hamette, P. Lukowicz, G. Tröster, T. Svoboda, "Fingermouse: A Wearable Hand Tracking System," UbiComp 2002: Adjunct Proceedings of the 4th International Conference on Ubiquitous Computing, September 2002, pages 15-16. [PDF, WEB]

[75]   J. L. Crowley, F. Bérard and J. Coutaz, "Finger Tracking as an Input Device for Augmented Reality", IWAGFR '95 - International Workshop on Gesture and Face Recognition, Zurich, June 1995. [PDF, WEB]

[76]   Ramesh Raskar, Paul Beardsley, Jeroen van Baar, Yao Wang, Paul Dietz, Johnny Lee, Darren Leigh and Thomas Willwacher, "RFIG Lamps: Interacting with a self-describing world via photosensing wireless tags and projectors," ACM SIGGRAPH 2004. [PDF, WEB, PPT, MOV]

[77]   Ramesh Raskar, Paul Beardsley, Paul Dietz, and Jeroen van Baar, "Photosensing wireless tags for geometric procedures," *Communications of the ACM*, September 2005. [PDF, WEB]

[78]   Eric Foxlin and Leonid Naimark, "VIS-Tracker: A Wearable Vision-Inertial Self-Tracker," IEEE VR2003, March 22-26, Los Angeles. [PDF,AVI]

[79]   Leonid Naimark and Eric Foxlin, "Circular Data Matrix Fiducial System and Robust Image processing for a Wearable Vision-Inertial Self-Tracker," IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2002), Darmstadt, Germany. [PDF]

[80]   Leonid Naimark and Eric Foxlin, "Miniaturization, Calibration and Accuracy Evaluation of a Hybrid Self-Tracker," IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR 2003), Oct. 7-10, 2003, Tokyo. [PDF]

[81]   B. Boufama and A. Habed, "Registration and tracking in the context of Augmented Reality," [PDF, WEB]

5.5 Visualization

[82]   Gun A. Lee, Geard J. Kim, and Mark Billinghurst, "What you experience is what you get(WYXIWYG)," *Communications of the ACM*, Vol. 48, No. 7, pp 76-81, 2005. [PDF]

[83]   Lee, G.A., Nelles, C., Billinghurst, M., and Kim, G.J., "Immersive authoring of tangible Augmented Reality applications," Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2004), 2-5 Nov. 2004 Page(s):172 -181. [PDF, WEB]

[84]   Mark Billinghurst, Raphael Grasset, Julian Looser, "Designing Augmented Reality Interfaces," *Computers and Graphics*, Feb. 2005, pp 17-22. [PDF]

[85]   Mark Billinghurst, "ARToolKit project," http://www.hitl.washington.edu/artoolkit/

[86]   MacIntyre, B., Gandy, M., Dow, S., and Bolter, J. "DART: A toolkit for rapid design exploration of Augmented Reality experiences," In Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (Santa Fe, NM, Oct. 24–27). ACM Press, New York, 2004, 197–206.

[87]   T. Ohshima et al., "RV-Border Guards: A Multi-Player Mixed Reality Entertainment," *Trans. Virtual Reality Soc. Japan*, vol. 4, no. 4, 1999, pp. 699-705.

[88]   S. Feiner et al., "A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment," Proc. 1st Int'l Symp. Wearable Computers (ISWC '97), IEEE CS Press, Los Alamitos, Calif., 1997, pp. 74-81.

[89]   J. Rekimoto and M. Saitoh, "Augmented Surfaces: A Spatially Continuous Workspace for Hybrid Computing Environments," Proc. ACM SIGCHI 99, 1999, ACM Press, New York, pp. 378-385.

[90]   A. Butz et al., "Enveloping Users and Computers in a Collaborative 3D Augmented Reality," Proc. 2nd Int'l Workshop Augmented Reality (IWAR 99), IEEE CS Press, Los Alamitos, Calif., 1999, pp. 35-44.

[91] Zs. Szalavári and M. Gervautz, "The Personal Interaction Panel—A Two-Handed Interface for Augmented Reality," Proc. 18th Eurographics, Eurographics Assoc., Aire-la-Ville, Switzerland, 1997, pp. 335-346. [PDF, WEB]

[92] H. Kato et al., "Virtual Object Manipulation of a Table-Top AR Environment," Proc. Int'l Symp. Augmented Reality 2000 (ISAR 00), IEEE CS Press, Los Alamitos, Calif., 2000, pp. 111-119. [PDF, WEB]

[93] T. Ohshima et al., "AR 2Hockey: A Case Study of Collaborative Augmented Reality," Proc. IEEE VR Ann. Int'l Symp. (VRAIS 98), IEEE CS Press, Los Alamitos, Calif., 1998, pp. 268-275.

[94] M. Billinghurst, H. Kato, and I. Poupyrev, "The MagicBook—Moving Seamlessly between Reality and Virtuality," *IEEE Computer Graphics and Applications*, vol. 21, no. 3, May/June 2001, pp. 2-4. (related links: MagicBook project: http://www.hitlabnz.org/index.php?page=proj-mb-top)

[95] Kaiser, E., Olwal, A., McGee, D., Benko, H., Corradini, A., Li, X., Feiner, S., Cohen, P. "Mutual Dissambiguation of 3D Multimodal Interaction in Augmented and VR." In Proceedings of the Fifth International Conference on Multimodal Interfaces (ICMI 2003). Vancouver, BC. Canada. November 5–7, 2003. p. 12–19. [PDF, WEB]

[96] J. Stauder, "Augmented Reality with Automatic Illumination Control Incorporating Ellipsoidal Models," *IEEE Trans. Multimedia*, vol. 1, no. 2, June 1999, pp. 136-143.

[97] Y. Mukaigawa, S. Mihashi, and T. Shakunaga, "Photometric Image-Based Rendering for Virtual Lighting Image Synthesis," Proc. 2nd Int'l Workshop Augmented Reality (IWAR 99), IEEE CS Press, Los Alamitos, Calif., 1999, pp. 115-124.

[98] P. Debevec, "Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-Based Graphics with Global Illumination and High Dynamic Range Photography," Proc. SIGGRAPH, ACM Press, New York, 1998, pp. 189-198.

[99] S. Julier et al., "Information Filtering for Mobile Augmented Reality," Proc. Int'l Symp. Augmented Reality 2000 (ISAR 00), IEEE CS Press, Los Alamitos, Calif., 2000, pp. 3-11. [PDF, WEB]

[100] B. Bell, S. Feiner, and T. Hollerer, "View Management for Virtual and Augmented Reality," Proc. ACM Symp. User Interface Software and Technology (UIST 2001), ACM Press, New York, 2001, pp. 101-110. [PDF, WEB]

[101] A. Fuhrmann et al., "Occlusion in Collaborative Augmented Environments," Computers and Graphics, vol. 23, no. 6, Dec. 1999, pp. 809-819.

[102] Georg Klein and Tom Drummond, "Sensor Fusion and Occlusion Refinement for Tablet-Based AR," Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'04), pp. 38-47, 2004. [PDF, WEB]

[103] Cakmakci, O., Yonggang Ha, Rolland, J.P., "A compact optical see-through head-worn display with occlusion support." Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2004), 2-5 Nov. 2004, pp. 16-25. [PDF, WEB]

[104] V. Lepetit and M.-O. Berger, "Handling Occlusions in Augmented Reality Systems: A Semi-Automatic Method," Proc. Int'l Symp. Augmented Reality 2000 (ISAR 00), IEEE CS Press, Los Alamitos, Calif., 2000, pp. 137-146.

[105] K. Kiyokawa, Y. Kurata, and H. Ohno, "An Optical See-through Display for Mutual Occlusion of Real and Virtual Environments," Proc. Int'l Symp. Augmented Reality 2000 (ISAR00), Los Alamitos, Calif., 2000, pp. 60-67. [PDF, WEB]

[106] H.L. Pryor, T.A. Furness, and E. Viirre, "The Virtual Retinal Display: A New Display Technology Using Scanned laser Light," Proc. 42nd Human Factors Ergonomics Soc., Santa Monica, California, 1998, pp. 1570-1574. [DOC, WEB]

[107] Oliver Bimber, Gordon Wetzstein, Andreas Emmerling and Christian Nitschke, "Enabling View-Dependent Stereoscopic Projection in Real Environments," ISMAR 2005.

[108] R. Raskar, G. Welch, and W-C. Chen, "Table-Top Spatially-Augmented Reality: Bringing Physical Models to life with Projected Imagery," Proc. 2nd Int'l Workshop Augmented Reality (IWAR 99), IEEE CS Press, Los Alamitos, Calif., 1999, pp. 64-71.

[109] Raskar, R., Brown, M.S., Yang, R., Chen, W., Welch, G., Towles, H., Seales, B., and Fuchs, H., "Multiprojector displays using camera-based registration," Proc. of IEEE Visualization, pp. 161-168, 1999.

[110] Cotting, D., Naef, M., Gross, M., and Fuchs, H., "Embedding Imperceptible Patterns into Projected Images for Simultaneous Acquisition and Display," Proc. of IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR'04), pp. 100-109, 2004.

[111] Low, K-L., Welch, G., Lastra, A., and Fuchs, H., "Life-Sized Projector-Based Dioramas," Proc. Symp. VR Software and Technology, IEEE CS Press, pp. 93-101, 2001.

[112] H. Hua et al., "An Ultra-light and Compact Design and Implementation of Head-Mounted Projective Displays," Proc. IEEE VR 2000, IEEE CS Press, Los Alamitos, Calif., 2000, pp. 233-240. [PDF, WEB]

[113] Jochen Ehnes, Koichi Hirota, Michitaka Hirose, "Projected Augmentation II - A Scalable Architecture for Multi Projector Based AR-Systems based on 'Projected Applications'," ISMAR 2005. [PDF, MOV]

5.6 Links

[114] International Symposium on Mixed and Augmented Reality (ISMAR)
http://www.augmented-reality.org/ismar/

[115] On the Resources – Augmented Reality,
http://www.hitl.washington.edu/projects/knowledge_base/aug.html
(people, research organizations and projects, and commercial projects)

[116] Rockwell Scientific, http://hci.rockwellscientific.com/

[117] NOMAD display system, http://www.mvis.com/nomadmilitary/index.html

[118] Microoptical SV-6 display,
http://www.microopticalcorp.com/Products/vga.html#SV6

[119] Virtual Retinal Display, http://www.hitl.washington.edu/research/vrd/

[120] A Introduction to Virtual Retinal Display (VRD) Technology [PDF],
http://www.cs.nps.navy.mil/people/faculty/capps/4473/projects/fiambolis/vrd/vrd_full
.html

[121] ProView XL40/50 STm Display
http://www.vrlogic.com/html/keo/proview_xl_35xl_50.html

[122] Datavisor SeeThrough by nVision Industries
http://www.nvisionindustries.com/nVisionIndustries/index.htm

[123] VideoVision Display, http://www.worldviz.com/videovision/index.html

[124] AddVisor™ 150
http://products.saab.se/PDBWeb/ShowProduct.aspx?ProductId=1097

[125] HMD/headset/VR-helmet Comparison Chart by Christoph Bungert,
http://www.stereo3d.com/hmd.htm

[126] Logitech 3D mouse (Acoustic), http://www.vrdepot.com/vrteclg.htm

[127] MotionAnalysys Tracking System, http://www.motionanalysis.com/

[128] Vicon Tracking System, http://www.vicon.com/

[129] HiBall Tracking System, http://www.3rdtech.com/HiBall.htm

[130] Wide-Area Tracking research project, http://www.cs.unc.edu/~tracker

[131] Flock of Bird tracking system, http://www.ascension-tech.com/products/flockofbirds.php

[132] A link about Motion Capture, http://vr.isdale.com/MotionCapture_Links.html some links to tracking research groups and Commercial Vendors.

[133] The Gesture Recognition Home Page http://dream1.leeds.ac.uk/~ajh/tracker.html

[134] FingerMouse, http://www.wearable.ethz.ch/fingermouse.0.html

[135] Global Positioning System, http://www.colorado.edu/geography/gcraft/notes/gps/gps_f.html

[136] RFIG: Radio frequency identity and geometry, http://www.merl.com/people/raskar/Sig04/

[137] InterSense InertiaCube2, http://www.isense.com/products/prec/ic2/index.htm

[138] STARMATE project (Jan. 2000~Mar. 2003), http://vr.c-s.fr/starmate/

[139] ARVIKA Project(7/1999-6/2003), http://www.arvika.de/www/e/home/home.htm

[140] Metaio Augmented Solutions, http://www.METAIO.COM

[141] Mixed Reality Systems Laboratory Inc. http://www.mr-system.com/incdata/index_e.html .

[142] MicroVision Applications on car maintenance, http://www.mvis.com/nomadexpert/app.html

[143] KARMAR project, http://www1.cs.columbia.edu/graphics/projects/karma/

6 REFERENCES

Abbeel, P. and A. Ng (2004). Apprenticeship learning via inverse reinforcement learning. Proceedings of the Twenty-first International Conference on Machine Learning.

Allbeck, J., K. Kipper, et al. (2002). ACUMEN: Amplifying Control and Understanding of Multiple ENtities. Autonomous Agents and Multi-Agent Systems, Bologna, Italy.

Ambady, N., F. Bernieri, et al. (2000). Towards a Histology of Social Behavior: Judgmental Accuracy from Thin Slices of Behavior. Advances in Experimental Social Psychology. M. P. Zanna: 201-272.

Atkeson, C. and S. Schaal (1997). Robot learning from demonstration. Proc. International Conference on Machine Learning, Morgan Kaufmann.

Azuma, R. T. (1997). "A Survey of Augmented Reality." Presence: Teleoperators and Virtual Environments 6(4): 355-385.

Badler, N., R. Bindiganavale, et al. (2000). A parameterized action representation for virtual human agents. Embodied Conversational Agents. J. S. J. Cassell, S. Prevost, and E. Churchill. Cambridge, MA, MIT Press: 256-284.

Badler, N., C. Erignac, et al. (2002). "Virtual Humans for Automating Maintenance Task Validation." Communications of the ACM.

Badler, N., M. Palmer, et al. (1999). "Animation control for real-time virtual humans." Communications of the ACM 42(8): 65-73.

Badler, N., B. Webber, et al. (1990). Animation from Instructions. Making Them Move: Mechanics, Control, and Animation of Articulated Figures. B. B. N. Badler, D. Zeltzer. San Mateo, Morgan-Kaufmann: 51-93.

Bartenieff, I. and D. Lewis (1980). Body Movement: Coping with the Environment. New York, Gordon and Breach.

Bindiganavale, R. and N. Badler (1998). Motion abstraction and mapping with spatial constraints. Workshop on Motion Capture Technology, Geneva, Switzerland.

Bindiganavale, R., W. Schuler, et al. (2000). Dynamically Altering Agent Behaviors Using Natural Language Instructions. Autonomous Agents, AAAI.

Buchsbaum, D. and B. Blumberg (2005). Imitation as a First Step to Social Learning in Synthetic Characters: A Graph-based Approach. Symposium on Computer Animation, Los Angeles, CA, ACM & Eurographics.

Cassell, J., C. Pelachaud, et al. (1994). "Animated conversation: rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents." Computer Graphics: 413-420.

Chi, D., M. Costa, et al. (2000). The EMOTE Model for Effort and Shape. ACM SIGGRAPH, New Orleans, LA, ACM Press.

FLTK (2006). Fast Light Toolkit. http://www.fltk.org.

Gleicher, M. (2001). "Comparing Constraint-based Motion Editing Methods." Graphical Models 63(2): 107-134.

Kiyokawa, K., Y. Kurata, et al. (2000). An Optical See-through Display for Mutual Occlusion of Real and Virtual Environments. Proc. Int'l Symp. Augmented Reality 2000 (ISAR00), Los Alamitos, CA.

Razzaque, S., Z. Kohn, et al. (2001). Redirected walking. Eurographics, Manchester, UK.

Seligmann, D., S. Feiner, et al. (1996). Augmented reality in architectural construction, inspection and renovation, Anaheim, CA, ASCE.

Siskind, J. (2001). "Grounding the Lexical Semantics of Verbs in Visual Perception Using Force Dynamics and Event Logic." Journal of Artificial Intelligence Research 15: 31-90.

Vallino, J. (2006). Introduction to Augmented Reality. http://www.se.rit.edu/~jrv/research/ar/introduction.html.

Webber, B., N. Badler, et al. (1995). "Instructions, intentions and expectations." Artificial Intelligence 73: 253-269.

Zhao, L. and N. Badler (2005). "Acquiring and Validating Motion Qualities from Live Limb Gestures." Graphical Models 67(1): 1-16.

Zhao, L., S. Lu, et al. (2001). Structures and styles: Motion acquisition from video. Third International Workshop on Intelligent Virtual Agents, Madrid, Spain.