TWO 3-D SENSORS FOR ENVIRONMENT MODELING AND VIRTUAL REALITY: CALIBRATION AND MULTI-VIEW REGISTRATION

S.F. El-Hakim, J.-A.Beraldin, G. Godin, and P. Boulanger
Institute for Information Technology, National Research Council
Ottawa, Ontario, Canada

Commission V, Working Group 1

KEY WORDS: Calibration, Modeling, Registration, Graphics, Scanner, Real-time, Three-dimensional, Virtual-Reality.

ABSTRACT

Virtual environments (VEs) also known as virtual reality (VR) are increasingly being considered for industrial, medical, and educational / training applications, to name a few. A VE provides real-time interaction with 3-D models when combined with a display technology that gives the user immersion in the model world and direct manipulation of objects. In many of the applications, the truthful representation of the environment and the accurate manipulation and navigation in the virtual world are crucial. In this paper we examine the potential use of laser range cameras and digital photogrammetry in the accurate creation of VE models of real scenes and in the tracking of the user for precise interaction with models.

I. INTRODUCTION

Virtual environments are defined as the real-time graphics interaction with three-dimensional models, when combined with a display technology that gives the user immersion in the model world and direct manipulation (Bishop and Fuchs, 1992.) The technology will radically change the way people interact with computers and allow them to act as if they were in places they are not. Obviously, the entertainment industry is the leading market, however many other applications do exist. For example, training such as flight simulators, industrial design and prototyping (Hedberg, 1996), medical (Goble, et al, 1995), and military (Polis et al, 1995) applications are now employing the technology. VE is currently advancing at a very rapid pace in both research organizations and industry. There are many useful publications that may serve as introduction to the topic and document the state of the technology (e.g. Adams, 1993, Azuma, 1995, Bishop and Fuchs, 1992, Phillips-Mahoney, 1995, Sturman and Zelter, 1994, and Stevens, 1994). We will only give a brief overview here.

1.1. Overview of VE Technology

Virtual exploring of real places and environments, either for leisure, engineering design, simulations, or tasks in remote hazardous environments, is more effective and useful if geometrical relationships and dimensions in the virtual model are accurate. Also, since in VE the rendering of images must respond immediately to one's movements, the relationship between the viewer's head and hands and the 3-D environment must be continuously and accurately known. This is also true for interacting with and manipulating objects in that environment. The degree of accuracy of the modeling and positioning will widely vary with applications. Even within an application the accuracy requirements may vary. For example, the accuracy of the spatial location and orientation of doors and openings through which the viewer or moving platforms will go, is higher than other details.

Figure 1 summarizes the main components of a VE system. First, the 3-D world has to be created. The "computer-generated" environment can be a truthful representation of the "real" environment if precise, well-calibrated, laser range cameras are used to digitize the latter to create the former. However, for several reasons such as availability and cost, most models are built by either using standard geometric primitives, libraries of pre-modeled objects, or manual digitizing of every point. Building such a model graphically for a detailed environment takes enormous efforts and time and may look unrealistic. On the other hand , digitizing the environment with laser range cameras is an excellent alternative to graphically creating the model. It saves time and effort while providing a more realistic model. Real-world 3-D image-based VE can advantageously complement or replace artificially created VE in many endeavors. Currently, creating such models of the real world remains an obstacle for this technology and is a limitation to the implementation in a wide range of useful applications. Other components of the VE system include the head trackers, the image rendering engine, and the 3-D display (either head mounted or one or more computer screen). The rate of all the processes must be fast enough to update the display at 20 Hz or faster.

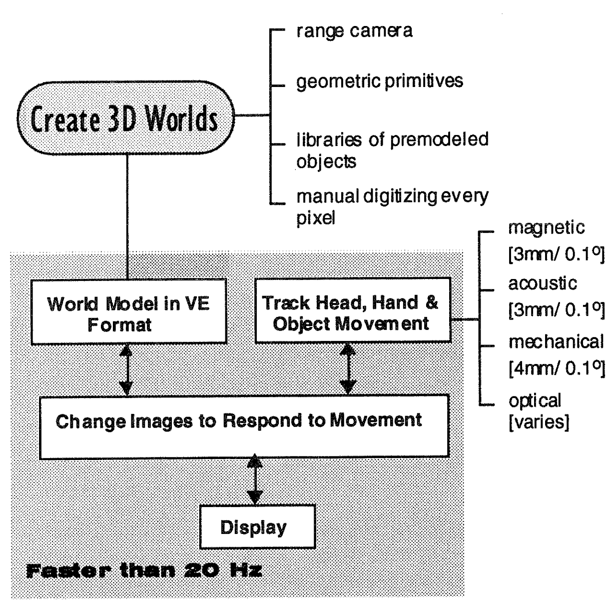


Figure 1: The main components of a VE system

Models suitable for virtual environments differ from those designed for other graphical environments, such as CAD systems (Phillips-Mahoney, 1995). The main constraints are the real-time requirement and how fast the rendering engine is capable of producing an image on the display. These constraints will decide on the maximum number of polygons in the model, thus the level of details that can be handled. The model must also include information on object behavior (response to user action), a script, the hierarchical relationship between objects, and other attributes such as texture and sound. The model must also include a multi-resolution hierarchical representation of all its components. For example, when objects are further than a given distance from the user, it is sufficient to display them at a decreased resolution up to a distance where the objects are not displayed at all. Figure 2 shows the elements of a virtual-world data base (model) and how this model relates to the other processing components of the system.

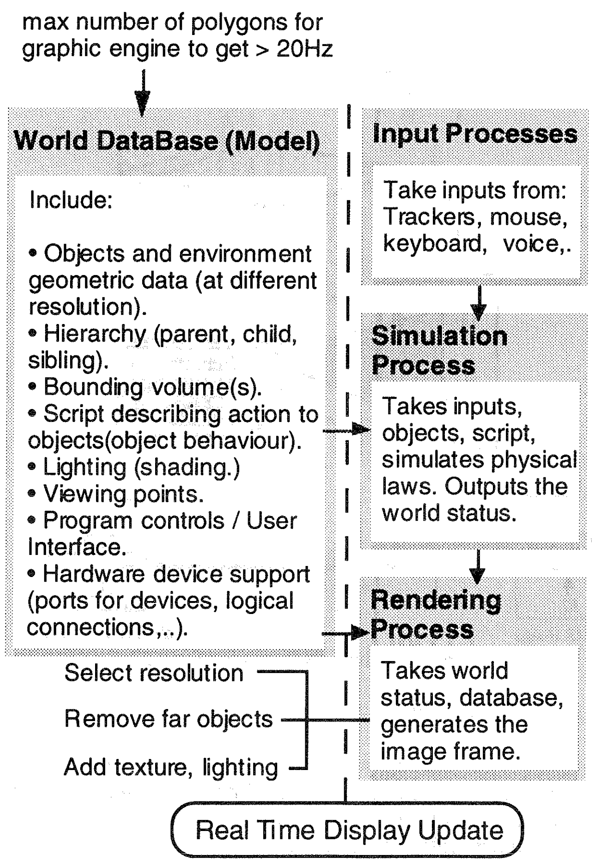


Figure 2: Software for VE: Models and Processes

1.2. Sensors for Creating the VE World

Since VE interacts with 3-D environments, relaying on video images to generate a full non-structured environment is usually not sufficient to capture all surfaces and features. Sensors that directly produce 3-D images are better suited for the creation of the geometric model. However, video images are useful for adding texture (Haeberli and Segal, 1993.) In addition, photogrammetric techniques can give accurate coordinates for certain features and assist in the registration of the 3-D images generated by laser range cameras.

The construction of most environment models will likely require a large number of views, due to the limited field of view of sensors as well as the complexity of the scene itself. Each view taken with a range camera provides an ordered set of 3D points in a camera-centered Cartesian coordinate system. In order to reconstruct the model, the rigid transformation linking the different views to a common Cartesian coordinate system must be recovered. This process is known as registration. The position and orientation of the camera at each acquisition station could be accurately measured with extrinsic means (such as inertial systems), which are costly and may be impractical, or by intrinsic means from the data itself.

1.3. Tracking in VE

Photogrammetry, when performed in real time, can be used for tracking the position of the viewer's head and hands. Currently, tracking is based on magnetometer sensors, ultrasonic sensors, or mechanical devices. The first loses accuracy in the presence of conductive metals and, to a lesser extent, stainless steel, and the second is affected by interference sound sources which make these technologies not suitable for applications in some environments such as factory floors. The mechanical devices are bulky and not convenient for most applications. The current accuracy levels of these sensors and devices are 3 - 4 mm in position and 0.1^{o} in orientation and get worse as the distance or noise / interference signal level increases.

1.4. Objectives and Scope of the Paper

Creating VE models from laser range cameras is the main objective of this paper. Another objective is to explore the potential of photogrammetry for this emerging technology. The paper deals only with the visual part of VE. It does not get into other senses such as hearing and sensation of force.

We first describe two laser range cameras developed at the National Research Council of Canada which are useful for creating 3-D environment models. Then, two important issues are addressed; the calibration of the sensors in order to produce accurate data, and the registration of multiple views in order to arrive at one unique model of the environment and objects. The paper then describes the use of photogrammetry for head tracking followed by a description of the VE facility at our laboratory and finally some concluding remarks.

2. DESCRIPTION OF THE RANGE CAMERAS

Range cameras based on structured light can generate complete image data of visible surfaces that are rather featureless to the human eye or a video camera. Among the advantages of using a laser source one finds larger depth of fields, compared with what is achievable with incoherent light, and better ambient light rejection for accurate measurement.

2.1. The Autosynchronized Laser Scanner

In the basic geometrical principle of optical triangulation, the light beam generated by the laser is deflected by a mirror and scanned on the object. A camera, composed of a lens and a position sensitive photodetector, measures the location of the image of the illuminated point on the object. The X, Z coordinates of the illuminated point on the object are calculated by simple trigonometry . The error in the estimate of Z is inversely proportional to both the separation between the laser and the position detector and the effective focal length of the lens, but directly proportional to the square of the distance. Unfortunately, the separation cannot be made as large as desired. It is limited mainly by the mechanical structure of the optical setup and by shadow effects. A synchronized geometry provides a way to alleviate these tradeoffs. Rioux, 1984, introduced a synchronized scanning scheme, with which large fields of view with small triangulation angles can be obtained without sacrificing precision. With smaller triangulation angles, a reduction of shadow effects is inherently achieved. The intent is to synchronize the projection of the laser spot with its detection. As depicted in Figure 3, the instantaneous field of view of the position detector, defined by P and f_0, follows

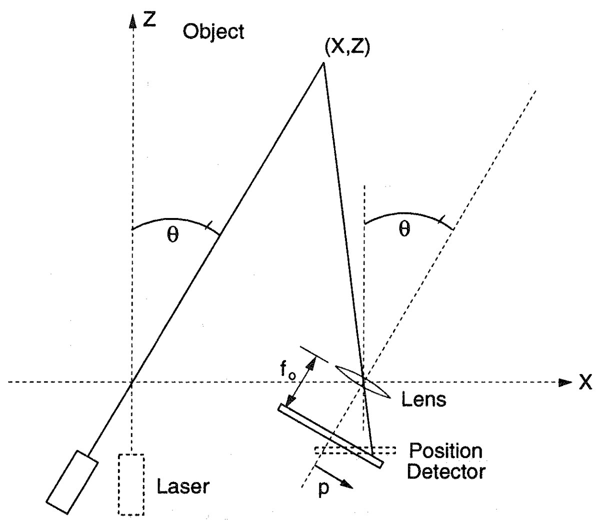


Figure 3: Synchronized scanner approach

the spot as it scans the scene. The focal length of the lens is therefore related to the desired depth of field or measurement range and not to the field of view. Implementation of this triangulation technique by an auto-synchronized scanner approach allows a considerable reduction in the optical head size compared to conventional triangulation methods. Figure 4 displays schematically the basic components of a dual-axis auto-synchronized camera. A 3-D surface map is obtained by (1) scanning a laser beam onto a scene with two oscillating mirrors mounted orthogonally from one another, (2) collecting the light that is scattered by the scene in synchronism with the projection mirrors, and (3) focusing this light onto a linear position-sensitive photo-detector. Beraldin et al., 1993, give the functions of 3-D coordinates computation for this implementation.

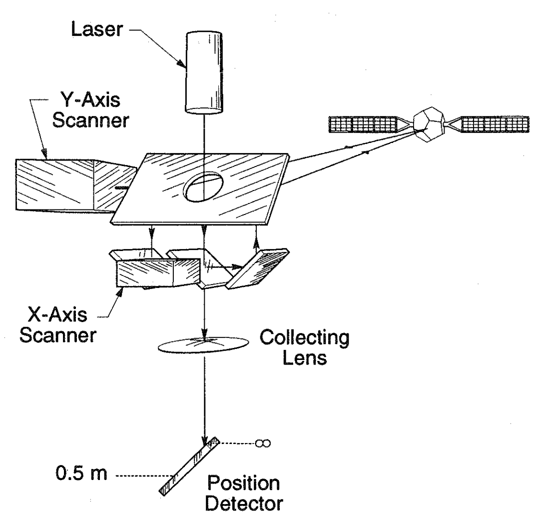


Figure 4: Auto-synchronized scanner approach: dual-axis synchronized scanner.

2.2. The BIRIS Camera

The BIRIS range camera was developed at NRC to work in difficult environments where reliability, robustness, and ease of maintenance are important. The optical principle of BIRIS is shown in Figure 5. The main components are a mask with two apertures, a camera lens, and a standard CCD camera. In a practical implementation, the double aperture mask replaces the iris of a standard camera lens (hence the name bi-iris). A laser line, produced by a solid state laser diode and a cylindrical lens, is projected on the object and a double image of the line is measured on the CCD camera. The separation between the two imaged lines is proportional to the distance between the object and the camera and provides direct information about the shape and dimensions of the object. For example, in Figure 5, the line separations b1 and b2 represent the ranges Z1 and Z2 respectively. Details of the mathematical model and the calibration can be found in Blais et al, 1992.

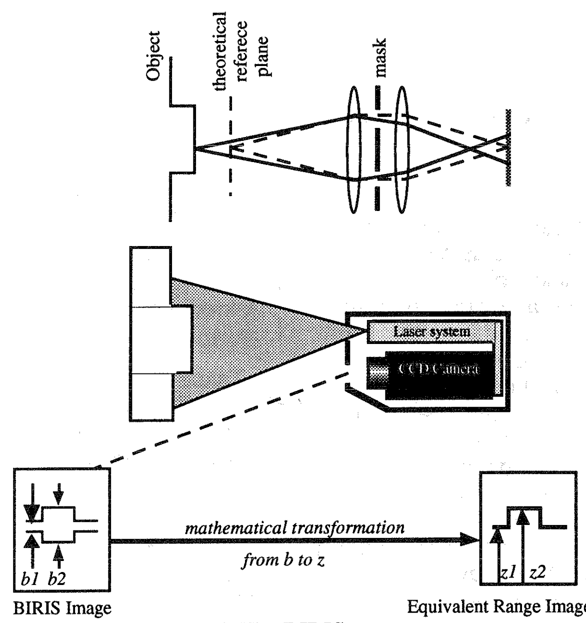


Figure 5: The BIRIS range camera

3. CALIBRATION

Rigorous calibration is needed for both internal and external camera parameters, and, when multi-sensors are used, for the registration of the data acquired by the various sensors. Specifically, it must recover:

- interior sensor parameters, including distortion parameters,
- position and orientation of all sensors (figure 6-b).

The calibration requires points of precisely known positions in the object space coordinate system. For the set up shown in figure 6-a, which is designed for environment modeling, a set of well-defined targets mounted at various heights on three orthogonal sides, is employed. The targets are centered on blocks with flat surfaces. The main requirements for this field of control point is the dimensional stability and the high accuracy of measuring the target locations after they have been built. This accuracy must be significantly higher than the expected accuracy of the system. Therefore, care must be exercised in measuring these targets with manual surveying equipment or close-range digital photogrammetry. The number and distribution of these targets must be designed to completely cover the expected viewing volume of all the sensors. The points must vary in height by at least 10% of the stand-off distance to cameras.

The calibration procedure is performed by placing the platform where the cameras are mounted in an initial position where all the targets are visible. All the sensors are to be adjusted at this stage since no further adjustment (such as refocusing) will be permitted once they are calibrated. The calibration is performed using the appropriate software (Beraldin et al, 1993, and El-Hakim et al, 1994 and 1995.)

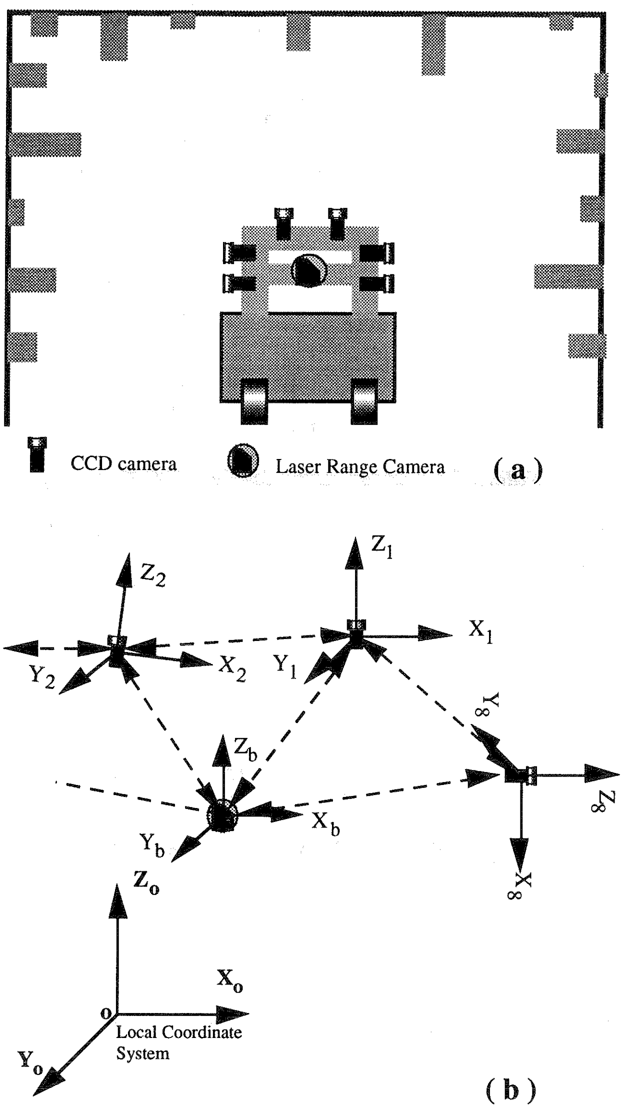


Figure 6: The calibration set up

4. MULTI-VIEW REGISTRATION

The construction of models of large environments usually requires a large number of views, due to the limited field of view of sensors as well as the complexity of the scene itself. Each view taken with a 3D sensor provides an ordered set of 3D points in a camera-centered Cartesian coordinate system. In order to reconstruct the model, the rigid transformation linking the different views to a common Cartesian coordinate system must be recovered. This process is known as registration. The position and orientation of the camera at each acquisition station could be accurately measured with extrinsic means (e.g. inertial systems, theodolite,...), but these methods are costly (in time or money) or impractical in some situations. We have experimented with methods which rely on the measurements in the range image in order to compute the rotation and translation of the camera with regard to a reference coordinate systems. Of course such an approach requires that there is some overlap between the portions of scene viewed in the different images, and that there are no residual degrees of freedom (e.g. if only one plane is shared between the views). One way to address the second issue is to also use, when available, the intensity values measured by the range sensor as a way to constrain the matching of surfaces between views (Godin et al, 1994). The intensity measurement provided by range sensors are a function of the distance, orientation and reflective properties of the surface element imaged in the sensor. Since the distance is directly measured by the range sensor, and the local surface orientation can be numerically estimated, it is possible, under certain conditions (Baribeau et al, 1991,) to recover an invariant reflectivity estimate for each point. A simpler technique involves the classification of image points based on differential methods (e.g. curvature classes) or projection of color space (e.g. computing hue). The labeling of points is then used to identify common portions of surfaces between images in order to compute registration.

The method proposed in (Godin et al, 1994) uses an iterative minimization of the least-squares distance between points in different images. When viewpoint-invariant properties (reflectance, curvature, etc.) are available, only points of similar property are paired in the minimization process. This is useful to accelerate convergence, especially in the first steps of the method, and to remove potential residual degrees of freedom when surface markings provide an unambiguous match. Depending on the context of the model building problem, the algorithm can be applied incrementally as each new image is scanned, or globally to minimize in parallel the transformation matrices for all images. The incremental approach may be useful for interactively controlling and validating the acquisition process, but the error accumulation may become a serious issue. It is therefore necessary to perform a global minimization to re-balance the errors over all the images of the scene.

Another method for the determination of rigid transformations between range images takes advantage, when available, of cooperative targets and photogrammetric techniques in order to find precisely the position of the targets seen in more than one image. This assumes that the geometric relationship between the video cameras and the range measurement system, as discussed in Section 3, has already been established. If the pairings of targets between stations are known, then the rigid transformation linking the different views is readily computed. Information from the range image can be used in two ways: in case of ambiguities in the pairing of targets imaged at different stations, the whole range image can serve a posteriori as a sanity check by computing the error between overlapping parts of the images to confirm the hypothesis; or a registration method can be applied first on the data in order to identify common targets. Extrinsic positioning information (e.g. mechanically from dead reckoning devices) can also be used to provide an approximate starting point for the registration.

Once the range data from different viewpoints are registered, the creation of a usable computer graphics model requires the additional steps of: removing redundancy between different views, compressing the model by representing it by geometric primitives at different levels, prepare the texture maps from the measured intensities. A methodology for the creation of triangulated, texture-mapped models from a set of range images is presented in (Soucy et al. 1996).

5. OPTICAL TRACKING

Popular tracking devices for VE, as mentioned above, have many disadvantages. The accuracy is limited and many systems require uncomfortable devices to be worn by the user. Therefore, an optical tracking system, based on photogrammetry, is proposed. Two digital cameras, each outfitted with light source and matching filter (figure 7), are used for the real-time measurement of the coordinates of retroreflective targets mounted on the head and / or the hand of the user. Since the user is moving while the images are being acquired, CCD cameras which allow full frame transfer (rather than the standard interlaced two fields separated by 16.6 ms), such as digital cameras, are recommended. Also, shuttering is required to decrease the integration period from frame time to only 1 ms or less, thus providing stop-action effect.

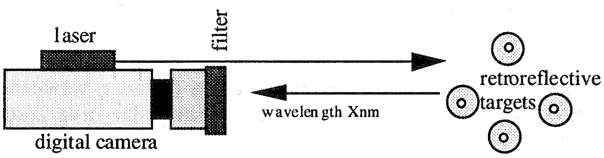


Figure 7: Camera-light-filter design for tracking

A fast image processing board is used for the real-time extraction of targets. The matching and computation of 3-D coordinates of the targets, using photogrammetric algorithms are described in El-Hakim and Pizzi, 1993.

6. VE FACILITY AT NRC

A facility dedicated to VE research and application has been established at the Institute for Information Technology at NRC. The goal is to develop a 3-D electronic visualization test bed that will integrate technologies in the field of VE, Real-time imaging, and 3-D range sensing in order to display and interact with a digital model of an environment in a realistic manner. The main objectives of the current project are:

1- To demonstrate, in a virtual environment, the realism of reconstruction produced from the digital model generated;

2- to help in the development of new digital modeling scheme to improve the realism of the digital model produced;

3- to experiment and develop various devices to interact with the virtual environment;

4- to assess the usefulness of such systems in realistic applications; and

5- to acquire an expertise in the field of VE and 3-D interactive graphic systems.

The facility, which is a 10 m x 6 m x 3 m (h) room, includes the following equipment:

- One high speed projector, model Marquee 9500 from Electrohome;
- One 2.25 m by 3 m rear projection screen;
- Liquid crystal glasses and a controller for large rooms;
- Two digital video camera, one PC-Pentium and two Matrox frame grabbers;
- SGI graphic workstation- Infinite Reality; and
- Two electromagnetic trackers.

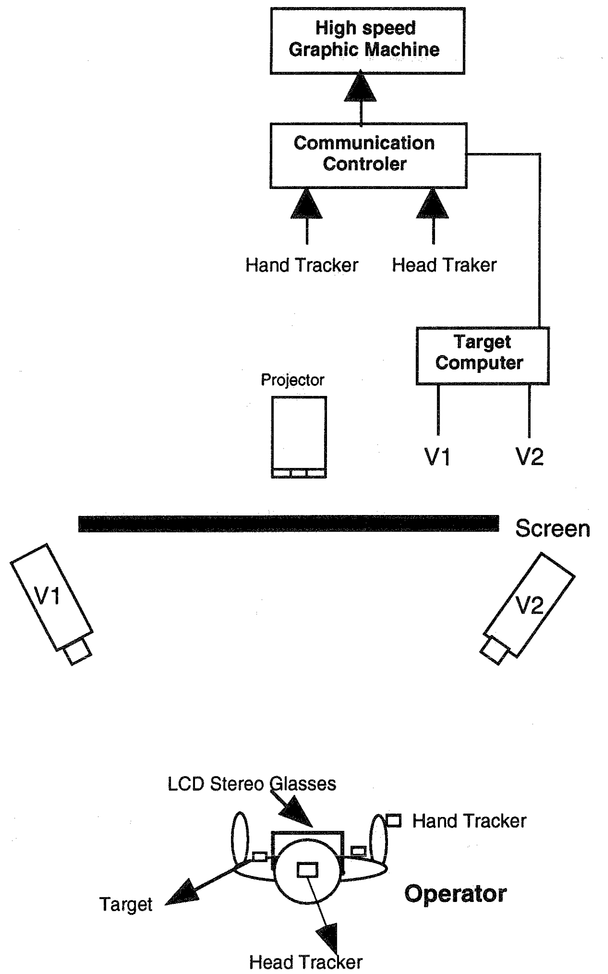


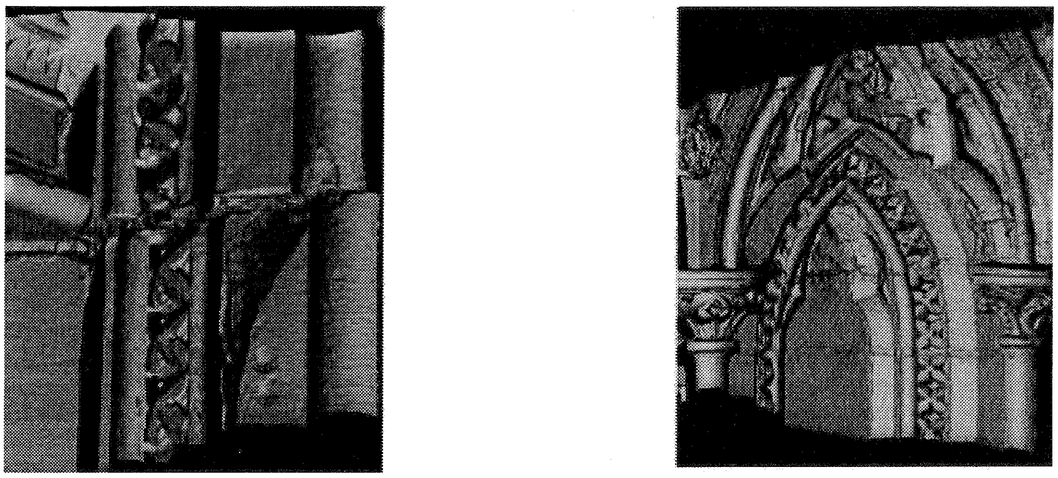
Figure 8: The VE facility at NRC (top view)

Figure 8 shows a top view of the set up, named ViEW (Virtual Environment Wall).

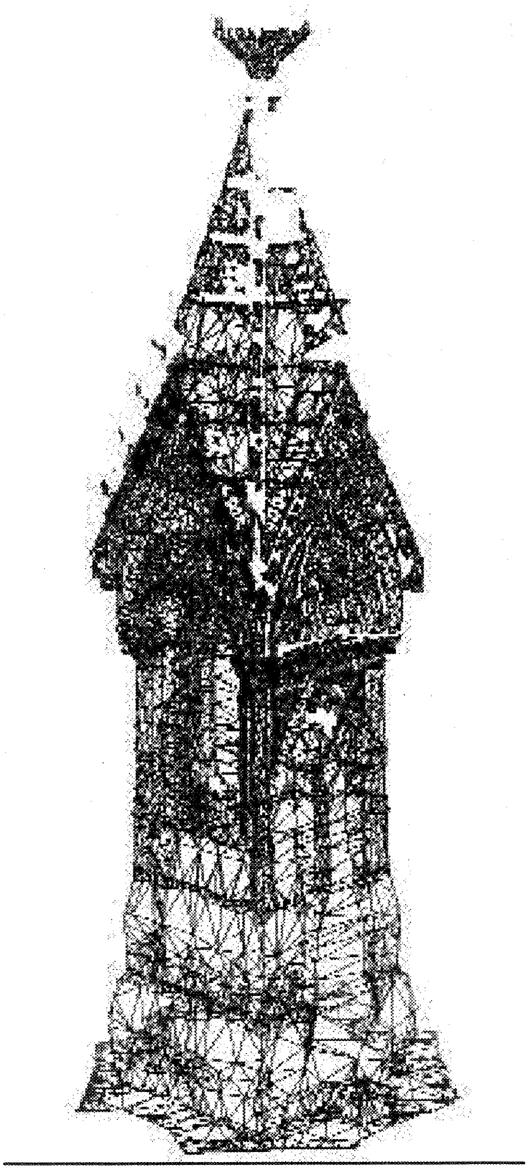
Figure 9 shows an example of model created by over 50 views taken by the autosynchronized laser range camera and displayed by ViEW. Figure 9-a displays two of the scanned views while the registered images forming the complete geometric model, represented by a triangulated mesh, is shown in figure 9-b. The model with added shading is shown in figure 9-c.

7. CONCLUDING REMARKS

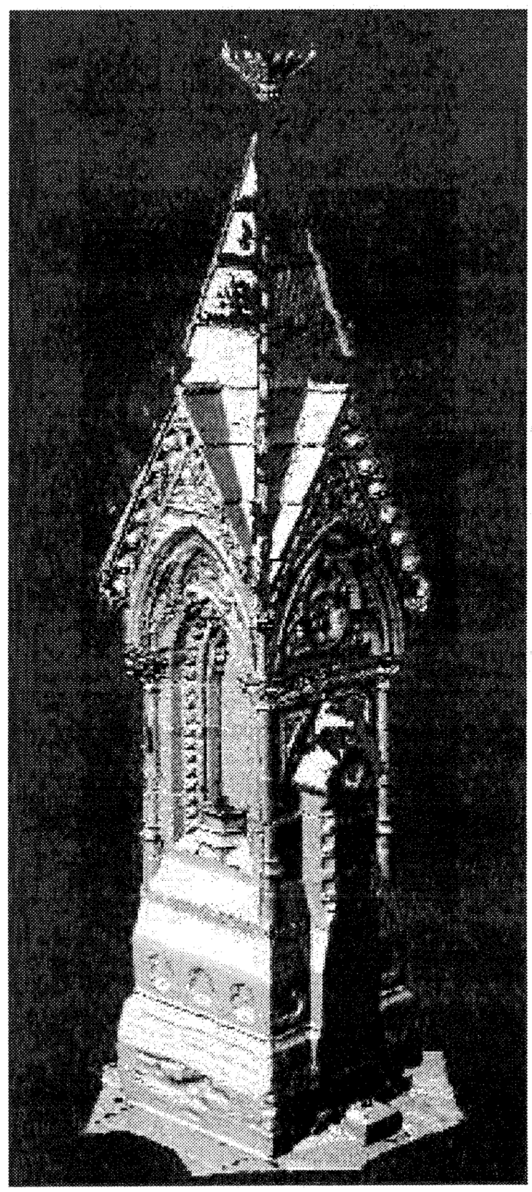
• A virtual environment, where people are immersed, navigate, and interact, can provide ideal solutions for many applications such as; training, design for architecture and industrial products, medical such as planning for surgery, and leisure and entertainment.



(a) Two single 3-D images (full resolution)



(b) The geometric model (polygons only)



(c) The complete model with added shading

Figure 9: Constructed model of historic building

• Although many of the applications are under development, only some have advanced beyond proof-of-concept or prototype stages. In fact many serious practical problems do exist, two of which we offered solutions to in this paper. These are:

1- The creation of realistic VE models.

2- The accurate non-intrusive tracking of the user's head and hands.

• Advanced laser range cameras, such as the two presented in this paper, can be fast and accurate tools for the creation of VE models. However, they must be properly calibrated and the acquired images must be perfectly registered. We have briefly presented the calibration approach and the registration technique developed in our laboratory. These methods have been extensively tested and applied to a variety of environments and objects and gave excellent results. The calibration facility and the software tools for registration are currently being used for several collaborative projects with industrial partners, such as virtual museums and rapid prototyping.

• The use of video images (gray-scale or color), which are registered with the 3-D images, is valuable in adding texture maps to the geometric models to provide visually realistic environments. Feature extraction from the video images, along with their 3-D coordinates computed with photogrammetric techniques, provide useful information to assist in the registration of multi-views, particularly in large and complex environments.

• The accurate tracking of user's head and hands can be accomplished optically by real-time digital photogrammetry. Our specially designed photogrammetric tracking system can currently update the targets positions and orientations at a rate of 15-20 Hz using Pentium PC and image processing hardware. This speed is sufficient for only some applications where the user is not moving too fast. For many applications, however, a much faster speed (about 3 to 6 times faster) is required so that the user will not experience problems related to latency. We are currently working on increasing the tracking speed using prediction methods, such as Kalman filter, to find the position between system outputs. Also the ever increasing speed of PCs will contribute to ultimately achieving the desired tracking rate at a reasonable cost. On the positive side, the accuracy achieved by our optical tracking system is an order of magnitude better than any existing tracking device currently used for VE. The system also does not require uncomfortable head or hand gear to be put on by the user.

REFERENCES

Adams, J.A., 1993. Virtual reality is for real. IEEE Spectrum, 30(10) October, pp. 22-29.

Azuma, R., 1995. A survey of augmented reality. Course Notes 9: Developing Advanced Virtual Reality, SIGGRAPH'95, ACM SIGGRAPH, N.Y., N.Y.

Azuma, R. and Bishop, G., 1995. A frequency domain analysis of head-motion prediction. In: SIGGRAPH'95, L.A., Cal., Aug. 6-11, ACM Computer Graphics Proceedings 95, pp. 401-408.

Baribeau, R., Rioux,M., and Godin,G., 1991. Color reflectance modeling using a polychromatic laser range sensor. IEEE Trans. Pattern Anal. Machine Intell., 14(2), pp. 263-269.

Beraldin,J.-A., El-Hakim,S.F., and Cournoyer,L., 1993. Practical range camera calibration. In: Videometrics II, Proc. SPIE 2067, pp. 21-31.

Bishop, G. and Fuchs, H., 1992. Research directions in virtual environments. Computer Graphics, 26(3) August, pp. 153-177.

Blais, F., Lecavalier,M., Domey,J., Boulanger,P., and Courteau,J., 1992. Application of the BIRIS range sensor for wood volume measurement. NRC/ERB-1038, 19 pages.

El-Hakim,S.F. and Beraldin,J.-A., 1994. On the integration of range and intensity data to improve vision-based three-dimensional measurements. In: Videometrics III, Proc. SPIE 2350, pp. 306-321.

El-Hakim,S.F., Beraldin,J.-A., and Blais,F., 1995. A comparative evaluation of the performance of passive and active 3-D vision systems. In: St. Petersburg Great Lakes Conference on Digital Photogrammetry and Remote Sensing, St. Petersburg, Russia, June 25-30, Proc. SPIE 2646, pp. 14-25.

El-Hakim, S.F. and Pizzi,N., 1993. Multicamera vision-based approach to flexible feature measurement for inspection and reverse engineering. Optical Engineering, 32(9), pp. 2201-2215.

Goble, J.C., Hinckley,K., Pausch,R., Snell,J.W., and Kassel,N.F., 1995. Two-handed spatial interface tools for neurosurgical planning. IEEE Computer, 28(7) July, pp. 20-26.

Godin, G., Rioux, M., Baribeau, R., 1994. 3-D registration using range and intensity information. In:Videometrics III, Proc. SPIE 2350, pp. 279-290.

Haeberli,P. and Segal,M., 1993. Texture mapping as a fundamental drawing primitive. In: Proc. of the Fourth Eurographics Workshop on Rendering, June, Paris, France.

Hedberg, S.R., 1996. Virtual reality at Boeing: Pushing the envelope. Virtual Reality Special Report, 3(1), Jan,-Feb., pp. 51-55.

Phillips-Mahoney, D., 1995. Modeling for Virtual Reality. Computer Graphics World, 18(10) October, pp. 45-50.

Polis,M.F., Gifford,S.J., and McKeown Jr.,D.M., 1995. Automating the construction of large-scale virtual worlds. IEEE Computer, 28(7) July, pp. 57-65.

Rioux,M., 1984. Laser range finder based on synchronized scanners. Applied Optics, 23, pp. 3837-3844.

Soucy, M., Godin, G., Baribeau, R., Blais, F., Rioux, M., 1996. Sensors and algorithms for the construction of digital 3-D colour models of real objects. To appear in: Proc. IEEE Int Conf on Image Processing, Lausanne, Switzerland, September 16-19.

Sturman,D.J. and Zelter,D., 1994. A survey of glove-based input. IEEE Computer Graphics & Applications, 14(1) January, pp. 30-39.

Stevens, L., 1994. Virtual Reality Now. MIS Press, N.Y., N.Y., USA.