Windows on the World: An example of Augmented Virtuality

Kristian T Simsarian *kristian@sics.se*
Karl-Petter Åkesson *kalle@sics.se*

Swedish Institute of Computer Science
S-16428 KISTA
Stockholm, Sweden
fax: +46 (8) 751-7230
tel: +46 (8) 752-1570

Abstract

This work presents a system for creating augmented virtuality. We present a tool and methodology that can be used to create virtual worlds that are augmented by video textures taken of real world objects. These textures are images automatically extracted from images of real world scenes. The idea is to construct and update, in real time, a representation of the salient and relevant features of the real world. This idea has the advantage of constructing a virtual world that has the relevant data of the real world, but maintaining the flexibility of a virtual world. One advantage of the virtual-real world representation is that it is not dependent on physical location and can be manipulated in a way not subject to the temporal, spatial, and physical constraints found in the real world.

Keywords: virtual environments, distributed collaborative environments, augmented reality, autonomous robotics, computer vision, augmented reality, texture mapping, immersive virtuality

1 Introduction

Through interactive augmented virtuality we provide the ability to interactively explore a remote space. The object of this work is to create an augmented virtual world that contains real world images as object textures. This allows a user to explore a virtual representation of a real space. The textures are taken from objects that exist in the real world and which have dual (mirror) objects in the virtual world. This has the advantage of making a virtual world appear as the real world, but maintaining the flexibility of the virtual world, in this respect the augmented virtual world could be viewed as an instantiation of immersive 3D video. Objects look like their real counterparts but can be manipulated in a virtual setting. The virtual world can also be used as a control interface for manipulating objects in the real world. An advantage of the virtual world is that it is not dependent on physical location and can be manipulated in a way not subject to the temporal, spatial, and physical constraints of the real world. It is also has the advantage that irrelevant parts of the real world can be left out. Thus the "interesting" parts of the world can be extracted and made salient. In this way a custom view of the world can be created, forming an instantiation of selective vision. We call the instantiation of this system *Windows on the World*, or WOWs.

The purpose of using video images in the virtual world is manifold. The first is the same as the purpose of using textures, to give a richness to the virtual world that contains information and immediate access through visual memory to object identification. Another advantage is to furnish a method for viewing visual information from the real world, for example to visually monitor real world devices in real time via the virtual environment.

The applications for this technology are mainly in the area of telepresence and tele-exploration. An example of such an application is a security system where a guard virtually roams a domain that is composed of fresh video images taken of the scene. The advantage of this is that the 3D model of the scene is made explicit, this should have advantages over 2D video remote monitoring systems.

On the continuum from totally virtual to fully real scenes by incorporating real video into a virtual setting and vice-versa this application is located somewhere on the real side of a fully virtual world. In this paper we first present related research and then present our approach by offering an application example and then describing the basic methodology along with an overview of the system.

2 Related Research

It is not difficult to see how the capability of sending autonomous robots into hazardous and remote environments (e.g. space, sea, volcanic, mine exploration, nuclear/chemical hazards), can be useful. Robotics can be built to stand higher environmental tolerance than humans, they can be built to perform specialized tasks efficiently and they are expendable. To this end there has been much research on autonomous and teleoperated robots that can navigate into an area, perform actions such as taking samples, performing manipulations, and return the information to base controllers without human assistance, such as in space [10], and in the battlefield[1].

We are using an immersive virtual environment as the interaction paradigm with the real world and the robot which carries the camera. Specifically our work is an application in the SICS Distributed Interactive Virtual Reality (DIVE) system[2],[4]. We are incorporating on-board video from the robot into the virtual world. We see the video as a medium that can be enhanced and augmented to communicate information to the operator. This is quite similar to work in *Augmented Reality*, which is at its base the concept of combining video with graphic overlays. Augmented reality techniques have been demonstrated in a variety of practical applications. Examples of these are displaying graphical information in printer repair operations[3]. Or displaying information from a 3-D model base on a real world artifact [11], [9], and enhanced displays for teleoperated control[8]. Here we are looking at the reverse situation, augmenting the virtual environment with real images. Even so the work in Augmented Reality is important to our application.

The work in this paper can also be viewed as an instantiation of work in the general area of immersive 3D video, or 3DTV. In immersive video, a user is not restricted to the 2D plane for video but can interactively choose choose views and explore video regions. We have not yet gone so far as to create full live video environments. Rather, for our purposes, we see the WOW application

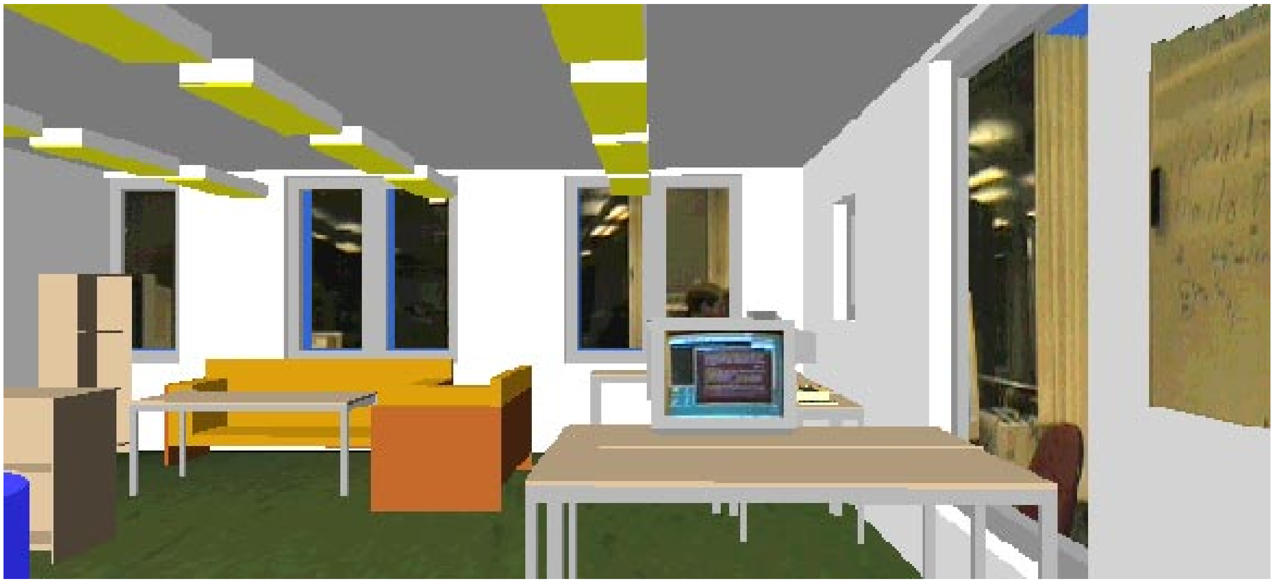

Figure 1: This figure shows a view from a virtual environment with a number of WOW textures automatically placed into the scene. These textures can be seen on the glass windows, the computer monitor on the table as well as the whiteboard on the right side wall.

as a means of filtering out non-essential details from a potentially cluttered scene in the real world. Work in the field of 3D video has also been done at UCSD by Ramesh Jain's group[6], and also at CMU by by Takeo Kanade's group[5]. Both of those applications have concentrated more on creating a database that can be interactively accessed. In this work we have created a tool that can be used for real-time remote investigation.

3 Approach

3.1 Application Example

One application of this technology would be for Security monitoring. Currently security "guards" observe a suite of monitors and watch for abnormalities and security situations. The main limitations in this system are the great extent of space the guard must observe. Often to overcome this situation a number of monitors are used by the guards, where each monitor rotates through a sequence of video inputs. A complimentary practice of security professionals is to perform walking tours of the monitored space. A virtual model of the real space can be built and used in combination with the application described in this paper. Cameras situated around the space could monitor the security space and apply textures to the virtual model. Thus the security guard could instead monitor a space by navigating through a virtual world instead of observing 2D video cameras. By adding intelligent cameras and simple image processing, changes in the scene could be flagged and brought to the attention of the security guard.

In current implementations of remote surveillance the 3d model of the environment is inside the head of the security guard. Here the model is instantiated offering both a clear view of the structure of the environment and of the permissible navigation. The main goal is to set up a natural interface where the choices and scene visualization and structure is more obvious.

In the next section we will discuss the basic methodology and system we have used to demonstrate the WOW concept.

3.1.1 Methods

In figure 2, we present what we have termed the *monitor* metaphor for displaying video. We contrast the monitor metaphor with the method presented here which we call *windows on the*

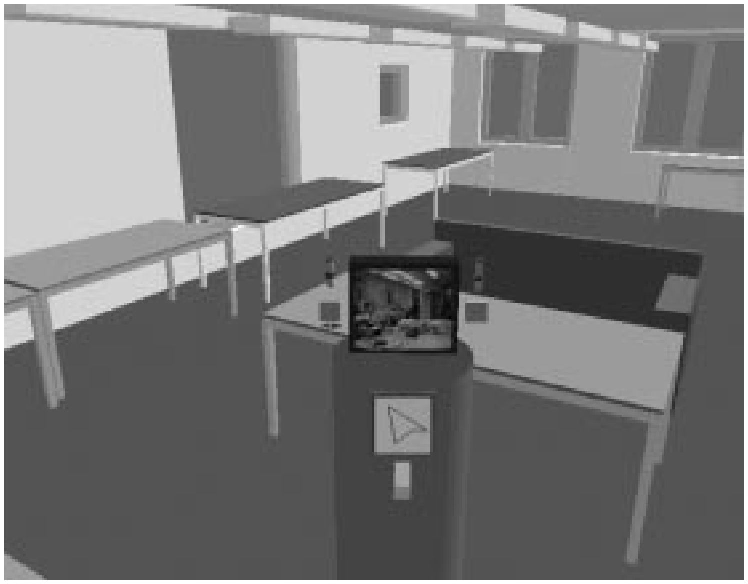

Figure 2: This figure displays a view of the immersive environment. It is a model of the laboratory space in which the robot which carried the camera is situated. In the center, the virtual robot representation can be seen. On the robot sits a live camera view into the real world. This image shows the use of the "monitor" metaphor for presenting the remote video.

world (see figure 1). We introduce the new method as a more immersive, and compelling way to view video from real world scenes than the monitor method.

The implementation of the system depends on the virtual camera objects in the virtual world. The camera objects contain both real and virtual object positions and camera geometry. The objects in the virtual world that mirror real world objects are called WOWs, *Windows On the World*. The virtual object that represents the camera is a special object that contains the *intrinsic* camera parameters obtained by calibration. The pose of the virtual camera is synchronized to the pose the real camera, thus the *extrinsic* camera parameters can be obtained.

In the work reported here we use a mobile robot as a means of moving the camera around a remote scene. Thus the position information to synchronize the virtual and real cameras is provided by a link between the robot positioning system and the virtual robot and camera objects.

3.2 Augmented Reality and Augmented Virtuality

One branch of research in computer generated immersive environments that we are exploring through this work is the boarder between real and virtual scenes. The interface presented in this paper takes a model of the world and with the aid of calibration, lays graphics onto the video image and then takes a piece of the video image and applies that as a texture on the virtual object. As mentioned in [9] this boarder is actually a continuum from real to virtual.

Given this notion of augmented reality, we are mainly concentrating on performing the compliment operation, that is embellish the virtual environment with real-world images. With the proper, possibly user-guided, feature extraction and image warping we decorate the world with much of the richness of the real world. Thusly, images are captured in real-time and texture mapped to corresponding synthetic surfaces. An operator can then interact within a world which contains objects that can be recognized in familiar and potentially useful ways. This could be viewed as a method for creating a world of 3D visual mneumonics. In this respect we use the term *augmented virtuality*.

3.3 System Overview

The first step in the process is to calibrate the camera. The calibration process can be achieved by one of the known processes. Methods that we have employed in this work are based on [12, 7].

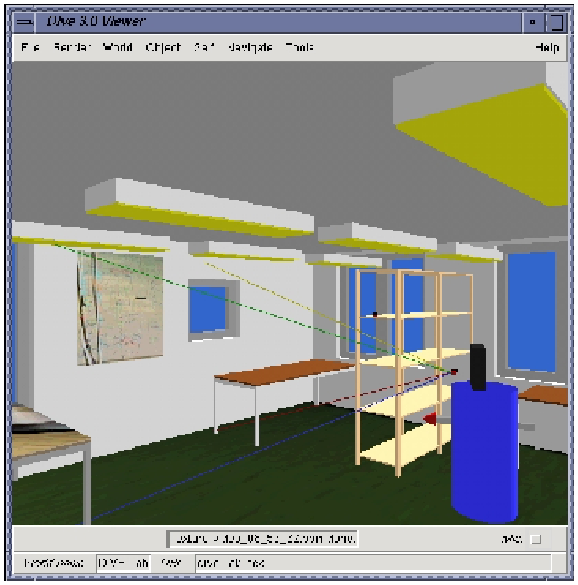

Figure 3: In this This is a visualization of the the camera's field of view on the scene. The rays cast out from the camera on the robot show the limitations of the field of view. On the wall was a Wow texture that was waiting for the camera to pass over and send it the image of the whiteboard.

The cameras we employ are active in the sense that they sit on a pan/tilt head or a mobile platform. While navigating around the environment, the camera acquires images of the world and the controlling process is capable of extracting video textures from the camera signal. The image-textures are acquired either by the event of camera movement, user choice or periodically. Any of these events triggers the camera to produce textures for each WOW within its field of view. The camera records its own position and the position of the WOWs and uses that information to extract data from the camera signal (images). The camera process then uses these images as textures. This process is accomplished by using the virtual object geometery coupled with the camera parameters (extrinsic + intrinsic) and projecting the relevent object surface onto a virtual camera image plane matching the real camera. Using this 3D to 2D projection, the relevent portion of the camera image can be extracted as a texture and used in the virtual world. Thus as the camera pans around a room, WOW objects receive textures. As the camera moves around the environment the virtual world fills up with textures and comes to resemble more closely the real world scene. The quality and accuracy of the textures depend on the following factors: calibration, the distance the camera is from the object, lighting and the angle the camera makes with the objects surface.

Another key component of the system is warping the texture onto the object. Because of, for example, camera skew, an extracted camera image segment and texture may not contain right angles. Such images will first go through a warping stage to fit onto the WOWs polygonal surface. Another possible problem with extracted textures is if the entire WOW surface is not visible from a particular camera view point. If there are segments of a WOW that lie outside the field of view of the camera, the texture will be transparent. In the event a whole surface of an object is not seen from a camera, the remaining part of the WOW will be transparent. Thus, a layer of WOWs with textures from different cameras can be placed on the surface of the object. In this way segments of a WOW that are not seen by one camera will show underlying WOW textures and the result will be a an overlapping surface composed of a number of video textures.

3.4 Mobile robotics

The robot that we are using for this system is a Real World Interface B21 robot with on-board processing and sensing. The robot can perform basic tasks such as navigate around obstacles, recognize objects to the best of its ability and take its high level commands from a human, thus displacing the artificial high-level planner with a human one. The system is not fully autonomous,

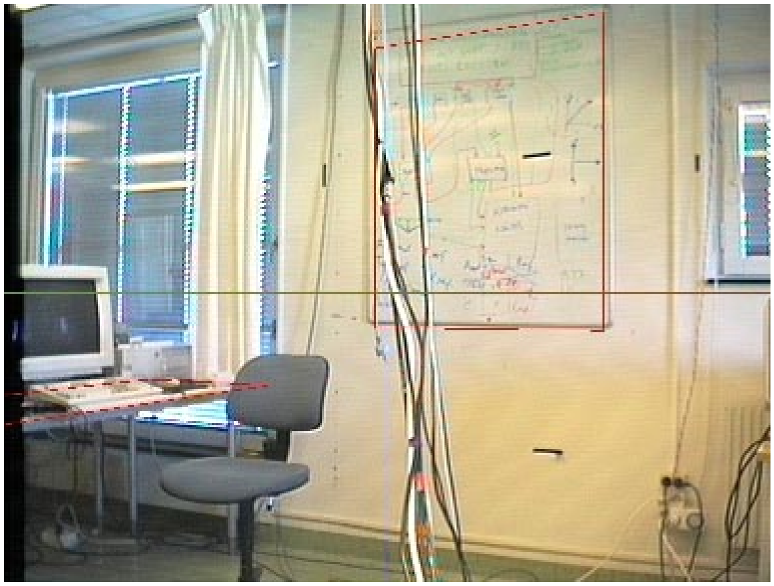

Figure 4: This image shows the same scene as in the previous figure, only taken from the directly from the camera on top of the robot. The red outlines area is the part of the image to be extracted.

but is seen as collaborating with its operator.

The robot can perform basic point-to-point navigation tasks and avoid basic obstacles while negotiating the indoor structured environment in which is situated. Having access to a model of the environment as well as access to the operator's knowledge and assistance gives a great deal of leverage on the harder problems of robot navigation, such as localized sensor pitfall situations.

The main application theme for such hybrid human-machine systems follows the metaphor of the machine working as an assistant. The main application described here allows an operator to perform remote operations something we are calling *virtual presence*. Many situations exist where it is dangerous or difficult for humans to go, i.e. toxic waste spill, nuclear accident, space exploration, deep sea exploration, volcanic exploration, and wild fire fighting. In such environments one would like to have a machine proxy in place of a human.

4 Summary and Future work

We have presented a description of a system that implements a concept we are calling Augmented Virtuality. The system is able to roam around a real space while sending texture updates based on video to a virtual world model of the real space. Using the virtual model a user can perform an off-line tour of the remote space in a form of tele exploration. Using this application the user can decouple his actions from the actions of the robot in the remote space.

Future work centers on improving the quality of the automatic extraction by improving the initial calibration of the camera as well as adding methods for automatic detection of objects by using simple image processing routines.

These techniques could likewise be used in worlds where no model has been supplied. There there will be a reliance on sensors and their ability to generate models of the world. An example here would be to use range sensors to sense solid objects. When such obstacles are detected a object could be instantiated in the location with a WOW that then extracts the texture. This information would prove useful to the operator for identifying the object that triggered the response. A more sophisticated sensor system could also be constructed with a basic image processing techniques coupled to the camera and callibration process. In this case, some structure, e.g. edges, could be determined from the scene, possibly with the user's assistance and then placed in the scene.

References

[1] W.A. Aviles, T.W. Hughes, H.R. Everett, A.Y. Martin, and A.H. Koyamatsu. Issues in mobile robotics: The unmanned ground vehicle program teleoperated vehicle (tov). In *SPIE Vol. 1388 Mobile Robots V*, pages 587–597, 1990.

[2] Christer Carlsson and Olof Hagsand. DIVE – a platform for multi-user virtual environments. *Computers and Graphics*, 17(6), 1993.

[3] Steven Feiner, Blair MacIntyre, and Doree Seligmann. Knowledge-based augmented reality. *Communications of the ACM*, 36(7):52–62, July 1993.

[4] Olof Hagsand. Networked virtual environments in the dive system. *IEEE Computer: Special issue on Virtual Worlds, forthcoming*, 1995.

[5] T. Kanade, P.J. Narayanan, and P. Rander. Virtualized reality:concepts and early results. In *IEEE Workshop on the Representation of Visual Scenes*, page http://www.cs.cmu.edu/Groups/VirtualizedR/, Boston, MA, June 1995.

[6] Arun Katkere, Saied Moezzi, Don Kuramura, Patrick Kelly, and Ramesh Jain. Towards video-based immersive environments, http://vision.ucsd.edu/papers/virtual_world. Technical Report VCL-95-105, Visual Computing Laboratory, University of California, San Diego, May 1995. A version to be published in the ACM-Springer Multimedia Systems Journal: Special Issue on Multimedia and Multisensory Virtual Worlds, 1996.

[7] M. Li. 'camera calibration of the kth head-eye system. Technical report, Computational Vision and Active Perception Lab., Dept. of Numerical Anaysis and Computing Science, Royal Institute of Technology (KTH), March 1994.

[8] M. Mallem, F. Chavand, and E. Colle. Computer-assisted visual perception in teleoperated robotics. *Robotica (10)*, pages 99–103, 1992.

[9] Paul Milgram and David Drascic. Enhancement of 3-d video displays by means of superimposed stereo-graphics. In *Proceedings of the Human Factors Society 35th Annual Meeting*, pages 1457–1461, 1991.

[10] NASA. *Proceedings of the NASA Conference on Space Telerobotics*. JPL Publication 89-7, Vol 1-5, Pasadena, Ca, 1989.

[11] Eric Rose, David Breen, Klaus H. Ahlers, Chris Compton, Mihran Tuceryan, Ross Whitaker, and Douglas Greer. Annotating real-world objects using augmented vision. Technical report, European Computer-Industry Research Center GmbH, Arabellastrasse 17 D-81925 Munich, 1994.

[12] R. Tsai. A versatile camera calibration technique for hifg-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of robotics and automation*, RA-3(4):323–344, August 1987.