JAN 7 1998

SANDIA REPORT

SAND98–2803
Unlimited Release
Printed December 1998

Advanced 3D Sensing and Visualization System for Unattended Monitoring

Jeffrey J. Carlson, Charles Q. Little, Cynthia L. Nelson

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Prices available from (615) 576-8401, FTS 626-8401

Available to the public from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd
Springfield, VA 22161

NTIS price codes
Printed copy: A03
Microfiche copy: A01

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

SAND98-2803
Unlimited Release
Printed December 1998

Advanced 3D Sensing and Visualization System for Unattended Monitoring

Jeffrey J. Carlson and Charles Q. Little
Intelligent Systems Sensors and Controls Department

Cynthia L. Nelson
Security Technology Department

Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-1003

Abstract

The purpose of this project was to create a reliable, 3D sensing and visualization system for unattended monitoring. The system provides benefits for several of Sandia's initiatives including nonproliferation, treaty verification, national security and critical infrastructure surety. The robust qualities of the system make it suitable for both interior and exterior monitoring applications. The 3D sensing system combines two existing sensor technologies in a new way to continuously maintain accurate 3D models of both static and dynamic components of monitored areas (e.g., portions of buildings, roads, and secured perimeters in addition to real-time estimates of the shape, location, and motion of humans and moving objects). A key strength of this system is the ability to monitor simultaneous activities on a continuous basis, such as several humans working independently within a controlled workspace, while also detecting unauthorized entry into the workspace. Data from the sensing system is used to identify activities or conditions that can signify potential surety (safety, security, and reliability) threats. The system could alert a security operator of potential threats or could be used to cue other detection, inspection or warning systems. An interactive, Web-based, 3D visualization capability was also developed using the Virtual Reality Modeling Language (VRML). The interface allows remote, interactive inspection of a monitored area (via the Internet or Satellite Links) using a 3D computer model of the area that is rendered from actual sensor data.

Intentionally left blank

Contents

1. **Introduction** 5

2. **Rapid World Modeling** 6
   - 2.1. Background 6
   - 2.2. Traditional Modeling Approach 7
   - 2.3. New Modeling Approach 7
     - 2.3.1. Registration 8
     - 2.3.2. 3D Data Collection 8
     - 2.3.3. Filtering 8
     - 2.3.4. Editing and Segmenting 8
     - 2.3.5. Triangulation 8
     - 2.3.6. Decimation 8
     - 2.3.7. Primitive Geometry Fitting 9
   - 2.4. Results and Discussion 9
     - 2.4.1. Registration 9
     - 2.4.2. 3D Data Collection 9
     - 2.4.3. Filtering 10
     - 2.4.4. Triangulation 10
     - 2.4.5. Decimation 11
     - 2.4.6. Comments on Data Reduction 11
     - 2.4.7. Updating the World Model 11
     - 2.4.8. Primitive Geometry Fitting 12
   - 2.5. Change Detection Using 3D Range Data 12
     - 2.5.1. Method 12
     - 2.5.2. Experiment 12

3. **3D Video Motion Detection** 15
   - 3.1. Background 15
   - 3.2. Material Monitoring 16
   - 3.3. Personnel Monitoring 16
   - 3.4. Sensor Calibration 16
   - 3.5. Hardware Implementation 20
   - 3.6. Prototype System Description 20

4. **Virtual Reality Modeling Language (VRML) Interface** 21
   - 4.1. Background 21
   - 4.2. WEB-based 3DVMD Graphical User Interface 23
   - 4.3. Display of 3DVMD Data 24
   - 4.4. Benefits of a 3D Graphical User Interface 26

5. **Conclusions** 27

Intentionally left blank

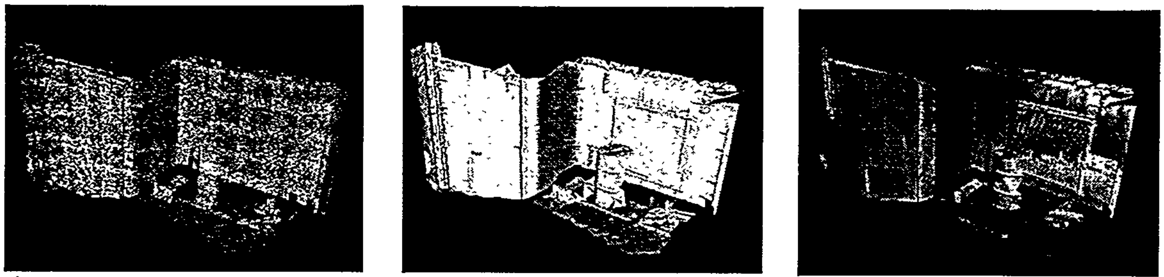
1. Introduction

This project was the result of integrating two separate, but synergistic, idea proposals submitted to VP 5000 for discretionary funding: Static and Dynamic 3D Monitoring System for Detecting Surety Threats, and Advanced 3D Visualization for Remote Site Monitoring. Our combined goal was to create the 3D sensing and visualization technologies to enable unattended 3D site monitoring and remote, interactive 3D site inspections.

Current 3D sensing systems are based on both passive and active sensing modalities. Active systems have been dominated by laser time-of-flight and structured lighting techniques, whereas the principal passive techniques have been stereo ranging and depth from motion, which employ more than one 2D image and make comparisons between them. There has also been work on passive techniques to derive depth information from single 2D images, notably shape-from-shading and shape-from-texture. Currently, none of these approaches provide a cost-effective or practical solution for monitoring, in real time, the 3D characteristics of both static and dynamic objects – especially in exterior applications. The objective of this project was to create both cost effective and reliable 3D sensing and visualization capabilities necessary for real-time, unattended monitoring. The sensing technique should be feasible in both interior and exterior monitoring applications and should provide a high level of stealth, such that the presence of the sensor is difficult to detect by a skilled adversary.

The monitoring system we envisioned relies heavily on real-time 3D sensing and interpretation of 3D sensor data. Portions of monitored areas can remain relatively static while other portions, including those occupied by humans or machines, can be very dynamic. Relatively static components can be modeled on the order of every few minutes. Dynamic components require 3D mapping at a much higher rate. To function in exterior applications, the sensing system also needs to maintain an adequate level of insensitivity to natural changes in lighting and other environmental factors including temperature and humidity. In order to maintain a high degree of stealth, the use of active sensing modes needs to be carefully controlled. Ease of application necessitates automatic sensor calibration and setup. Finally, an intuitive and easy to use Graphical User Interface, or GUI, is necessary to interactively visualize and interpret the sensor data.

Our technical approach was to create a tightly integrated system that leverages current Sandia research and expertise in 3D sensing and 3D modeling. Using a series of video cameras and lasers placed around a site, we created accurate 3D site maps using Sandia's Laser Mapper (LAMA) technology and monitored the 3D shape, location, and motion of humans and moving objects using Sandia's 3D Video Motion Detection (3DVMD) technology. The LAMA sensor works by projecting lines of eye-safe laser light into the scene and measuring their positions with cameras. It can provide updates about once every minute with range resolution down to a millimeter. Figure 1.1 illustrates various stages of the 3D mapping capability provided by LAMA.



a) b) c)

Figure 1.1. Processing stages using LAMA: a) 3D point cloud generated in scanning stage, b) Surface map generated from 3D point cloud, and c) Texture map generated using actual video data.

Unlike LAMA, the 3DVMD system is totally passive. The system monitors the scene from several points of view and triangulates to locate motion in 3D. It provides the sub-second updates necessary for monitoring humans and moving objects. Figure 1.2 shows a graphical portrayal of crawling and standing intruders detected by the 3DVMD sensor in an interior monitoring application. The block figures indicating the position and orientation of intruders are overlaid on a graphical rendition of the monitored area. This graphical rendition, or world model, is important in relating the 3DVMD detection data to objects within the monitored area. Prior to this project, the world model used by the 3DVMD was generated by hand and took several weeks to create. If the environment changes, the model had to be recreated. By integrating LAMA with the 3DVMD, the model can be continuously updated using actual sensor data.

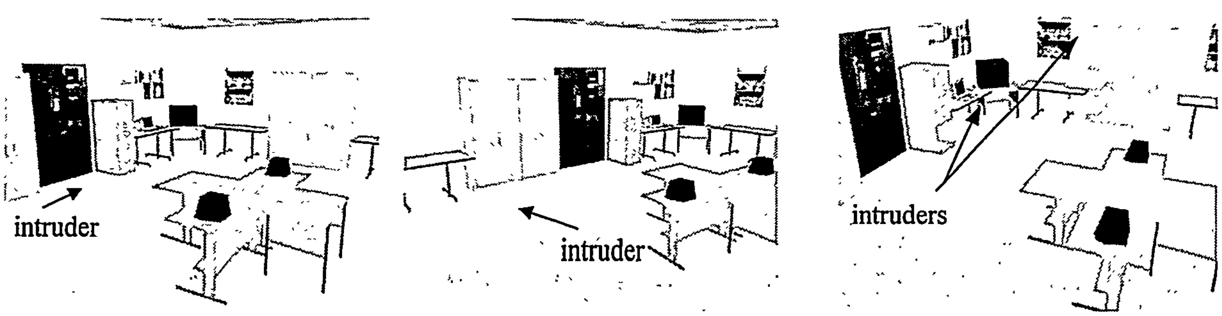


Figure 1.2. Graphical portrayal of intruders detected by the 3DVMD sensor.

2. Rapid World Modeling

2.1. Background. The ability to use an interactive world model, whether it is for simulation or most other virtual graphical environments, relies on the user's ability to create an accurate world model. Typically this is a tedious process, requiring many hours to create 3-D CAD models of the surfaces within a workspace. One of the goals of this project is to develop usable methods to rapidly build world models of real world workspaces. This brings structure to an unstructured environment. To accomplish this, LAMA is deployed to capture surface data within the workspace. This data is transformed into surface maps, or models. A 3D-world model of the workspace is built quickly and accurately, without ever having to put people in the environment.

World modeling is defined as the process of creating a numerical model of a real world environment, or workspace. This can be graphically displayed to provide the user with a 3D surface model of the workspace for simulations, analysis and task planning. In particular, world modeling enables 3D monitoring of the physical structure of a controlled workspace. This is critical for unattended monitoring applications.

An objective of this project is to demonstrate a suite of hardware and software capable of collecting and processing 3D range data, and then transforming that data into surface maps. These maps, or models, will be added together to form a 3D world model of the monitored workspace. Some application areas could include:

- Workspace updates, due to undocumented design changes, additions, deletions, etc.
- Unattended monitoring
- Facility mapping
- Underground storage tank mapping
- Under-sea operations

2.2. Traditional Modeling Approach. World models are traditionally generated using CAD tools. Modelers work from drawings where available, but frequently the design does not match construction, or modifications and additions are needed. Therefore the modeler must enter the workspace and carefully record distances and locations, then build up the model piece by piece. CAD modeling of a monitored workspace is a manual, time consuming operation. A single object in a workspace, such as a barrel, can take hours to days to model. This traditional method breaks down when objects are difficult or impossible to CAD model, such as dirt or loose surfaces, or crushed or dented objects. Even after good models are created, they must be accurately registered into the world model to be useful. Any movement of objects will require updating, forcing the modeler to re-measure distances and locations. Using the traditional approach, *creating the world model takes too long.*

2.3. New Modeling Approach. To overcome the problems of past approaches, and generate usable models quickly and accurately, we are developing tools to aid in model registration and in the creation of new or updated models. The primary modeling hardware uses video cameras and 3D range sensors. Selected range or distance measurements are used in updating and registering existing CAD models. Where traditional CAD modeling is not appropriate, dense range data is collected over an area and transformed directly into surface maps, or models. As additional areas are scanned and modeled, they are added together to form a 3D world model of the monitored workspace. An added advantage to this approach is that it can be done remotely, exposing sensors rather than people to the environment.

The following sections describe the components used in this new approach.

2.3.1. Registration. A very real problem in collecting data from 3D range sensors is registering the data to the real world it came from. To be useful in a world model, the sensors need to be calibrated to give accurate range information relative to the sensor. Just as important is knowing where the sensor was at the time the data was collected. Given these two things, the proper transforms can be generated to translate the data to the coordinate system of the world model.

2.3.2. 3D Data Collection. LAMA is used to capture surface data over an area. There are also several other 3D range sensors available. Each provides a list of range values of detected surface points scattered over the scan area. With sufficient information about the range sensor, this data can be transformed into an x-y-z coordinate system. Given the position and orientation of the sensor, this data can then be registered to the real world target workspace.

2.3.3. Filtering. Filtering is often needed to reduce the quantity of data and enhance the quality. 3D range sensors are capable of generating enormous amounts of data. Current models may output range images of 256 by 256 readings or more (over 64,000 points per scan). This is often more data than is needed, especially in areas that have little detail, such as flat floors and walls. Resolution limits can add a false texture to truly flat surfaces. 3D range data also has noise and perhaps artifacts. These effects can often be reduced by filtering.

2.3.4. Editing and Segmenting. Editing and segmenting of 3D scattered data gives the user more control of subsequent processing. Usually, there is data within the scan that is not of interest. Frequently, false data points appear, perhaps due to sensor limitations with highly reflective surfaces. It is often advantageous to remove this extraneous data. Segmenting of the data allows subsections to be handled independently. This can be useful in dividing the workspace into logical objects. Both user assisted and automated techniques are needed.

2.3.5. Triangulation. The next step is to connect these scattered data points. The data is processed to form a triangulated surface, or a series of triangles. Often the vertices are the original scattered data points. This process may be trivial where data is collected in a uniform grid. Other techniques are available to handle truly scattered data, such as Delaunay triangulation.

2.3.6. Decimation. Due to the usually large number of scattered input data points, the resulting triangle list can be quite large; often too large to be useful in the world model. The next step is data reduction. One method is to reduce the triangle count by removing vertices, using a technique called decimation. As with most data reduction, decimation lowers resolution. The desire is to keep high resolution data where surfaces are changing rapidly, while giving up data where surfaces are not changing.

2.3.7. Primitive Geometry Fitting. Another avenue of reducing data is through primitive geometry fitting. Segments of the scattered data can be used to fit a simple geometry (spheres, cylinders, boxes...). The simple geometry object becomes the model, replacing the scattered data segment. Fitted geometries have much lower data requirements than triangulated surfaces. This technique has great promise in creating models of many surfaces, such as man-made objects (pipes, tanks, walls, floors, etc.).

2.4. Results and Discussion. The results, along with some commentary of the work performed in the development of the components mentioned above, follows.

2.4.1. Registration. Data registration from LAMA is accomplished by calibrating all video and laser systems after they are installed. We have used surveyed positions for initial calibration points. This allows us to simultaneously calibrate the sensors and back out the sensor positions for proper registration of their data output.

A problem we often face is having valid models of objects in the work cell, but incorrect locations because they have been moved around. One way we resolve this is by using stereo imaging techniques to locate pre-selected fiducials, such as corners. By comparing these points to their counterpoints in the models, valid transforms can be generated to register the model with its real world position.

2.4.2. 3D Data Collection. The LAMA sensor is based on structured lighting. This system consists of a light source and a camera detector, with geometric processing to determine range. A high intensity, tightly focused light (a laser) is steered over the target area. The reflected signal is detected through a camera. A-priori knowledge of the camera and laser (position, orientation, optics) allow triangulation techniques to map out surface range information. The beam pattern is a line. Targets larger than the beam width or field of view of the camera can be covered in multiple passes. Unlike laser range imaging sensors that provide a dense 2D range grid or image, the output of our structured lighting sensor is a string of irregularly spaced points, with little implicit connectivity.

Figure 2.1 is an image of a six foot bowl that sits in one of our labs, from one of several video cameras mounted around the lab. The contents include a 30 gallon ribbed drum (with a large dent) imbedded in a vermiculite 'soil'. A flat wood block was also included. Scattered data was collected from one of several overhead 3D sensors of the scene of Figure 2.1. The scanning gathered 17,067 x-y-z data points.

The time it takes to gather data is based on the desired density and the number of sweeps needed to cover the target area. This scan made 5 sweeps and took about 5 minutes. In addition, we are also developing stereo imaging techniques to gather sparse range data. This will allow us to find surface points in the scan area along high contrast areas, such as edges and corners. This is complementary to the data gathered by typical 3D range sensors, which do best on broad surface areas and poorly at edges. The data will be fused to enhance coverage.

2.4.3. Filtering. From the beginning we often have too many points to work with, so we have developed a spatial filter to grossly reduce data counts. The filter is designed for the irregularly spaced data of our structured lighting system. This also allows it to be applied to any 3D sensor data, and to multiple scans of data at once.

The initial data (17,067 points) was passed through the spatial filter to reduce data by eliminating neighbors within one inch. Additional data that fell on the floor surrounding the bowl was also removed by manual editing. This reduced the point count to 2,290 points. At this point, the data reduction is 86%. Figure 2.2 shows a view of the 3D range sensor data, after filtering.

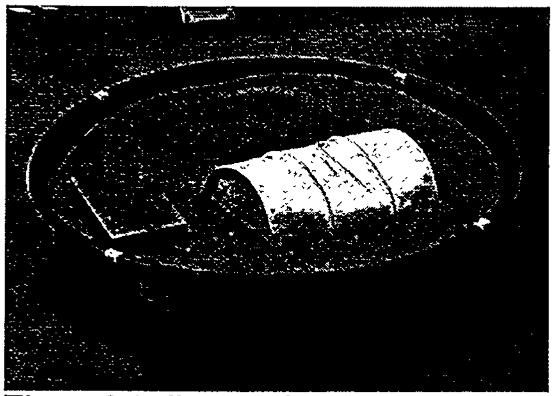


Figure 2.1. Image of target workspace.

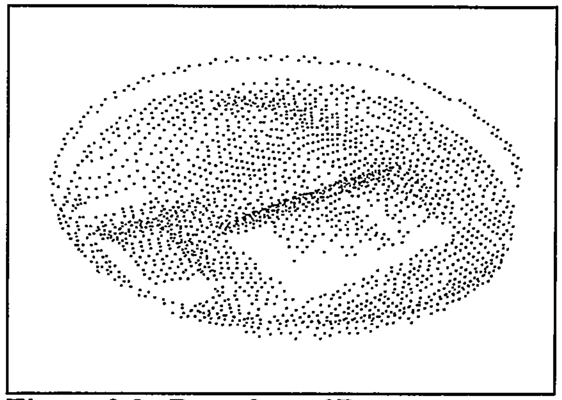


Figure 2.2. Data from 3D range sensor.

2.4.4. Triangulation. The next process is to connect the dots to form a triangulated surface. Our current process is based on a Delaunay triangulation. This grids in x and y, but tracks the z value so irregular 3D sheets are generated.

Figure 2.3 is the result of triangulation of the data shown in Figure 2.2. The polygon count is 4,528. Notice in particular the ribs and the dent in the barrel are easily seen. This technique is computationally intensive, but at these data counts takes only a few seconds on our equipment (an SGI Indigo II). Although we can triangulate, the present algorithm has limitations when fitting data from multiple views. We are working on a full 3D triangulation algorithm.

It can be seen from Figure 2.2 that data was not collected over all surfaces. In particular, the area from the rim of the bowl to the fill, and at the base of the barrel, are blank. These areas fell in the shadow of the 3D sensor. A decision must be made, whether to create surfaces that span the holes or leave them blank. We usually choose to span them.

Another potential concern is the blanketing effect: raw surfaces tend to come out looking like a snow covered vista. It is difficult to tell when one part ends and another begins. This can be improved by editing and segmenting, and perhaps color coding of various pieces. Another way to increase user perception is by mapping video images onto the surface through texture mapping.

2.4.5. Decimation. We have included a decimation method as part of our approach. Parameters have been added to maintain areas of high detail, such as edges, while allowing low detail surfaces, such as the flat areas between edges, to be reduced.

Figure 2.4 is a decimated version of the original surface model shown in Figure 2.3. Some fine detail is lost, but the salient features remain. The data count is 2,190 polygons, which is a 52% polygon reduction. This process takes a few seconds on our present system. Total data reduction is better seen comparing the original scanned data (17,067 vertices) to the final (1,121 vertices), or a 93% reduction.

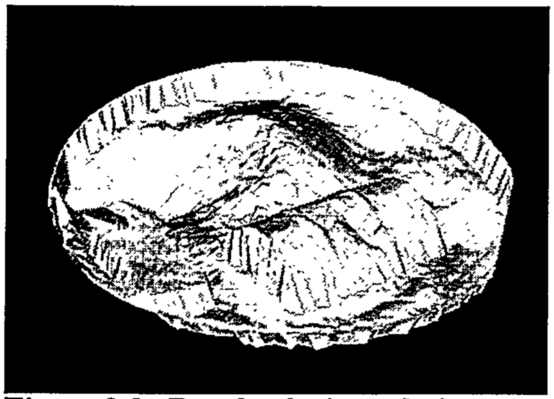


Figure 2.3. Result of triangulation.

Figure 2.4. Result of decimation.

2.4.6. Comments on Data Reduction. One of our requirements of the world model is that it be accessible in real time. Even with high-speed display hardware, there is a limit to the number of polygons that can be re-displayed per second, which directly translates to the total polygon count for the world model. Performance is also affected in change detection, where a current model is compared to a reference model on a polygon by polygon basis. This places a penalty for models with high polygon counts. On the other hand, operator perception benefits from high resolution. Therefore, parameters for data reduction often depend on the application needed.

2.4.7. Updating the World Model. The final step is to place the newly created surface into the world model. At this point, the registration issues have been resolved, and the world model becomes one step closer to representing the real workspace. The process continues until a sufficient level of modeling is reached.

The next two figures show an example of a robot manipulating the workspace. A robot tool was tasked to move to the dented portion of the barrel and perform a swiping motion. Figure 2.5 is a view from the world model. The robot arm, the floor, and the bowl models where created by CAD modeling. The surface map from above was then added. The location and orientation of the target spot on the barrel was taken directly from the surface model. Figure 2.6 is a camera view of the robot tool performing the swiping motion operation. The robot can now interact with the model, both to move to an object, and to detect collisions before they happen.

To add some perspective, the bowl took about 4 hours to model, using traditional CAD modeling, with all measurements taken by hand (there where no drawings). The surface map took about 6 minutes from start to finish.

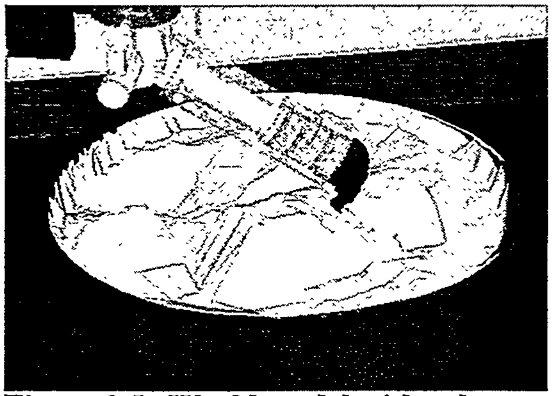


Figure 2.5. World model with robot.

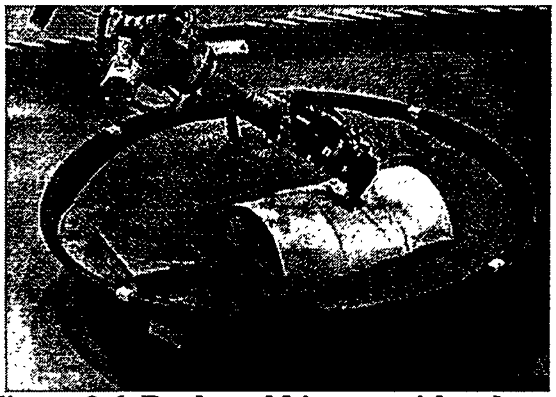


Figure 2.6. Real world image with robot.

2.4.8. Primitive Geometry Fitting. We have experimented with primitive geometry fitting, using quadratic surfaces. An object, such as the barrel in Figure 2.3, can be replaced with a primitive geometry model of a cylinder. In some analysis tasks, such as planning the removal of the barrel, the cylinder model may be better than the more detailed barrel surface. It provides more information of the overall shape of the assumed object, even though fine detail, such as the dent and ribs, are lost.

2.5. Change Detection Using 3D Range Data

2.5.1. Method. This method uses 3D surface models to detect changes in a volume of space. LAMA is first deployed to capture information within the volume. The range sensor captures positional information of the surfaces it scans. The raw data consists of a point cloud, where each point is a x-y-z position in space. By connecting the points together, the surface can be recreated. To detect change, the sensor is deployed a second time, and the data is compared to the original. Missing, moved, or added objects show up as differences between the data sets.

In this application, the change detection was done by surfacing the first data set and comparing the points in the second data set to the surface in the first. This is more complicated than a dot-by-dot comparison, but many range sensors do not guarantee a dot-dot correspondence between subsequent range scans. This also allows the sensor to be moved between scans, where dot-by-dot correspondence would certainly be lost.
A distance threshold is used to flag the change detection. This value is set above the surface variation expected between scans, based on the inherent noise in the system. A second issue is the level of registration between scans; the points must be in the same coordinate system before the comparison is viable.

2.5.2. Experiment. An experiment was done as proof-of-concept. LAMA was used to monitor the front of a metal lab cabinet with two shelves opened. Figure 2.7 shows the point cloud data from the initial scan. Figure 2.8 is a photo of the cabinet. A cardboard box on the lower shelf was moved about a half inch forward. Figure 2.9 shows a photo of the cabinet after the change. Obviously, the change is not apparent. Figure 2.10 shows the results of the algorithm. The gray sheeting is the surfaced version of the first data set. The green objects are highlighted points in the second scan that exceeded the detection threshold.

A second experiment shows a missing object. In this case, a pair of pliers was removed from the lower shelf, on the left side. Figure 2.11 shows the scene with the pliers in place. Figure 2.12 is a photo of the same scene with the pliers gone. Figure 2.13 shows the results in this case. The green objects again are highlighted points in the second scan that were detected as different from the initial scan.

Figure 2.7. Point cloud from the initial scan.

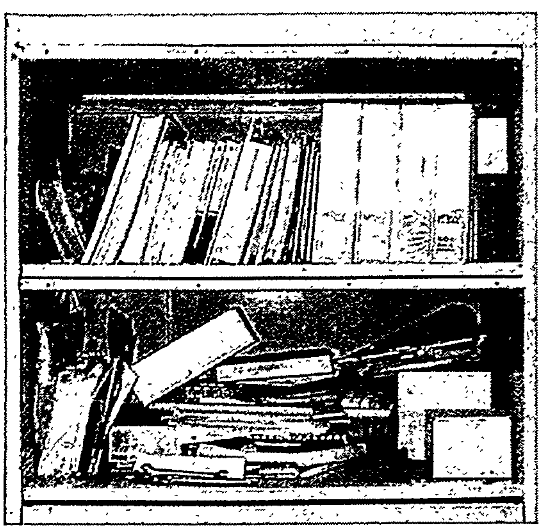


Figure 2.8. Photo of the cabinet.

Figure 2.9. Cabinet after the change.

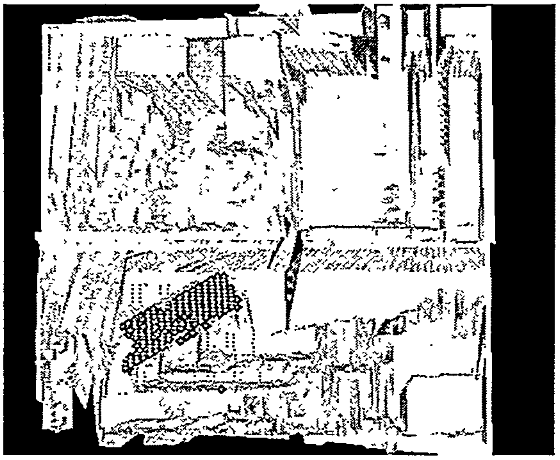


Figure 2.10. Change detected from scene 1 to 2.

Figure 2.11. Scene with pliers.

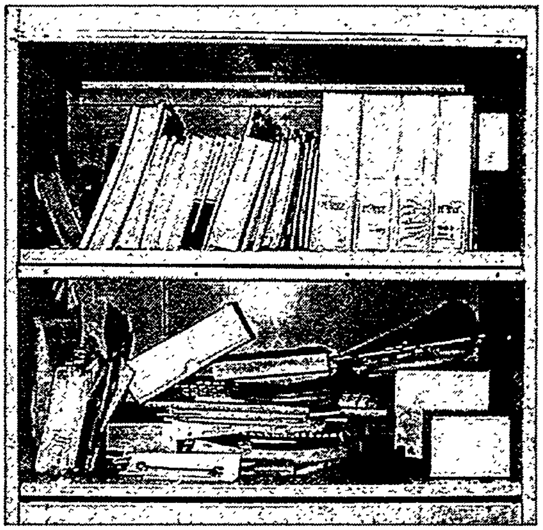


Figure 2.12. Scene with the pliers gone.

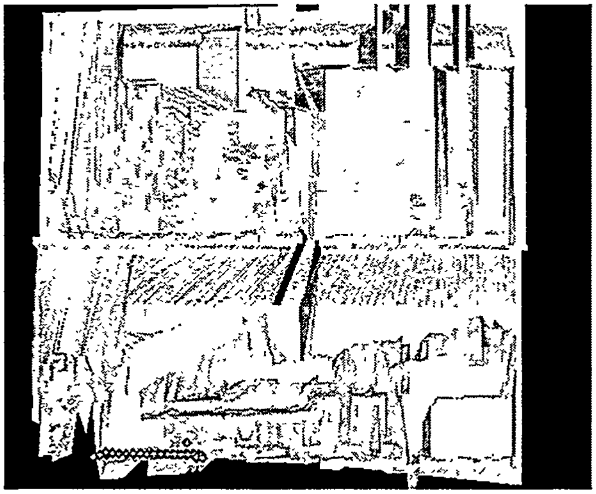


Figure 2.13. Change detected from scene 2 to 3.

3. 3D Video Motion Detection

This section highlights the operation, hardware implementation, and automated calibration procedure for the 3D Video Motion Detection (3DVMD) system. The communication and display of sensor information is also addressed.

A goal of this research and development project was to provide a new, state-of-the-art sensor technology for improving detection, assessment, and response capabilities of security systems.

This work capitalizes on the ability of 3DVMD technology to provide three-dimensional (3D) information about the position, shape, and size of moving objects 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 the Virtual Reality Modeling Language (VRML) to display information relating to the sensors and the sensor environment. VRML enables an operator, or security guard, to be immersed in a 3D graphical representation of the remote site containing the video sensors. 3DVMD data can be transmitted from the remote site via ordinary telephone lines, the internet, or any other low-bandwidth communication link, and can then be displayed in real-time.

There are several benefits to displaying 3DVMD 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 3DVMD 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. The 3D graphical display also provides the potential for fusing and visualizing information from sensors other than the 3DVMD – all from within the same graphical user interface. 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.

3.1. Background. The primary goal of this project has been 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.

The 3DVMD technology 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 3DVMD 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 3DVMD technology: 1) material monitoring and 2) personnel monitoring. The similarities and differences in these applications are discussed in the following sections.

3.2. Material Monitoring. In the material monitoring application, arbitrary zones, occupied by materials, 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.

3.3. 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 later.

3.4. Sensor 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 has been developed. The calibration process involves three steps: lens calibration, camera orientation calibration, and zone masking.

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 large non-linear distortions in the video images as shown in Figure 3.1. The distortion is particularly evident near the perimeter of each image. The problem is further complicated by variations among lenses having the same specifications.

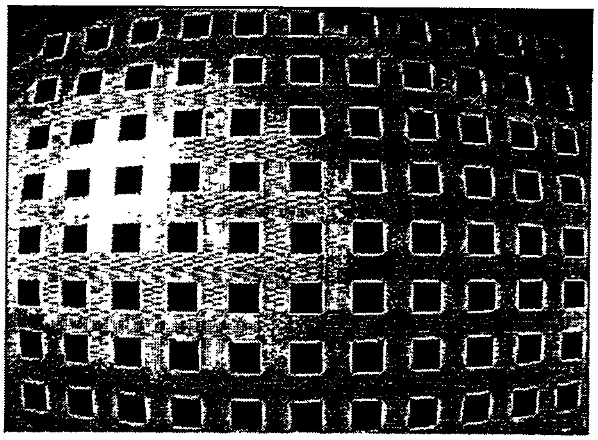


Figure 3.1. Distortion resulting from a wide-angle lens.

The calibration procedure begins by calibrating each camera and lens. A calibration board is placed in front of the camera, as portrayed in Figure 3.2, with the lens focused at infinity.

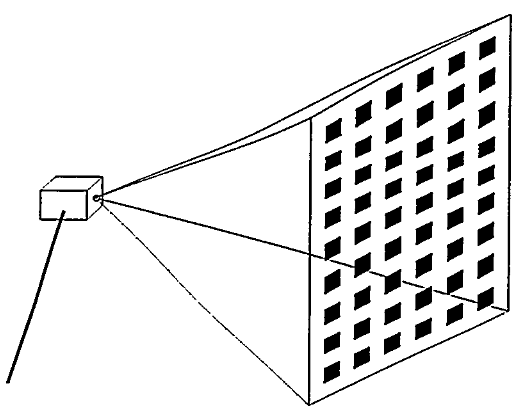


Figure 3.2. Portrayal of camera positioned for calibration.

Next, the coordinates of pixels mapping to the corners of the squares are located, as illustrated in Figure 3.3, and the pointing angles for these pixels, relative to the optical axis of the camera, are established.

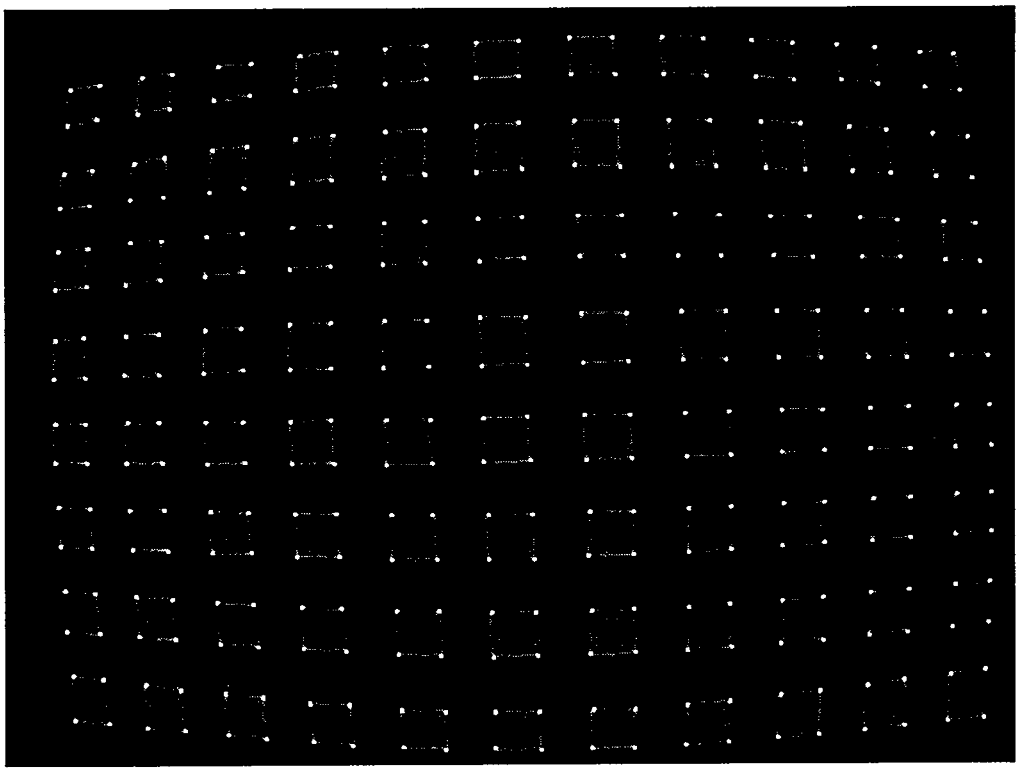


Figure 3.3. Pixels located at vertices of the squares.

An affine transformation is then used to interpolate the pointing angles for all other 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. Two images representing the relative azimuth and elevation angles for each pixel are shown in Figure 3.4.

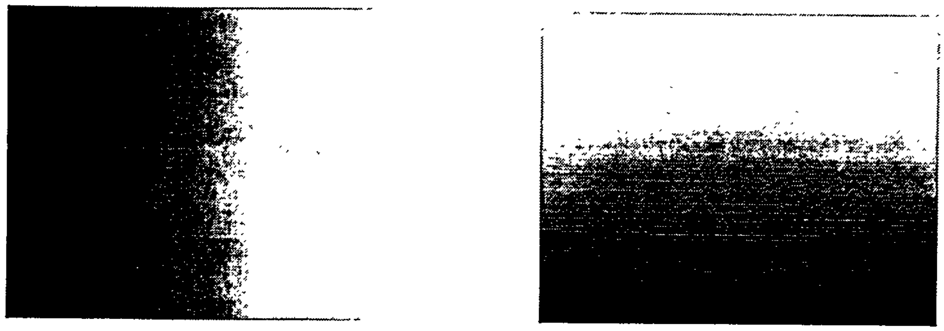
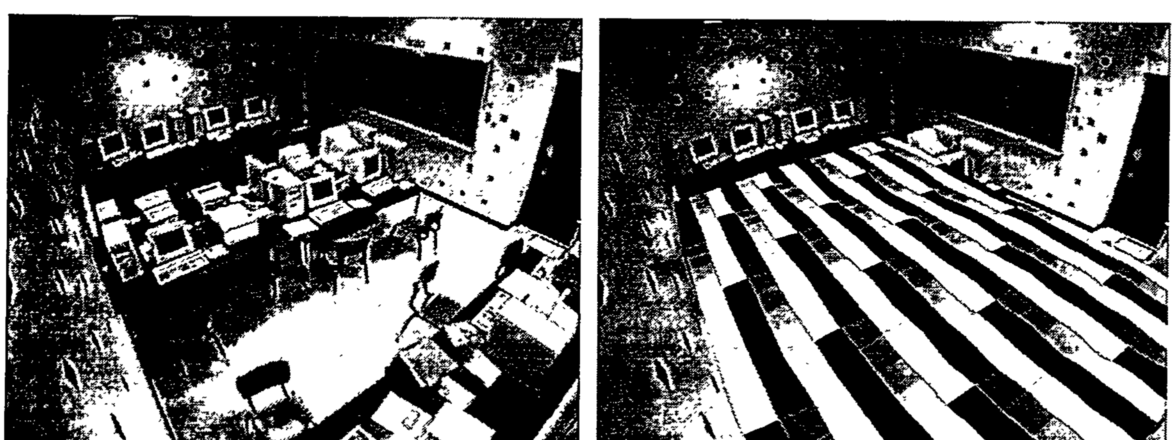


Figure 3.4. Relative azimuth and elevation angles for image pixels.

The next step in the calibration procedure involves mounting the cameras in the area to be monitored and obtaining their 3D positions. For interior applications, the camera positions can easily be 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 the 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. Figure 3.5 shows a camera view of the monitored area used in developing the calibration procedure and a partition and masking of pixels to one-foot square zones at the floor level. The zones are shaded in the figure for the purpose of illustration. In actual practice, each zone is assigned a unique shade, or code.



a) b)

Figure 3.5. Camera view of monitored area (a) and corresponding partition of pixels to one square foot zones at the floor level (b).

3.5. Hardware Implementation. The 3DVMD 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.

The 3DVMD system used a single Pentium computer system as a host with a Peripheral Component Interconnect (PCI) bus and multiple PCI bus framegrabbers. The host provides both image processing and user interface functions. The PCI bus provides a high bandwidth communications channel for transferring 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).

3.6. Prototype System Description. Our prototype system used 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 prototype system was 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 3.6 shows the room as seen by each of the four cameras.

Figure 3.6. Surveillance volume as seen from each camera.

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.

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 VRML to display information relating to the sensors and the sensor environment. VRML enables an operator, or security guard, to be immersed in a 3D graphical representation of the sensor environment via the Internet. The 3D graphical representation depicts essential details within and around the protected volume in a natural way for human perception. The VRML component also includes voice annunciation. The voice annunciator provides audio information about important security events occurring within the protected volume

4. Virtual Reality Modeling Language (VRML) Interface

4.1. Background. Data display is a key element of a remote monitoring application. An interactive graphics viewer is highly desirable to show the full effects of the 3D surface data captured with the range sensors, along with the 3D geometric information provided by the other sensors. For this application we have used the WEB based display format VRML 2.0 (Virtual Reality Modeling Language). This allows the data to be archived as a web site that could be transmitted across a local area network or across the Internet using 128-bit encryption. To view the data you would use an Internet browser (such as Internet Explorer or Netscape) and a VRML browser plug-in, which allows the user to load, view and interact with 3D worlds. Figure 4.1 shows a 3D surface map of a lab.

Figure 4.2 is the same data with texture mapping. Note that both figures are screen shots of a standard Netscape browser. The VRML format handles both geometry and image data. Any user with the standard browser (and proper access) can download the data and then view the data from any angle.

Additionally, Java scripts can be downloaded as well to provide customized interaction with the data sets. In this way, a history of motion could be replayed, specific objects could be animated, or other related data could be incorporated into the data set that would be viewable when a user selected an area of the screen.

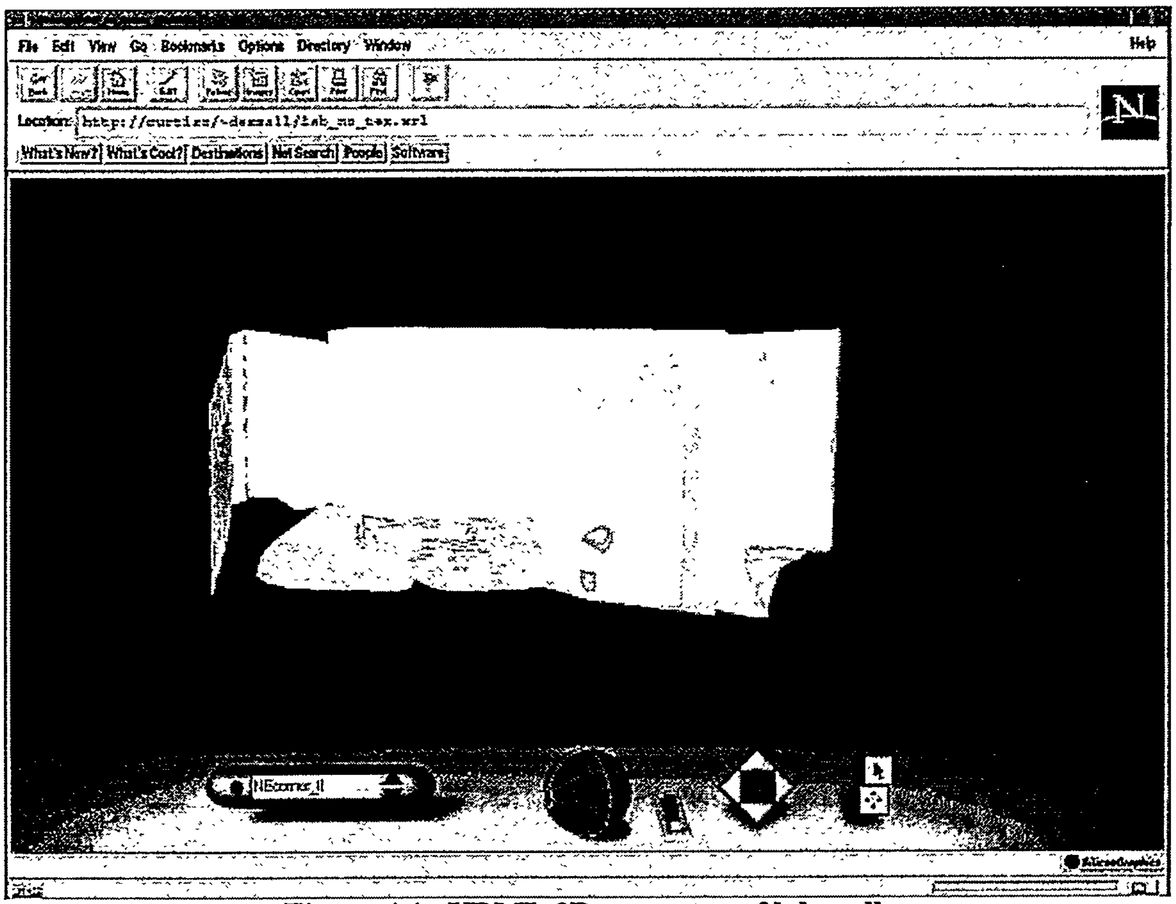


Figure 4.1. VRML 3D geometry of lab wall.

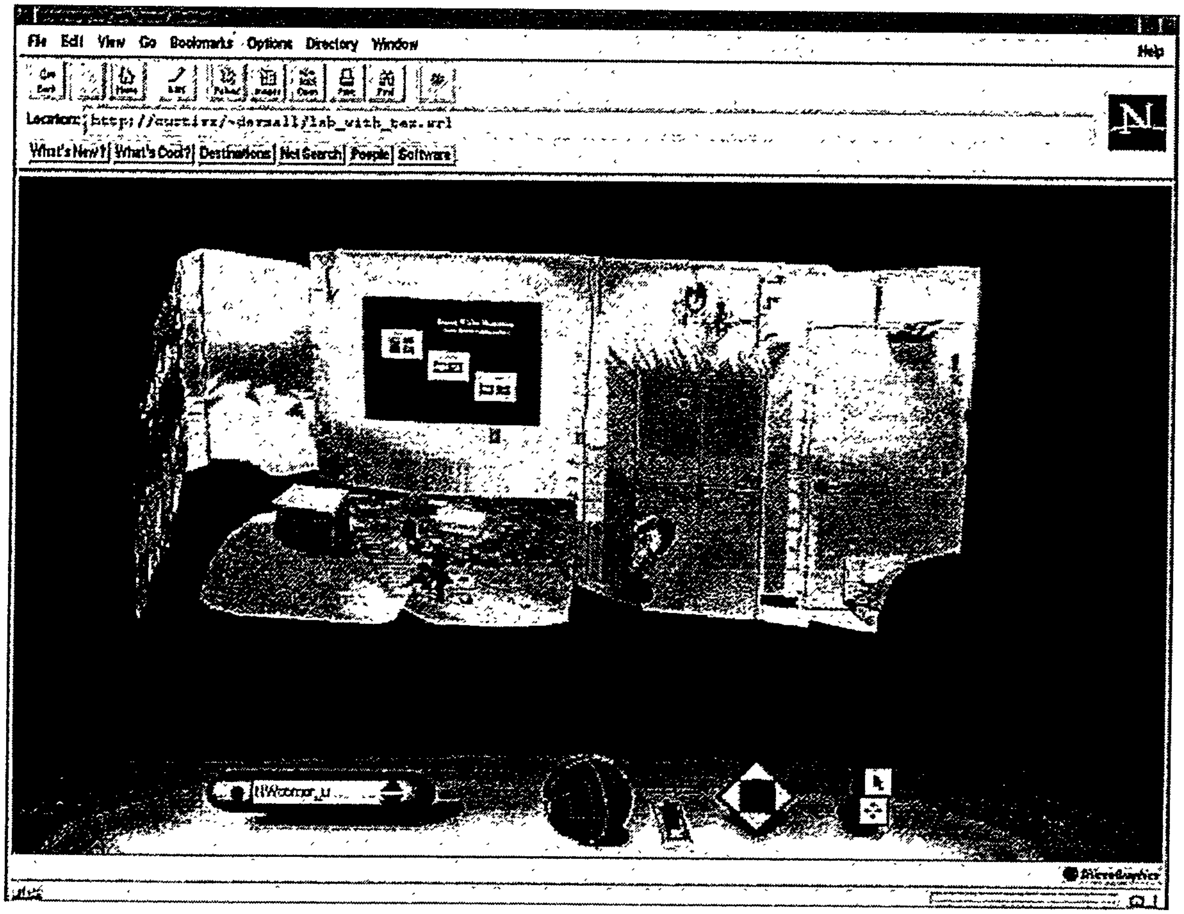


Figure 4.2. VRML model with texture mapping.

4.2. WEB-based 3DVMD Graphical User Interface. VRML is used to display 3DVMD 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 can be created from physical measurements and photographs of the actual sensor environment or from a 3D-range sensor such as LAMA. As illustrated in Figure 4.3, the 3D graphical representation depicts essential details within and around the protected volume in a natural way for human perception.

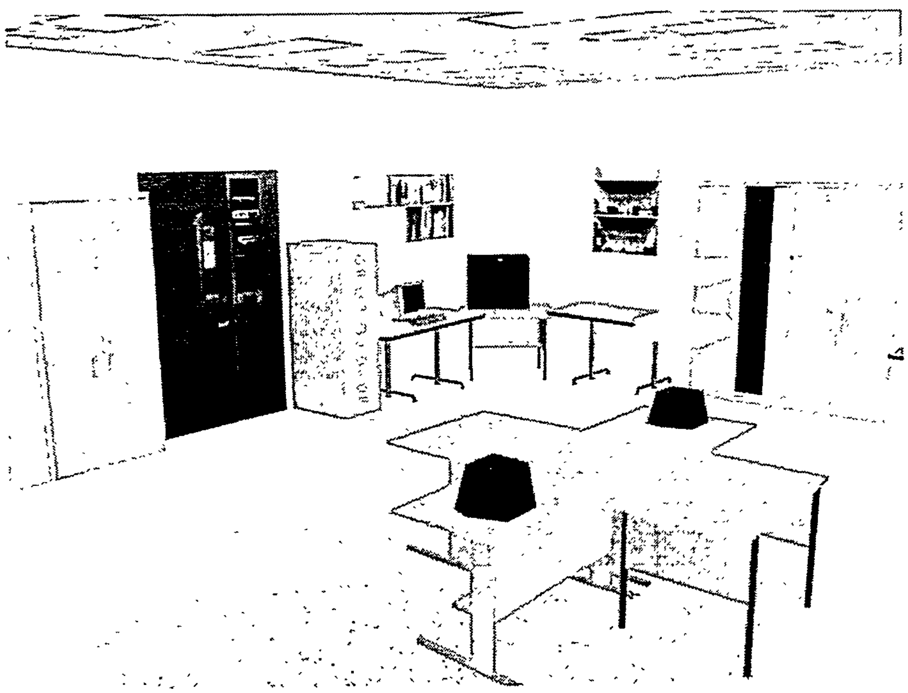


Figure 4.3. 3D graphical representation of the sensor environment.

4.3. Display of 3DVMD Data. 3DVMD data is overlaid on the world model in real-time to indicate personnel movements. This is illustrated in Figure 4.4 which shows a graphical rendition of a crawling intruder displayed within the world 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 3DVMD data, the position, shape, and height of personnel are indicated, as is the status of protected assets.

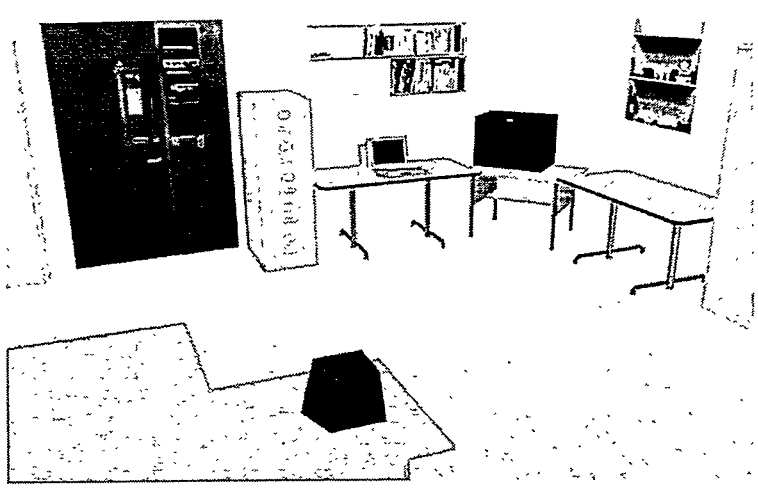


Figure 4.4. Graphical rendition of a crawling intruder.

Figure 4.5 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 3DVMD.

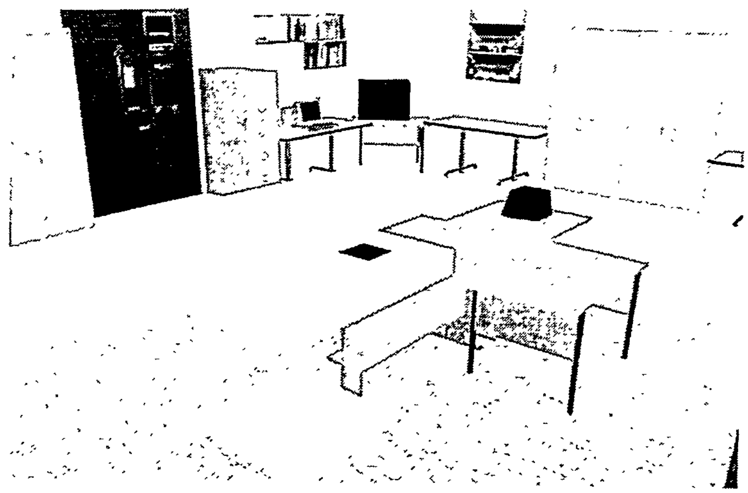
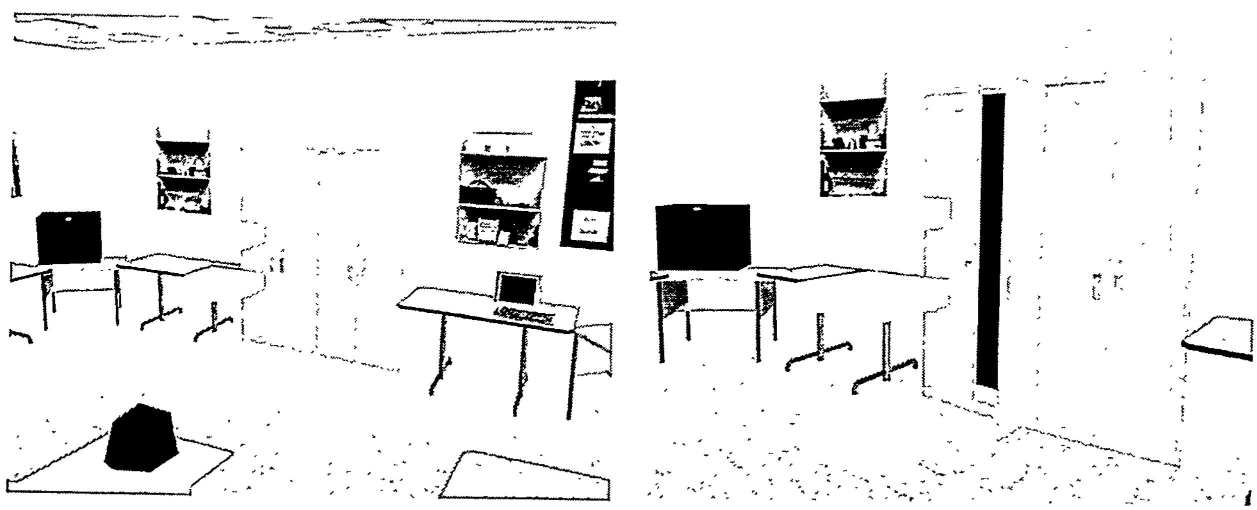


Figure 4.5. Graphical rendition of protected asset violated by a standing intruder.

Sensor information other than 3DVMD data can also be monitored and displayed within the 3D model of the protected area. Figures 4.6 and 4.7 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 VRML interface and is appropriately displayed within the 3D model.



Figures 4.6 and 4.7. An intruder before and after entering a secured cabinet.

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

Using VRML, 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 4.8 and 4.9, 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 magnetic tracking devices, or a simple mouse interface.

Figures 4.8 and 4.9. Different viewing perspectives of two people within the room.

4.4. Benefits of a 3D Graphical User Interface. In the current system, 3DVMD and other sensor data can be transmitted from the remote site via ordinary telephone lines or the Internet and displayed in real-time. 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 of the 3DVMD is the extremely low bandwidth requirement for transmitting detection data. The 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 3D graphical display. 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. One of our objectives was 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. 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.

5. Conclusions

The use of 3D sensing technologies for remote unattended monitoring shows great promise. However, the exploration of potential 3D sensing applications for safeguards and security has only just begun. There is agreement that 3D sensing is a valuable security technology and that continued research and development will undoubtedly lead to improved capabilities over conventional security sensor technologies. A broad area of application exists in the general area of surveillance. Exterior applications, such as battlefield surveillance, are currently under development. The application of VRML has also demonstrated that the Internet is a useful tool for unattended monitoring applications.

Distribution

10	Jeffrey Carlson, 9611	MS 1003
1	Charles Little, 9611	MS 1003
1	Rush Robinett, 9611	MS 1003
1	Cynthia Nelson, 5838	MS 0780
1	Steve Ortiz, 5838	MS 0780
1	LDRD Office, 4001	MS 0188
1	Central Technical Files, 8940-2	MS 9018
2	Technical Library, 4916	MS 0899
1	Review & Approval Desk, 15102 For DOE/OSTI	MS 0619