

The Two-User Responsive Workbench: Support for Collaboration Through Individual Views of a Shared Space

Maneesh Agrawala*

Andrew C. Beers*
Ian McDowall†

Bernd Fröhlich*
Mark Bolas†

Pat Hanrahan*

*Stanford University
Stanford, CA

†Fakespace, Inc.
Mountain View, CA

Abstract

We present the two-user Responsive Workbench: a projection-based virtual reality system that allows two people to simultaneously view individual stereoscopic image pairs from their own viewpoints. The system tracks the head positions of both users and computes four images - one for each eye of each person. To display the four images as two stereo pairs, we must ensure each image is correctly presented to the appropriate eye. We describe a hardware solution to this display problem as well as registration and calibration procedures. These procedures ensure that when two users point to the same location on a virtual object, their fingers will physically touch. Since the stereo pairs are independent, we have the option of displaying specialized views of the shared virtual environment to each user. We present several scenarios in which specialized views might be useful.

CR Categories and Subject Descriptors: I.3.7 [Computer graphics]: Virtual Reality, I.3.1 [Computer graphics]: Three-Dimensional Displays.

Keywords: Virtual environments, Responsive Workbench.

1 Introduction

Many tasks require people to work together and there is great interest in using technology to improve the effectiveness of group activities. Groups, unlike individuals working alone, communicate and exchange information. For example, a whiteboard provides a single shared drawing surface that facilitates such collaborative interaction. Users can communicate by voice, gestures and by writing on the shared surface. A natural question is whether virtual environments can similarly provide a shared space in which collaborative interactions occur as easily and smoothly as they would in the real world.

Projection-based virtual reality systems such as the CAVE [4] and the Responsive Workbench [8] [7] are close analogues to whiteboards. In such systems, stereoscopic images are projected onto a large screen allowing groups of people to simultaneously view a virtual environment. As with whiteboards, the whole group is focused on the same display screen and can interact with one another in a natural manner. These systems track the head position of a single

Figure 1: Two users simultaneously view a shared virtual environment on the Responsive Workbench. Calibration ensures that when they point to the same feature on the virtual cube, their fingers touch. Note that the image on the Workbench is rendered for the point of view of the camera.

user and compute a stereoscopic image pair for that person's point of view. Unfortunately, when non-tracked users see these images, the virtual environment appears distorted. This distortion makes it difficult for non-tracked users to visually examine and refer to virtual objects.

In this paper, we describe an extension to the Responsive Workbench that allows *two* users to see individual stereoscopic images from their own viewpoint. Our two viewer display system is based on the common single viewer stereo display technique of *frame interleaving*. The system tracks the head positions of both users and computes four images - one for each eye of each person. The four images are displayed in sequence and we ensure each image is correctly presented only to the appropriate eye. Since both users see correct stereo images for their own point of view, they both see an undistorted, perspective correct view of the virtual environment. With proper calibration, when the users point to the same location on a virtual object their fingers will touch, as shown in figure 1.

At times users are interested in different aspects of a shared space. The ability to display two independent views offers the intriguing possibility of presenting different information in each view. As an example, consider a scenario in which an electrician and a plumber are discussing the plans for a new building. They must ensure that there is no interference between the wiring and plumbing. With a two-user Responsive Workbench, the electrician and plumber could see specialized views. Both would see the basic structure of the house, but only the electrician would see a detailed representation of the wiring and only the plumber would see a detailed representa-

(a) (b) (c)

Figure 2: (a) The view seen by a user standing on the left side of the table. (b) The view seen from the right side. (c) The view a user standing on the left side of the table would see if generated from the position of the view on the right. Even though the image on the Workbench is the same in (b) and (c), the cube appears sheared to the non-tracked user in (c).

tion of the water pipes. Regions in which the plumbing and wiring interfere would be shown in detail to both contractors. There are many interesting variations of this idea, and in this paper we discuss scenarios in which these specialized views might be useful.

The main contribution of our work is the development of a system that provides two users with perspective correct views of the same virtual environment on the Responsive Workbench. In sections 3 and 4, we describe the hardware and calibration required to create this shared space. In section 5, we present several scenarios that might benefit from specialized views of a virtual environment. We begin with a discussion of the problems with face-to-face collaboration in current projection-based virtual reality systems.

2 Support for Face-To-Face Collaboration

Face-to-face collaboration is difficult in current projection-based virtual reality systems because they do not provide a visually consistent shared space for all users. Before we analyze this problem in detail, let's consider what face-to-face collaboration entails.

Direct verbal communication is the most common type of interaction amongst a group of people. In a face-to-face situation, facial expressions and hand gestures provide important backchannels which create a sense of awareness and involvement within the group [2]. Other visual cues such as lip motions make it easier to understand what a speaker is saying. When a group performs a task in a shared physical space, gestures are often used to refer to objects. Such gestural interactions are important in establishing a shared context for the group [11]. When someone refers to an object by pointing to it, the object becomes the focus of the group. It is because of these visual cues that people often find it easier to communicate and collaborate when they are face-to-face than across a voice-only medium like a telephone [1]. Thus, face-to-face collaboration requires that participants can see some representation of one another and is best achieved when they share a single, visually consistent environment

In projection-based virtual reality systems, users can see each other and therefore communicate face-to-face. However, current systems do not give the users a single visually consistent environment to discuss. Since the stereo images are perspective-correct only for the point of view of the tracked user, non-tracked users will notice two types of distortions (visual inconsistencies) in the stereo images:

- **Point of View Distortion:** For non-tracked users, virtual objects appear to shear or lean in some direction because they are not viewing the environment from the point of view for which it was rendered.

As shown in figure 2, a non-tracked user would see a distorted

image when standing to the left of a tracked user. In the appendix, we analyze the point of view distortion in more detail.

- **Motion Distortion:** Since the stereoscopic image is computed for the tracked-user's head position, the image changes whenever the tracked user moves. To the non-tracked user who cannot predict the motions of the tracked user, these changes seem haphazard and cause disorientation.

Despite this distortion, it is possible for the users to talk about the general features of the virtual environment. The shear caused by the first distortion requires non-tracked users to stand close to tracked users to get an approximately correct view of the environment. If a tracked user points at a certain point on a virtual object, the non-tracked user will see the tracked user pointing at a different point in the virtual model, even though both see the tracked user pointing to the same point in physical space. For the users this is unintuitive and reduces the feeling of a shared space.

The second type of distortion can be jarring if the tracked user moves quickly. In normal use, tracked users will continuously move their heads by small amounts as they are examining the virtual objects. This motion causes the environment to continually swim for non-tracked users. We can solve both of these distortion problems by giving the users individual, viewpoint dependent, perspective-correct stereo image pairs.

3 Two-Viewer Display Method

The most common display technique for single viewer stereoscopic image displays uses two different frame buffers to store images computed for the left and right eyes. The display hardware alternates scanning out the two buffers at a typical rate of 120Hz, 60 images per eye per second. Shutter glasses are used to ensure that each eye sees only the appropriate image.

In order for two viewers to see individual stereoscopic image pairs, four different images must be rendered and displayed. We extended the single-viewer approach described above to two viewers by using four different frame buffers, one buffer for each eye of each user. As shown in figure 3, our current hardware consists of a Silicon Graphics Onyx2 workstation with four R10000 processors and two Infinite Reality graphics pipelines. Each pipeline generates two of the four images in parallel with the other pipeline. The generated frames are merged using custom hardware that interleaves two genlocked analog video streams. Our frame interleaving hardware supports a variety of resolutions and refresh rates, including a high refresh rate configuration with a screen size of 1280x492 at 144Hz, and a higher resolution configuration with a screen size of 1024x768 at 120Hz. The bandwidth of our current projector prevents us from

using higher refresh rates or larger screen sizes than the ones described above.

We have also modified the design of the single viewer shutter glasses for use with this system. In addition to the two standard one eye open, one eye closed states, we added a third state in which both eyes are closed. This state is required whenever the images for the other viewer are displayed.

The main drawbacks of this two viewer approach are that we cut the display frame rate in half for each user compared to the single viewer approach, and that the images had a slight but noticeable flicker at the frame rates we were able to display. Also, half of the frames seen by each user are black while the views for the other user are shown. These black frames will reduce the perceived brightness of the images seen by the user, since the users see only half the amount of light as they would in a single user system.

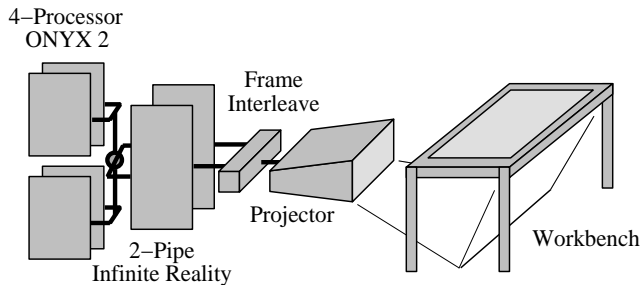


Figure 3: Two View Responsive Workbench System Overview.

The four images required for two users can be displayed in two fundamentally different sequences. Consider the four frame buffers and their corresponding images, which we name as follows: L1—left eye first viewer, R1—right eye first viewer, L2—left eye second viewer, R2—right eye second viewer. The two sequences are then:

- **Viewer Sequential:** ... L1 R1 L2 R2 ... The images for each viewer are displayed one after the other.
- **Viewer Interleaved:** ... L1 L2 R1 R2 ... The images of the two viewers are interleaved.

The viewer interleaved method exposes each viewer to an image at every other frame time, whereas with the viewer sequential method, each viewer receives two consecutive images, and then no image for two frame times. As expected, we found the flicker less perceptible for the viewer interleaved method.

We explored another display technique for a two user system that uses anaglyphic (i.e. red/blue) stereo, and is suitable for a prototyping system. One user wears a red filter over the left eye and a blue filter over the right, and the other does the opposite. Both users additionally wear shutter glasses. The shutter glasses ensure that only one of each user’s eyes can see an image each frame, and the different colored filters ensure that the users see their own view. Note that one could combine this anaglyphic approach with the approach above to provide independent stereo views for four users.

4 Registration and Calibration

The methods of the previous section provide each user with an independent view onto a shared environment. In order for them to interact with this environment, we must calibrate the display and tracking systems so that a user can simply reach out and directly interact with virtual objects. This requires that the coordinate systems for the various devices are properly aligned. Our system uses three different coordinate systems: The user and the table top display exist within the real world, and therefore we can describe their positions within a *physical coordinate system*. Similarly, virtual objects are defined in

a *virtual coordinate system*. The user’s head and hand motions are tracked with a tracking system, which reports coordinates in its own *tracker coordinate system*. Registration connects these three coordinate systems together to provide a common reference frame. We choose the physical coordinate system as our reference system, and affix it to the table top so that its origin is in the middle of the table, z is up and x and y are aligned with the edges of the table.

Accurately registering these three spaces to each other provides each user with a perspective correct image, and ensures that when both users point to the same virtual object, they also point to the same point in physical space above the tabletop. The three coordinate systems are registered in two steps: First, the virtual coordinate system is registered to the physical coordinate system, and then the tracker coordinate system to the physical coordinate system.

4.1 Registering the physical and virtual systems

We begin by aiming and aligning the projector so that the displayed image is centered on the table top and aligned parallel to the edges of the table. We then carefully linearize the display to ensure that virtual objects do not change size and shape when they are moved across the table. These two steps accurately register the virtual image plane with the physical xy-plane. Finally, we define our perspective projection so that the virtual and physical coordinate systems are aligned and use the same physical units. This projection depends on the physical size of the displayed image, which we measure with a ruler.

4.2 Registering the physical and tracking systems

The simplest way of registering the tracker and physical coordinate systems is to determine the transformation mapping the coordinate frame of the tracking system onto the coordinate frame of the physical table. This can be accomplished by measuring the location of three known points on the physical table using the tracking system. The corresponding points define a change-of-basis matrix between the two coordinate frames.

This method works well when there are no distortions in the tracker measurements. However, all tracking devices suffer distortions, and those present in magnetic systems like the Polhemus Fastrak can be quite large [10]. We have developed a look-up table approach for correcting these distortions, which is similar to methods developed by Bryson [3], Ghazisaedy et al. [6], and Livingston [9]. We begin by measuring points on a grid in physical space using a Polhemus 6DOF stylus. For each of these points, we save the corresponding location reported in tracker space. These measurements are samples of a function mapping physical space into tracker space. We compute the inverse of this function on a grid in tracker space, and then use trilinear interpolation to map positions reported by the tracker into our physical coordinate system.

We measured absolute position error in the space above the table surface and in front of the table where users normally stand. Using the above method, our position measurements have an average error of approximately 0.5cm and a maximum error of less than 3cm over the operating space of the Workbench, as compared to an average error of 3cm and a maximum error of 14cm or more without our distortion correction.

Combining the two registration steps yields accurate stereoscopic images for both users, and facilitates the interaction between the users and the virtual objects.

5 Specialized Views of a Shared Environment

In face-to-face collaborations, participants generally assume that the objects in the environment are visible to all the participants. If I see an object in the real world I can safely assume that you see it as well. Most of the examples we have designed for the two-user Responsive Workbench conform to this real world assumption and present exactly the same virtual environment to both users.

Sometimes, however, collaborators are interested in different aspects of a shared environment. One advantage of displaying individual stereo pairs with the two-user Responsive Workbench system is that each user can see a specialized view of the environment. We can independently display and highlight the information that is most important to each of them. This may keep the users from being overwhelmed by extraneous information and help to focus their attention on the most relevant details of the environment.

We have developed three scenarios that make use of such specialized views. The scenarios are described in figures 4–6 and are meant as simple demonstrations of situations in which specialized views might prove useful. The scenarios are based on more general strategies for partitioning information among multiple viewers of a shared environment. The strategies are not mutually exclusive and we expect applications to use different combinations of them.

Layer Partitioning. Many virtual models, simulations, or environments contain distinct layers of information that can be viewed independently or superimposed in combination to be viewed all at once. Typically each layer contains a very specific type of information and when all the layers are viewed at once, the spatial information density is so high that it is impossible to interpret or analyze the data. Moreover, in many cases each user will only be interested in a few layers of the model. Instead of displaying all the layers to every user, we can present individually specialized views of only the layers each user is most interested in. In figure 4, we develop a scenario that uses this layer partitioning strategy to facilitate a construction **Contractors’ Meeting**. The layer partitioning technique does require some knowledge of the kinds of information users will be interested in during the modeling process. The approach is less useful for a user that needs information which is spread across many layers.

Given a model made up of information layers it is also possible to present different users with different representations of the same layer. In our contractors scenario (see figure 4), a bricklayer might see a detailed representation of each individual brick while the roof is represented as a single large slab. Simultaneously a roofer might see individual roof tiles while the walls are represented as slabs. Both users see both the walls and roof information layers, but the level of abstraction in the presentation of the layers is dependent on what each user is most interested in.

Spatial Partitioning. Some virtual environments can be spatially partitioned so that each user only sees a small region of the environment in exhaustive detail. We can deemphasize non-focus regions by presenting them at a lower level of geometric detail than the focus region. This type of partitioning works especially well for large display surfaces like the Responsive Workbench. If users stand close to the display surface, it provides a large field of view. Users tend to focus on subregions of the surface rather than viewing the entire surface. We develop an **Air Traffic** scenario in figure 5 that makes use of spatial partitioning.

Private Information. In some environments, some information is relatively independent of the shared space and of interest only to a single user. Such private information is not considered part of the general virtual environment, but may be very useful to a particular user. In figure 6, we describe a scenario in which a teacher uses private notes while giving an **Anatomy Lesson** to a student. Tang [11] describes meetings in which participants take notes or sketch ideas on a private notepad which they later make public and present to the rest of the group. Such private information spaces could be useful for developing one’s own ideas in a group setting.

Carried to an extreme, specializing views can lead to a situation in which each user sees a completely different environment and there

is no longer a shared space. With two-user technology, it is possible for each user to be engaged in a completely different application. In such cases, having two simultaneous users is probably not very useful. When displaying specialized information, the challenge is to ensure that the notion of a shared space is not lost.

6 Conclusions and Future Work

The two-user Responsive Workbench is a projection-based virtual reality system that supports tight, face-to-face collaborative interaction between two users. We have developed a hardware setup for enabling a two-user frame interleaved display, allowing us to present each user with a perspective-correct view of a shared virtual environment.

We are currently exploring methods for improving the display hardware and examining the feasibility of scaling this approach to support more than two users. We have encountered three limitations in our two-user display hardware: flicker, crosstalk and reduction of brightness. Flicker is caused by our current refresh rate of 144Hz, yielding only 36Hz per eye, per user. Crosstalk occurs because the CRT phosphors do not completely decay between the display of consecutive frames. Next generation video projectors promise higher refresh rates which should reduce flicker, but will necessitate faster phosphors or different technology. Higher refresh rates may also enable three or four user display systems. However, the ratio of time for which each user actually sees an image decreases as support for two or more users is added to the system. Users will experience this as a reduction in the brightness of the display.

As we discussed, specialized views is an interface paradigm that might be useful for reducing information overload or maintaining a boundary between public and private spaces. However, it is still unclear whether real-world applications supporting collaborative work would benefit from specialized views. Currently we are trying to identify such applications. We hope to assess the effectiveness of specialized views in aiding collaborative interactions within these applications.

Giving users the ability to view the same virtual environment while standing near one another allows them to communicate about the environment by voice and gesture. We are just beginning to explore the types of software tools and interfaces that might further assist such collaborative interaction. Cutler et al. [5] have shown that two-handed interaction tools can be useful in the Responsive Workbench environment. In a two user system, there are three or four hands possibly acting in a coordinated fashion. We are investigating how to extend their methods for this situation.

In this paper, we have demonstrated that a frame-interleaved stereo display on a projection-based system is possible for two simultaneous viewers. We expect that several simultaneous viewers will be possible in the near future. We believe that the advent of such technology will provide a great opportunity for developing tools, interfaces and paradigms that strengthen face-to-face collaborative interaction in a shared virtual environment.

7 Acknowledgments

We would like to thank Oliver Riedel of Fraunhofer for help with the initial hardware design. GMD, Germany’s National Research Center for Information Technology, developed part of the basic workbench software that was used for this project. The air traffic data set is courtesy William F. Eddy and Shingo Oue. We would like to thank Larry Cutler and Yeva Fineberg for contributing ideas and for help with modeling objects. Jeff Feldgoise created the house model. The Stanford Instructional Television Network provided technical assistance with capturing video and images of the workbench. This work was supported by Interval Research Corporation.

(a)

(b)

(c)

Figure 4: **The Contractors' Meeting.** Several contractors are responsible for designing and constructing various sections of a house. Each section is a different layer of information in our house model. One layer contains the wall support structure for the house, another contains the roof structure, still another contains the wiring and so on. (a) Two contractors are discussing modifications to a wall that supports part of the roof above. One contractor is responsible for the wall support structure, and the other for the roof. Both contractors see a basic model of the house, plus an additional information layer specific to their responsibilities. Areas where their work may come into conflict are shown to both contractors. (b) The wall contractor sees all walls that require modification in green and part of the roof that will be affected by the modifications. (c) The roofing contractor sees the entire roof and the wall that is being discussed for modification in green.

(a)

(b)

(c)

Figure 5: **The Anatomy Lesson.** (a) A teacher is using a virtual skeleton to teach anatomy. Both the student and teacher see the virtual skeleton, but only the teacher sees custom notes about the lesson. These notes include tags that appear when the teacher points to a bone. Here the teacher asks the student about a bone, indicating which one by pointing to it. (b) The student sees the indicated bone highlighted, but cannot see the notes about the bone. (c) In this case, the teacher privately see his notes about the pelvis.

(a)

(b)

(c)

Figure 6: **Air Traffic.** (a) Two users monitor air traffic over the continental United States using an air traffic visualizer. They divide the responsibility by spatially partitioning the US into eastern and western regions. Under this spatial partitioning, both users see flights that cross over both regions. (b) Only the user responsible for the western half of the country sees flights that stay in the western region. (c) Only the user responsible for the eastern half sees those that stay in the eastern region. Flights that stay within a region leave purple trails while those that cross between regions leave blue trails. Note that (b) and (c) are overhead views.

Note: The images on the Workbench in the three establishing shots on the left were generated for the point of view of the camera.

References

- [1] Sara A. Bly. A use of drawing surfaces in different collaborative settings. In *Proceedings of ACM CSCW'88 Conference on Computer-Supported Cooperative Work*, Synchronous Communication, pages 250–256, 1988.
- [2] Susan E. Brennan. Conversation as direct manipulation: An iconoclastic view. In Brenda Laurel, editor, *The Art of Human Computer Interface Design*, pages 393–404. Addison-Wesley, 1990.
- [3] Steve Bryson. Measurement and calibration of static distortion of position data from 3D trackers. In *Virtual Reality course notes, SIGGRAPH 93*, 1993.
- [4] Carolina Cruz-Neira, Daniel J. Sandin, and Thomas A. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the cave. In James T. Kajiya, editor, *Computer Graphics (SIGGRAPH '93 Proceedings)*, volume 27, pages 135–142, August 1993.
- [5] Lawrence D. Cutler, Bernd Fröhlich, and Pat Hanrahan. Two-handed direct manipulation on the responsive workbench. *1997 Symposium on Interactive 3D Graphics*, 1997.
- [6] Morteza Ghadisaedy, David Adamczyk, Daniel J. Sandin, Robert V. Kenyon, and Thomas A. DeFanti. Ultrasonic calibration of a magnetic tracker in a virtual reality space. In *Proceedings of the IEEE Virtual Reality Annual International Symposium*, pages 179–188. IEEE, 1995.
- [7] Wolfgang Krüger, Christina-A. Bohn, Bernd Fröhlich, Heinrich Schüth, Wolfgang Strauss, and Gerold Wesche. The responsive workbench: A virtual work environment. *IEEE Computer*, pages 42–48, July 1995.
- [8] Wolfgang Krüger and Bernd Fröhlich. The responsive workbench. *IEEE Computer Graphics and Applications*, pages 12–15, May 1994.
- [9] Mark A. Livingston and Andrei State. Magnetic tracker calibration for improved augmented reality registration. *To appear in Presence*, October 1998.
- [10] Kenneth Meyer, Hugh L. Applewhite, and Frank A. Biocca. A survey of position trackers. *Presence*, pages 173–198, 1992.
- [11] J. C. Tang. *Listing, drawing and gesturing in design: A study of the use of shared workspaces by design teams*. Phd thesis, Mechanical Engineering, Stanford University, California, April 1989. Also available as research report SSL-89-3, Xerox Palo Alto Research Center, Palo Alto, California.

Appendix: The Point of View Distortion

The Responsive Workbench system displays images on the tabletop, which is therefore the image plane in our virtual world. Each image is a projection of the virtual environment onto this plane. Tracking the user's head position ensures that a virtual object always keeps the same shape and position independent of the user's viewpoint. However, to a non-tracked user viewing a stereoscopic image generated for a tracked user, the virtual world appears distorted. The perceived shape and position of virtual objects is dependent on the positions of the tracked and non-tracked users. The distortions can be large and if they are extreme it may not even be possible for the non-tracked user to fuse the stereoscopic images.

In Figure 7a, we analyze the distortion for a simple 2D situation and find that it is indeed a shear. The general 3D case as shown in Figure 7b is more complicated. Two images presented to

the non-tracked viewer generally cannot be interpreted as the projection of a single 3D scene. For the Responsive Workbench, we have found that users almost always manage to fuse two such non-corresponding images and create a sensation of depth. If the tracked and the non-tracked users are standing next to each other and are about the same height, the 3D case reduces approximately to the simple 2D case and the untracked viewer experiences a sheared virtual world.

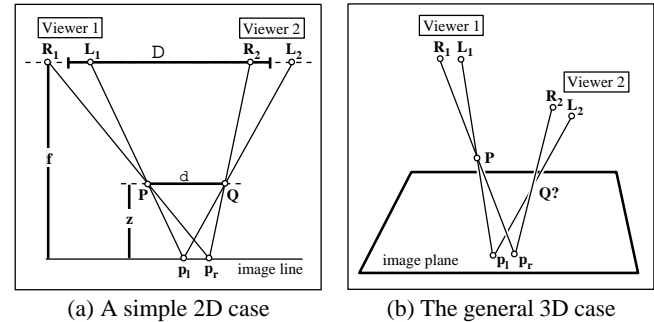


Figure 7: The shearing distortion. Part (a) shows the distortion for a 2D case. Here a 2D world projects onto an image line instead of an image plane. We analyze the special case, in which the eye points of two viewers are on a line parallel to the image line, but the results for the general 2D case are similar. L_1 , R_1 , L_2 and R_2 denote the left and right eye positions for viewer 1 and 2 and we assume the interocular distance for both viewers is the same. Point P as seen by viewer 1 creates the projections p_1 and p_r on the image line. Viewer 2 reconstructs point Q from looking at these projections. If z and f are the distance of P and the eyes from the image line, a distance of Δ in head positions results in a shift of δ in the perceived position of P . Using similar triangles we see that δ is related to Δ by the following simple formula: $\delta = \frac{\Delta}{f} z$, which is a shearing transformation. Part (b) shows the 3D case. In general the lines connecting the second viewer's eye points with the projection of point P do not intersect. Therefore no corresponding point Q exists.