Seeing Eye to Eye: a shared mediated reality using EyeTap devices and the VideoOrbits Gyroscopic Head Tracker

Felix Tang, Chris Aimone, James Fung, Andrej Marjan, Steve Mann
University of Toronto, 10 King's College Road, Toronto, Canada
{tangf,aimone,fungja,amarjan}@eyetap.org, mann@eecg.utoronto.ca

Abstract

We present a system which allows wearable computer users to share their views of their current environments with each other. Our system uses an EyeTap: a device which allows the eye of the wearer to function both as a camera and a display. A wearer, by looking around his/her environment, "paints" or "builds" an environment map composed of images from the EyeTap device, along with head–tracking information recording the orientation of each image. The head–tracking algorithm uses a featureless image motion estimation algorithm coupled with a head mounted gyroscope. The environment map is then transmitted to another user, who, through their own head-tracking EyeTap system, browses the first user's environment solely by head motion, seeing the environment as though it were their own. As a result of browsing the transmitted environment map, the viewer builds and extends his/her own environment map, and thus this is a data–producing head–tracking system. These environment maps can then be shared reciprocally between wearers.

1. Introduction

The shared mediated reality system allows users wearing EyeTap reality mediators to share their visual experience. The EyeTap reality mediator is a wearable computing device that enables the user's vision to be computationally processed and modified in real–time [2].

People often scan their environment with head rotations. As a result, a large percentage of EyeTap video can be described by a camera moving about a fixed center of projection in a static scene [2]. Using this simplification and applying the VideoOrbits algorithm to projectively transform the images, images from arbitrary camera orientations can be synthesized by forming a composite of spatially relevant neighbouring images.

By using head–tracking, users can build a visual map of their environment and also browse another's environment by simply looking around. Users continue to build their environment map even as they browse another's environment, and thus, continually add information to be shared with others. This reciprocity allows the creation of visual interactions between users that share each other's reality. In this interaction users see "eye to eye", to create a shared mediated reality.

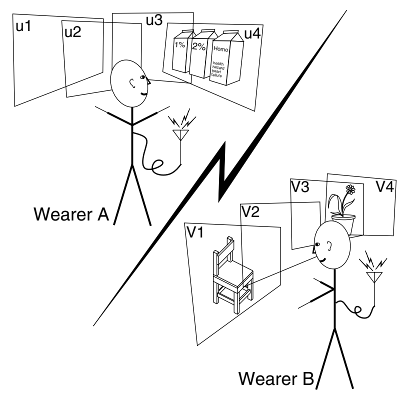

Figure 1: Wearer A, in a grocery store, is building an environment map $(u1, u2, u3, u4)$ that includes an image of milk cartons. Wearer B is browsing wearer A's environment map at home. The images $(u1, u2, u3, u4)$ are displayed in the positions $(v1, v2, v3, v4)$, replacing wearer B's view of the chair and plant. At the same time, wearer B is incidentally constructing an environment map of the home, which would allow wearer A to view the chair and plant from Wearer B's position.

2. Head Tracking

The VideoOrbits/Gyroscopic Head Tracking system (VOGHT) tracks absolute position of a camera in terms of the projective coordinate transformation (PCT) between the current frame of video and a base frame [1]. It also estimates the camera's absolute 3–D rotational position.

In the proposed shared mediated reality system, the PCT based camera position is used for image registration and the creation of composite views. The rotational position is used to orient the camera views in an environment map as well as for browsing another user's environment.

2.1 Reference Frame Database

The Reference Frame Database is a spherically indexed database of reference images of the environment that are continuously collected and updated through normal system

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)
0-7695-1781-1/02 $17.00 © 2002 **IEEE**

Figure 2: Outdoors: the environment map generated by wearer A, and the views of that map synthesized by wearer B.

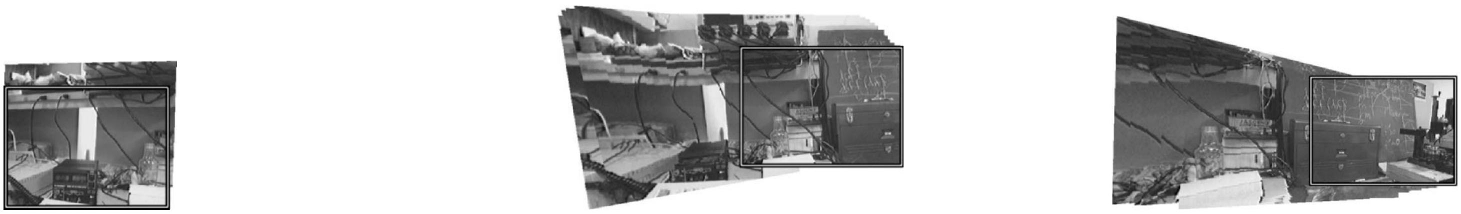

Figure 3: Indoors: the environment map that wearer B generates while browsing wearer A's outdoor map.

use. Each image in the database includes a time stamp, a 3–D rotation matrix describing the camera orientation and a PCT describing its projective position with respect to the base frame. Reference frames are selected for use with VideoOrbits based on the current position estimate generated by the gyroscope.

The sphere of reference images is mapped to a two dimensional matrix, indexed by equal increments of azimuth and elevation. In experimentation, four degree increments were used.

3. Seeing Eye to Eye: a shared mediated reality

Sharing of the mediated reality is accomplished by transferring this reference frame database to wearer B (see Figure 1), who can use a second VOGHT and a browser program to synthesize views of wearer A's environment map from an arbitrary viewing direction. Synthesis of views is performed by taking the current estimate of the rotational position of wearer B's VOGHT and using its equivalent PCT to re–project wearer A's reference images to wearer B's viewpoint.

The images that are projected are selected by the absolute position of the VOGHT. This estimate is used to calculate a region in which to apply the projective coordinate transform. Then using the projected images an image composite is formed. From this composite the correct view is cropped (see Figure 2).

Thus the head tracking of wearer B is used to provide a perspective with which to view the environment map generated by wearer A. As wearer B rotates his/her head, new perspectives are supplied to the browser program, which cause the projection and synthesis of new portions of the environment map. All these projections are relative to wearer B's current head position, thus all navigation is performed solely with head movement.

The size of the region to search is a tunable parameter, a small parameter is useful on a resource constrained wearable computing system. If the search area is large then multiple images from the map can be projectively transformed and registered (see figure 2). This registered image allows for the creation of a complete image for the desired perspective even if there is no image in the map at that exact perspective. However, if there are no images nearby then no view can be generated.

4. Conclusion

The shared mediated reality system allowed users to exchange their current environments through the EyeTap reality mediator. The views are generated and controlled through the users' head motion using the VOGHT. This was demonstrated through the use of the system between two wearable computer users, one outdoors and one indoors. The views generated had the correct perspective and allowed each viewer to see the other's environment as if it were their own. This resulted in the users seeing Eye to Eye in a projectively stabilized shared mediated reality.

References

[1] C. Aimone, A. Marjan, and S. Mann. Eyetap video-based projective motion estimation assisted by gyroscopic tracking. To appear in International Symposium on Wearable Computing, 2002.

[2] S. Mann. *Intelligent Image Processing*. John Wiley and Sons, November 2 2001. ISBN: 0-471-40637-6.

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)

IEEE COMPUTER SOCIETY