Collaborative Visualization and Interaction for Detailed Environment Models

Borgeat, Louis; Godin, Guy; Lapointe, Jean-François; Massicotte, Philippe

Publisher’s version / Version de l'éditeur:

The 10th International Conference on Virtual Systems and Multimedia [Proceedings], 2004

National Research Council Canada
Institute for Information Technology

Conseil national de recherches Canada
Institut de technologie de l'information

NRC · CNRC

Collaborative Visualization and Interaction for Detailed Environment Models *

Borgeat, L., Godin, G., Lapointe, J.F., and Massicotte, P.
November 2004

* published at The 10th International Conference on Virtual Systems and Multimedia. Softopia Japan, Ogaki City, Japan. November 17-19, 2004. pp. 1204-1213. NRC 47400.

Canada

Collaborative Visualization and Interaction for Detailed Environment Models

Louis BORGEAT, Guy GODIN, Jean-François LAPOINTE, Philippe MASSICOTTE
National Research Council of Canada
Ottawa, Ontario, Canada
<*firstname.lastname*>@nrc-cnrc.gc.ca

Abstract. As the tools for the construction of detailed models of environments evolve, the need for visualization and interaction in a shared context is emerging. This paper presents DIMENSION, a new adaptive technological platform that allows several distant groups of users to dynamically interact, in a multimedia fashion, around detailed digital 3D models of objects and environments. It also describes a set of technologies based on this infrastructure and specifically applied to the collaborative exploration of models of cultural heritage sites: multi-resolution representation and display of scene geometry, video communication that exploits graphics hardware, foveated stereoscopic display, and laser-based interaction with a large screen display.

1. Introduction

Recent years have seen a rapid increase in the number of objects and sites that have been digitally modeled, using an expanding palette of tools. A driving force for this trend is the increased availability of optical 2D/3D sensing techniques, which facilitate the construction of highly detailed models of existing objects and sites. Such models are often used for visualization and analysis. But the emergence of computer-supported collaborative work technologies open new types of applications that combine shared 3D model viewing, audio-visual communications and interaction techniques such as distributed model construction and review, guided visits and virtual tourism, and scholarly examination of artefacts and sites by individuals or groups at different locations linked by a network infrastructure. In practice, the deployment of such tools will have to accommodate a range and asymmetry of needs, workstation powers and bandwidth, support different visual output devices (desktop screens, immersive stereo displays, head-mounted displays, etc.), and provide new tools enabling several users to efficiently interact with a complex model.

This paper presents a set of technologies developed towards these goals. We first describe the infrastructure on which the system is built. We then outline the efficient multi-resolution representation and rendering method that is required for the interactive display of large geometric models. In the virtual environment, participants control their local view of the shared 3D model, with avatars representing the position of other viewers. It is also necessary to provide audio and video linkage between the participants at remote locations: we describe our implementation of an efficient technique for video transcoding that takes advantage of the power and programmability of modern graphics cards; we also propose a method for feeding the images rendered on one machine to another one through the video stream. The flexibility of the software infrastructure allows configuring the visualization subsystem as a foveated stereoscopic display: this mode provides a resource-efficient tool for enabling enhanced appreciation of the details of the model in a wall display. A laser pointer technique for natural interaction with a large screen is also integrated into the environment. The proposed infrastructure supports modification of the 3D model: a tool for annotating models during multi-user sessions is shown as an example. We conclude with a discussion of future developments and applications.

2. DIMENSION: a 3D Collaborative Environment

The DIMENSION (DIstributed Multimedia ENvironment for Synchronous Interaction Over a Network) project at the National Research Council of Canada aims at creating and developing an advanced 3D collaborative visualization environment for the sharing, exploration and dynamic transformation of detailed models of 3D worlds. The DIMENSION project integrates research work in the fields of telecollaboration, 3D sensing and modeling, advanced visualization techniques, distributed computing and multi-modal interfaces.

At the heart of DIMENSION lies a generic telecollaboration infrastructure designed to be easily adapted to different application domains: even if DIMENSION has been designed with 3D applications in mind, it is not limited to the sharing of 3D information, and can thus seamlessly integrate various types of information streams that are shared between nodes participating in a collaborative session. In the context presented here, complex multi-resolution models built from large 3D datasets are shared and displayed in a resource-adaptive fashion. This represents an advantage over existing Collaborative Virtual Environments (CVE) such as [11] that are based on VRML (Virtual Reality Modeling Language) or its successor X3D (Extensible 3D), in which sharing large and evolving 3D datasets is limited by the simple level-of-detail schemes available and where extension outside of the 3D context requires the use of exception mechanisms. DIMENSION includes a H.323 layer to allow audio and video streams to be shared between nodes, providing the usual teleconference-style linkage between participants, but also a channel to feed lightweight clients with a video stream showing renderings of the scene, instead of downloading complete models. This solution can provide interoperability with existing H.323 telecollaborative environments such as NetMeeting or GnomeMeeting.

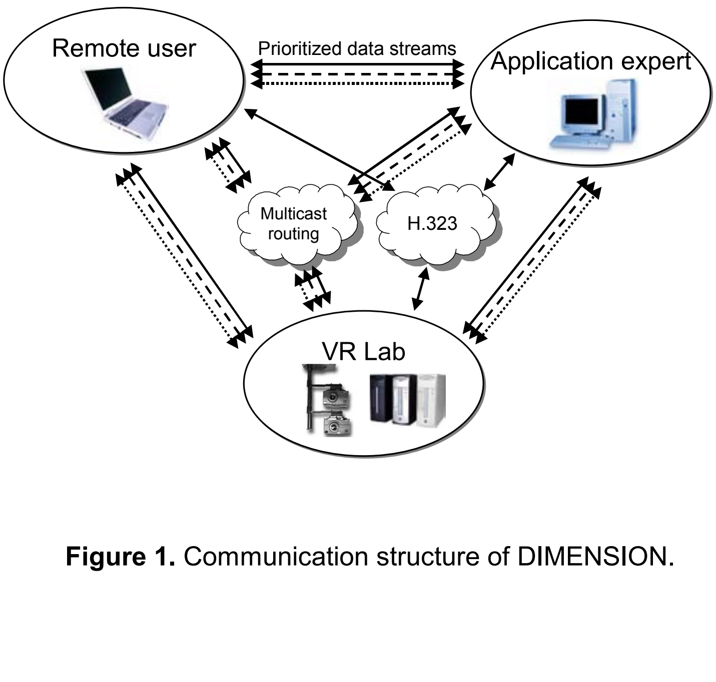


Figure 1. Communication structure of DIMENSION.

Developing synchronous collaborative software is an inherently complex task that involves managing real-time constraints, resource limitations and complex synchronization problems. Within DIMENSION, these issues are addressed by combining a generic communication framework with advanced code generation techniques in order to rapidly produce collaborative applications that can be easily tailored to specific contexts. First, a basic collaborative application skeleton is automatically generated from one or many XML Document Type Definitions (DTD) corresponding to the description of the dataset to be shared. The resulting C++ code provides optimized differential encoding of changes in a dataset for efficient synchronization between nodes. The framework allows for easy overriding of all default functionalities in the generated code to facilitate the implementation of application-specific optimizations. The generated objects can then be synchronized between hosts through a high performance communication framework that includes several advanced features required by real-time collaborative applications: multi-threading, asynchronous streaming, priority and interest management, reliable multicasting, and session management (late joining and early leaving). The DIMENSION 3D environment is based on the X3D DTD, and on a set of application-specific DTDs defining extensions: to support efficient collaborative drawing on the 3D scene, to integrate the multi-resolution rendering algorithms (Section 3) in the collaborative environment, and to manage the configuration of the physical display and VR set-up.

The DIMENSION 3D application also provides great flexibility when it comes to choosing the physical stations on which each participant accesses the collaboration in a given application context. For example, the required hardware at different nodes can scale from a simple portable computer to a cluster of PCs for increased computational power (Figure 1). The same platform enables various display configurations such as tiled wall displays, dual-head passive stereo systems, CAVE-type multiple wall environments and high-resolution foveated displays (Section 5). The infrastructure can also accommodate a variety of human-computer interfaces, in the form of physical or virtual devices. The overall management of application configuration is achieved using a shareable dataset built on the generic framework.

3. Modeling and Display

The construction of 3D models from optical sensors has become an important tool in numerous fields of application, perhaps most notably for the documentation of cultural heritage objects and sites [8]. The size and complexity of these models continue to increase, driven by the steady improvement in performance of the 3D sensors and digital cameras, of the algorithms for transforming their data into models, and of graphical computers. In this paper, we will use as an example a detailed 3D textured model representing the Crypt of Santa Cristina in Carpignano Salentino, Italy; it was built using a combination of laser range sensing and digital photography (see [4] for more details).

Such models often exceed the memory capacity of most current graphics card (especially the texture images). In the context of model exploration and analysis, there is a need for an adaptive display technique that maximizes the available amount of visible details with an optimal use of the rendering power, while maintaining smooth motion in navigation. This suggests the use of a multi-resolution representation, where the model is rendered at a resolution controlled by the triangles' on-screen size and the desired frame rate.

Towards this goal, we have developed a multi-resolution representation and display method [5] that is a hybrid technique integrating aspects of edge contraction progressive meshes and of classical discrete Level of Detail (LOD) techniques. It aims at efficient multi-resolution rendering with minimal visual artefacts when displaying high-resolution scenes or object datasets derived from sensor data. The method involves complex preprocessing of the data but requires only minimal real-time resources at rendering time.

The first step of the preprocessing is to decimate the model into a series of discrete LODs using an algorithm based on vertex pair contraction. The low resolution LOD is then decomposed into a set of triangle groups. Each level is then recursively partitioned along the same borders as its lower resolution counterpart, and each group of the new level is subpartitioned to achieve the desired granularity. At the end of the process, we obtain a hierarchy of group subdivision spanning the whole sequence of LODs. Groups are shaped based on criteria such as compactness, common orientation, texture/viewpoint association, and desired granularity (number of primitives per group). Groups are individually converted into vertex-ordered triangle strips in order to maximize rendering speed. At run-time, LOD levels, group visibility, and geomorphing ratios between selected groups are computed for each frame. Border points between groups are geomorphed in order to maintain seamless continuity between neighbouring LOD groups at all time. All the geomorphing is actually performed by a vertex program on the graphics processing unit (GPU), and therefore most CPU resources are left available for other tasks such as video transcoding and application-specific computations. Morphing is applied to space and texture coordinates, normals and colors. The group structure allows for a more progressive transition than other geomorphing schemes. This approach therefore produces almost unnoticeable transformations during LOD transitions and gives especially good results with textured data such as the model shown here. Due to the additional real-time multi-resolution processing, the graphics performance is about half of what can be achieved by rendering an optimized static model on a typical gaming graphics card. But it should be noted that this penalty is very small compared to the overall gain obtained by the technique, which reduces the number of rendered primitives by orders of magnitude, especially when closely examining details of the model. Figure 2 shows the displayed geometry of the same model for different positions of the observer and the corresponding increase in levels of details as the viewer approaches the wall. The 3D display layer of DIMENSION is based on OpenGL Performer under Linux, and uses off-the-shelf PCs and graphics cards.

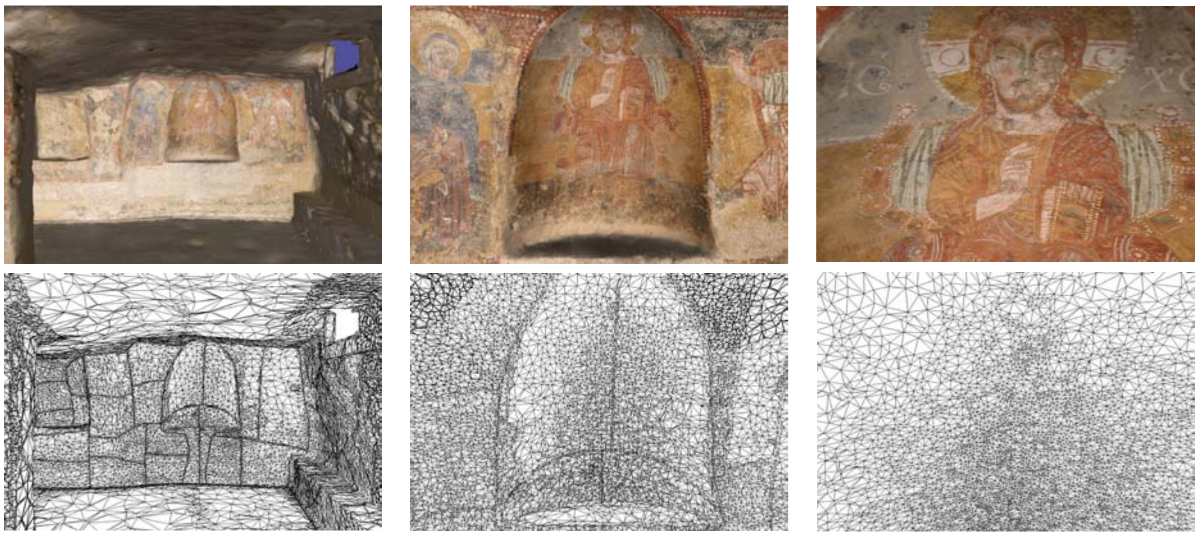


Figure 2. Multi-resolution display. Three different levels of details are shown as the viewer approaches the wall. The total drawing speed per frame remains almost constant.

4. Audio/Video Link

Users at different locations must be able to directly interact and discuss when exploring the model. For this purpose, DIMENSION integrates an audio and video infrastructure based on the H.323 protocol, which allows interoperability with existing desktop teleconference tools such as NetMeeting and GnomeMeeting. The audio stream is transcoded using one of the following codecs: G.711-PCM(64kbps), GSM-06.10(16.5kbps), or LPC10(3.46kbps). Different codecs can be used to compress the video data, such as H.261 (supported by NetMeeting and GnomeMeeting) or MPEG-4, the later having a much higher quality/size ratio. Depending on the chosen codec, the video resolutions can be: QSIF 160x120 /QCIF 176x144, SIF 320x240 / CIF 352x288, or 4SIF 640x480. The video communication is integrated into the 3D display environment: the images received on the stream can be mapped onto arbitrary 3D shapes, by texture mapping, or into a separate rectangular window. In typical usage, video windows are attached to avatars representing each of the participants to the session (Figure 6).

One particularity of our implementation is to transfer the burden of the last step of video decoding and display onto the graphics hardware. The conversion from YUV to RGB is performed as the video frame is rendered in the 3D world by texturing a surface: the fragments are fetched directly from the YUV image produced by the software decoder, and the linear color space transformation is applied before they are drawn in RGB on the screen. This fragment program is implemented in the Cg language. One additional performance advantage of this approach is that the conversion is applied only to pixels (fragments) as they are drawn on the screen, and not systematically to the entire video image.

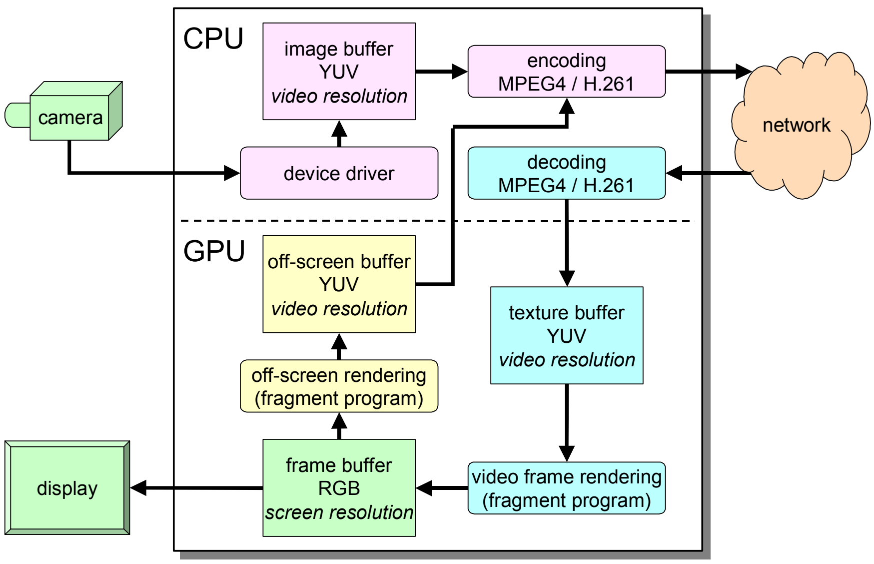


Figure 3. Video coding and decoding data paths, using GPU for conversion and resizing. Some intermediate storage is omitted for clarity.

The flexibility and power of current graphics hardware allows the implementation of another useful feature in our collaborative context: the images rendered on one station's 3D graphics pipeline can be streamed over a video channel to another one which does not have access to a copy of the 3D model of the environment, or the graphical capability to handle it. Through the use of standard protocols, any compatible usual videoconference client can be used instead of a DIMENSION node, albeit without the same gamut of tools. This concept is a simplified version of the functionality provided by SGI's OpenGL VizServer [1]. As is the case for the decoding process, our implementation of this encoding also uses OpenGL functions and Cg fragment programming to transform the frame buffer image rendered by the DIMENSION 3D graphics module into a YUV image, rescaled at the resolution requested by the video channel. The resulting image is then transferred from the graphics card to the CPU memory, where it is encoded and transmitted through the same operations as for the signal originating from a video camera. Figure 3 illustrates the data paths for encoding and decoding video from either a camera or the scene rendered in the frame buffer.

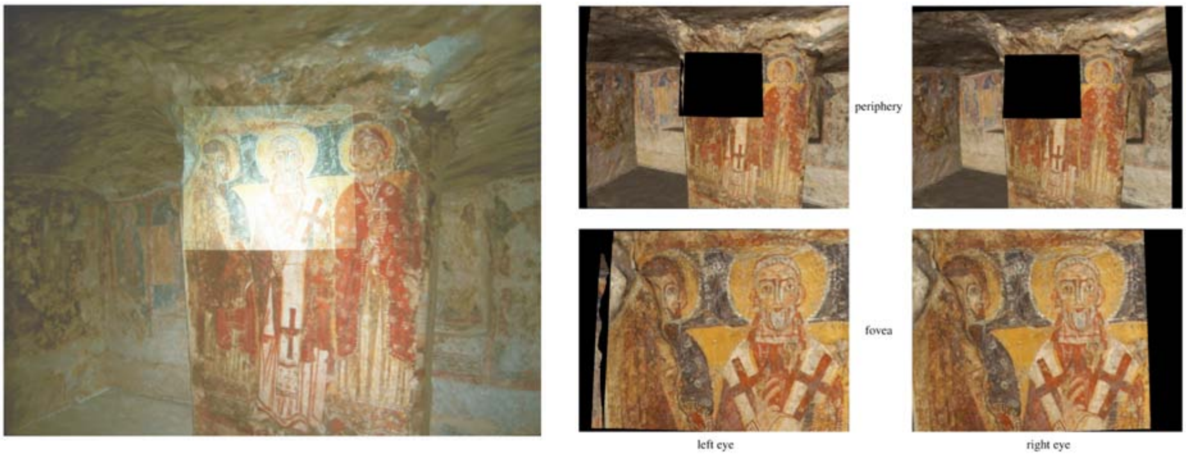
5. Foveated Stereoscopic Display

Large stereoscopic displays are the preferred visual output device for visualizing 3D models of sites. For this purpose, we employ a dual-head passive polarization system, where each eye's view is rendered on one computer, and displayed using a digital micro-mirror device projector fitted with a circular polarizing filter, on a rear-projection polarization-preserving screen. Refresh synchronization between left and right images is provided by the DIMENSION architecture, with communications over a Dolphin SCI low-latency interconnect.

In such a wall display, the size of screen pixels is typically several millimetres: this often becomes the limiting factor in the appreciation of the details in the model. Increasing the actual visual resolution of a display can be achieved by combining several units into a seamless unified display. Significant work has been accomplished in tiled displays built from off-the-shelf devices: a number of such projects are described in [7]. Recently, similar strategies have been extended to tiled stereoscopic displays [6][12]. Another approach to increased apparent resolution adds small high-resolution insets within a larger, lower-resolution field of view; the inset image is aligned and synchronized with the larger image to provide a unified display area The wide availability of digital projectors has considerably simplified the design and implementation of such dual resolution (or foveated) displays by removing the need for complex optical designs. For example, in the *focus+context* system [3], a high-resolution LCD panel display is surrounded by a very large screen on which a lower resolution image is projected; a pair of projectors is used in the *Escritoire* project [2] to create a desktop that incorporates a high-resolution area for improved document viewing.

Recently, we introduced a *stereoscopic* version of projector-based foveated displays [9][10]. In extending this dual-resolution approach to stereo, we identified a specific issue not present in monoscopic foveated displays: the visible boundary between the high-resolution inset and the low-resolution periphery creates a stereoscopic depth cue with a disparity which, in general, does not match that of the underlying scene. This creates a competition between two perceived layers of depth. We proposed an efficient solution for resolving this conflict by virtually displacing the boundaries within the footprint of the fovea. Resolving the stereo conflicts is dependent on the scene and viewpoint, and must therefore be performed at every frame. Our method requires only a single rendering pass, thus maximizing the level of details available for a given frame rate. It corrects for casually aligned projectors, in a manner similar to [14] but with a form of the warping matrix that optimizes usage of the depth buffer, and also applies the boundary corrections as a post-draw process at a low additional computational cost (see [10] for details).

Figure 4 shows a screen photograph of the foveated stereoscopic wall display. The brighter inset area contains the same number of pixels as the peripheral image (1024 x 768). The total screen size is 3 x 2.3 m: thus, for a viewer standing at 1.5 m, the 3 mm pixels subtend an angle of about 6.8 arcmin along the perpendicular viewing direction. The inset projection area is approximately 1 m wide, yielding a resolution of 2.3 arcmin that, while still not matching the limit of human visual acuity, significantly enhances the perception of details. Our current projector arrangement produces a fovea image of about one third of the display width, thus covering one ninth of the area. This value represents a compromise: obviously, reducing the screen size of the fovea would further increase the angular resolution; on the other hand, a smaller fraction of the scene would be available at once in the fovea for examination, and more navigation in the model would be required to position areas of interest in the high-resolution fovea. Conversely, a larger fovea reduces the relative gain in resolution, and increases the proportion of unused low-resolution image pixels.



(a) (b)

Figure 4. Foveated stereoscopic display. (a) photograph of the screen (only one eye's view is shown for clarity); (b) the four original images composing (a).

For the exploration and study of detailed models, a group of users gather in front of the screen, and can approach the fovea for details examination in stereo; the periphery provides a context (in the sense of [3]) over the entire model, which is particularly helpful for orientation and navigation. The periphery screen area can also be used as the locus for insertion of video avatars or other collaborative tools: it should be noted that in the DIMENSION framework, the entire screen is available as a unified display for all its components. As a direct consequence of the scene-adaptive positioning of the fovea boundary, the inset region is not perceived as a window floating over the scene, but rather as a surface-conforming patch of increased resolution and brightness reminiscent, in absence of occlusions, of a spotlight directed onto the model.

The foveated system is implemented through the same DIMENSION infrastructure as the basic stereoscopic system, using a local cluster of four linked computers instead of two. Each node renders either periphery or fovea of the left or right eye: per-frame communication between nodes is only required for events and refresh synchronization. Load balancing is achieved by adjusting the levels of resolution of the model, using the multi-resolution scheme described in Section 3, in order to maintain a target frame rate. With projectors, dual resolution displays are significantly simpler and cheaper to realize than tiled displays; obviously, the drawback is that the gain in resolution is limited to a subset of the display surface. However, this configuration presents important advantages: the central resolution provided by our set-up is comparable to that provided by a 3x3 tiled display. Yet, it requires only four PC+projector groups, instead of the 18 that would be required by a 3x3 tiled stereo set-up. Not only is there a reduction in equipment cost, but also more importantly, the reduced power consumption and heat dissipation simplifies the deployment of such systems in typical office environments without special needs for electrical supply and air conditioning.

6. Laser-based Interaction

Since the users collaborating from different physical locations can be alone or in group, it is necessary to allow them to interact in a collaborative manner on the same virtual environment through the use of wall-sized displays. Another design consideration is to provide them with the possibility of freely interacting with the screen, without the need for horizontal flat surfaces as is usually the case with computer mice. The sought interaction method should be easy to use and should not restrict user movements of the users in front of the display. Finally, it should be able to provide selection of objects on the screen as well as freehand drawing to highlight and/or annotate the virtual environment.

In order to achieve these goals, a laser pointing interface based on previous work [13] was modified to operate in the background, through the use of a configuration compatible with the wall-sized rear projection display used for DIMENSION. This new configuration improves the usability of the system, by allowing the users to move anywhere in front of the screen without interfering with the screen display nor the camera view necessary to implement the laser pointing interface. In order to allow the selection of icons and objects on the display, a wireless button has been coupled with the laser pointer in order to enhance its functionality while still preserving the freedom of movement and avoiding the need for a flat surface.

The operating principle of the laser pointing interface is illustrated in Figure 5. It is composed essentially of a high-speed monochrome camera linked to a computer that detects the laser spot on the screen and computes the corresponding screen pixel coordinates through the use of the planar homography-based auto-calibrated technique described in [13]. Those coordinates are then sent to the projection system through a socket-based communication protocol, before being used to display a corresponding on-screen action.

Although the laser pointing interface works both with a visible (red) or an invisible (infrared) laser pointer for monoscopic projection, the visible laser is not suitable for the stereo projection mode, since in the general case the spot will appear at the depth of the screen, which will differ in general from the apparent depth of the underlying scene (or be perceived as a double spot by users). We therefore select an infrared laser pointer when using the stereo projection mode.

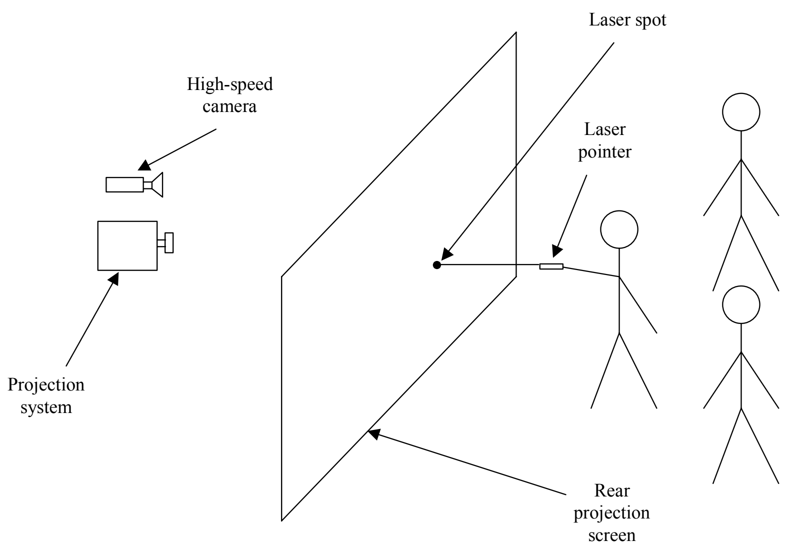


Figure 5. Laser-based interaction with wall screen.

7. Discussion

DIMENSION is currently being developed as part of an applied project in the field of virtualized cultural heritage. As a first application demonstration, we implemented a shared 3D environment where participants can upload and update models and avatars, add 3D annotations directly on the models, and guide or be guided by other participants with audio/visual linkage. The application uses the multiple data channels available within DIMENSION to insure that the transmission of large models does not affect the synchronization of real-time data such as avatars and light source positions. Figure 6 shows a video connection of a remote user, as described in Section 4, as well as the model annotation capability through the use of a 3D drawing system that writes directly on the 3D surface located under the cursor, controlled by the mouse or the laser pointer. This surface-based technique allows preserving the semantics of the annotation independently from the user's viewpoint. Furthermore, as an example of the customization features provided within DIMENSION, only the control points associated with the drawing are propagated and consistently maintained between the nodes, thus significantly reducing the information traffic. Experimental results show that the interactive 3D display system is not slowed down by the communication process even in the case of high latencies between hosts, and that the proposed video optimization significantly reduces the bandwidth and CPU requirements of video processing, an important gain when multiple feeds are being managed concurrently.

Figure 6. Snapshot of a DIMENSION session, showing a remote user with his avatar and video window, as well as scene annotations drawn by the local and remote users.

8. Conclusions

The DIMENSION system presented in this paper is a generic distributed multimedia telecollaboration system that integrates new visualization and interaction modalities; it allows easy integration of high-resolution 3D models with other standard data such as audio and video, and extension of the shared space with application-specific datasets and encoding schemes. The different components have been presented along with an example of their integrated use in the field of 3D modeling and visualization of virtualized cultural heritage. This research field is still in its infancy and several aspects require further investigation before the potential of this new kind of shared multimedia environments is fully exploited. Future work will involve new interaction modalities and technological improvements, in order to increase the efficiency of the system, in terms of computing resource and of user performance. This will require an assessment of its usability with actual users of the technology drawn from the application field, leading to further improvements of the platform to better suit their needs. Ultimately, a fully collaborative modelling and visualization system is envisioned to allow users to both build and exploit detailed models of real world environments in a natural manner.

Acknowledgments

We would like to thank our colleagues at the Visual Information Technology Group of the National Research Council of Canada, in particular J.-Angelo Beraldin, Michel Picard, Luc Cournoyer, Sabry El-Hakim and François Blais. We also thank Prof. Virginia Valzano, of the University of Lecce (Italy), for kindly granting us the permission to use images of the Crypt of Santa Cristina.

References

[1] "OpenGL VizServer™", http://www.sgi.com/software/vizserver/ , (as of October 2004).

[2] Ashdown, M., Robinson, P., "The Escritoire: a personal projected display", Proc. 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, 3-7 February 2003, Pilsen, Czech Republic, pp. 33-40.

[3] Baudisch P., Good, N., Stewart, P., "Focus plus context screens: combining display technology with visualization techniques", Proc. of the 14th Annual ACM Symposium on User Interface Software and Technology, 11-14 November 2001, Orlando, USA, pp. 31-40.

[4] Beraldin J.-A. *et al*., "Virtualizing a Byzantine crypt by combining high-resolution textures with laser scanner 3D data", Proc. Eight International Conference on Virtual Systems and Multimedia, 25-27 September 2002, Gyeongju, Korea, pp. 3-14.

[5] Borgeat, L., Fortin, P.-A., Godin, G., "A fast hybrid geomorphing LOD scheme", SIGGRAPH 2003 Sketches and Applications, 27-31 July 2003, San Diego, CA (electronic proceedings).

[6] Bresnahan, G. *et al*., "Building a large scale, high-resolution, tiled, rear projected, passive stereo display system based on commodity components", *Stereoscopic Displays and Virtual Reality Systems X,* SPIE Proc. Vol. 5006 (2003), pp. 19-30.

[7] Funkhouser, T., Li K., "Special issue: Large format displays". *IEEE Computer Graphics and Applications,* vol. 20, no. 4, July/August 2000.

[8] Godin, G. *et al*., "Active optical imaging for heritage applications", *IEEE Computer Graphics and Applications*, vol. 22, no. 5, September/October 2002, pp 24-35.

[9] Godin, G., Lalonde, J.-F., Borgeat, L., "Projector-based dual-resolution stereoscopic display". Proc. IEEE Conference on Virtual Reality 2004, 28-31 March 2004, Chicago, USA, pp. 223-224.

[10] Godin, G., Massicotte, P., Borgeat L., "Foveated Stereoscopic Display for the Visualization of Detailed Virtual Environments", Proc. Eurographics Symposium on Virtual Environments 2004, 8-9 June 2004, Grenoble, France, pp. 7-16.

[11] Hosseini, M., Georganas, N., "MPEG-4 BIFS streaming of large virtual environment and their animation on the web", Proc. of the Seventh Conference on 3D Web Technology, 24-28 February 2002, Tempe, AZ, pp. 19-25.

[12] Kresse, W., Reiners, D., Knöpfle, C., "Color consistency for digital multi-projector stereo display systems: The HEyeWall and the digital CAVE", Proc. 9th Eurographics Workshop on Virtual Environments / 7th Immersive Projection Technologies Workshop, 22-23 May 2003, Zurich, Switzerland, pp. 271-279.

[13] Laberge, D., Lapointe, J.-F., "An auto-calibrated laser-pointing interface for collaborative environments". Proceedings of the Ninth International Conference on Virtual Systems and Multimedia, Montreal, Canada, 15-17 October 2003, pp. 501-508.

[14] Raskar, R., "Immersive planar display using roughly aligned projectors", Proc. IEEE Virtual Reality 2000, 18 - 22 March 2000, New Brunswick, USA, pp. 109-115.