JanusLights: A Camera-Projection System for Telematic Omni-Presence with Correct Eye Gaze

Chris Hermans, Maarten Dumont, Philippe Bekaert
Hasselt University - Expertise Centre for Digital Media
Wetenschapspark 2, 3590 Diepenbeek, Belgium
{chris.hermans, maarten.dumont, philippe.bekaert}@uhasselt.be

Eric Joris
CREW
eric.joris@skynet.be

Abstract

We present a camera-projection system for single/multi-party tele-presence which inherently allows for correct eye gaze, and unlike standard videophony, provides a great deal of spatial context. The unique feature of our system is a combination of omnidirectional video capture and display from corresponding projection centers.

1. Introduction

Standard videophony fundamentally suffers from incorrect eye gaze. Because person A can either look at her computer screen (showing person B) or at the camera filming her, person A and B cannot have any eye contact. Correct eye gaze is something peculiar, and even small disturbances will be perceived as interruptive.

Also, because of the spatial nature of the viewing window, only a very limited amount of context information is available (fig.1a).

Fig.1b illustrates our goal. When looking head-on into a spherical mirror, a person looks into her own eyes (correct eye gaze), without losing a clear view of her environment. Exchanging such views between two (groups of) people creates a natural way of communication. This is the core idea behind our system.

Figure 1. (a) Traditional videophony; (b) Our goal: correct eye gaze + large amount of context information.

We try to achieve this goal by combining the capture of omnidirectional video (filming a spherical mirror), with the corresponding display by projection onto a diffuse white spherical screen. Both capture and display can be performed on a single sphere, forming a single communication device.

2. Conceptual Setup

Unfolding the environmental reflection filmed on a specular sphere yields omnidirectional video with a projection center approximately located at the center of the sphere [2]. During video capture, this provides us with the light intensities arriving at the center of the sphere, coming from all directions. As illustrated in fig.2a, light rays coming from direction $\vec{r_2}$ are observed at the intersection of $\vec{o}$ with the camera plane.

We can do a similar calculation on the viewer's side, mapping points on the diffuse sphere to the directions of reflected light. We have two options at our disposal, depending on the application:

1. mapping $\vec{r_2}$ to $\vec{n}$: each point on the half-sphere corresponds to the light coming from that direction.
2. mapping $\vec{r_2}$ to $\vec{r_1}$: the viewer observes the specular reflection of his conversation partner. This result can be seen as a sphere that has its specularities projected onto itself [3].

Because of this mapping, one automatically looks into the camera when watching the display. This accounts for the superior eye-gaze quality of our technique. As long as the direction of the viewer meets the center of projection (which it should, as the viewer looks at the sphere head-on), eye gaze should be correct.

By utilizing omnidirectional video, the full spatial context of a correspondent is unveiled. Compared to the limited information provided by a 2D viewing window, we believe this to be a significant enhancement.

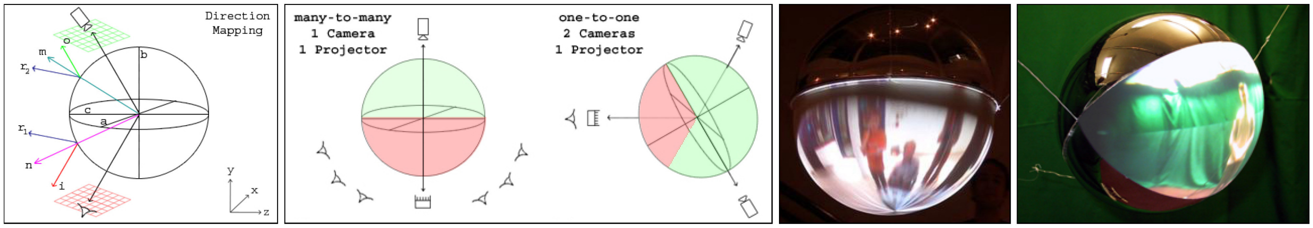

Figure 2. From left to right: (a) the basic vector mapping, (b) conceptual setup for both the many-to-many and the one-to-one scenario, (c)-(d) pictures of their respective prototypes.

As we are filming enviromental reflection, it is necessary to assume an affine camera model (parallel viewing rays). This means that a significant camera-sphere and viewer-sphere distances is required. Their practical dimension depends on the radius of the sphere.

3. Applications

- The combination of a diffuse and a specular half-sphere (fig.2b, left), mapping reflections on the camera side to normals on the projector side, generates sensible images from any point of view around the device. This allows for *multi-party many-to-many* telepresence (fig.2c).
- With a triangular setup (fig.2b, right), we can initiate *one-to-one communication*. The use of 2 cameras results in improved image quality, as our view of the needed intensities is available at a better angle (fig.2d). Both mapping methods can be used.
- JanusLights may be used as *decorative lighting devices*, offering a point of view into (network) linked spaces (e.g. pubs, halls, theatres, events,...).

4. Implementation

For our prototypes, we use of-the-shelf half-spherical mirrors, the kind typically used for shop surveillance. Image quality suffers because these mirrors are not perfectly spherical nor specular. We tried to compensate for these spatial deformations by modeling our 'sphere' as a spheroid. Image quality is also restricted by the resolution of the projector and cameras. Note that these restrictions are hardware related, and do not implicate flaws in our system.

All image manipulations employed are straightforward 2D/3D operations. This means that they can either be precalculated ahead of the rendering step (e.g. most of the mapping code can be calculated as a preprocessing step, and then be stored in a reference table), or they are well suited for implementation on a GPU (e.g. during many-to-many communication, users can easily rotate the view by performing a simple 2D rotation on the captured scene).

5. Results

Results from our many-to-many prototype system are displayed in fig.2c. As can be seen, projection on the edges of the display side is suboptimal. However, this can be resolved by placing the projection device *inside* the sphere [1]. Also noticeable is the vanishing point in the center of the projected image. This artefact occurs because of the poor quality of the information gathered from the edge of the captured sphere. Fortunately, in a typical setup, the sphere is placed over a desk, thus minimizing discomfort.

Our one-to-one setup (fig.2d) does not suffer from these artefacts, at the cost of having to merge results from two different views. The main issue here concerns careful image alignment, though soft blending helps to mask artefacts.

Overall, our system seems to give promising results. Our experiments have shown that the concept works, even though there are still technical difficulties. When pixel resolution was high enough to see the eyes of the correspondent, we have witnessed eye gaze to be accurate.

6. Future Work

In the previous paragraphs, we have assumed an affine camera model. Even though this is acceptable for our proof-of-concept prototype (and is perfectly suitable for the many-to-many case), it will pose problems in a practical setup of our one-to-one scenario. We are therefore looking into the work of Svoboda[4] to solve alignment problems, and getting more accurate results through the use of epipolar geometry.

References

[1] ARC. Omniglobe: A self-contained spherical display system. ACM Siggraph 2003 Emerging Technologies, 2003. 2

[2] P. E. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of SIGGRAPH98*. ACM, 1998. 1

[3] R. Raskar, G. Welch, K. Low, and D. Bandyopadhyay. Shader lamps: Animating real objects with image-based illumination. 1999. 1

[4] T. Svoboda. *Central Panoramic Cameras, Design, Geometry, Egomotion*. PhD dissertation, Czech Technical University, Sept. 1999. 2