Jedi Training: Playful Evaluation of Head-Mounted Augmented Reality Display Systems

Christopher S. Özbek, Björn Giesler and Rüdiger Dillmann

Institute for Computer Design and Fault Tolerance (IRF)
Chair for Industrial Applications of Informatics and Microsystems (IAIM)
Universität Karlsruhe (TH), D-76128 Karlsruhe, Germany

ABSTRACT

A fundamental decision in building augmented reality (AR) systems is how to accomplish the combining of the real and virtual worlds.[1] Nowadays this key-question boils down to the two alternatives video-see-through (VST) vs. optical-see-through (OST). Both systems have advantages and disadvantages in areas like production-simplicity, resolution, flexibility in composition strategies, field of view etc. To provide additional decision criteria for high dexterity, accuracy tasks and subjective user-acceptance a gaming environment was programmed that allowed good evaluation of hand-eye coordination, and that was inspired by the Star Wars movies. During an experimentation session with more than thirty participants a preference for optical-see-through glasses in conjunction with infra-red-tracking was found. Especially the high-computational demand for video-capture, processing and the resulting drop in frame rate emerged as a key-weakness of the VST-system.

Keywords: augmented reality, video see through, optical see through, laser training, training remote, Sony Glasstron, Trivisio ARVision, ARtoolkit, NDI Polaris, end-user evaluation

1. INTRODUCTION

Augmented reality (AR) is the concept of altering the user's perception of the real world with virtual elements, such as annotations, instructions, 3D models, images and others. One of the most popular ways to achieve this goal is to use head-mounted stereoscopic displays. These can be within one of two categories.

'Optical see-through' (OST) systems offer a direct view of the environment with the simulation elements projected into the view using a semi-transparent mirror. OST systems use the eye's optical system, giving the user a natural view of the environment, but suffer from a trade-off between see-through quality and contrast of the simulated elements, which appear semi-transparent to the user. OST systems also have to be calibrated before each use, are difficult to construct, and few are available on the market so far. Especially occlusion is hard to achieve since the material has to be able to opacify selected pixels.[2] Since they always retain the view of the real world OST systems are especially appealing for critical environments like for instance medicine application INPRES,[3] where a disconnected cable and a black screen might be catastrophic.

'Video see-through' (VST) systems use traditional, closed head mounted displays (HMDs) together with stereo cameras; the augmentation is combined with the camera images and then projected into the display. VST systems typically need to be calibrated only once, and the optical impression is usually much better than with OST systems. They are also readily available or can be constructed without much effort from TV-replacement glasses with added cameras. On the other hand, the user sees the world through the cameras' optical system, which is sufficiently different from the eyes' one to hamper the user's 3D perception. Since VST systems already capture the video-image as seen by the user it also is available for other purposes. Most notably are registration as the method to bring real and virtual world into alignment and the tracking of objects.

Examples for high dexterity and accuracy tasks are typing on a keyboard, assembly-line work, most sports like table tennis and video-games. All of these require swift, semi-automatic moves of the hands to precise locations in 3D-space. The main focus of these tasks is performance rather than information. An augmented tourist system might serve as an example for the reverse setting, where the main focus is to improved information content and not to uphold movement speed. Evaluation of the two system alternatives in production settings with high dependence on performance is therefore assumed as important.

Figure 1. A participant engaged in the experiment using the video-system. The picture has been modified by adding the laser saber and an incoming laser ray to imitate what the participant saw at the very moment. The screen content visible on the left and right monitors were projected to his left and right eye respectively.

2. METHODOLOGY

The experiment was targeted to provide insight into differences between optical-see-through and video-see-through systems regarding the following issues:

Performance How large are the performance differences that occur with high dexterity and accuracy tasks? This is the main focus of the research and essential for applications of AR in production settings that rely on fast and accurate movements.

3D-feeling Are both systems able to induce depth of focus in a sufficient manner? This is essential to operate most moving related tasks in our three-dimensional world.

Tracking Which advantages can be attributed to different tracking mechanisms and how much influence does this have on the performance of the participants? Since VST systems provide a video-image that can be used to perform tracking other mechanisms would be made redundant, thereby significantly reducing the required hardware setup. What limitations are introduced by this scheme?

Comfort Do the systems differ in perceived comfort when performing the task? Production settings usually require long durations for the equipment to be worn, which might not be possible if comfort is not sufficiently ensured.

Glasses How are people affected who are wearing glasses or contact lenses? Some HMD-glasses require very close attachment to the head which might not be possible for people wearing glasses. Another limitation in this concern can be introduced by the head-mounting strap.

Image quality and real/virtual merger Which system is superior when it comes to presenting images from the real and virtual world? How seamless is the merger of both images?

Component	Video system (VST)	Optical system (OST)
Glasses	Trivisio ARVision-3D HMD	Sony Glasstron PLM-700E
Tracking	AR-Toolkit	NDI Polaris
Rendering	Dual window renderer	Shutter stereo

Table 1. Overview over the main differences between the two system which are outlined in the text.

Acceptance How suitable are the systems when removed from the experimental setting? This question tries to move the focus to the non-technical aspects like social or cultural acceptance which might prevent successful introduction into work and daily-life.

Preexisting knowledge Are there certain skills like sports or computer games experience that enhance the performance with the systems? The motivation for this issue is the prospect of certain training areas which yield improvements in handling and efficiency using the systems.

In table 1 the main characteristics of the two systems are summarized. For more detailed information see the next two sections.

2.1. Video-See-Through (VST)

The video approach used the glasses 'ARVision-3D HMD' by Trivisio. The glasses feature two 6 mm interlaced $640 * 480$-pixel video cameras and two $800 * 600$-pixel LCD-screens. Video images are captured at 60 Hz, augmented by the AR-application and fed back to the glasses using a dual-head Radeon 9500 graphics-card by ATI. Only the setup with this video-card and the drivers from SchneiderDigital[4] was able to provide the required 3D-accelerated mode. Attempts with a Matrox G450 and *Xinerama*-extension were not successful. The glasses are powered by two separate battery packs which also relay video-signals from the PC to the glasses. The convergence between the cameras can be adjusted using a little knob.

Video capturing was performed using two Pinnacle 'PCTV Pro'-frame grabbers. Images were accessed using the Video4Linux API.[5] Since video capturing proved to be one of the main bottlenecks to achieve acceptable frame rates the suggested capture-algorithm for video-frames was modified (alg. 1) to achieve 50 percent higher fps at a possible cost of greater lag.

```
1: /⋆ Setup ⋆/
2: VIDIOCMCAPTURE(0)
3: while true do
4:    VIDIOCMCAPTURE(1)
5:    VIDIOCSYNC(0)
6:    /⋆ Process Frame 0 while frame 1 is captured ⋆/
7:    VIDIOCMCAPTURE(0)
8:    VIDIOCSYNC(1)
9:    /⋆ Process Frame 1 while frame 0 is captured ⋆/
10: end while
```

```
1: /⋆ Setup ⋆/
2: VIDIOCMCAPTURE(0)
3: VIDIOCMCAPTURE(1)
4: i ← 0
5: while true do
6:    VIDIOCSYNC(i)
7:    memBuffer ← videoBuffer(i)
8:    VIDIOCMCAPTURE(i)
9:    i ← 1 - i
10:   /⋆ Process data from memBuffer while capturing
      frame 0 and 1 ⋆/
11: end while
```

Algorithm 1: The suggested algorithm from the API-documentation[5] on the left and the accelerated version to the right.

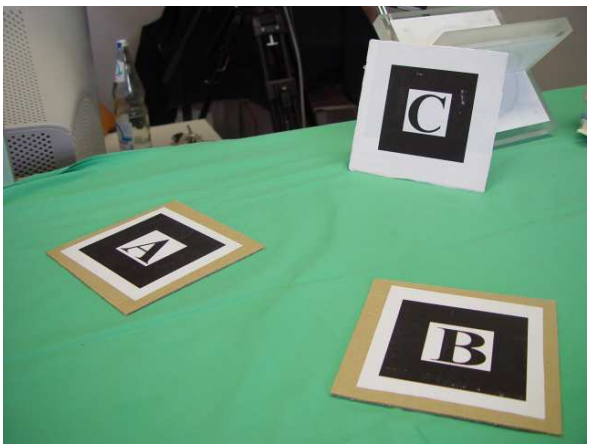
The camera-based ARtoolkit[6] was used for tracking head- and input-device-position. The ARtoolkit software library recognizes markers in camera images and calculates their three dimensional translation and rotation in world-space. Figure 2(a) shows three sample markers that the system was trained to recognize. An improved calibration for the ARtoolkit was used,[7] that comes into effect when several markers constitute the same object. By least-square correlating the estimates for each individual marker a better result for the main object can be achieved.

A plug-in* was developed that renders pictures on the two LCD-screens. Two windows are used to take up the two GL-contexts that represent the scene from the position of each eye. Those two windows are moved to cover the left and the right monitor respectively makes the output available to the left and right eye of the participant. A main hurdle for the implementation of this mechanism was the separation of GL-contexts and caching strategy using OpenInventor, which was used as high level graphic API. Aside from saving and restoring the Inventor-stacks, the function-calls to remember in this concern are `SoGLRenderAction::setCacheContext(int)` and `glutSetWindow`.

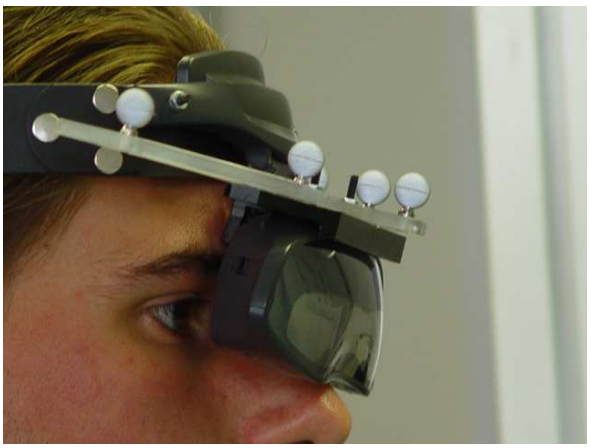
2.2. Optical-See-Through (OST)

Sony's 'Glasstron PLM-700E' were used as optical-see-through glasses. Even though the product has been discontinued, it is still very popular with augmented reality research. The original head-mounting-strap has been replaced to ensure better stability for the augmented reality mode. The glasses were operated using a stereoscopic shutter-mode at $800 * 600$ at 85 Hz. The required quad-buffer-extension for this stereo-mode was found in a 'Quadro 4'-graphics card by NVidia. Quad-buffering refers to the video-mode where two pairs of regular double-buffered framebuffers exist and the video-card switches between them in sync with the output-frequency, thereby multiplexing two video-buffers to the same monitor-out. The glasses then time-demultiplexe the frames on the right and left screen. The translucency of the glasses was set to maximum real world sight. The Sony-system is powered by a preprocessor unit, which only needs to unify power and video-in, therefore requiring three cables.

The optical-see-through configuration used the Polaris infra-red tracker by NDI, which stereoscopically tracks retroreflective targets. The operational range within free line of sight is limited to $2m^3$ underneath the tracker system which is fixed to the wall or a metal frame. The system can be taught to recognize new marker configurations as long as those don't exhibit rotational symmetry. Figure 2(b) shows the Sony-HMD with the employed markers.



(a) Three ARtoolkit markers used by the video-system to track the real 3D-space



(b) Polaris-markers attached to the Glasstron used by the OST system to determine the position of the user's head.

Figure 2. Marker types

Since optical-see-through does not require an interweave of video-images with 3D-pictures a simpler plug-in is possible. A single call to an API-call directly switches to shutter-stereoscopic mode, which is then directly configurable from the OpenInventor GUI.

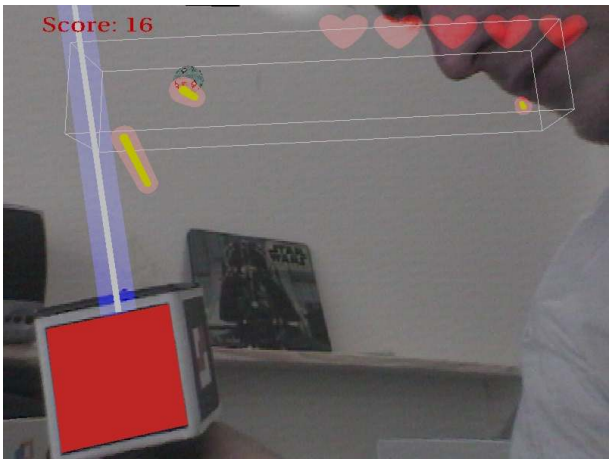
*The employed AR framework KARMA (see section 2.3) uses a plug-in architecture to encapsulate interfaces to hardware devices.

2.3. Gaming Environment

To perform the evaluation of these hardware-setups with concern to the key-questions from above a gaming environment was developed for the Karlsruhe Augmented Reality Multi-Purpose Application (KARMA). This augmented reality framework is written in C++ on the Linux platform and enables modular composition of trackers, applications, render- and input-devices using a plug-in architecture. The framework uses shared libraries, the Common Object Request Broker Architecture (CORBA) and a settings-file to avoid re-compilation of the whole application on each change configuration of devices and applications.

(a) The light saber used as an input-device by the user with attached markers valid for both systems.

(b) The robot remote which fires virtual laser-rays at the user during the game, reconstructed from a photo by Cerney.[8]



(c) Screenshot during game-play with full details of the virtual world. This is how the participant saw the world during play. The observer in the right top corner is not able to see any of the virtual elements, like the laser rays, the blade of the laser-sword, the heads-up-display, the drone and the rectangle that denotes a recognized AR-toolkit marker.

Figure 3. The gaming environment.

The game was developed to resemble the light saber training from the movie 'Star Wars - A New Hope'. In this game, a small spherical 'robot remote' (fig. 3(b)) is displayed in the user's field of view. This moving

drone fires 'laser rays' at the users, who must fend them off with their 'light saber'. The 6DOF AR-input-device 'Magic Wand' with integrated BlueTooth build at our institute was used for this purpose (fig. 3(a)). The score was increased by one for successful defense and one hit-point subtracted for failure to stop a ray. After losing 5 hit-points the game-session ended. Speed, variability of speed and frequency of fire increased exponential each ten points. This game heavily relies on the user's ability to estimate 3D object positions, complicated by the fact that the objects are moving. The game can run on both OST and VST systems, is compelling enough to keep a user interested for more than half an hour, and provides a score which can be used for objectively evaluating a player's performance. Therefore, the game is ideally suited for evaluation of the respective advantages and disadvantages of OST and VST systems when worn over extended periods of time.

2.4. Experiment procedure

An in-between subject design was employed to gather data. After having received sufficient instructions each volunteer played three games in a row on each system while scores were recorded. After each game the participants were prompted to type their name using a conventional keyboard. Help was provided to adjust the glasses and to understand the nature of the game. Instructions were standardized using a FAQ sheet to ensure minimal influence by the experimenter. After a total of six runs, participants were asked to answer a questionnaire of 14 questions. The first 4 questions gathered information about the experience of the participants with AR, sports, computer games and novel technology. Question 5-12 asked subjective impressions contrasting both system. Answers to these first 12 questions had to be given on a scale from 0 to 10. The following two question asked the participants to describe the main points for improvement for each system and collected suggestions for further development. Those last two questions were posed in an open-ended way.

3. DATA

The experiment was conducted with 40 students from the CS department of the University of Karlsruhe (TH), Germany. The datasets of 9 participants could not be included into the results because of cable defects (5), LCD-malfunction (3) and abort by participant (1). Of the remaining 31 valid datasets 27 were produced by male and 4 by female participants. The population contained 13 persons wearing spectacles. 15 participants started with the video system and 16 with the optical system. The average scores by participant group and run is shown in figure 4 and the answers from the questionnaire in the table 2.

Participants were requested to comment on the necessary improvements for each system (multiple answers were allowed). Suggestions are summarized in table 3 .

4. RESULTS

All gathered performance indicators and subjective questionnaire responses show the clear verdict in favor of Glasstron and Polaris. Video-based augmented reality struggled hard to reach half the score on average as the OST system. Participants were considerably more tired and gave worse grades without any exception after using the Trivisio/ARtoolkit combination.

Performance Using the OST system with the Polaris tracker proved to be considerably more easy and average scores for the VST system are around half as low as for the OST system. Only 5 of 31 participants were able to achieve a higher score using the VST system than with the optical version. With both systems there was a markedly learning effect during the first three sessions being more pronounced for the optical system.

3D-feeling Since video-based AR lacks the possibility to change focus, the user is stuck in the predefined vergence depth. Since the real-world and the virtual world are affected by this restriction, this might explain why participants have ranked the VST system lower in comparison to the optical-system ($\bar{X}_{video} = 4,39$, $\bar{X}_{optical} = 5,68$). This was aggravated by the fact that the game itself required a constant change in focus depth. Without eye-tracking or anticipatory focusing there is no way to remedy this problem. One troublesome remark was made by one of our participants who reported, that he did not use the 3D-information to play the game, but rather saw it as a variant of 2D-pong, where the player places the racket in the path of the ball.

Self-assessment of the participants	$\bar{X}$	s	Min	Max				
1. Previous experience with AR	0,52	1,03	0	4				
2. Experience with sports that require similar skills compared to the game	4,55	2,79	0	10				
3. Experience with computer-games	5,19	2,89	0	9				
4. Experience with adaptation to new technological devices and environments	7,16	1,46	5	10				
	Video-System				Optical-System			
Questions to compare both systems	$\bar{X}$	s	Min	Max	$\bar{X}$	s	Min	Max
5. Image quality of the virtual elements	5,35	2,42	2	10	7,03	1,47	4	10
6. Image quality of the real elements	5,00	2,39	2	10	5,90	1,89	2	10
7. Quality of real/virtual merger	4,10	2,30	0	10	4,84	1,98	1	9
8. 3D-feeling	4,39	2,35	1	10	5,68	2,34	0	10
9. Comfort of the head mounting strap	5,10	2,20	0	9	6,58	1,80	3	10
10. Eye strain caused by the systems	3,19	1,94	0	7	7,13	1,41	4	9
11. Every-day usability	1,65	1,64	0	7	3,32	2,31	0	8
12. Job-related usability	4,58	2,73	0	9	7,19	1,76	2	10
13. Comfort related to spectacles	3,19	1,94	0	7	7,13	1,41	4	9
14. Keyboard handling	3,23	2,17	0	8	7,23	2,35	2	10

Table 2. Questions for the participants and results. Note that participants on questions 10 and 12 to 14 significantly favored the optical system.

VST		OST	
21	frame rate and lag	9	3D-feeling
16	Tracking	7	Tracking
13	Field of view	5	Contrast
11	3D-feeling	4	Real world too dark
7	Eye strain	3	Field of view
6	Screen-border difficult to see	3	Calibration
4	Screen-resolution	2	frame rate and lag
3	Contrast	1	Eye strain
3	Head-mounting	1	Screen-resolution
3	Problems with contact lenses		

Table 3. Suggested areas of improvement provided by the participant for each system with multiple answering. Note that participants gave more than double the amount of suggestions for the VST system than for the OST system.

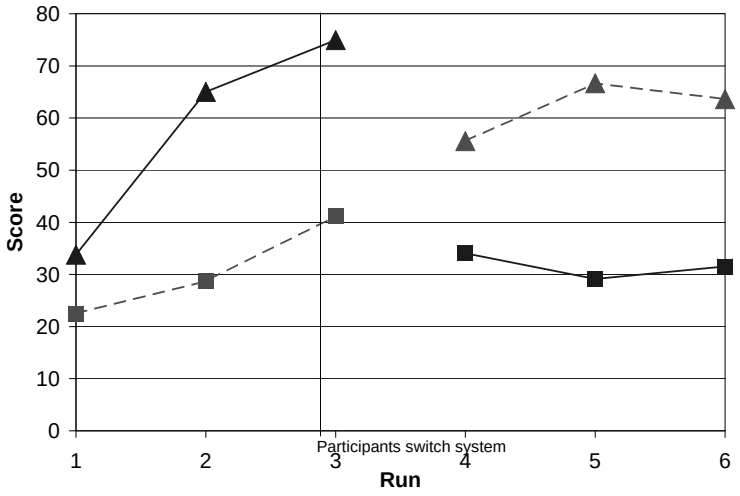


Figure 4. Average scores of the six runs. The dashed graph shows participants starting with the video-system (square) while the solid graph denotes participants starting with the optical system (triangle). Players switched systems after three runs. Note that participants showed a learning effect with both systems during the first runs. The optical system outperformed the VST system both in the first and second phase of the experiment.

Tracking A question concerning the tracking mechanism was unfortunately missing from the questionnaire, so that no quantitive data could be gathered. Sixty percent of all participants (21 of 31) on the other hand mentioned ARtoolkit-tracking as a target for improvement. It proved impossible to play without removing the world marker that determines the position of the HMD in the world. Since the position of the remote and the laser-rays was relative to the player the inaccuracies of the camera-based method introduced such considerable jitter that the game was rendered useless. Due to performance constraints also only one of the two cameras was used to determine the position of the input-device. The light-saber had to be held constantly and completely in the field of view of this camera. The latter being introduced by the ARtoolkit, which is only able to recognize markers that are completely visible in a video-picture. Since also performance dropped drastically under certain and especially changing lighting conditions, the use of the ARtoolkit in version 2.61 for this kind of tracking is rather discouraged.

Glasses Since 13 of the 31 participants wore glasses during their sessions a comparison of their performance against the remaining 18 became possible. Especially the video-system causes a major degradation in scores for those participants in comparison to the group without glasses ($\bar{X}_{videoglasses} = 22,61$, $\bar{X}_{videonoglasses} = 37,85$). The optical-see-through system also leads to lower scores but the difference was not comparably large ($\bar{X}_{opticalglasses} = 50,85$, $\bar{X}_{opticalnoglasses} = 66,61$). It appears that especially the direct contact between video-screens and glasses as caused by the Trivisio HMD is problematic. Three wearers of contact-lenses complained about strain with their eyes while using the video-system. They reported that the intense staring reduced their blinking frequency, which caused dryness with their eyes.

Image-quality and real/virtual merger Both systems operated in the same medium range when visual quality and combination of real and virtual images were concerned. It might surprise that the video-system did not score higher than the optical system on the latter question. This might indicate that the influence of resolution (and thereby pixel-jaggedness) is rather limited when it comes to the seamless merger of realities.

Acceptance When asked whether the systems were ready for deployment in daily-life or work settings participants rejected both systems for daily usage sharply ($\bar{X}_{videodaily} = 1,65$, $\bar{X}_{opticaldaily} = 3,32$). Work settings seem more feasible, and the Sony Glasstron/Polaris system gets acceptable grades ($\bar{X}_{videowork} = 7,19$) while Trivisio/ARtoolkit remains sub medium ($\bar{X}_{videowork} = 4,58$). The low grades in daily usage seem to be caused by the tiring effects the system had on the participants both for their eyes and heads.

Preexisting knowledge None of the gathered measures for preexisting knowledge showed any correlation to the attained score. Neither experience with sport, computer games nor adapation to novel technologies were valid indicators for task performance. Preexisting knowledge about AR was too low in general to draw any conclusion from the dataset.

Keyboard handling Participants had severe problems entering their name at the end of each run using the video-system. The considerable lag made it difficult to coordinate movement and resulting feedback. The amount of problems clearly demonstrates that video-systems with low frame rates and high system delay are not at all suitable for life critical environments. The OST solution performed considerably better, however we identified around half of all the participants with attempts to circumvent the HMD-screen by looking underneath the glasses to type their name instead of looking through the glasses. When prompted to give reasons for their behavior, participants exclusively mentioned the darkness of the real world view.

5. CONCLUSIONS AND FUTURE WORK

The evaluation of the two different systems shows a clear preference in all concerns for the optical see-through system in combination with the unobtrusive infra-red-tracker. The open-ended questions and comparative study design especially yielded practical directions for future research and development both in software and hardware which are summarized in the following list:

1. Insufficient **frame rate** is the main complain of participants using the video-system. Especially the image transfer from reality onto the video-display has to operate with minimal delay and frame rates of at least twenty-five. Effective video-capturing and a high-speed bus seem to be at the core of this problem.
2. The AR-toolkit **tracker** has to be improved in many ways. Especially jittering and dependence on lighting led to decreased performance and dissatisfaction with the participants.

   Objects should not be tracked using the same video-image that the participant sees. An approach using wall-mounted cameras seems to be more natural.
3. The **3D sensation** was described as inferior and needs explicit improvement. While this is partially a software problem that can be alleviated by fusing 'real' and virtual reality in better ways (e.g. using depth mapping[9] or shadows), it appears that the main focus should be the human eye. If optical prism based hardware solution can solve the problem (for instance[10]) or if active eye-tracking and perspective correction is required remains unresolved.
4. Both the **field of view** and the **screen-size** are to be increased. The display resolution remains secondary and should rather be kept small to not interfere with efforts to increase the frame rate.

Our study clearly demonstrates that optical systems are superior to VST-systems for high dexterity and accuracy tasks in both performance and subjective measures. However we do strongly believe that tracking problems and frame rate are both more important issues in this context. In contrast to findings by Ware et al.[11] that head-tracking frame rate is unimportant for reaching objects with relative stable head-position, the study in this article strongly indicates that when moving the head this lag leads to considerable dizziness and sickness.

REFERENCES

1. R. Azuma, "A survey of augmented reality," in *Computer Graphics (SIGGRAPH '95)*, pp. 1–38, August 1995.
2. R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE Computer Graphics and Applications* **21**(6), pp. 34–47, 2001.
3. T. Salb, J. Brief, T. Welzel, B. Giesler, S. Hassfeld, J. Mühling, and R. Dillmann, "INPRES (intraoperative presentation of surgical planning and simulation results) – augmented reality for craniofacial surgery," in *SPIE Electronic Imaging. International Conference on Stereoscopic Displays and Virtual Reality Systems*, J. M. et. al., ed., Januar 2003.

4. Schneider Digital, "ATI Driver Downloads." http://www.schneider-digital.de/html/download_ati.html.
5. F. Gleason, "The Video 4 Linux API." http://www.informatics.ed.ac.uk/teaching/modules/sdp/creative/v4lapi.html.
6. M. Billinghurst and H. Kato, "Collaborative mixed reality," in *Proceedings of the International Symposium on Mixed Reality (ISMR '99)*, pp. 261–284, March 1999.
7. J. Minar, "Positionskalibrierung von passiven Landmarken mit dem ARToolKit," Master's thesis, Universität Karlsruhe, 2003. Studienarbeit.
8. F. Cerney, "The Training Remote." http://www.hobbymatrix.com/models/remote/remote.html.
9. J. Ahlmann, "Erkennen von Verdeckungen durch Fremdkörper in der Intraoperativen Visualisierung," Master's thesis, Universität Karlsruhe, 2003. Studienarbeit.
10. A. Takagi, S. Yamazaki, Y. Saito, and N. Taniguchi, "Development of a stereo video seethrough HMD for AR systems," in *Proc. ISAR*, 2000.
11. C. Ware and R. Balakrishnan, "Reaching for objects in VR displays: lag and frame rate," *ACM Transactions on Computer-Human Interaction* **1**(4), pp. 331–356, 1994.