# MIXED REALITY TECHNOLOGY APPLICATIONS IN CONSTRUCTION EQUIPMENT OPERATOR TRAINING

Xiangyu Wang, Phillip S. Dunston, and Mirosław Skibniewski
School of Civil Engineering, Purdue University,
550 Stadium Mall Drive, West Lafayette, IN 47907-2051, USA
mirek@purdue.edu

Abstract: This paper provides information on Mixed Reality (MR), and more specifically Augmented Reality (AR) and their potential applications in heavy construction equipment and operator training. Mixed Reality involves the use of special display and tracking technologies that are capable of seamlessly merging digital (virtual) content into a real environment. Conceptual designs and application scenarios of two AR systems for construction are presented: an AR-based equipment management system (AR EMS) for maintenance, and repair of heavy equipment fleet, and an AR-based operator training system (AR OTS) that trains the novice operators in a real worksite environment populated with virtual materials and instructions.

Keywords: mixed reality, augmented reality, machine operator training, heavy construction equipment management.

## 1. INTRODUCTION

While construction equipment operators usually develop their skills on the job, it is generally accepted that formal training provides more comprehensive skills. The nature of construction equipment and the materials they may be used to handle makes such training expensive. Research has shown that the key to acquiring the necessary motor skills to control complex systems, such as a backhoe excavator, is hands-on and coached training [3]. Some research has considered the need of such hands-on experience by developing physical training systems. For example, Bernold et al. [2] developed an Internet-based prototype training system for backhoe operators with which the remotely located novice can manipulate a laboratory excavator via a joystick.

Personnel involved in equipment management also monitor equipment condition, maintain and repair equipment components by referring to technical specifications, and generate more information for future access. These individuals often face the problem of not finding the right information in a timely manner. Furthermore, critical information exists on different platforms (e.g., paper-based specification, standalone-computers) characterized by a lack of mobility. This practice makes it difficult to process information at any time and place.

Two innovative AR-based systems to visually merge digital information into the current workspace are conceptually developed for construction equipment operator training and equipment management.

## 2. MIXED REALITY TECHNOLOGY

### 2.1 Mixed Reality: general description

Mixed Reality (MR), defined by Milgram et al. [5], is a special class of Virtual Reality-related technologies for creating environments wherein real world and virtual world objects are presented together in a single display. Augmented Reality (AR), a sub-mode of MR, can provide an augmented workspace by inserting content from the virtual space in which we store and interact with digital content into a predominantly physical space where we work. The need for good operator training and for access to large amounts of equipment management information creates conditions making use of AR techniques most promising for engaging construction personnel into augmented workspaces. AR allows a user to work with the real 3D environment while visually receiving additional computer-generated or modeled information about the task at hand. Such displays can enhance the user's perception of the real environment by showing information the user cannot directly acquire unaided.

### 2.2 Mixed Reality: technological components

Augmented Reality is typically based on four major technological components discussed as follows:

**Media representation:** AR systems typically use more varied media for presenting the digital content than VR. The identified major classes of media representations from abstract (schematic) to realistic (3D) are text, indicators, platform (tablet and screen), 2D image, 3D wireframe, 3D data, and 3D object. It

is apparent that realistic representations convey more intuitive information that can assist an individual's comprehension thus augmenting the cognitive process and activity. Hybrid representations (e.g., texts + 3D object) can be very useful in situations where different cognitive channels need to be provided.

*Input device:* An input device is used to manipulate the digital information displayed over the real environment. Most of the input devices used in VR can also be applied in AR systems. A poor input metaphor may create a number of problems for the user. For example, intermediate devices such as joysticks actually place themselves between users and environments, requiring significant cognitive mapping to perform tasks. Aural and haptic input can be integrated to increase input capability of the system.

*Display device:* The display can be generally classified as visual display, acoustic display, tactile display etc. Head mounted displays (HMDs) were developed for visual rendering, 3D localized sound systems for aural rendering, force feedback devices for haptic rendering, etc. Visual displays are key to AR systems with the other types of displays as possible supplements. The wrong display may burden the user's physical and cognitive processes and reduce usability.

*Tracking technology:* Position and orientation tracking, are used where the orientation and the position of the user's head or hands, and sometimes eyes are required. Accurate registration—positioning and alignment—of virtual objects in the real environment requires accurate tracking from the user's viewpoint and sensing of the locations of other objects in the environment. The tracking technology must update position and orientation values in real-time as the user may move around in the real environment. Each class of tracker has its own advantages and disadvantages. For detailed comparison of tracking technologies, readers are referred to [4]. Tracking systems that might meet the stringent requirements of AR for construction will probably be hybrid systems, such as a combination of inertial and optical technologies [1] because hybrid systems may compensate for the shortcomings of a single technology by using multiple sensor types to produce robust results.

# 3. CONCEPTUAL AR SYSTEM DESIGN FOR EQUIPMENT MANAGEMENT

Accessing large amounts of information in equipment management operations can be facilitated by the dynamics of AR that complement human associative information processing and memory. The proposed AR system is called Augmented Reality-based equipment management system (AR EMS).

## 3.1 Scenario 1: site monitoring of equipment operation

An equipment manager is standing in an earth-moving construction site wearing video-see-through HMD with a portable computing unit in his backpack and a data glove on his hands (see Figures 1 and 3). When the manager turns on the headset display, a radio frequency identification (RFID) system is activated with read/writable RFID tags mounted onto equipment sending signals to readers connected to the manager's computer. All the equipment located within, say, 46 meters of the manager can be identified and the major working status information is queried from the tags on the equipment. The equipment information is periodically and wirelessly updated by a central server, and the retrieved information can be displayed in the real view of the manager. The displayed information could either be archived data that was previously documented in server and loaded to the RFID tags (e.g., equipment type, manufacturer and capacity etc.) or real-time generated/updated data (e.g., productivity, haul distance, engine temperature, gas consumption, temperature, humidity, etc.).



Figure 1. Hypothetical scenario for site monitoring of working equipment

## 3.2 Scenario 2: close-up inspection of equipment

In a related scenario, the manager can inspect any equipment by physically approaching the equipment. Then, the user is able to browse spatially correspondent digital information about a specific equipment in the course of performing a specific service (e.g., maintenance), or can access interpretive annotations posted by other domain experts or crew members. The user can see virtual icons via the HMD for different functions like identification of equipment components, maintenance procedures, repair procedures, management (performance and

maintenance) history, etc. Upon choosing one icon, for example, maintenance, virtual tags of all components (e.g., tire, engine, etc.) related to the maintenance program for the equipment are displayed right beside their real counterpart over the actual view. Complete records for each component of equipment regarding age, usage, fuel consumption, maintenance and repair costs, down time, and time between component replacements are kept in a database, which can be retrieved by clicking the tag of interest. Also a crew-to-expert communication can be established through wireless LAN (see Figure 3). The maintenance expert sitting in the office can observe what the on-site crew inspects through video and aid in trouble-shooting. He can either instruct

the crew via audio communication or send the updated data (pictures, video, and parameters) to the crew's local database through wireless network.

A maintenance scenario that can be implemented by AR EMS is described as follows: The engine of a hydraulic excavator is shown in Figure 2 with varying annotation that is generated in response to the state and context of a simulated 4-step maintenance procedure. The maintenance crew's task is to test hydraulic **oil from S.O.S. sampling valves for preventative maintenance**. The task sequence exemplifies several of the benefits of AR that were previously described.



(a) Identification of engine access panel

(b) Recognition of open panel causes identification of components and instruction to remove cap

(c) Recognition of cap removal causes new instruction to press test button.

(d) Additional information about position of occluded filter bypass structure in response to user's query

Figure 2. Hypothetical illustration of sequence of automatic instructions and information for maintenance task

A virtual tag highlights the position of the engine access panel to be removed in step 1 (see Figure 2a). The recognition of the open panel causes the identification of components in step 2 (see Figure 2b). Virtual tags appear to identify each component (see Figure 2b). Annotations appear automatically in the camera view, the easy access to information increasing usability. Virtual tags are statically and spatially arranged in the task area for a particular view. Virtual line segments connect each virtual tag to its nearby associated real component. This can elicit the user's attention. Virtual tags relate task information to the context of the work area (callouts identifying hydraulic lines, valves etc.), and guidance on the next action to be performed. The step 3 is to remove the dust cap as indicated by the virtual tag

(see Figure 2b). The design of the scene is intended to reduce error by (1) influencing attention and (2) supporting the users' encoding (learning for later recall) of layout and content.

The AR EMS system senses the state of the assembly (the user removed a dust cap in Step 3) and responds automatically with a new instruction: "Step 4: Press to test" (see Figure 2c). Thus, the content of the array directs performance and the interactivity inherent in the system's production of a new instruction contributes to the user's motivation.

Interactive queries are indicated by user selection of an area (see wire frame in Figure 2d) to reveal and identify additional hidden structures related to the

task. This form of interaction demands that the user attend to his or her view of the engine.

*3.3 Conceptual design of system architecture*

As a test bed to evaluate the above two scenarios, a technology platform, seen in Figure 3, has been conceptualized with a total of seven separate modules shown in Figure 4. The following sections discuss the mechanisms in these modules.

***Representation module:*** The media representation in AR EMS is text and/or 2D-image/video. AR EMS can be further enhanced by adding a hypermedia system [7] that provides contextual information in the form of labels and links. Thus, more detailed database information can be retrieved via links as with HTML webpage navigation.
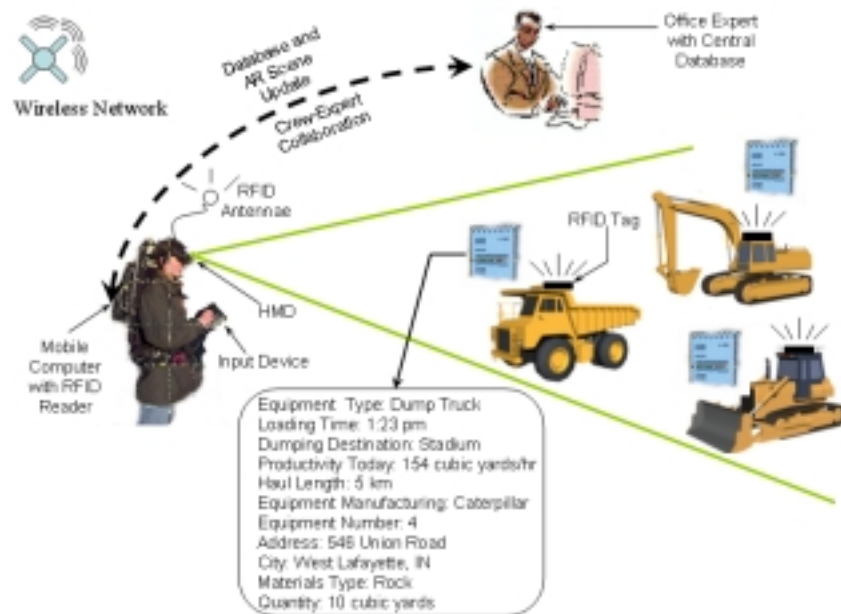


Figure 3. Graphical illustration of AR EMS system application in on-site equipment management

***Remote Server Module:*** Remote server module consists of several experts and a central database server. A service expert sits in an office and provides instructions to on-site management crew whenever necessary via wireless communication. Also, the real-time mixed scene that the on-site crew is viewing can be transmitted to the remote office where the expert can experience the same digitally augmented inspection in the format of a video stream. The data synchronization between databases in RFID tags, and the central database that monitors equipment working condition or service information is made in a real-time manner. Besides the archived data and interpretive annotations posted by domain experts, the real-time produced data (e.g., productivity, haul distance) can also be updated in the database located in RFID tags from the central server. For example, a GPS-based real-time productivity monitoring system can be integrated with the central database and productivity status can be sent to the database associated with each RFID tag and displayed in the user's augmented real world view.

***Identification Module:*** This module uses RFID technology to identify equipment. An RFID tag is mounted onto equipment and signals can be received by readers on the mobile computing unit. The proposed tags are read/writable (R/W) so they can be readily updated.

***Display module:*** ARvision-Stereo HMD, a prototype manufactured by Trivisio Company, is used for display. This HMD is equipped with two built-in color video cameras that are used to capture the real world scene for a video-based tracking algorithm in the close-view inspection scenario. The cameras act as the eyes of users and their position and orientation are tracked by a tracker module. An audio feature that gives warnings or directs attention allocation can also be used.

***Mobile Computer Module:*** This module is located on a high-performance, light-weight laptop and consists of a local database, RFID antenna with reader, and AR software. The *local database* can be accessed by Sybase SQL anywhere as the database query. All of the management data can be stored in the Sybase database. Technical imagery and data about the equipment fleet should be collected in order to compile a comprehensive, equipment-specific database. The *RFID antenna and reader*

embedded in the laptop collects and transmits the RF signal. The *AR program* is the core technological component of the rendering pipeline for the whole system. In addition to OpenGL and the GLUT graphics library, there are a few software development toolkits (SDK) such as ARTooKit (http://www.hitl.washington.edu/artoolkit/), and an AR software development library—STARlib1.0 [6], etc., which can be readily used as a function library for developing the AR program. The AR program is configured to switch between two functional modes depending on the work scenario: overall site view and equipment close view. The overall site-view mode corresponds to scenario 1 where the user has a comprehensive view of the working equipment fleet and only major working status information is displayed. In this mode, the AR program only accepts the tracker data from a real-time kinematic differential Global Positioning System (RTK-DGPS), suitable for large-area positioning. On the other hand, the equipment close-view mode refers to scenario 2 introduced above where the user is able to closely inspect or observe certain components of the equipment of interest. In this mode, the AR program will switch to fiducial tracking and only reads the tracking information the fiducials provide.
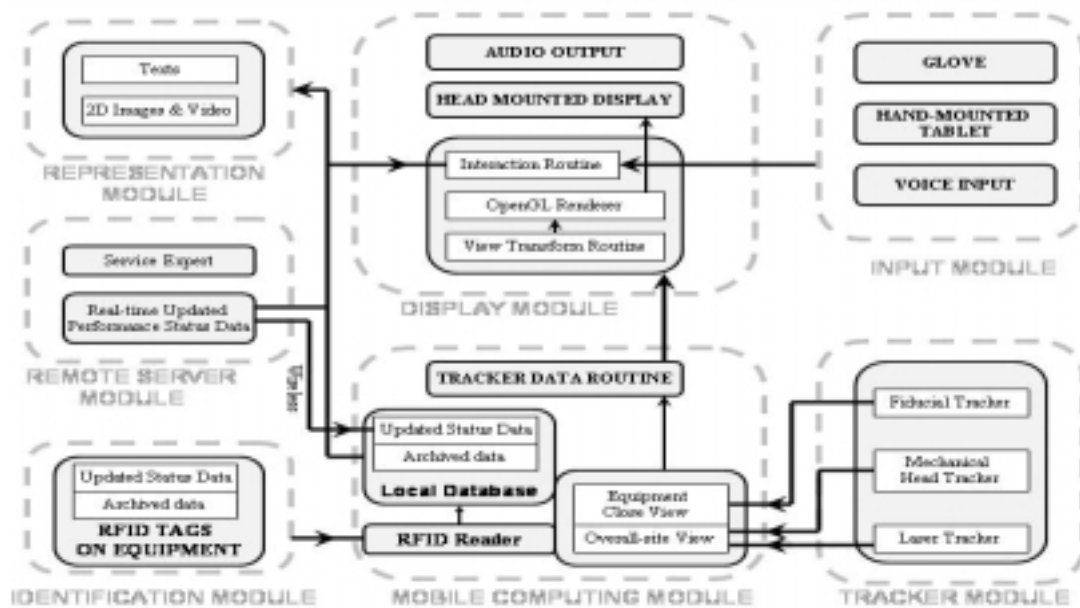


Figure 4. Generic system architecture and data flow of AR EMS

***Tracker Module:*** AR EMS requires a means to track the user and equipment's locations within the construction site. We propose a multiple-tracker solution to realize the two functional modes discussed above. In the overall site-view mode, AR can exploit RTK-DGPS combined with Unscented Kalman Filtering (UKF) techniques [8] to obtain accuracies in the cm-dm range to determine the positions of users and identified equipment. A mechanical tracker fixed the user's body can be used to determine the orientation of the camera on the HMD (the viewing direction). These two trackers can work within the same reference frame and continuously and spatially register the digital information with the video stream of the real scene of the equipment. Instead, in the equipment close-view mode, fiducial recognition can be used for tracking. These fiducials may be LEDs, square markers, dots with a circular pattern, colored dots, or color multi-rings. These routines assume that one or more fiducials are visible at all times; without them, the registration can fail. The way that the system adds labels is that each fiducial beside an object/component identifies the component. After the digital information is loaded, the position and viewing direction of the user are continuously updated by the data obtained from the tracker. Simultaneously, the digital information image is transformed using location and direction matrices to produce an image that aligns with the objects in the real view.

***Input module:*** This module handles manipulation of the displayed digital content (e.g., clicking hypertext to browse details) and annotates the comments. An annotation function coded into the AR program can enable the user to record comments via a hand-mounted tablet for future reference.

## 4. CONCEPTUAL AR SYSTEM DESIGN FOR OPERATOR TRAINING

One of the increasingly important application areas of AR is in training and education, where simulation can be used to illustrate concepts and provide exercises that allow the learner to train in a realistic

environment. Compared to real exercises, AR offers reduced costs and hazards and unlimited training conditions/scenarios. This is particularly attractive in heavy construction equipment training. For example, a computer-generated virtual stockpile or piping materials will incur only computational resource costs. In addition, AR allows the creation of unlimited training scenarios (e.g., various terrain, shapes and sizes of virtual stockpiles) providing flexibility without increasing costs. The safety advantage can be realized in many situations such as the simulated fall of elevated virtual materials or equipment tipping due to lifting errors. The proposed training system is called Augmented Reality operator training system (AR OTS), and this technology can be used to increase the number, frequency, and realism of training simulations and offers the means to rehearse future heavy construction equipment operations for actual construction sites.

4.1 Scenario

A front loader operation training scenario using AR OTS is described as an example of how the generic system may function. A novice operator sitting in a real front loader cabin can manipulate the equipment to move a virtual soil stockpile on the ground (see Figure 6a) to a dumping container (Figure 6c) a certain distance away. First, the novice can see virtual tags surrounding their real counterparts and learn the functions of each item on instrument panel (see Figure 5). The novice can follow step-by-step starting procedure instructions or choose to see a pre-recorded training video. Safety instruction will automatically come up with reminders such as wearing a hard-hat, checking the machine and the surrounding area before moving, etc.



Figure 5. Hypothetical illustration of augmented instrument panel with corresponding virtual tags registered to fiducials markers on the panel

Then the novice can start the equipment and move to the virtual stockpile, then position the machine to load soil. The bucket edge position is tracked by

laser sensors mounted on it and a wireframe boundary is drawn to represent the bucket in the virtual world understandable by computer (see Figure 6b). The novice controls the loader and the harder he swings the bucket, the faster the bucket passes through the virtual stockpile and a greater volume of soil is loaded. The amount of soil can be controlled by the curvature shaped by the moving wireframe boundary of the bucket while loading. A critical problem to solve here is how to determine the quantity of soil that is taken from the stockpile by the individual bucket pass. Heaping, spillage, and the resistive force from the particular soil type should be accounted in estimating the soil volume. The force-feedback and weight of loaded materials may be "felt" in the novice's hand via a computer integrated haptic device. After the loader finishes one load, the novice maneuvers the equipment to the container on the other side of the work area. Laser sensors on the edge of the container provide data for a virtual wireframe rectangle representing the container's rim in the virtual world. When the novice positions the loader, he or she can start to dump the virtual soil. When the bucket is tilted to a certain angle, the algorithm will understand that dumping should start and then the virtual soil pours from the bucket. If the dumping action yields too much soil falling outside the wireframe rim of the container (missing the target), the system will make comments indicating the poor performance. A performance score could be given as feedback.

4.2 Conceptual design of AR OTS architecture

The experimental test-bed should be established in an indoor facility because current medium-range, high accuracy trackers require equipment setup that may be left in place indoors. Six separate modules were configured and Figure 7 shows the flow diagram for the whole system. The following sections discuss the mechanisms in these modules.

Representation module: Three types of digital content are defined—panel tags, static object, and dynamic object. Panel tag is the descriptive text registered to controls in the cabin instrument panel. The media representation should be text and/or 2D-image/video. A static object is defined as a stationary entity stocked on site such as an untouched soil stockpile. A dynamic object is defined as a moving entity such as loaded soil, the virtual loader bucket edge, etc. Both static and dynamic objects should be rendered in a rich manner because the realistic rendering can provide more cues for essential spatial judgment. Otherwise, the system will fail stimulate proper operator skill development.

Display module: The visual and audio display devices used are similar to AR EMS. A computer-integrated haptic device can provide force feedback

to the operator, which along with the force-control algorithm can enable the novice to feel the simulated force from the "weightless" virtual materials.
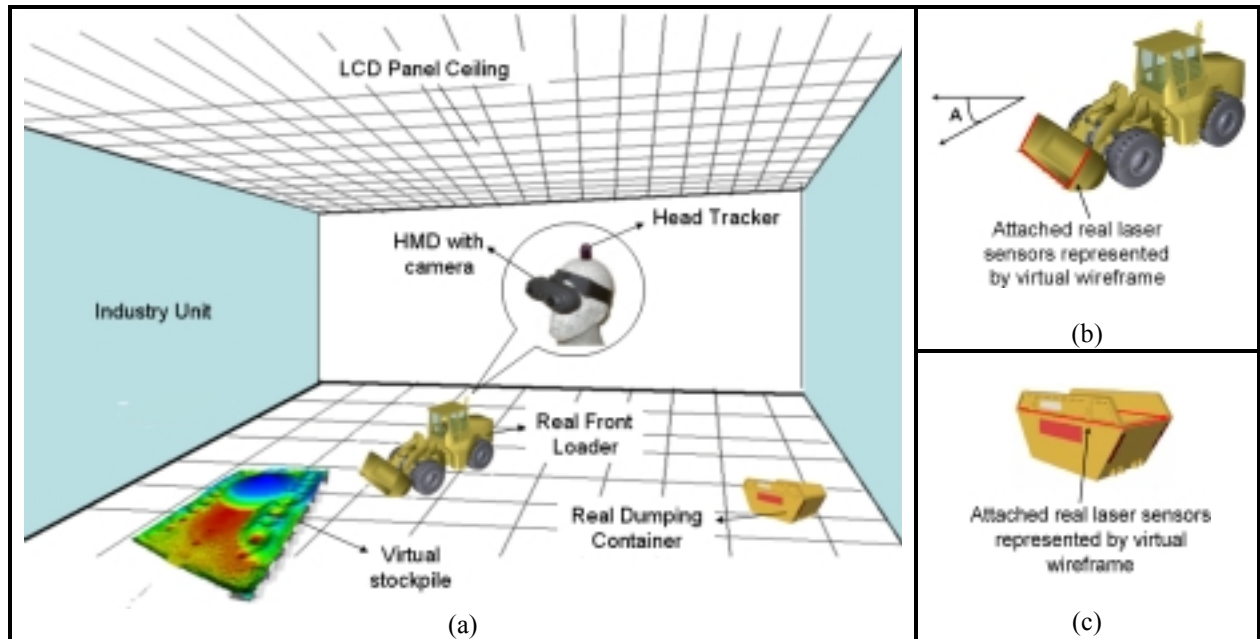


Figure 6. Hypothetical concept illustration of (a) AR system application in loading operation training (b) sensor locations and virtual wireframe representation of real front loader bucket, and (c) sensor location and virtual wireframe representation of real dumping container

**Mobile Computer Module:** A customized AR program is the core technological component of rendering pipeline for the whole system. More details have been covered in the section of AR EMS.

**Tracker Module:** The panel tags, static objects, and dynamic objects require multiple-trackers to be combined. Fiducials are to be used for registering panel tags. User's position and orientation (camera) can be tracked by a precise head mounted LED device. With this tracker, the static object can be registered to the location relative to the user. A laser sensor can be added to track the position of dynamic objects (see Figure 6b and c). All of these trackers should be configured according to the same reference, making the static object(s) interact with the dynamic object(s).
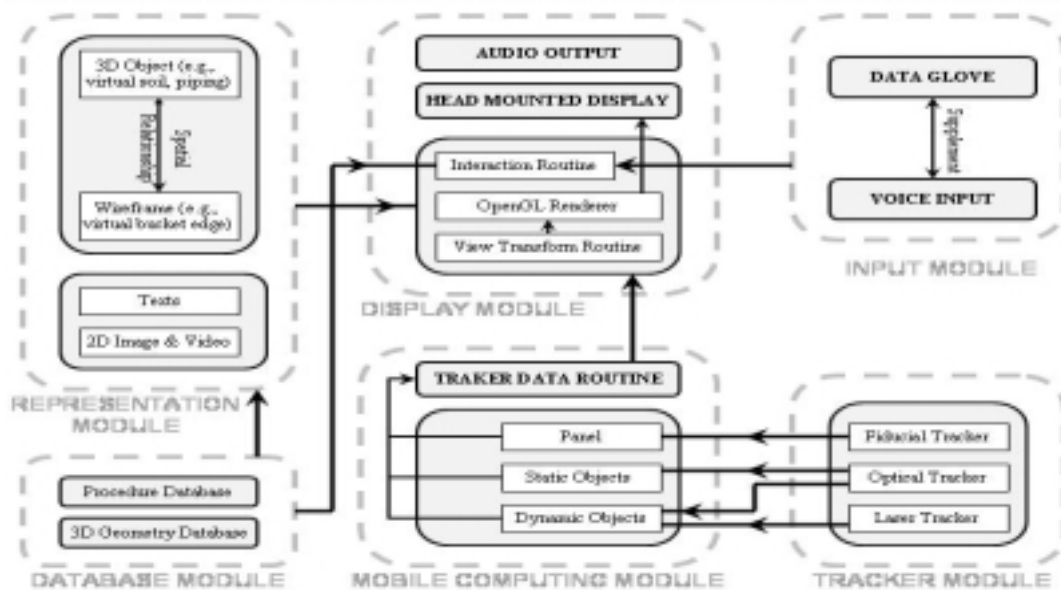


Figure 7. System Architecture Diagram of AR OTS

***Database Module:*** The the local database, similar as in AR IMS, provides detailed training information and procedures for the instrument panel controls.

***Input module:*** A data glove is used to manipulate a virtual 3D cursor and yet not negatively alter equipment operation. Speech recognition can function as a supplement input device for simple command input.

## 5. CONCLUSIONS

MR is a promising advanced technology for the construction industry that can be integrated into heavy equipment management and operator training. AR can reduce the cost, time, and hazard levels by augmenting these activities with digital content. Two conceptual AR systems (AR EMS and AR OTS) were presented in this paper with detailed hardware configurations discussed, and several selected application scenarios for these two concepts were elaborated.

## Acknowledgments

## REFERENCES

[1] Azuma, R. T., "A Survey of Augmented Reality", *In Presence: Teleoperators and Virtual Environments*, MIT Press, Vol. 6, Issue 4, pp. 355-385, 1997.

[2] Bernold, L. E., Lloyd, J., Vouk, M., "Equipment Operator Training in the Age of Internet2", *Proceedings of 19th International Symposium on Automation and Robotics in Construction (ISARC 2002)*, NIST Special Publication 989, Washington D.C., pp. 505-510, 2002.

[3] Cuqlock-Knopp, V. G., Wilkins, C. A. and Torgerson, W. S., "Multiple Cue Probability Learning and the Design of Information Displays for Multiple Tasks", D.L. Damos, Ed. *Multiple-Task Performance*, Taylor & Francis, London, pp. 139-152, 1991.

[4] Holloway, R. and Lastra, A., "Virtual Environments: A Survey of the Technology", *SIGGRAPH'95 Course*, Vol. 8 (A), pp. 1-40, 1995.

[5] Milgram, P. and F. Kishino, "A Taxonomy of Mixed Reality Visual Displays", *IEICE Trans. on Information and Sytems (Special Issue on Networked Reality)*, Vol. E77-D (12), pp. 1321-1329, 1994.

[6] Neumann, U., and Cho, Y., "A Self-tracking Augmented Reality System", *Proceedings of the ACM International Symposium on Virtual Reality and Applications*, pp. 109-115, 1996.

[7] Sinclair, P., Martinez, K., Millard, D., and Weal, M., "Links in the Palm of Your Hand: Tangible Hypermedia Using Augmented Reality", *Proceedings of the Thirteenth ACM Conference on Hypertext and Hypermedia HT'02*, pp. 127–136, 2002.

[8] Wan, E. A., Van der Merwe, R., "The Unscented Kalman Filter for Nonlinear Estimation", *Proceedings of Symposium 2000 on Adaptive Systems for Signal Processing, Communication and Control (AS-SPCC)*, IEEE, Alberta, CA, 2002.