SAND96-2781C

Real-time 3D visualization of volumetric video motion sensor data

CONF-961113--2 SAND--96-2781C

Jeffrey Carlson, Sharon Stansfield, Dan Shawver

Sandia National Laboratories
Albuquerque, New Mexico 87185

Gerald M. Flachs, Jay B. Jordan, and Zhonghao Bao

New Mexico State University
Las Cruces, New Mexico 88003

ABSTRACT

This paper addresses the problem of improving detection, assessment, and response capabilities of security systems. Our approach combines two state-of-the-art technologies: volumetric video motion detection (VVMD) and virtual reality (VR). This work capitalizes on the ability of VVMD technology to provide three-dimensional (3D) information about the position, shape, and size of intruders within a protected volume. The 3D information is obtained by fusing motion detection data from multiple video sensors. Other benefits include low nuisance alarm rates, increased resistance to tampering, low-bandwidth requirements for sending detection data to a remote monitoring site, and the ability to perform well in a dynamic environment where human activity and motion clutter are commonplace. The second component of this work involves the application of VR technology to display information relating to the sensors and the sensor environment. VR technology enables an operator, or security guard, to be immersed in a 3D graphical representation of the remote site containing the video sensors. VVMD data is transmitted from the remote site via ordinary telephone lines and displayed in real-time within the virtual environment. There are several benefits to displaying VVMD information in this way. Often, raw sensor information is not in a form that can be easily interpreted and understood—especially when taken out of the context of the sensor environment. Because the VVMD system provides 3D information and because the sensor environment is a physical 3D space, it seems natural to display this information in 3D. Also, the 3D graphical representation depicts essential details within and around the protected volume in a natural way for human perception. Sensor information can also be more easily interpreted when the operator can "move" through the virtual environment and explore the relationships between the sensor data, objects and other visual cues present in the virtual environment. By exploiting the powerful ability of humans to understand and interpret 3D information, we expect to 1) improve the means for visualizing and interpreting sensor information, 2) allow a human operator to assess a potential threat more quickly and accurately, and 3) enable a more effective response. This paper will detail both the VVMD and VR technologies and will discuss a prototype system based upon their integration.

Keywords: volumetric, video, motion, detection, virtual, reality, immersive

This work was supported by the United States Department of Energy under Contract DE-AC04-94AL 85000.
Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy.

1. VOLUMETRIC VIDEO MOTION DETECTION

Sandia National Laboratories has been funded by the Department of Energy's Office of Safeguards and Security (OSS) to develop and advance image-processing technologies for monitoring personnel and materials inside controlled access facilities. The objective is to provide security technologies that minimize interference with work activities yet maintain an effective deterrent against insider threats. Our focus is to provide continuous monitoring within a dynamic environment, such as a working vault or during a weapons dismantlement process.

One technology being developed to meet OSS requirements is a volumetric video motion detector (VVMD). This detector uses multiple cameras to monitor activity in predefined, three-dimensional (3D) zones. Each zone is a partition of the total common volume of video surveillance. The total common volume is defined by the intersection of viewing frustums from combinations of two or more cameras in the system. If any part of the volume under surveillance is not seen by at least two cameras, then it is not a part of the total common volume. A 3D zone is defined by masking the pixels from each camera that intercept that zone. Activity in a zone is detected when a specified number of cameras sense changes in those pixels associated with that zone. The resolution of the VVMD depends on camera resolution, lens focal length, range from cameras to zones, and the number of cameras used in the system. Two similar applications have been developed using VVMD technology: 1) material monitoring and 2) personnel monitoring. The similarities and differences in these applications are discussed in the following sections.

1.1 Material monitoring

In the material monitoring application, arbitrary zones, occupied by assets, are masked from each camera's perspective. These zones are typically disjoint, variable in size, and distributed nonuniformly throughout the total common volume of video surveillance. In this application, the number of assets (zones) monitored is typically small—probably fewer than one-hundred and most likely less than ten. Since the number of zones monitored is small, a manual masking, or calibration, approach is adequate. A relatively simple interface is required for system setup and to convey detection information to the security operator.

1.2 Personnel monitoring

In the personnel monitoring application, predefined zones are masked from each cameras perspective. Unlike the material monitoring application, the zones are the same size and are uniformly distributed throughout the total common volume of surveillance. In this application, there are potentially numerous zones; therefore, an automated calibration and setup procedure is desirable. The uniformity of zones used in this application facilitates an automated calibration process. In addition to setup functions required for zone masking, the personnel monitoring application also requires setup for display of detection information. This can be a tedious and time-consuming task. The procedure is discussed in more detail in section 2.

1.3 Calibration

The purpose of calibration is to inform the system of the locations of protected materials and zones. The process amounts to masking each zone from each camera's perspective. For material monitoring, masking is accomplished manually by using a trace-and-fill process. In the "trace" stage of the process, a mouse or other pointing device is used to trace a closed boundary around each zone as seen in digitized images taken from each surveillance camera. The closed boundaries are "filled," or colored, to designate image pixels associated with each zone. The fill color is used to identify a particular zone (i.e., each zone has a unique color). The trace-and-fill process is repeated for every zone seen by each camera. The result is a set of mask images (one for each camera) that designates which pixels from each camera map to a given zone.

For personnel monitoring, which potentially involves an extremely large number of zones, an automated calibration process is used. The automation of this process involves two steps: lens calibration and camera orientation calibration.

The calibration process can be very tedious and can lead to significant errors if not properly performed. The problem is compounded by the use of wide-angle lenses. These lenses introduce a large non-linear distortion as shown in Figure 1. The distortion is particularly evident near the perimeter of the image. The problem is further complicated by variations among lenses having the same specifications.

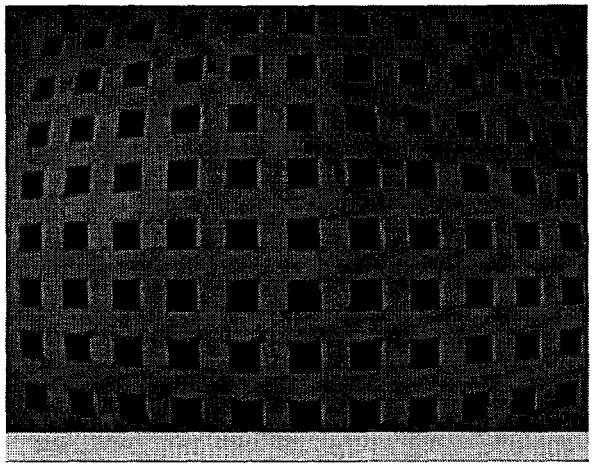


Figure 1. Distortion apparent in a wide-angle lens.

The calibration procedure begins by calibrating each camera and lens. A calibration board is placed in front of the camera with the lens focused at infinity as shown in Figure 2. The image coordinates of the corners of the dark squares are located and their pointing angles relative to the optical axis of the camera are established. An affine transformation is used to interpolate the pointing angles for all image pixels. These pointing angles are stored in two matrices defining the azimuth (pan) and elevation (tilt) for each pixel relative to the optical axis of the camera.

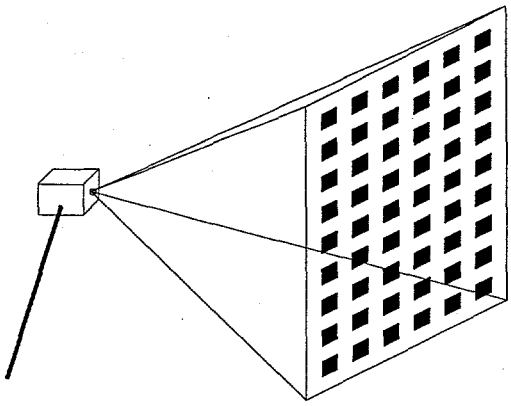


Figure 2. Portrayal of camera positioned for calibration.

The next step in the calibration procedure involves mounting the cameras and obtaining their 3D positions. The camera positions can be easily established by physical measurements which eliminates a significant source of error in the calibration procedure. Establishing the orientation of the optical axis of each camera is, however, a very tedious task. There are three angles that must be determined. These are the pan and tilt of the optical axis of the camera and the roll angle of the camera. Two surveyed points in the view of each camera are sufficient to establish these angles; however, more than two points can provide significant improvement in angle estimates. The final step in the calibration procedure uses information acquired in the previous two steps to generate a set of mask images. The mask images define a mapping from camera pixels to 3D zones.

1.4 Hardware implementations

The VVMD consists of custom software integrated with commercially available hardware. Several hardware implementations are possible. The simplest and least expensive implementation includes two or more video cameras, a video framegrabber and multiplexer, a video monitor, and an IBM-compatible personal computer running MS-DOS. We recommend at least a 50 MHz 486. The cost of a four-camera system running on a 50 MHz 486 is under $3000.

Our initial hardware configuration consisted of a single host computer with an Industry Standard Architecture (ISA) bus, a video multiplexer, and an ISA bus video framegrabber. Although this is probably the simplest and least expensive configuration, performance is limited. The video multiplexer allows use of a single framegrabber, but limits video access to one camera at a time. Acquisition of one frame takes roughly 1/30th of a second. For a four-camera system, 4/30th's, or approximately 1/8th, of a second is spent acquiring video data. After acquisition, the large quantity of video data is transferred via ISA bus to PC memory where it can then be processed. The slow ISA bus represents another significant bottleneck. We have used both a 50 MHz 486 host and a 166 MHz Pentium host with the hardware configuration just described. The performance improvements achieved with the Pentium host were almost imperceptible compared to the performance achieved with the 50 MHz 486 host. Performance in this case was measured in terms of total frame processing rate (i.e., the number of video frames processed per second). There are at least two options for improving the hardware configuration.

The first option is to use a single Pentium computer system as a host with a Peripheral Component Interconnect (PCI) bus and multiple PCI bus framegrabbers. The host would provide both image processing and user interface functions. The PCI bus is the new, high-speed replacement for the old ISA bus. The PCI bus will significantly speed up the transfer of video data to the host's memory where it can be processed. Using multiple framegrabbers, video data acquisition can be initiated simultaneously from each camera. The total time required to acquire video data is then about 1/30th of a second—no matter how many cameras are used in the system. The only requirement is a PCI slot for each framegrabber. This configuration is probably limited to a four-camera system because of the cost and complexity of an extended PCI bus (every set of three PCI slots requires a bus extender and associated electronics).

Another option is a distributed system using PC104-based image processing components and a PC, Pentium based user interface, or host. The PC104 standard consists of a small form factor PC board (approximately 4 x 4 inches) with a bus that is electrically equivalent to an ISA bus. The only difference in the two bus architectures is the pin-out arrangement. In this configuration, each camera in the system is provided a dedicated PC104 processor and PC104 framegrabber for image processing purposes. The processor and framegrabber would be co-located with the camera. Detection data would be sent by either RS-422, Ethernet, or RF back to the PC-based host. The host would combine detection data from each independent PC104 video system to determine if activity has occurred in a zone. The host would also handle user interface functions such as setup and display. There are advantages and disadvantages to this configuration. Advantages include: 1) more than a four-camera system is realizable, 2) image processing is offloaded from the host, and 3) image processing for all cameras is essentially accomplished in parallel—that is, the total image processing time required for processing each frame from each camera is equal to the time required to process a single frame from a single camera. Disadvantages of this implementation include the added cost and complexity of communicating information between multiple computing systems.

1.5 Prototype system description

Our prototype system uses four cameras to monitor a room of 24 x 30 x 10 feet; however, the extension to many more cameras is straightforward. A 16- or 32-camera system is feasible with today's computer technology. Each camera in the current system is equipped with a 2.8 mm lens. This gives a field-of-view close to 90 degrees. A large portion of the room is visible to each camera by placing the cameras in the four corners of the room, close to the ceiling, with a downward look angle. Figure 3 shows the room as seen by each of the four cameras.

For personnel monitoring purposes, three levels, or planes, of zones were defined. The levels are at the floor, at table-top height, and at four feet from the floor. The purpose of multiple levels is to distinguish crawlers from walkers, walkers from chairs and so forth. Excluding the areas on the perimeter of the room which are occupied by tables, the room is divided into an area of nine by thirteen zones at each level. The zone dimensions are approximately 2 x 2 feet x 4 inches. In addition to zones defined for personnel monitoring, zones for monitoring materials can also be defined.

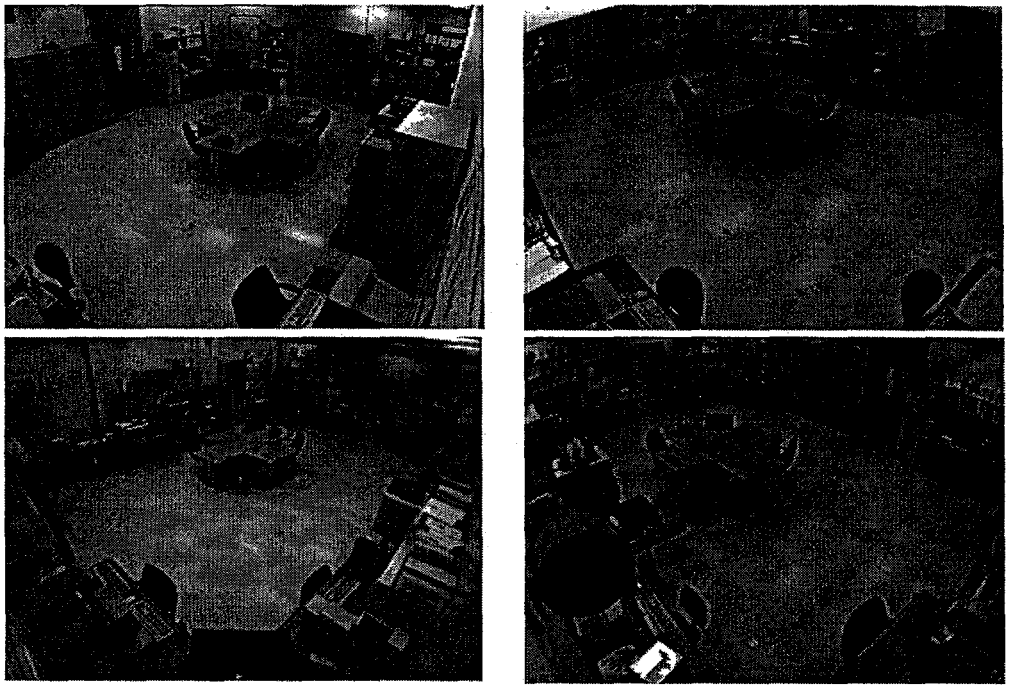


Figure 3. Surveillance volume as seen from each camera.

The system uses video motion detection algorithms to detect pixel activity from each camera. Pixels with detectable levels of activity are classified as change pixels. Corresponding pixels from the set of mask images are used to determine if activity has occurred within a zone. Two or more cameras need to sense changes in pixels from a given zone for activity to be declared in that zone. This provides a significantly increased resistance to nuisance alarms as compared to conventional video motion detection systems that use just a single camera.

A data logging utility has been incorporated into the prototype system to maintain a history of all zone activity. Zones can also be prioritized in terms of consequence. For example, a high-consequence zone might contain a valuable asset. Whenever activity is detected in a high-consequence zone, a snapshot from each camera is logged to the system disk. We have defined four of these zones in our present installation. Two are located on the tables near the center of the room; two more are just inside the room by the door. The two zones on the table are for monitoring high-value assets. Designating these as high-consequence zones enables identification of authorized and unauthorized intruders from the logged snapshots. Having two high-consequence zones just inside the door allows identification of people entering and leaving the protected area.

The prototype system uses virtual reality (VR) technology to display information relating to the sensors and the sensor environment. VR technology enables an operator, or security guard, to be immersed in a 3D graphical representation of the sensor environment. The 3D graphical representation depicts essential details within and around the protected volume in a natural way for human perception. The VR component also includes voice annunciation. The voice annunciator provides audio information about important security events occurring within the protected volume. The next section discusses the VR component of this work.

2. VIRTUAL REALITY INTERFACE

2.1 Modeling the sensor environment

VR technology is used to display VVMD data within a 3D graphical model of the sensor environment. The model shows the layout of the room, indicating the position of permanently located objects such as tables, cabinets, doors, and windows. The locations of protected assets are also indicated. The model is created from physical measurements and photographs of the actual sensor environment. As illustrated in Figure 4, the 3D graphical representation depicts essential details within and around the protected volume in a natural way for human perception. There are several commercially available software tools that can create the model. As mentioned previously, the modeling process can be tedious and time consuming. Fortunately, it is generally only done once for each installation.

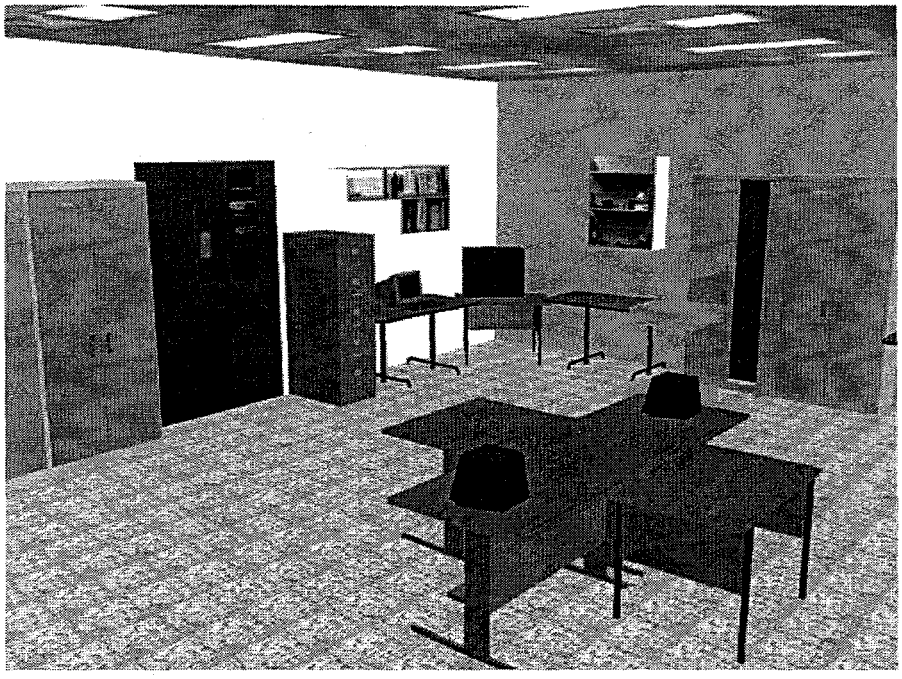


Figure 4. 3D graphical representation of the sensor environment.

2.2 Overlaying sensor data

VVMD data is overlaid on the 3D display in real-time to indicate personnel movements. This is illustrated in Figure 5 which shows a graphical rendition of a crawling intruder displayed within the 3D model of the sensor environment. The crawler is shown as a rectangular shape on the floor in the center of the figure. In the graphic portrayal of VVMD data, the position, shape, and height of personnel are indicated, as is the status of protected assets.

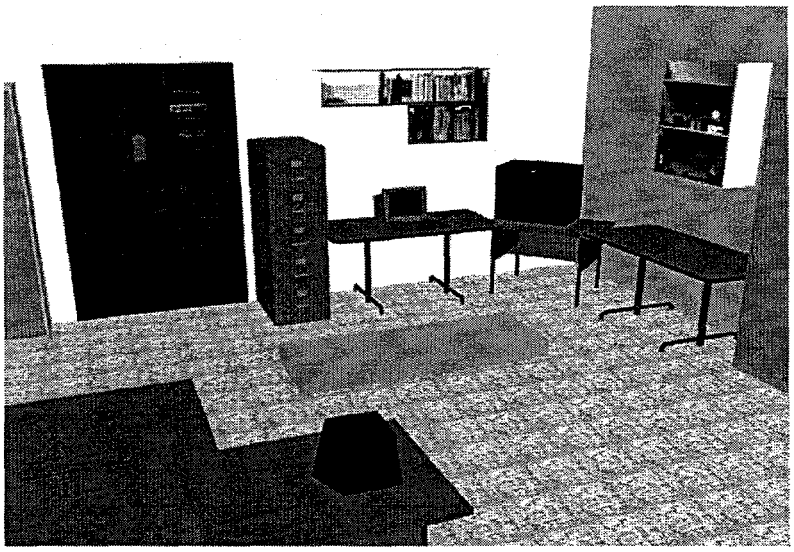


Figure 5. Graphical rendition of a crawling intruder.

Figure 6 shows a protected asset being violated by an intruder. The dark objects on the table represent the protected assets. The three-block figure in the foreground represents the space occupied by an intruder as detected by the VVMD.

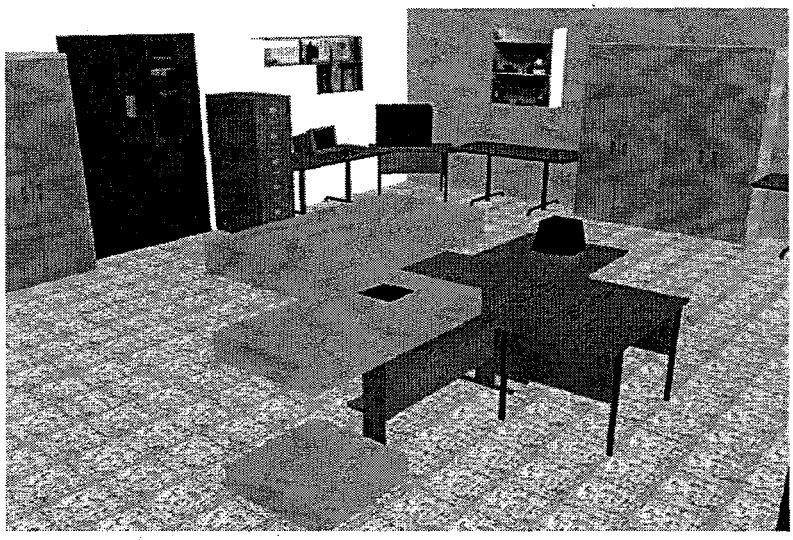


Figure 6. Graphical rendition of protected asset being violated by a standing intruder.

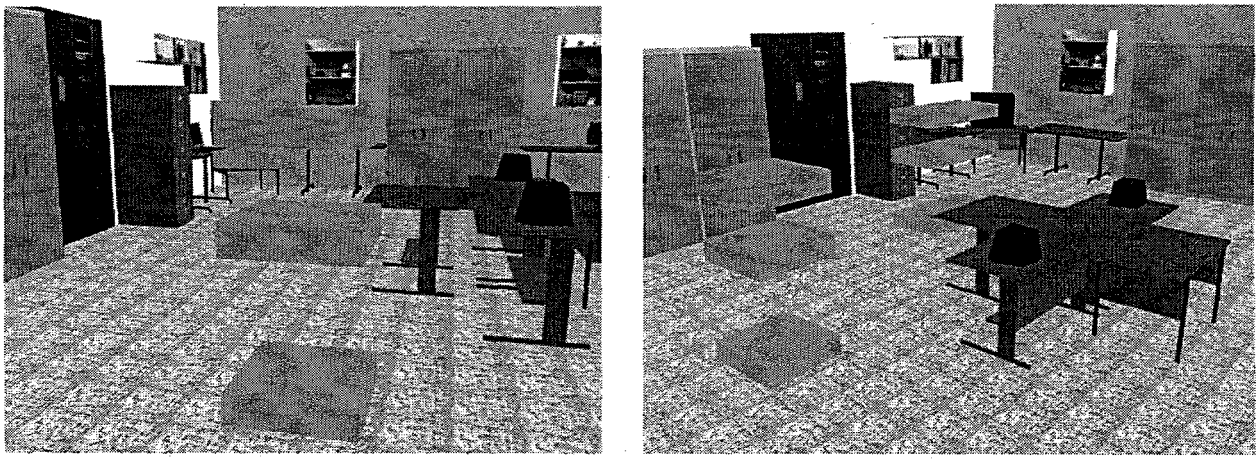
Sensor information other than VVMD data can also be monitored and displayed within the 3D model of the protected area. Figures 7 and 8 illustrate an intruder before and after entering a secured cabinet. A magnetic switch is used to monitor the status of the cabinet door. Similarly, the status of the main room door (open or closed) is conveyed to the VR interface and is appropriately displayed within the 3D model.

Figures 7 and 8. An intruder before and after entering a secured cabinet.

The VR component also includes voice annunciation. The voice annunciator provides oral information about important security events occurring within the protected volume. The oral information is presented concurrently with the visual display of sensor information.

2.3 Multiple viewing perspectives

Using VR technology, the remote operator is able to "move" through the 3D model and explore the relationships between the model and sensor information from multiple viewing perspectives. If a particular viewing perspective is undesirable, the operator can move to a more desirable one. This capability is illustrated in Figures 9 and 10 which show two different views of two people within the room. The operator can move close to an intruder displayed in the virtual environment to assess position, size, and intent. As the intruder moves, the operator can follow along within the virtual environment. A variety of technologies exist that allow the operator to navigate through the 3D graphical model. Some examples include boom-mounted displays, head-mounted displays with integral magnetic tracking devices, or a simple mouse interface.



Figures 9 and 10. Different viewing perspectives of two people within the room.

2.4 Benefits of the VR display interface

In the current system, VVMD and other sensor data are transmitted from the remote site via ordinary telephone lines and displayed in real-time within the virtual environment. The communication of important security events to security personnel via a single, intuitive, high resolution, 3D, color display is a key benefit of the system. Another benefit is the extremely low bandwidth requirement for transmitting detection data. The current system requires a communications bandwidth of only ten bytes/second. This is nearly three million times less than the bandwidth required for a single, uncompressed, Red-Green-Blue (RGB), digital video signal.

There are several other benefits to displaying sensor information within a virtual environment. Often, raw sensor information is not in a form that can be easily interpreted and understood—especially when taken out of the sensor environment context. One of our objectives is to demonstrate that joint sensor information (i.e., the combination of information from a variety of sensors) can be more easily interpreted when the operator can move through the virtual environment and explore the spatial and temporal relationships between the sensor data, objects and other visual cues present in the virtual environment. By exploiting the powerful ability of humans to understand and interpret 3D information, we expect to improve the means to visualize and interpret sensor information, allowing a security operator to assess a potential threat more quickly and accurately and to respond in a more effective manner.

3. CONCLUSIONS AND FUTURE WORK

Enhancements to the current system are being planned. These include the ability to intuitively interact with the sensor environment from within the VR model of that environment. Examples of this capability could include the remote activation of delays or barriers (e.g., cold smoke or sticky foam), or pointing a remote surveillance camera.

Another enhancement will be the addition of a sound field simulator. Audio signals, obtained from an array of microphones distributed throughout the protected volume, will be processed in real-time using low cost, but extremely powerful digital signal processing hardware. One of our goals is to present an audio signal to a listener located within the virtual environment (via headphones) which accurately represents what the listener would hear if she/he were actually at the corresponding location within the protected volume.

We are also improving the 3D resolution of the volumetric video motion detector. Using the current setup and using the current hardware configuration, we believe we can achieve a 3D zone resolution on the order of 3 x 3 x 3 inches. In the future, we hope to achieve a more realistic graphic portrayal of an intruder. As discussed previously, the hardware implementation of the VVMD component is also being upgraded.

The use of VVMD technology for personnel and material monitoring shows great promise. However, the exploration of potential applications of VVMD concepts to safeguards and security has only just begun. There is agreement that VVMD is a valuable security technology and that

continued research and development will undoubtedly lead to improved capabilities over conventional video motion detection systems. A broad area of application exists in the general area of surveillance. Exterior applications, such as battlefield surveillance, are being considered. The use of VVMD technology in a non-intrusive, biometric, identity verification system is also under consideration.

4. ACKNOWLEDGMENTS

We would like to acknowledge the U.S. Department of Energy's Office of Safeguards and Security for funding this project. We would also like to acknowledge the support and contributions of the following individuals:

Steve Ortiz
David Skogmo
Tricia Sprauer
Monica Prasad
Denise Carlson
James Singer
Debbie Lewis
Theresa Bourne