Depth Perception – A Major Issue in Medical AR: Evaluation Study by Twenty Surgeons

Tobias Sielhorst[1], Christoph Bichlmeier[1], and Sandro Michael Heining[2], and Nassir Navab[1]

[1] Chair for Computer Aided Medical Procedures (CAMP), TU Munich, Germany
{sielhors, bichlmei, navab}@cs.tum.edu
[2] Chirurgische Klinik und Poliklinik - Innenstadt, LMU Munich, Germany
sandro-michael.heining@med.uni-muenchen.de

Abstract. The idea of in-situ visualization for surgical procedures has been widely discussed in the community [1, 2, 3, 4]. While the tracking technology offers nowadays a sufficient accuracy and visualization devices have been developed that fit seamlessly into the operational workflow [1, 3], one crucial problem remains, which has been discussed already in the first paper on medical augmented reality [4]. Even though the data is presented at the correct place, the physician often perceives the spatial position of the visualization to be closer or further because of virtual/real overlay.

This paper describes and evaluates novel visualization techniques that are designed to overcome misleading depth perception of trivially superimposed virtual images on the real view. We have invited 20 surgeons to evaluate seven different visualization techniques using a head mounted display (HMD). The evaluation has been divided into two parts. In the first part, the depth perception of each kind of visualization is evaluated quantitatively. In the second part, the visualizations are evaluated qualitatively in regard to user friendliness and intuitiveness. This evaluation with a relevant number of surgeons using a state-of-the-art system is meant to guide future research and development on medical augmented reality.

1 Introduction

Real-time in-situ visualization of virtual data overlaid on its real counterpart is getting increasing attention. Augmented Reality (AR) for intra-operative visualization and navigation has been a subject of intensive research and development during the last decade. Birkfellner et al. have realized an optical see-through, [3] in-situ visualization system based on an operating binocular that is used for maxillofacial surgery. Sauer et al. [5] realized in-situ visualization with a video see-through technology head mounted display (HMD). Edwards et al. have developed an operating microscope with in-situ visualization capabilities for brain surgery [6]. These research groups report in their late works that their systems meet the desired accuracy in terms of tracking and overlay. Edwards and his colleagues report unstable and ambiguous depth perception despite accurate system calibration [7]. They report misperceptions due to multiple transparent

surfaces with textures. Already in the first paper on medical AR, where Fuchs et al. described a video see-through system for overlaying echography images, the authors discussed the issue of depth perception in the scene. They have super-imposed the virtual images on the real image. Unfortunately, this visualization appears to be in front of the patient, because the image occludes the patient. Therefore, the depth cue of occlusion tells the brain that it is in front of the patient even though the images are correctly rendered in terms of position. Psychologists distinguish in the field of cognition and perception [8] sixteen different depth cues that are exploited by the human visual system for depth perception. These depth cues have different persuasive power, they have different precision and they interact in different ways with each other. Furthermore, some depth cues have absolute quality while others only have relative quality. This means that visual system can estimate from the accommodation of the eye an absolute value for the distance of an object. However, the depth cue of occlusion can only tell, which one of two objects is in front. Since the depth cues have such a complex behavior, a straightforward construction of a perfect visualization is not possible. Apart from the correct depth perception, a compelling visualization should allow for easy interaction. That means that surgeons can understand the visualized data quickly and change the visualization easily according to the desired information. Our evaluation intends to bring more knowledge into the community about the influence of visualization techniques on depth perception. Among the tested visualizations, we have integrated the ones mentioned above from other research groups, and we also suggest new kinds of visualizations.

2   Methods

We have compared seven different visualizationmodes relative to each other with the same hardware setup. In the following paragraphs, we describe our hardware setup, the evaluation procedure and the visualization modes that we have evaluated

2.1   Description Technology

Figure 1 gives a complete overview about an AR system in use. The hardware setup is similar to the one proposed by Sauer et al. [2] for medical augmented reality. The augmentation is provided by a stereoscopic video see-through HMD (figure 3, bottom) worn by the surgeon. The virtual images are merged with two color cameras; one for each eye. Video see-through technology offers for our experiments a broader range of visualization possibilities than optical see-through, since virtual objects may be displayed opaque.

For superior registration quality, the system uses two synchronized tracking systems. The single camera inside-out tracking system allows for a high rotational precision, which is necessary for tracking of the HMD. The optical outside-in tracking system with four cameras fixed to the ceiling covers a large working area. Both of the systems use the same kind of retroreflective fiducial markers offering a registration free transformation from one tracking system to
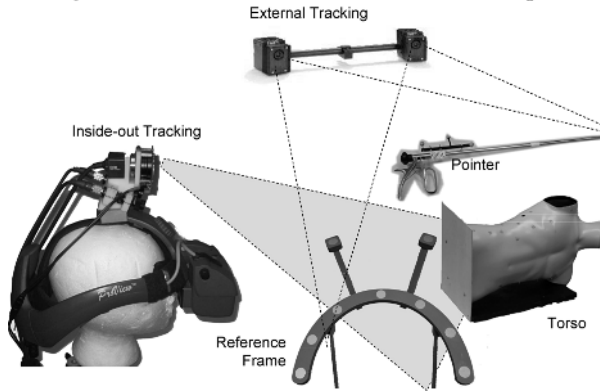
Fig. 1. Technical overview of the experimental setup

the other. These fiducial markers are attached to surgical instruments as well as on the body phantom (see figure 1). In order to recover the six degrees of freedom of a rigid body, the external optical tracking system requires four rigidly attached markers. As a third tracked object in the scene, there is a common reference target that facilitates the fusion of the inside-out and the outside-in tracking systems with no need for hand-eye coordination. The off-the-shelf computer used to render 3D graphics, to compute and include tracking data, to synchronize and combine imagery of virtual and real entities is an Intel Xeon(TM), CPU 3,20 GHz, 1,80 GB RAM, NVIDIA Quadro FX 3400/4400.

2.2   Description of the Evaluation Procedure

The evaluation of the visualization methods is divided into two parts. In the first part, we want to evaluate, which visualization offers the most reliable depth perception. In order to get quantitative results, we let the participants fulfill a task on a phantom and measure the accuracy of the performance and the time to fulfill the task.

In the second part, we survey the user acceptance. The participants fill out a questionnaire regarding the usability of the system. This questionnaire has been filled out directly after the first part of the evaluation. Thus we expect the experience with the AR system to be very present and pristine. The group of participants consists of 20 surgeons of our partner hospital.

2.3   Description of the Tests

All participants had to fulfill the same task with different visualization techniques. The task and the visualization modes can be viewed in the video[1]. The task consists of a pointer that has to be moved to a spot inside a body phantom.

---

[1] http://campar.in.tum.de/files/publications/Sielhorst2006DepthPerception.video.avi

The body phantom consists of a plastic surface of a torso and a spinal cord model inside. The data model has been recorded with a CT scanner before.

During the test, each participant has to move a pointer to randomly located spots on the surface of the spine of the phantom as shown in figure 2. Each participant had to find twenty points for each visualization. For the tests, the model of the spine has been taken out of the body phantom in order not to get haptic feedback. The participants have to rely on their mere vision. The phantom guarantees that the participants do not have a direct view to the point of interest, which is also true for minimally invasive surgery. The participants have been asked to move the pointer to the target spot and tell if they are done. The spot has been marked with a small black point in order not to interfere with the current visualization (figure 2).

For each participant we provided two minutes during he or she got used to the HMD and AR visualization. After this, the participants were asked to fulfill the tasks above. The average time for the whole test was 16.4 minutes. This exceeds the expected usage time of the visualization system in the operating room.

Fig. 2. Augmented surgical instrument serves as the pointer device. The small black point is visualized. The surgeons were asked to touch the point with its tip in the virtual position.

2.4   Evaluated Visualization Modes

The following visualization modes have been tested in a random order in order to average out effects of learning and exhaustion. The table in figure 4 contains the visualizations in the same order as figure 3. Explanation of these follow in the order of figure 3. The visualization speed is denoted in brackets in frames per second.

Triangle mesh. The triangle mesh (see figure 3.4) is a representation of the surface of the bone structure in the scene. The surface has been segmented from the CT scan before the evaluation. The surface is stored in the computer as a list of triangles. In this mode only the edges of the triangles are displayed. Edges
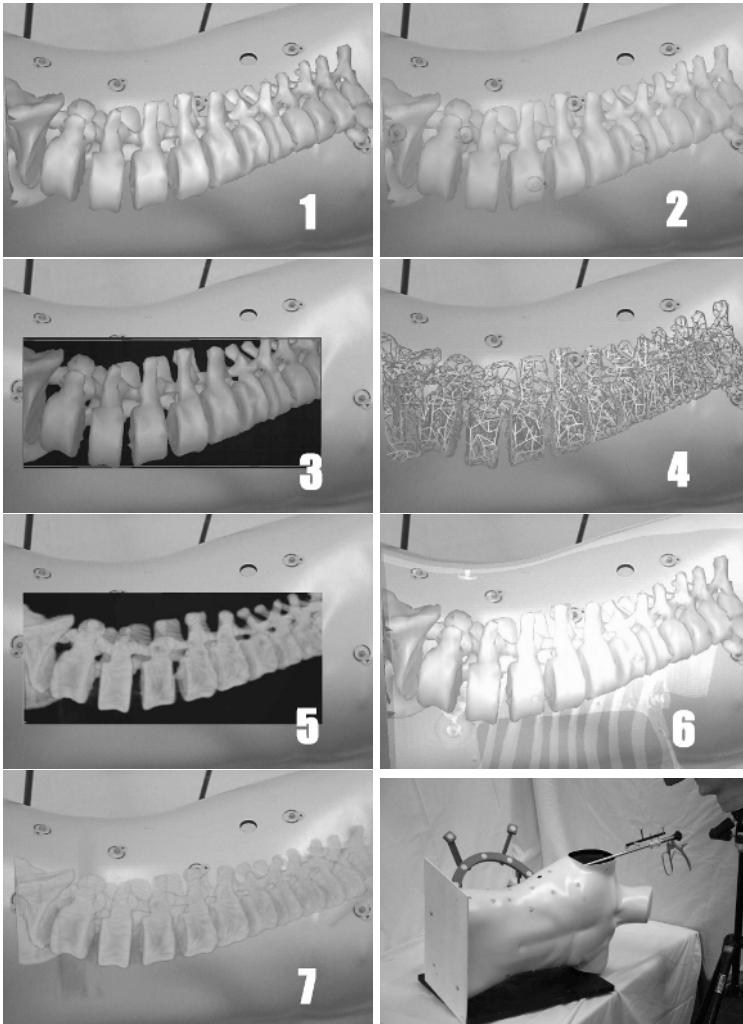
Fig. 3. The seven evaluated visualizations; bottom right: The setup

in a higher distance to the viewer are displayed darker than closer ones in order to give a strong depth impression.

Surface rendering. The surface rendering (see figure 3.1,2,3,6) is as the triangle mesh a representation of the surface of the bone structure in the scene. The surface has been also segmented from the CT scan. The surface is visualized with untextured, but shaded solid triangles.

Volume rendering. Volume rendering (see figure 3.5,7) represents the whole volume rather than a surface. The data need not be segmented or prepared. Each voxel is rendered with a certain transparency according to its value. The transfer function that relates the value to its color and transparency has been chosen in

	Depth perception	Effective-ness	Questionaire: Experience of Perception (1 = good, 5 = bad)	Frame-rate in our system
Surface rendering opaquely superimposed	o	+	2,2381	30 fps
Surface rendering transparently superimposed	+	+	2,5238	30 fps
Surface rendering through a virtual window in the skin	+	+	2,4762	30 fps
Triangle mesh	-	+	3,4211	30 fps
Volume rendering model through a virtual window in the skin	-	-	2,2857	9-10 fps
Surface rendering with a glass effect of the skin	-	-	2,8571	7-9 fps
Volume rendering superimposed	-	-	2,381	5-6 fps

Fig. 4. Table of tested visualizations including simplified results

a way that bone structure in the spine model is emphasized. The rendering is performed with the support of 3D texture hardware as suggested by Hastreiter et al. [9]

Glass effect. This effect is only applied to the surface of the skin, but not the vertebrae (see figure 3.6). The skin is rendered transparently and achromatically. Only reflections of a virtual light source to the skin are rendered. By this means the skin looks like glass. The idea is to provide a visual impression that the skin is in front of the bone model, since the reflections of the glass skin occlude slightly the bone model or influence its colour at certain areas.

Virtual window. This effect is also only applied to the skin(see figure 3.3,5). In this visualization add-on we define a region on the skin, which we call the virtual window. The visualization of the bone can only be seen through the virtual window, but not outside. The window enables the effective depth cues *occlusion* and *motion parallax* because its frame occludes partially the spinal column in the background and the projections on the retina of these stationary objects move relative, which is caused by observer movement.

3   Results

We have evaluated an overall number of 2398 3D points that had to be touched in the model. We would like to point out that this evaluation is not intended to find out the accuracy of the system. Especially the exaggerated length of the pointer (64 cm) and its crude tip with a diameter of 9 mm has been designed to provoke errors due to interaction. A shorter pointer that is peaked at the end yields far more accurate results, but it would emphasize errors of the system and unwanted movements like tremor, which we would like to totally disregard.

The volume rendering visualization proved to have a performance that can be displayed in a current PC based AR system even with a large data volume. However other kinds of volume visualizations that need prior segmentation showed

a faster and more precise interaction. Even though the surgeon liked the volume rendering and reported a good perception of depth (see figure 4).

The triangle mesh mode did not perform as well as expected. The reason might be that the visualized model has a complex structure of many self occluding surfaces that confuse the viewer when rendered with transparent triangles.

Figure 5 (top right) shows the performance in the course of time. The participants were able to work with the video see-through HMD without losing performance during the whole experiment that took 16.4 minutes in average. The graphs show no indication of exhaustion during the test. The accuracy and speed of interaction improved over the whole experiment. In the beginning of the experiment the participants trade accuracy for interaction speed (see figure 5, top right). We interpret this as indicator for getting used to the test and to the unfamiliar system.
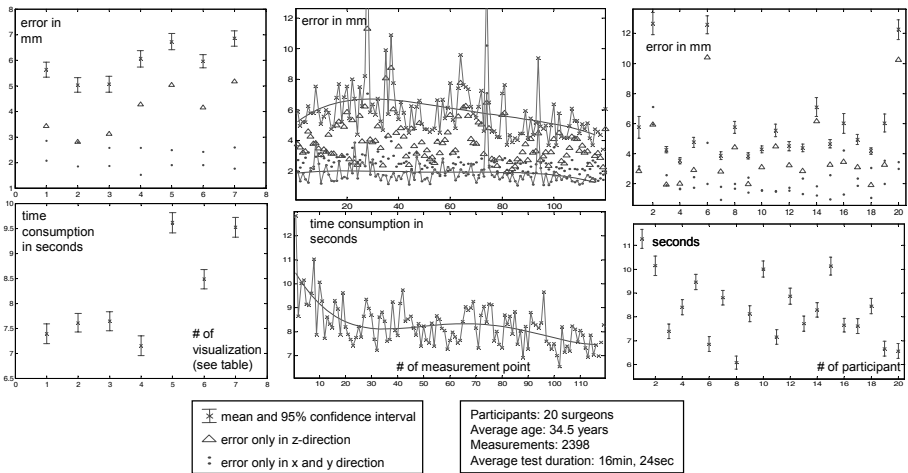

Fig. 5. Figures of the evaluation. Visualization 2 and 3 conveys depth best(left), fast visualizations cause fast user performance; users try to compensate missing depth with lateral accuracy, user trade speed for accuracy during the first 30 points i.e. relaxing after new situation (middle); even after 110 points ($-\frac{1}{4}$ hour) participants get still faster while becoming more accurate i.e. concentration does not drop; deviations of the partipicants: Fast interaction does not necessarily mean bad results and vice versa (right)

Figure 5 (top left) shows the performance of each visualization mode. The numbers are according to figure 3 and 4. The different visualization modes have a clear impact on the interaction precision and speed. In our experiments the frame rate seems to be an important factor for the performance. The participants complained about slow visualization modes and had slower and less precise performance using these. The graph shows the average error as well as the error splitted up into x and y and the z-axis that is parallel to the viewing direction. The fast modes using surface rendering clearly show better results than the fast modes using triangle mesh. It also performs better than the slow modes using volume rendering or surface rendering. Therefore, we suggest using the two best

visualization modes for interacting with a 3D model in medical AR. This is the transparent surface rendering or the one with the virtual window. It would be quite exiting to find out, if a combination of these could further improve the performance.

Figure 5 (bottom left) shows the performance of each individual surgeon. Comparing the graph depicting speed and the accuracy, we can see that these do not correlate with each other. That means that the success depends on the skills and not on the speed of the performance.

4   Conclusion

We conclude that the kind of visualization makes a serious difference in terms of effectiveness with the interaction. These findings do not even include the fact whether a visualization offers the desired medical information. This is application dependent. We have only been interested in the depth perception with different visualizations for stereoscopic augmented reality in the medical context. Due to hardware developments in the recent five years we can offer new visualization techniques that have been compared to the classic ones. One of the new visualization could outperform the other tested visualizations in terms of depth perception and effectiveness. This supports furthermore our thesis that visualization and interaction with augmented reality have to be better understood for medical application to take dvantage of its power.

References

1. King, A.P., Edwards, P.J., Maurer, et al.: A system for microscope-assisted guided interventions. IEEE Trans. Med. Imag. **19** (2000) 1082–1093
2. Sauer, F., Khamene, A., Vogt, S.: An augmented reality navigation system with a single-camera tracker: System design and needle biopsy phantom trial. In: MICCAI 2002. Volume 2489. 116–124
3. Birkfellner, W., Figl, M., Huber, K., et al.: A head-mounted operating binocular for augmented reality visualization in medicinedesign and initial evaluation. IEEE Trans. Med. Imag. **21** (2002) 991–997
4. Bajura, M., Fuchs, H., Ohbuchi, R.: Merging virtual objects with the real world: seeing ultrasound imagery within the patient. In: Proceedings of the 19th annual conference on Computer graphics and interactive techniques, ACM Press (1992) 203–210

5. Sauer, F., Khamene, A., Bascle, B., Vogt, S., Rubinob, G.J.: Augmented reality visualization in imri operating room: System description and pre-clinical testing. In: Proceedings of SPIE, Medical Imaging. Volume 4681. (2002) 446–454
6. King, A., Edwards, P., Maurer, et al.: Stereo augmented reality in the surgical microscope. Presence: Teleoperators and Virtual Env. **9** (2000) 360–368
7. Johnson, L.G., Edwards, P., Hawkes, D.: Surface transparency makes stereo overlays unpredictable: The implications for augmented reality. In: Medicine Meets Virtual Reality (MMVR). IOS Press (2002) 131–136
8. Cutting, J.E., Vishton, P.M.: Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In: W. Epstein & S. Rogers (Eds.), Perception of Space and Motion. (1995) 69–117
9. Hastreiter, P., Rezk-Salama, C., Tomandl, B., Eberhardt, K., Ertl, T.: Fast analysis of intracranial aneurysms based on interactive direct volume rendering and cta. In: MICCAI '98: 660–669