IPT-EGVE Symposium (2007)
B. Fröhlich, R. Blach, and R. van Liere (Editors)

Spinnstube®: A Seated Augmented Reality Display System

J. Wind[1] and K. Riege[1] and M. Bogen[1]

[1]Fraunhofer Institute for Intelligent Analysis and Information Systems, Germany

Abstract

We describe the Spinnstube®, a new projection-based Augmented Reality display system for collaborative work environments. By simply attaching the system to a desk, real objects are augmented with virtual objects by using a half-silvered mirror. The display system is designed in a way that its components do not interfere with a table. So a user can freely move his/her hands within the interaction space. A user is able to interact with the system by grabbing and manipulating the real parts under the half-silvered mirror or using task specific input devices to control an application. By arranging up to 4 of these systems around a table, a collaborative work environment can be created.

Categories and Subject Descriptors (according to ACM CCS): H.5.1 [Multimedia Information Systems]: Artificial, Augmented and Virtual Realities I.3.1 [Hardware Architecture]: Three-dimensional Displays I.3.7 [Computer Graphics]: Virtual Reality I.5.5 [Pattern Recognition]: Interactive Systems

1. Introduction

In different phases of a design or production process, small teams often work together sitting around a table, using it as a platform for sharing ideas and material. This cooperation and communication method can also benefit from using Augmented Reality (AR) technology by being provided with a transparent way of accessing and visualizing complex datasets directly within a workspace.

We developed the Spinnstube®, a new projection-based Augmented Reality display system for collaborative work environments that can be attached to a normal desktop. A user on a chair in one of the Spinnstube® seats sees real objects on the table through a semi-transparent mirror in front of him/her as well as virtual objects projected on a rear projection screen above his/her head.

To allow free hand movements within an interaction space, the display system is designed in a way that its components do not interfere with the table. This is achieved by putting the whole construction above and behind a user. This allows direct-hand interaction with objects on top of the table and keeps the desktop free, ensuring its usability as a natural collaboration space. Users are able to work with objects located in front of them as well as with objects located in front of other users at the same table. The Spinnstube® thus provides small groups a platform to act and interact as a team.

To guarantee maximum flexibility, the system is constructed freely movable and adjustable to a work environment and to a user. The see-through mirror acts as a window that combines real objects with virtual information. The half-silvered mirror is moveable, allowing a user to align the augmented area with the region of interest. To track the position and orientation of the window and a user's head, infrared tracking is used whereas a user's interaction space is observed by an optical tracking system. Originally developed within the ARiSE project [BWG06] [ARi07] as an AR system for school environments, the Spinnstube® is a low-cost solution; so only conventional hardware is used. Recently used as an e-Learning platform in a summer school, this display system offers a lot of potential for other application areas too.

In this paper we present the design of the Spinnstube®. The next section discusses related work and compares the system with similar systems. In section 3 we describe the display setup, its construction and calibration. The tracking capabilities used to recognize a user's point of view, the motion of the see-through mirror as well as the augmented objects are described in section 4. On interaction and collaboration we elaborate in section 5.

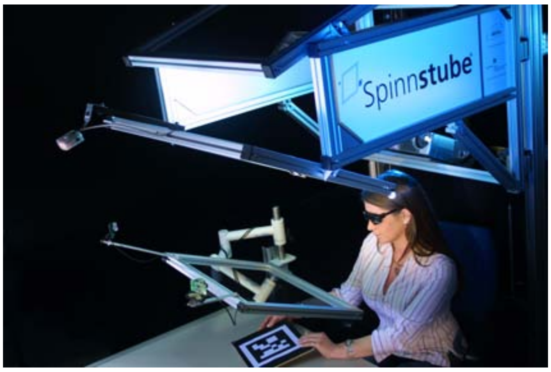

Figure 1: *A user sitting on an office chair using the Spinnstube® for reviewing CAD models. The virtual models are visualized on the board in the user's hands.*

2. Related Work

There is a lot of technology available for realizing Augmented Reality applications. Bimber and Raskar [BR06] made a categorization of Augmented Reality displays from retinal to optical see-trough displays. Because we wanted to build up an extension of a table-like work environment, a stationary display was chosen appropriate. To also keep the calibration effort for new users low, we found a projection-based approach according to the group of spatial optical see-through displays to be the most suitable. Therefore a short overview of the most relevant displays in this category is given below.

The Virtual Showcase [BFSE01] is a projection-based platform that enables up to 4 users to examine a real object placed inside a frustum-like showcase that consists of half-silvered mirrors. Monitors placed before users display a stereoscopic image which is reflected by the semireflective surfaces, so that a user can see a combination of a real object enhanced by virtual information. As the name already suggests, the Virtual Showcase was mainly designed for exhibition halls and museums. As the real object is placed inside the showcase no direct-hand interaction is supported. The same applies to the Extended Virtual Table [BEB01] where only interaction with the virtual part is possible.

An approach of a VR display supporting collaboration of two users seated at a desk was presented by Arthur et al. [APT*98]. The so-called Protein Interaction Theater is based on an upright L-shape combination of two rear projection screens attached to a desk. Each user is located in front of one screen. A shared interaction space is created in the area where both visualisations overlap each other.

The main idea of the Swedish Reachin display [Rea07] is to introduce a possibility to interact with a virtual scenario by hand movement directly underneath a mirror on which a user is looking. So a user interacts with objects he/she can see in a more natural and direct way. Because the whole construction is located on the desktop in front of a user, the space left for user interaction is limited.

With their Personal Space Station (PSS), Mulder and van Liere [MvL02] presented an approach for a near-field Virtual Reality display located mainly above a user to keep the desktop free for direct-hand interaction. In contrast to the Reachin display a user's head is tracked to present a perspectively correct view onto the virtual scenario. The construction that holds the projection system is built like a cuboid with two open sides; therefore collaboration while arranging a couple of these displays around a table is not supported as it is difficult to see the other users. In addition, within an AR system based on this approach, the space left for placing objects in the augmentable area would be restricted too, which makes it hard to work with bigger objects where only a part of should be augmented on.

3. Display Setup

We adapted the idea of the Reachin display [Rea07] and the PSS [MvL02], which enables a user to physically reach under the mirror to directly interact with real and virtual objects. To enable a user to see the virtual as well as real objects we replaced the conventional mirror with a half-silvered mirror consisting of a glass plate with reflecting foil on the front side similar to the Virtual Showcase [BFSE01]. The system is constructed in a way that keeps the table free for user interaction (see Figure 2). In addition, the users should also be able to interact in a social environment for guidance and collaboration. Therefore the construction must not enclose the user or cut down his/her actions by blocking his/her physical reach or his/her ability to look around and see other participants. In the next two sections we will describe the technical construction and the calibration of the display setup we developed.

3.1. Construction

The construction is made of standard aluminium profiles to guarantee maximum stability and robustness. Castors mounted on the chassis make the system easily movable. The display system has three main components: the visualization unit (described in this section) and two tracking units described in detail within sections 4.1 and 4.2.

The visualization unit of the Spinnstube® consists of an active stereo capable video projector that is projecting the stereo image onto a rear projection screen above a user's head. As conventional video projectors can not create a sharp image on short distances without special (and expensive) lenses, a first surface reflector is used to extend the way of light while keeping the construction compact.

The relative position of the projector, first surface mirror, and rear projection screen is computed in a way that the

Figure 2: *With its compact and mobile construction the Spinnstube® can be attached to a desk. The augmented area is created on top of a table.*

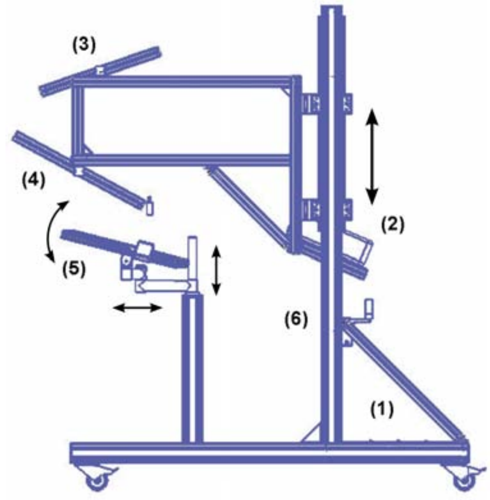

Figure 3: *Side view of the display system: (1) space for computer, (2) stereo video projector, (3) first surface mirror, (4) rear projection screen, (5) half-silvered mirror, (6) hydraulic actuator. Arrows symbolise moving directions of adjustable parts.*

image on the rear projection screen has no keystone distortion and is fixed during normal operation of the system. The height of the visualization unit is hydraulically adjustable to a user's needs and to the physical properties of the work environment the system is attached to (see Figure 3).

A half-silvered mirror through which the user is looking at the desktop reflects the image projected on the rear projection screen, creating the impression that the stereo image is located in front of him, behind the half-silvered mirror. This mirror is fixed on a jointed arm on the chassis of the system which allows a user to freely position it. Additionally a user can tilt the mirror. This way the reflection of the stereo image can be adapted to a user's region of interest.

The position of the half-silvered mirror is tracked and this information is then used to update the display setup in realtime.

3.2. Calibration

As described in the previous section, a user sees the reflected image of the rear projection screen in the half-silvered mirror in front of him/her. If the positions of a user's head and of the mirrored image of the rear projection screen are known, we can calibrate the system with the simplified assumption of using the reflected image of the screen as a real stereo screen. A schematic overview of the transformations involved can be found in Figure 4.

To compute the mirrored screen position, we have to compute the world coordinates of the half-silvered mirror (M_{World}) first:

$$M_{World} = O_M^{-1} I O_{Webcam}^{-1} W^{-1}$$

where O_M is the position of the infrared camera relative to the half-silvered mirror, I is the position of the infrared camera relative to the infrared pattern, O_{Webcam} is the position of the webcam relative to the infrared pattern and W is the position of the webcam relative to the calibration pattern.

Next we compute the center of the rear projection screen in the coordinate system of the half-silvered mirror (S_M):

$$S_M = O_S (O_M I)^{-1} = O_S I^{-1} O_M^{-1}$$

where O_S is the position of the rear projection screen relative to the infrared pattern. It is important to notice that for the computation of the reflection only the plane in which the half-silvered mirror lies is important, so that the infrared camera offset O_M can be relative to any point on the half-silvered mirror.

We can now reflect the screen transformation S_M along the Up-Axis of our mirror. The graphics API we are using has a right-handed coordinate system with Z facing up, so that the mirroring becomes

$$S'_M = S_M \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

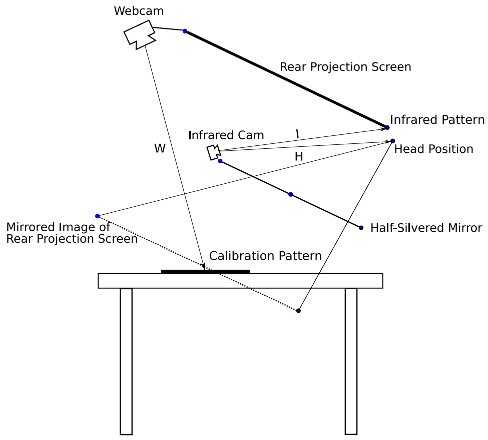

Figure 4: *Schematic overview of the transformations used for calibration.*

With this we can now compute the reflection of the rear projection screen in world coordinates:

$$S'_{World} = S'_M M_{World}$$

The position of a user's head in world coordinates can be computed as follows:

$$H_{World} = H O^{-1}_{Webcam} W^{-1}$$

where H is the position of the head relative to the infrared pattern.

With this information, having the reflection of the rear projection screen and the position of the head in the same coordinate system, it is easy to compute the display setup for a standard graphics library.

Calibration matrices W and I are constantly measured, so that the display setup can be recalibrated in realtime and anytime.

4. Tracking

In the following subsections we describe the tracking systems used and their integration into the display setup.

4.1. Tracking of Head and Half-Silvered Mirror

Head tracking is used to recognize from which point of view a user is looking onto a scene. This input is required to recalculate the visualization of virtual objects depending on a user's perspective. Mirror tracking is used to observe any movement of the half-silvered mirror to adapt the augmentation to the region a user is actually seeing. Combining both, the system is able to augment on the actually visible area and to render the virtual parts from the perspectively correct view of the user.

To keep the hardware equipment as simple as possible we decided to implement a marker-based tracking system which does both simultaneously. For this we used an approach based on simple dot patterns as described by Mulder et al. [MJvR03]. Instead of using FireWire cameras two low-cost infrared cameras from NaturalPoint [Nat07] were attached to the system. This reduces the effort for recognizing markers to a simple blob detection in a black-white image. A related approach was, amongst others, introduced by Ribo et al. [RPF01].

To calculate the position and orientation (6 DOF) of the half-silvered mirror, a small passive markerboard is attached to the frame of the rear projection screen. This markerboard consists of retroreflective circular labels arranged in a regular $M \times N$ grid pattern. Both cameras are trying to detect this infrared pattern by first identifying M blobs along N lines and secondly verifying their linear dependencies. Because distance and orientation of both IR cameras are fixed with respect to each other the 3D position and orientation of the $M \times N$ pattern relative to the cameras can be calculated. This information is used for recalibrating the system as described in section 3.2 and is done once on application start and each time the mirror is moved by a user during runtime. Each movement of the mirror that changes the visible area results in a recalculation of the visualization. This offers the ability to use the half-silvered mirror as an interactive window to the augmentable area, which is adapted at runtime.

To determine the point of view, one additional retroreflective marker attached to a user's glasses is tracked by both cameras, too. By knowing the location of each camera with respect to the marker's location and the relative transformation from each camera to the surface of the half-silvered mirror, which is also fixed, we are able to determine the position of the stereo glasses relative to the reflected image of the screen. We make the simplified assumption that a user is looking at the center of the reflected stereoscopic image. Therefore the use of just one retroreflective dot is sufficient to recognize the position of a user's head (3 DOF). The arrangement of both IR cameras was chosen in a way that the user's head is tracked within the area limited by the width of the half-silvered mirror.

Figure 5 shows the relation between a camera's location and the marker positions. Both markers (the $M \times N$ markerboard and the single marker) are seen by both cameras. So they can be used for head tracking as well as tracking the mirror's position at the same time. To extend the camera's field of view without using wide-angle lenses (to avoid capturing strong distorted images) they are attached to the frame by using an arm.

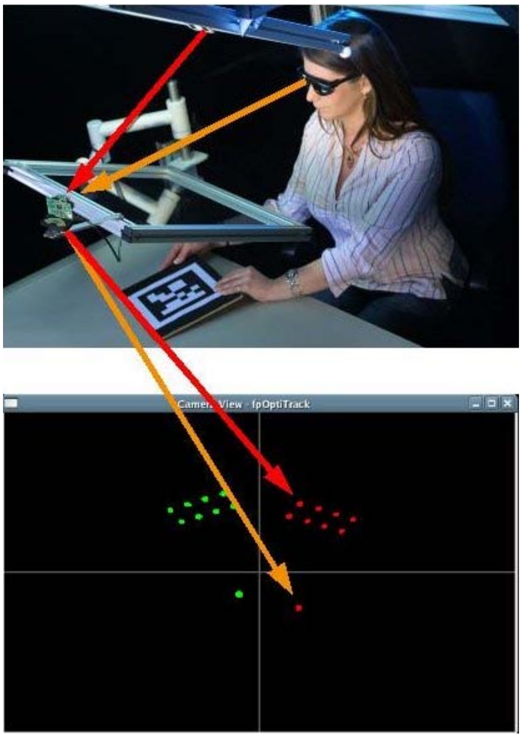

Figure 5: *The upper image shows a user reviewing a CAD model within the Spinnstube®. The model is displayed on the location where the paper marker is detected. Below the captured images of both cameras are shown (red: right IR camera, green: left IR camera).*

4.2. Object Tracking

Object tracking is used to track positions of objects and interaction devices in the interaction area of the display. This area is defined as the space where real and virtual objects are colocated. A FireWire webcam is fixed to the frame of the rear projection screen, looking down at the area in front of and half beneath the half-silvered mirror. Depending on the relative height of the visualization unit the trackable area on the desktop is approximately $80cm \times 60cm$. The images captured by the camera are used to compute the positions and orientations of multiple fiducial id-encoded paper markers seen by the camera with respect to its location, based on the approach introduced by Kato and Billinghurst [KB99].

Due to the limited resolution of the webcam and the comparably small representation of the markers in the captured image, errors are introduced into the pose estimation which results in jittering. This problem is well known in the field of computer vision; a standard approach is to use a Kalman filter to smooth the measurement results [WB01]. At the moment we use a simple approach which filters on the estimated position and orientation directly. This reduces the jittering considerably, with the side effect of introducing latency into the pose estimation. It is expected that a better approach would be to use the Kalman filter in the image recognition process, which would reduce the jittering even more while keeping latency low. We are currently evaluating this approach against others for object tracking improvement (see section 7) .

5. Interaction

The tracking techniques explained in the sections above lead to interaction metaphors which we desribe in the two following subsections.

5.1. Viewpoint Control

By changing the position of the half-silvered mirror, a user can control the location of the augmentation area. This way the display system can be used to augment on objects that are much bigger than the augmentable space would be with a fixed mirror setup. As described in the previous section, the system adapts the calibration of the display setup in such a way that the perspective projection for a user is always correct (see for example [BKJP05a]).

5.2. Direct Manipulation

The object tracking described in section 4.2 is used for direct object manipulation [BKJP05b], which gives a user the possibility to arrange objects freely within the augmentable space. It is also used for application control like menu selection or pointing interaction techniques.

6. Collaboration

As described in section 1, support for user collaboration was a major design goal for the Spinnstube® display setup. Users sitting at the same table have a very natural way of collaborating: they see each other, talk to each other and hand over things like sheets of paper, pens or other objects. The way the Spinnstube® is built allows this kind of interactions by freeing the desktop of construction parts that might block the natural movement of the hands or the direct line of sight to other users. Thus multiple devices can be arranged around a table, supporting these features.

6.1. Sharing Objects

For collaboration, users have to be able to talk and reason about the same objects. By giving a user the possiblity to choose the augmentable area, the Spinnstube® enables him/her to share his/her augmentable space with other users sitting in their own display devices at the same table. This means that in this area a real object can be augmented from several users' perspectives. To support this in the display software, multiple Spinnstube® seats can be calibrated with the same calibration pattern (see section 3.2), creating a common world coordinate system for colocated users.

Another way for the users to collaborate is to pass objects they are working with around, letting other users work with them.

6.2. Specialized Views

Users in a group might have different skills or interests. The possibility to share physical objects can also be used to let different users have different augmentations on them, offering them the possibility of specialized views on the object (as described for colocated multi-user VR displays in [ABM*97]).

7. Conclusion and Future Work

In this paper we described the Spinnstube®, a projection based augmented reality system for collaborative work environments. We have developed a display setup that is scalable, allowing a group of users to augment on a shared collaboration space. Furthermore the display is robust and adaptable to a user's needs. Being constructed with the help of commodity components, it is also comparably low-cost.

First tests with user groups consisting of pupils of different ages have shown that the students fully concentrate on the application and are not aware of the technical setup after a short time. Our assumptions regarding interaction and collaboration were verified. Detailed user studies in the near future shall help us to shed more light on this.

The resolution of the infrared cameras used to track the users head and the half-silvered mirror is restricted to 355 x 288 pixels. This is comparably low, but meets our requirements as the area that has to be tracked is limited, too.

The Spinnstube® has been built with standard aluminium profiles to make it robust and easy to customize. After this proof of concept, the arrangement of the elements of the visualization unit (video projector, first surface mirror and rear projection screen) is known. Keeping these characteristics, further iterations of the Spinnstube® can have different designs or can be made from other materials.

We are currently exploring methods for improving the accurary of the webcam-based tracking system. One of them is the possiblity of using a second webcam, effectively doubling the camera resolution in one dimension. Other methods are improvements in the tracking algorithm and better lighting conditions. One of the most promising approaches is the combination of the optical tracking system with an inertia tracking system [YNA99]. This would also help with another issue of the marker-based optical tracking: obstruction of the markers.

The video projectors we use at the moment are active stereo capable DLP-projectors with a resolution of 800x600 pixels. We will replace them with projectors providing higher resolutions as soon as they are available.

8. Acknowledgements

The ARiSE project is funded by the European Commission as part of the IST programme within the 6th framework (contact number IST-027039).

References

[ABM*97] AGRAWALA M., BEERS A. C., MCDOWALL I., FRÖHLICH B., BOLAS M., HANRAHAN P.: The two-user Responsive Workbench: support for collaboration through individual views of a shared space. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1997), ACM Press/Addison-Wesley Publishing Co., pp. 327–332.

[APT*98] ARTHUR K., PRESTON T., TAYLOR R., BROOKS F., WHITTON M., WRIGHT W.: *Designing and Building the PIT: a Head-Tracked Stereo Workspace for Two Users*. Tech. rep., Chapel Hill, NC, USA, 1998.

[ARi07] ARiSE: Augmented Reality in School Environments. http://www.arise-project.org, last access 15 May 2007.

[BEB01] BIMBER O., ENCARNAÇÃO L. M., BRANCO P.: The Extended Virtual Table: An Optical Extension for Table-Like Projection Systems. *Presence 10*, 6 (2001), 613–631.

[BFSE01] BIMBER O., FRÖHLICH B., SCHMALSTIEG D., ENCARNAÇÃO L. M.: The Virtual Showcase. *IEEE Comput. Graph. Appl. 21*, 6 (2001), 48–55.

[BKJP05a] BOWMAN D. A., KRUIJFF E., JR. J. J. L., POUPYREV I.: *3D User Interfaces: Theory and Practice*. Addison-Wesley, 2005. pp. 217–219.

[BKJP05b] BOWMAN D. A., KRUIJFF E., JR. J. J. L., POUPYREV I.: *3D User Interfaces: Theory and Practice*. Addison-Wesley, 2005. pp. 158–159.

[BR06] BIMBER O., RASKAR R.: Modern approaches to augmented reality. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Courses* (New York, NY, USA, 2006), ACM Press, p. 1.

[BWG06] BOGEN M., WIND J., GIULIANO A.: ARiSE - Augmented Reality in School Environments. In *Innovative Approaches for Learning and Knowledge Sharing, First European Conference on Technology Enhanced Learning, EC-TEL 2006, Crete, Greece, October 1-4, 2006, Proceedings* (2006), pp. 709–714.

[KB99] KATO H., BILLINGHURST M.: Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing System. In *IWAR '99: Proceedings of the 2nd IEEE and ACM International Workshop on*

Augmented Reality (Washington, DC, USA, 1999), IEEE Computer Society, p. 85.

[MJvR03] MULDER J. D., JANSEN J., VAN RHIJN A.: An affordable optical head tracking system for desktop VR/AR systems. In *EGVE '03: Proceedings of the workshop on Virtual environments 2003* (New York, NY, USA, 2003), ACM Press, pp. 215–223.

[MvL02] MULDER J. D., VAN LIERE R.: The personal space station: Bringing interaction within reach. In *Proceedings of the Virtual Reality International Conference, VRIC 2002* (2002), pp. 73–81.

[Nat07] OptiTrack Flex3: by NaturalPoint Inc. http://www.naturalpoint.com, last access 15 May 2007.

[Rea07] The Reachin Display: by Reachin Technology AB. http://reachin.se, last access 15 May 2007.

[RPF01] RIBO M., PINZ A., FUHRMANN A.: A new Optical Tracking System for Virtual and Augmented Reality Applications. In *Proc. of IEEE Instrumentation and Measurement Technology Conference, IMTC* (2001).

[WB01] WELSH G., BISHOP G.: *Course 8 - An Introduction to Kalman Filter*. Tech. rep., SIGGRAPH 2001 Courses, 2001.

[YNA99] YOU S., NEUMANN U., AZUMA R.: Hybrid Inertial and Vision Tracking for Augmented Reality Registration. *vr 00* (1999), 260.