

different Projection-based VR systems. A first prototype of the approach presented in GMD's open house event in October'97, is also presented in this paper. The scientific approach requires camera calibration, in order to obtain the camera parameters, which are then used for integrating the stereo-video into the virtual space, while preserving the stereo-effect and the viewer's perspective. In addition, stereo image rectification is required when a nonparallel stereo camera system is used.

Taking advantage from the distributed spatial audio effects of our AVOCADO Software Framework, we are able to include live audio as a localized audio source linked to the video planes that represent the remote participant in the virtual world. The audio spatial feedback then corresponds to the visual effect of moving nearer or further away from the integrated image of the remote participant in the virtual world, providing an even more realistic feeling of immersive telepresence.

The results drawn by the use of this prototype have encouraged us to continue our research towards Immersive Telepresence in a number of different directions. In particular, we are working on solutions on a number of open issues such as the use of more than one camera-pair, the real-time rectification algorithm, incorporating more video and audio streams for representing other information, and different usage scenarios for different Projection-based VR systems and for different application areas.

7. REFERENCES

- [1] Ayache N. *Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception*. The MIT Press, 1991.
- [2] Breiteneder C., Gibbs S., Arapis C., TELEPORT- An Augmented Reality Teleconferencing Environment, Proc. 3rd Eurographics Workshop on Virtual Environments Coexistence & Collaboration, Monte Carlo, Monaco, February 1996
- [3] Cruz-Neira C., Sandin D.J., DeFanti T.A., Kenyon R., and Hart J.C, The CAVE, Audio Visual Experience Automatic Virtual Environment, Communications of the ACM, June 1992.
- [4] Dai P., Eckel G., Goebel M., Hasenbrink F., Lalioti V., Lechner U., Strassner J., Tramberend H., Wesche G., *Virtual Spaces - VR Projection System Technologies and Applications*, Tutorial Notes of the 1997 Eurographics Conference, Budapest, 1997.
- [5] Dechelle, F., DeCecco, M., *The IRCAM Real-Time Platform and Applications*, Proceedings of the 1995 International Computer Music Conference, International Computer Music Association, San Francisco, 1995.
- [6] Dornaika F. and Garcia C., Pose estimation using point and line correspondences. To appear in *International Journal of Real Time Imaging*, New York, February 1998.
- [7] Dornaika F. and Garcia C., Robust camera calibration using 2D to 3D feature correspondences, *Proceedings of the International Symposium SPIE --Optical Science Engineering and Instrumentation, Videometrics V*, Volume 3174, pages 123--133, San Diego, Ca., July 1997.
- [8] Falkenhagen L., Block-Based depth estimation from image triples with unrestricted camera setup, *IEEE*, 1997, pages. 280-285
- [9] Faugeras O., *Three-Dimensional Computer Vision: A Geometric Viewpoint*. The MIT Press, Cambridge, 1993.
- [10] Faugeras O. and Toscani G., Camera calibration for 3D computer vision. In *Proceedings of the International Workshop on Machine Vision and Machine Intelligence*, Tokyo, Japan, February 1987.
- [11] Haase H., Dai F., Strassner J., Goebel M., *Immersive Investigation of Scientific Data*, Scientific Visualization, IEEE Press, 1997
- [12] R. I.Hartley. Euclidean reconstruction from uncalibrated views. *Applications of Invariance in Computer Vision*, pages 237--256, Springer Verlag, Berlin Heidelberg, 1994.
- [13] Ishii H. and Kobayashi M., ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact, CHI'92, May 3-7, 1992.
- [14] Krueger W. and Bernd Froehlich B., *The Responsive Workbench*, IEEE Computer Graphics and Applications, May 1994
- [15] Lindemann E., Starkier F., Dechelle, F., *The IRCAM Musical Workstation: Hardware Overview and Signal Processing Features*, In: S. Arnold and G. Hair, eds. *Proceedings of the 1990 International Computer Music Conference*. San Francisco: International Computer Music Association, 1990.
- [16] Magnenat Thalmann N. and Thalmann D., *Virtual Worlds and Multimedia*, Wiley Professional Computing, John Wiley and sons, Chichester, England, 1993.
- [17] Papadimitriou D.V. and Dennis T.J., Epipolar line estimation and rectification for stereo images pairs. *IEEE Transactions on Image Processing*, 3(4):672-676, April 1996.
- [18] Tsai R., A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323-344, August 1987.
- [19] Wellner P., *DigitalDesk*, Communications of the ACM, Vol. 36, No. 7, July 1993.
- [20] URL <http://www.xerox.fr/ats/br/livead.html>

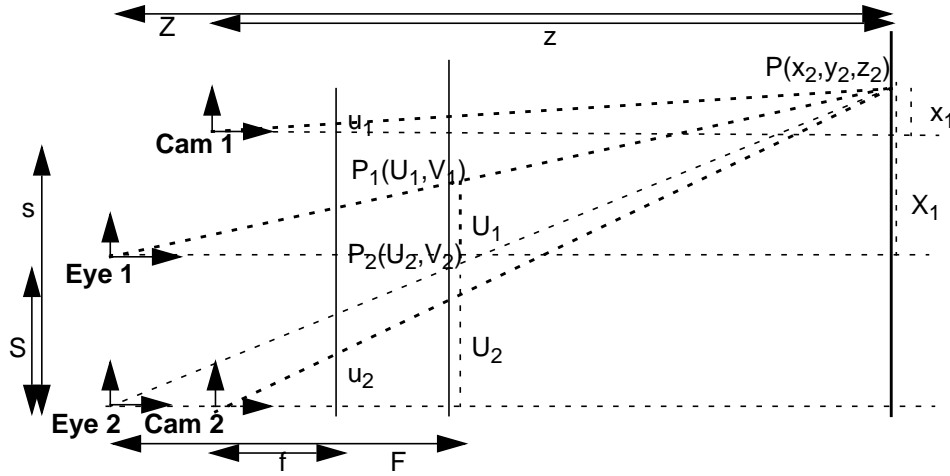


Figure 7 Eye and Camera Coordinate Systems

Similarly, we could find the new V_1 and V_2 . For simplicity we assume that the two cameras are at the same height h from the $(0,0,0)$ while the eyes of the viewer are at height H . The equations derived for V_1 and V_2 are as follows:

$$\left. \begin{array}{l} \frac{V_1}{Y_1} = \frac{F}{Z} \\ \frac{v_1}{y_1} = \frac{f}{z} \\ Y_1 = y_1 + (h - H) \end{array} \right\} \Rightarrow V_1 = v_1 \cdot \frac{F}{f} \cdot \frac{z}{Z} + \frac{F}{Z} \cdot (h - H) \quad (5)$$

$$\left. \begin{array}{l} \frac{V_2}{Y_2} = \frac{F}{Z} \\ \frac{v_2}{y_2} = \frac{f}{z} \\ Y_2 = y_2 + (h - H) \end{array} \right\} \Rightarrow V_2 = v_2 \cdot \frac{F}{f} \cdot \frac{z}{Z} + \frac{F}{Z} \cdot (h - H) \quad (6)$$

To summarize, given a point $P(x,y,z)$ and its equivalent points $P(u_1,v_1)$ $P(u_2,v_2)$ on the stereo camera's image planes equations (4) (5) and (6) calculate the new $P(U_1,V_1)$ and $P(U_2,V_2)$ points according to the viewer's perspective. Approaches such as [8], can produce a depth map out of a pair of video-images and then generate a 3-D model of the real object. The video-texture is then mapped onto this 3D model.

In the approach presented in this paper, we are focusing more on a real-time virtual meeting scenario which at the same time would not limit the complexity of the data visualized within Cyberstage. To achieve this we minimized the number of extra polygons needed for the representation of the stereo image by using the simple geometry of a plane. In order to also minimize the calculations needed for adjusting the stereo video images according to the viewer's perspective, we assume that every point of the real scene has the same depth. Therefore, instead of calculating the above equations for each point separately, we apply them to the geometry onto which the video images are mapped.

According to the equations (4), (5) and (6), to position the remote participant at an average distance of Z from the viewer, when the observed scene is located at an average

distance of z from the stereo camera's optical centers, we perform the following operations:

- define two planes of size G multiplied by the size of the video image plane in u and v , where $G=(F*z)/(f*Z)$
- position them in front of each eye (each eye's optical axis passing through the center of each plane) at a distance Z into the virtual world
- translate the plane by T along the X -axis and T' along the Y -axis of the CC, where $T=(F/Z)*(s-S)$ and $T'=(F/Z)*(h-H)$

5. SCHLOSSTAG'97

A prototype of the system was demonstrated at GMD's open-house event in October 1997 (Schlosstag'97). In this chapter we summarize the scenario and technical details of the demonstration. The scenario was asymmetric and involved two sites with different infrastructure, the Cyberstage and a blue room. In particular, participants and visitors were located at the Cyberstage site, while at the blue room a remote participant was located. The remote participant, was captured by a stereo-camera and the image was chroma-keyed and integrated into a 3-dimensional virtual environment, allowing a fully immersed 3D virtual meeting on the Cyberstage site.

An overview of the set-up, used during Schlosstag'97, is shown in Figure 8. The left and right video streams, generated by the stereo-camera in the blue room, were spatially composited into one, in real-time. The spatially combined stream, which contained the even fields of each signal, was then enhanced with an alpha channel, by the use of an Ultimatte keying system. The resulting video stream was used as an input to the Sirius boards connected to the two Infinite Reality pipes of an SGI Onyx machine which was powering the Cyberstage. The video stream was then integrated as a stereo video texture in the virtual scene, by the use of the AVOCADO Software Framework.

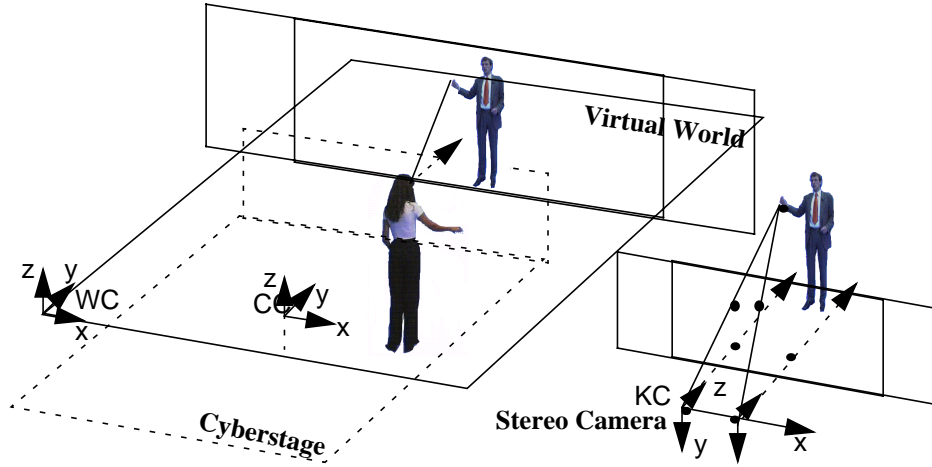


Figure 5 Virtual, Camera and Real World Coordinate Systems

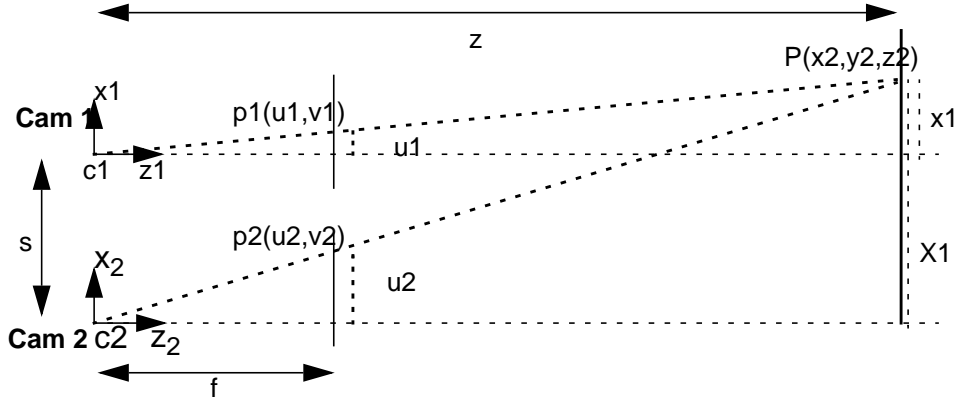


Figure 6 Two Camera Coordinate System

In Figure 6, the camera pair is shown after rectification. Let c_2 , the focal point of Cam2, be the center of the camera coordinate system, f the focal length and (u, v) the image plane of the cameras.

A point $P(x, y, z) = P(x_2, y_2, z_2)$ would be projected on the image plane of camera 2 at $p_2(u_2, v_2)$ and on the image plane of camera 1 at $p_1(u_1, v_1)$ respectively. Let s be the distance between the focal point of the two cameras, then:

$$\left. \begin{array}{l} \frac{u_1}{x_1} = \frac{f}{z} \\ \frac{u_2}{x_2} = \frac{f}{z} \end{array} \right\} \Rightarrow \begin{array}{l} x_1 = \frac{u_1}{f} \cdot z \\ x_2 = \frac{u_2}{f} \cdot z \end{array} \quad (1)$$

The disparity between the two images for point $P(x, y, z)$ is then:

$$d = u_1 - u_2 \left\} \Rightarrow d = \frac{f}{z} \cdot s \quad (2)$$

The same point $P(x, y, z)$ viewed by a person at distance Z , with eye-focal length F and distance S between the eyes, would appear at point $P_2(U_2, V_2)$ and $P_1(U_1, V_1)$ respectively. In the general case, shown in Figure 7, the distance s between the camera pair and the focal length f of the cameras are different from the distance S and focal length F of the eyes of the viewer. Then the relationship between the projection of point P on the eye image plane with the projection on the camera image plane is as follows:

$$\left. \begin{array}{l} \frac{U_1}{X_1} = \frac{F}{Z} \\ \frac{U_2}{X_2} = \frac{F}{Z} \end{array} \right\} \Rightarrow \begin{array}{l} X_1 = \frac{U_1}{F} \cdot Z \\ X_2 = \frac{U_2}{F} \cdot Z \end{array} \quad (3)$$

From Figure 7 we derive that $X_1 = x_1 + (s-S)$ and $X_2 = x_2$. Then by substituting X_1 and x_1 from equation (3):

$$\left. \begin{array}{l} \frac{U_1}{F} \cdot Z = \frac{u_1}{f} \cdot z + (s-S) \\ \frac{U_2}{F} \cdot Z = \frac{u_2}{f} \cdot z \end{array} \right\} \Rightarrow \begin{array}{l} U_1 = u_1 \cdot \frac{F}{f} \cdot \frac{z}{Z} + \frac{(s-S) \cdot F}{Z} \\ U_2 = u_2 \cdot \frac{F}{f} \cdot \frac{z}{Z} \end{array} \quad (4)$$

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = M \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (1)$$

where s is an arbitrary scale factor and M is a 3×4 matrix, called the *perspective projection matrix* or the *calibration matrix*. This matrix is defined up to a scale factor.

One can produce the following two equations that give the relation between the point calibration matrix P and its image point p :

$$\begin{cases} su = m_{11}x + m_{12}y + m_{13}z + m_{14} \\ sv = m_{21}x + m_{22}y + m_{23}z + m_{24} \\ s = m_{31}x + m_{32}y + m_{33}z + m_{34} \end{cases} \quad (2)$$

$$\begin{cases} (m_{11} - u m_{31})x + (m_{12} - u m_{32})y + (m_{13} - u m_{33})z + m_{14} - u m_{34} = 0 \\ (m_{21} - v m_{31})x + (m_{22} - v m_{32})y + (m_{23} - v m_{33})z + m_{24} - v m_{34} = 0 \end{cases} \quad (3)$$

where the m_{ij} are the coefficients of the matrix M . These two equations are linear and homogeneous in the coefficients m_{ij} . In order to solve them, m_{34} can be set to an arbitrary non-zero value (for instance, $m_{34}=1$). Therefore, the system (3) becomes:

$$\begin{cases} (m_{11} - u m_{31})x + (m_{12} - u m_{32})y + (m_{13} - u m_{33})z + m_{14} = u \\ (m_{21} - v m_{31})x + (m_{22} - v m_{32})y + (m_{23} - v m_{33})z + m_{24} = v \end{cases} \quad (4)$$

In order to estimate this matrix (i.e. 11 unknown elements), we need at least 11 equations. The linear system (4) provides two equations, therefore the matrix M can be linearly estimated if at least 6 pairs (p_i, P_i) are provided.

Once the coefficients of the matrix M have been determined by using standard linear algebra techniques (for instance, the Least Squares method), one can decompose it into an intrinsic parameters matrix (matrix C) and an extrinsic parameters matrix:

$$M = \underbrace{\begin{bmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_C \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \quad (5)$$

where

- α_u and α_v are the scaling factors along the u and v axes of the CCD plane. The focal length is derived from these scaling factors, given the number of pixels (namely n_u and n_v) and the size of the camera CCD plane (namely l_u and l_v): $f = \alpha_u * (l_u/n_u) \alpha_v * (l_v/n_v)$.

- (u_0, v_0) are the *principal point* or *image center* coordinates in pixels
- $R(r_{ij})$ and $T(t_x, t_y, t_z)$ are rotation and translation matrices from world coordinates to camera coordinates

This decomposition can be performed using a variation of the well known QR decomposition of matrices, namely RQ decomposition [12]. Matrices C and R can be recovered by decomposing the 3×3 submatrix of M into an upper triangular matrix and an orthogonal matrix. Even though the results are precise enough, we plan to make the calibration step more stable by using a non-linear optimization method [7] based on a quaternion formulation and derived from 2D to 3D point or line correspondences.

A special object has been constructed for camera calibration which is a white cube with black circular marks. These landmarks are precisely positioned and define a coordinate system attached to the object. Image points are extracted by segmenting the dark marks and by locating the center of gravity of each of these clusters. Then, pairs of 2D-3D point coordinates (p_i, P_i) are built automatically, establishing the 2D-3D correspondences. As mentioned above, at least 6 pairs of 2D-3D correspondences have to be used in our camera calibration algorithm.

After calibrating separately each of the cameras, external parameters such as optical center position, image plane orientation are gained and are used as input for the stereo image rectification step together with internal parameters such as real focal length (and not the nominal one). These internal parameters are also used when positioning and scaling the stereo images into the cyberstage.

4.3 Stereo-video integration into Virtual World

In order to merge the stereo video of a remote participant into the Cyberstage, the rectified stereo images are mapped onto virtual planes. Figure 5 shows the relation between Cyberstage, Camera and Virtual World coordinate systems. The video planes and virtual objects are positioned, in the virtual world coordinate system (WC). The $(0,0,0)$ of the Cyberstage coordinate system (CC) is at the center of the CyberStage room 1.5m from the floor. The viewer's location is given according to the CC and we know the mapping from WC to CC at any given time (e.g. the Cyberstage is moving around the virtual world). On the other hand we have the stereo camera coordinate system (KC) which for simplicity we assume that it is also the coordinate system of our real world.

When the mapped planes are projected onto the Cyberstage, the viewer must have the sensation of seeing the 3D image of the remote participant together with the 3D scene of the virtual world. For this purpose the video planes must be sized and translated to reflect the viewer's movement in Cyberstage. In the rest of the section we examine which are the parameters that determine the scaling and translation factors.

distributed sound effects of our AVOCADO Software Framework. For live stereo video, each image of the stereo camera is mapped onto a simple geometry representing a plane. The AVOCADO Software Framework then displays the right camera image to the right eye of the viewer and the left to the left eye respectively. However, these two planes have to be fully defined in terms of size, position, aspect ratio and orientation in order to provide the Cyberstage viewer with the best quality stereo-video and to respect the appearance of the remote person. In addition, the segmentation techniques of the TELEPORT system are also used in our approach. Therefore, we are able to determine the regions of the video signal that are of interest (i.e. the image of the remote participant) and combine this information into the original video signal, thus making the background transparent.

The Cyberstage viewer, must have the sensation of sharing the same virtual space with the remote participant. Therefore, when positioning the video planes we have to respect physical rules such as the size of the participants. In addition, when the image of a remote participant is positioned behind a virtual object the stereo impression for the Cyberstage viewer must confront to this (i.e. the stereo-effect should not result in the remote participant perceived as been within virtual objects). Thus, the stereo images have to be adjusted in terms of size, orientation and position, before being integrated into the virtual world. Automatic camera calibration is required to perform such a task. In case of a non-parallel stereo vision system, image rectification is needed, as described in the subsections that follow.

4.1 Stereo Image Rectification

We choose to use an unrestricted stereo vision system, consisting of a rig of standard video cameras placed all over the observed scene. In the current implementation, two cameras are used with the only constraints that they share a large field of view. In most of the applications using stereo vision, a parallel stereo vision system, like the one shown in Figure 4, is required and especially in the present case, where the produced images should be comfortably viewed by a human visual system.

It is a difficult task to mechanically set up a parallel stereo vision system. A way to use a non-parallel stereo vision system is to perform image rectification. This process aims at transforming the images taken by such a non-parallel system into images taken by a virtual parallel stereo vision system. Efficient methods for rectification may be found in [1],[9] and [17]. Our work improves and extends [1] given that we enforce explicitly all the constraints necessary and sufficient to derive a unique rectification matrix, and obtain the latter as the solution of a resulting system of 4 simultaneous linear equations. As shown in Figure 4, we compute the transformation from (C_1, x_1, y_1, z_1) to (C_1', x_1', y_1', z_1') , which is the new coordinate system for camera 1.

The rectification algorithm performs in real time since it uses only the parameters gained by the cameras calibration

and needs at least 6 operations per pixel. The images are rectified by projectively mapping all pixels of the virtual image plane into the original image plane and interpolating the intensity information using a lowpass filter. In order to perform rectification and to position the resulting rectified images in the cyberstage, a robust and full recovery of the camera parameters is needed. Camera calibration methods are designed to perform these parameters estimation.

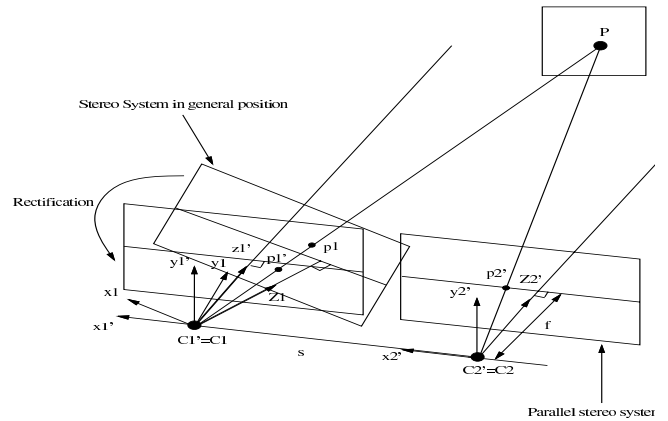


Figure 4 Rectification of a stereo vision system

4.2 Camera Calibration

There are two kinds of parameters associated with camera calibration [10],[18]: (i) the *intrinsic parameters* including optical and electronic properties of a camera, such as focal length, lens distortion coefficients, image center, scaling factors of the pixel array in both directions. Sometimes, manufacturers provide a partial set of these parameters, but they are not accurate enough. Also, some of these parameters may vary from time to time, while some of them may be calibrated once for all, depending on the stability of the mechanical and optical construction of the camera; (ii) the *extrinsic parameters* corresponding to the pose estimation (rotation and translation) of the camera system relative to a user-defined 3D world coordinate frame.

We take profit of our previous work [6][7], consisting of the fully automatic and simultaneous estimation of both the intrinsic and extrinsic camera parameters. The camera is described by the widely used pin-hole model in which the camera is supposed to perform a perfect perspective transformation of 3D space on a retinal plane. In the general case, we must also account for a change of world coordinates, as well as for a change of retinal coordinates, so that a generalization of the previous assumption is that the camera performs a projective linear transformation rather than a mere perspective transformation.

The coordinates of a 3D point $P = [x,y,z]^T$ in a world coordinate system and its retinal image coordinates $p = [u,v]^T$ are related (in homogeneous coordinates) by:

from the heavy load and inconvenience related to head mounted displays, increasing resolution and rendering speed enables VR for serious applications [11]. A common characteristic of the projection-based VR systems, is that they all extend the real space by a virtual space providing a common world coordinate system, where the local and the remote participants are part of, as shown in Figure 1.

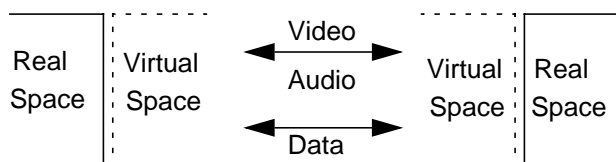


Figure 1 Projection-based VR Systems

The mapping of virtual space to real space allows us to characterize them as desk or room size installations. Currently desk and room size installations are available, like the Responsive Workbench^{TM1}, the CyberStage or the Teleport [2]. In the RWB concept [14] the user no longer experiences simulations of interesting procedures on the computer, but the computer is invisibly integrated into the user's world, Figure 2. The virtual objects and control tools, displayed as computer-generated stereoscopic images are projected onto the surface of a table. The user interacts with the virtual objects and manipulates them as if there were real.

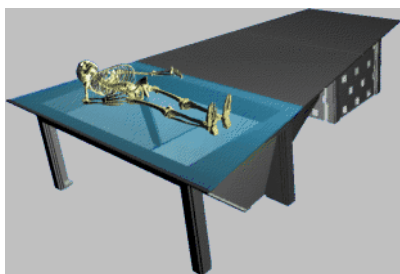


Figure 2 Responsive Workbench

CyberStage is a CAVE^{TM2} like [3] four-side room-size stereo display system installed at GMD, which creates the illusion of immersion within a computer generated virtual environment. Users see large virtual spaces and hear spatially distributed sound. Projection systems like CyberStage allow a direct and body centered human interaction within virtual worlds as well as team work. Three wall size rear projection systems are installed orthogonal to the floor projection, each with a size of 3x3 meters. An SGI 4 pipe Onyx 2 Infinite Reality generates

¹ CyberstageTM is a registered Trademark of the German National Research Center for Information Technology GMD

² CAVETM is a registered Trademark of the University of Illinois

eight user controlled images. The user position is tracked with Polhemus Fastrak sensors. Crystal Eyes shutter glasses are used for stereo image perception. The display resolution is 1024 x 768 pixels at 120 Hz for each of the four displays. The eight channel-surround-sound system is fed by IRCAM's room acoustic software Spatilisateur [5][15] and provides support for localized sound sources within the virtual environment. A significant characteristic of the Cyberstage is the acoustic floor which allows to generate the sense of vibrations.

TELEPORT is a synchronous collaboration system that provides high degree of co-presence [2]. The system is based around special rooms, called display rooms, where one wall is a "view port" into a virtual extension. The geometry, surface characteristics, and lighting match the real room to which it is attached. When a teleconferencing connection is established, video imagery of the remote participant (or participants) is composited with the rendered view of the virtual extension, Figure 3. Two techniques are used for *segmentation* (for determining the regions of the video signal where a participant appears) chroma-keying and delta-keying. The viewing position of the local participant is tracked, allowing imagery appearing on the wall display to be rendered from the participant's perspective. The current system uses a 3m x 2.25m rear-projected *video wall* attached to a 3m square room.



Figure 3 Telepresence session in the TELEPORT room

Our software framework for virtual environments, AVOCADO, was developed in parallel to the Cyberstage installation and became our main platform for research and development in all the above Projection-based VR systems. One of the main features of AVOCADO is its easy extensibility which facilitates the expansion of the system. Therefore, numerous effects have been added to AVOCADO including a set of texture based effects. These effects mainly deal with the dynamic exchange of the texture image and are capable of showing live video input, and play movie files from disk.

4. SCIENTIFIC APPROACH

In the approach presented in this paper, we integrate live stereo video and audio of a remote participant into Cyberstage using the texture based effects and the spatially

Virtual Meeting in CyberStage

Vali Lalioti

Research Scientist, PhD.

GMD - IMK.VE

53754 Sankt Augustin, Germany

+49.2241.14.2787

Vali.Lalioti@gmd.de

Christophe Garcia

Research Scientist, PhD.

FORTH-ICS

PO BOX 1385

Heraklion, Crete, 71110 Greece

+30.81.391701

cgarcia@csi.forth.gr

Frank Hasenbrink

Research Scientist

GMD - IMK.VE

53754 Sankt Augustin, Germany

+49.2241.14.2051

Frank.Hasenbrink@gmd.del

1. ABSTRACT

Today's technology and advances in networking and telecommunications stimulate a change in the way business is carried out, making it a globally distributed process, in which communication and collaboration of geographically dispersed groups is of vital importance. Virtual Reality systems are adapting accordingly, by providing not only a better man-machine interface, but also by facilitating human to human interaction and collaboration over distance. The approach presented in this paper, creates an environment where remote participants not only meet as if face to face, but also share the same virtual space and perform common tasks. Live stereo-video and audio, from a projection-based VR system are transmitted and integrated into the virtual space of another participant at a distant VR system, allowing geographically separated groups to meet in a common virtual space, while maintaining eye-contact, gaze awareness and body language. The scientific approach involves stereo-camera calibration and rectification, and use of the camera parameters for integrating the stereo-video into a virtual environment, while maintaining the stereo-effect and correct perspective for each participant. A prototype environment in CyberStage, is also presented in detail in this paper.

1.1 Keywords

distributed VR environments, VR software, VR applications

2. INTRODUCTION

Virtual Reality is widely accepted as a promising approach to a better man-machine interface, overcoming the present limitations of desktop systems and adapting more closely to the user needs. Projection-based VR systems are using metaphors, such as the blackboard or the desk for creating shared working environments that provide a more natural man-machine communication [14][19]. Today's technology and advances in telecommunications lead to sophisticated multimedia systems which combined with virtual reality can provide a high degree of co-presence and co-working for geographically dispersed groups [2] [16]. Therefore, new challenges are introduced in terms of multimedia integration in distributed virtual reality environments and interaction. It is not only a question of solving the technical problems of gathering and transmitting multimedia datastreams with sufficient quality and speed, but also a question of addressing the specific needs of human communication. For example, facial expression, body language and eye contact are an integral part of this communication.

Teleconferencing systems that provide high-degree of co-presence, such as [2], and collaborative co-presence systems such as [13][19][20], give enough evidences that projection-based VR systems when combined with telepresence facilities, can greatly facilitate the communication and collaboration over distance in a variety of application areas. The approach presented in this paper, creates an environment where remote participants not only meet as if face-to-face, but also share the same virtual space and perform common tasks, in order to reach a common goal. In particular, live stereo-video and audio of remote participants is integrated into the virtual space of another participant, allowing a geographically separated group of people to collaborate while maintaining eye-contact, gaze awareness and body language. Participants could be using a wide range of Projection-based VR systems [2][14][4], resulting symmetric or asymmetric collaboration scenarios. In the section that follows, we present some of our Projection-based VR systems, while in section 4 the scientific approach of integrating live stereo-video and audio in a Projection-based VR system is described. Section 5, summarizes the demonstration in Cyberstage, of a prototype environment. Finally, section 6 concludes this paper with some of the open issues and future research directions.

3. BACKGROUND WORK

Projection Display Systems are the state of the art in high end Virtual Reality Environments [4]. Releasing the user