

A Practical Approach for Intraoperative Contextual In-Situ Visualization

Christoph Bichlmeier¹, Maksym Kipot¹, Stuart Holdstock¹, Sandro Michael Heining², Ekkehard Euler², and Nassir Navab¹

¹ Department of Chair for Computer Aided Medical Procedures & Augmented Reality, TU Munich, Germany

² Trauma Surgery Department, Klinikum Innenstadt, LMU, Munich, Germany

Abstract. The composition of real and virtual entities has a major contribution to the usability and acceptance of an Augmented Reality (AR) Scene. The composition method highly depends on the accessibility of information in order to allow the user to intuitively perceive relative and absolute distances of objects within the AR scene. This paper proposes a novel, flexible and practical approach of creating a context preserving view onto the operation site. It exclusively employs information from video data of video see-through AR systems to compute context information.

1 Introduction

Providing a reasonable composition of real and virtual entities within AR environments is a crucial task and a hot subject of recent published research [1–7]. In particular, the composition becomes important when virtual objects are physically positioned behind real ones. This kind of object topology is given when medical imaging data such as CT or MRI is visualized in-situ, however, appears to be floating above the patient’s body [2, 8–12]. The previously introduced *Contextual In-Situ Visualization* [3, 4] for AR was inspired by *Focus & Context Visualization* methods having been previously introduced for virtual worlds [13–15]. These approaches use primarily the geometry of the patient’s skin generating context information for better perception of deeper seated anatomy. However, they are not practical and flexible enough for being intraoperatively applied in future AR systems. The majority of image guided surgeries do not use imaging data that is large enough to cover the skin geometry. Even if the skin geometry is available, we deal here with only static context information captured at a certain point in time. For this reason, the skin serving as the major context layer cannot adapt to geometry deformation or color changes.

We decided to follow the approaches of Kalkofen et al. [1], Stoyanov et al. [2] and Avery et al. [6] using exclusively information from video data of video see-through devices to generate context information. We extended these approaches with additional visual cues reacting on changing light and color conditions, deformation, and interaction with surgical instruments. This paper presents the rendering pipeline of a new version of *Contextual In-Situ Visualization* strongly optimized for surgical AR environments. In addition, we discuss our first approach of designing of an experimental setup

to qualitatively assess visualization methods that are intended for improving the spatial perception of medical AR environments.

2 Method

After the description of the used AR system, we introduce a new method for contextual in-situ visualization and our preliminary experimental setup up for evaluating the quality of depth perception.

2.1 In-Situ Visualization

The rendering approach has been tested with the RAMP system introduced by Sauer et al. [8] as well as with a new video see-through head mounted display (HMD), which is an NVIS NVisor SX having a 1280x1064, 60Hz display with 24-bit colors. A black and white 640x480 resolution PTGrey Flea camera is used for the infrared tracking. Two PTGrey Flea color cameras with a resolution of 1024x780 capture the view in front of the HMD and these images are displayed in the HMD simulating the view from a user's eyes. The HMD is tracked using both an inside-out tracking as well as an outside-in tracking system. The inside-out tracking system uses the infrared camera attached to the HMD to track a set of infrared markers fixed to an arc and this allows head pose estimation to be calculated. This tracking setup provides high rotational accuracy. The outside-in tracking system is from the company A.R.T. GmbH, Germany and we have four infrared cameras fixed on a ceiling mount with each camera positioned in a corner to enable a clear view of the infrared markers from all angles. Infrared markers attached to the HMD enable an outside-in tracking which is used as a backup system for head pose estimation and only in the case when the inside-out tracking system fails. However, an advantage of this setup is that the HMD user can move freely within the AR scene even when the infrared markers on the arc are out of the field of view of the HMD's infrared camera and it also provides for additional image stability. This dramatically increases the usability of the system even though the outside-in tracking cannot provide such high rotational precision [16]. In our preliminary but extendable AR setup an endoscopic instrument and an object simulating the patient's body are tracked.

2.2 Rendering Pipeline

Our visualization pipeline installs a transparent vision channel [17, 4] virtually onto the deepest context layer of the AR scene, which is in our case the skin, that is aligned with the line of sight of the HMD user. A transparency fading effect is applied at the border of the vision channel in order to generate a smooth transition from the original skin color of the video to the assigned transparency of the vision channel.

$$AF = ((1.0 - AL)/FR) * (CPVC - VCR + FR) + AL; \quad (1)$$

AF describes the alpha value within the region of transparency transition. It is calculated with the parameters of the initial global alpha value inside the vision channel

(AL), the width of the concentric disk describing the region of transition (FR), the current position of the pixel within the vision channel being used for the calculation of the alpha value (CPVC), and the vision channel radius (VCR). The proposed rendering pipeline for our method of image composition involves four main components using information from tracking data and video images. The transparency generating the context information within the vision channel is parameterized with the following functions on the video data:

- MATERIAL1: A direction invariant Sobel filter for the detection of edges in video images caused by surgical instruments and trocars, surgeons' hands, the open operation site itself, blood drops, birthmarks, scars, cover sheeting, artificially introduced marks by a sterile pen onto the skin.
- MATERIAL2: Detection of specular highlights caused by bright metallic light reflecting instruments, humid skin surface, fluids, bright/specific colors in general.

In addition, two combined shadow effects enhance the spatial appearance and provide interactive visual feedback:

- SHADOW1: The combination of MATERIAL1 and MATERIAL2 is projected shadow-like onto virtual objects behind the skin layer using OpenGL supported projective texture mapping.
- SHADOW2: Shadow mapping [18] applied to all virtual objects.

Video data coming from a video capturing device like an HMD, a laparoscope or an arthroscope camera is processed by a GLSL³ based shader. For modular computation and final composition, Frame Buffer Objects (FBO) for GPU accelerated off-screen rendering have been employed. Regarding MATERIAL1 (see Fig. 1(a)), the transparency (AE) value of each pixel of the transparency map is computed by comparing the gray values (GV) of both the original image pixel (OP) and the extracted edge pixel (EP) and adjusting it with a transparency level (AL):

$$AE = OP.GV/EP.GV + AL \quad (2)$$

MATERIAL2 thresholds the gray scale version of the video data to detect brightly colored regions (GV), which remain opaque while darker regions are coded with a high transparency value:

$$AS = AL + 1.0 - (1.0 - OP.GV)/(1.0 - ST) \quad (3)$$

The alpha value (AS) of a pixel of the video image uses the following parameters. OP is the original pixel value, AL is again the initial global alpha value inside the vision channel and ST determines the threshold for bright regions (see Fig. 1(b)). MATERIAL1 and MATERIAL2 are then combined as shown in Fig. 1(c). The result is then forwarded to another shader program performing projective texture mapping [18] to feed SHADOW1. SHADOW1 and SHADOW2 apply *percentage closer filtering* (PCF) [19], which is supported by the used graphics hardware⁴ to create smooth shadow and avoid

³ OpenGL Shading Language

⁴ Nvidia™ GeForce 8800 Ultra

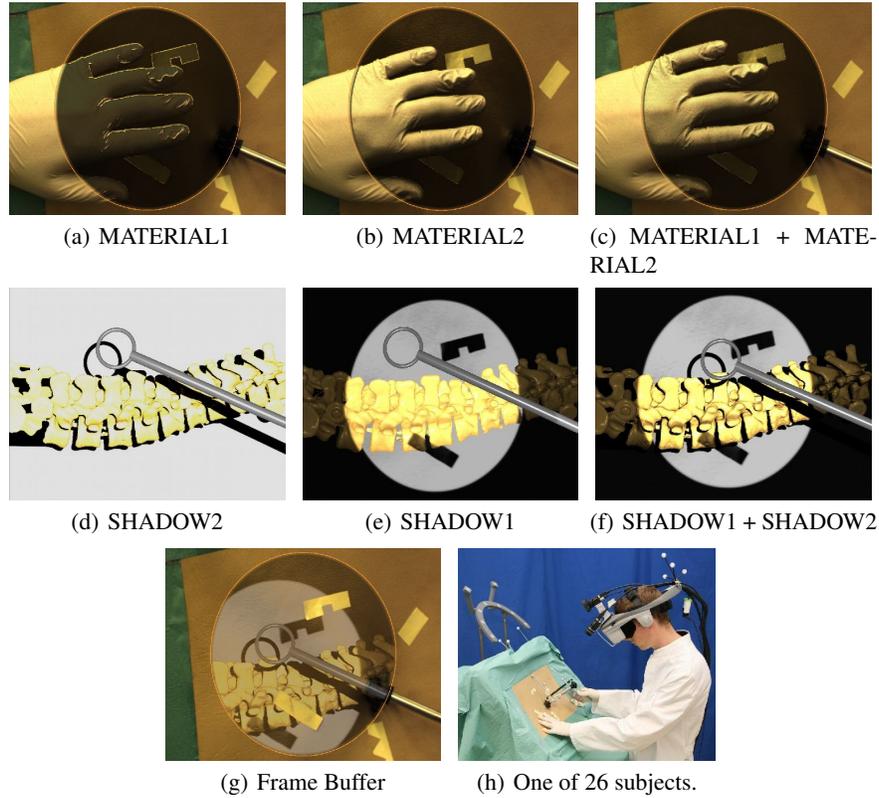


Fig. 1. Full pipeline described with screenshots of an exemplary AR scene.

aliasing artifacts (see Fig. 1(d)). The method of combining SHADOW1 and SHADOW2 follows a natural shadowing principle based on *diffusion* and *interference*. The blurriness of a shadow increases and the contrast decreases in proportion to an increase of the distance between a light source and a fixed object in space. According to our assumption that 3D objects are located behind the context layer, features filtered by MATERIAL1 and MATERIAL2 can be projected with SHADOW1 onto these objects so that their shadows will be sharper and have more contrast. Blurriness can be adjusted with PCF kernels and both, SHADOW1 and SHADOW2, can be weighted. However, blurring and weighting of SHADOW1 is based on an approximation of the distance between the skin surface and virtual objects behind the skin since the 3D pose of the skin is not known. The light for calculating SHADOW1 and SHADOW2 is positioned relative to the HMD of the user in our case 20cm to the right.

At the end of the rendering pipeline we deal with two textures called *Processed Video* and *Final Virtual Scene*. *Processed Video* represents the video data with a previously individually computed alpha value for each pixel within the vision channel. The other texture contains all virtual 3D information including the composition of SHADOW1 and SHADOW2. *Processed Video* and *Final Virtual Scene* are then composed with OpenGL

blending and rendered into the Frame Buffer resulting in the final AR scene (see Fig. 1(g)).

2.3 Assessing the Perceptive Quality

User studies are essential to qualitatively and quantitatively prove the advantage of a proposed method. However, designing an experiment that satisfies the quality criteria objectivity, reliability and validity is a difficult task. For assessing the perceptive quality to estimate the relative or absolute depth of objects in the designated medical AR scene, in particular the decision on the right measuring instruments is difficult. The hardware weight, lens configuration of the cameras, FOV of the cameras and display, tracking quality, possible stability, resolution, color depth and update rate, presence or absence of medical knowledge, vision abilities of subjects and many more parameters may bias the measuring data.

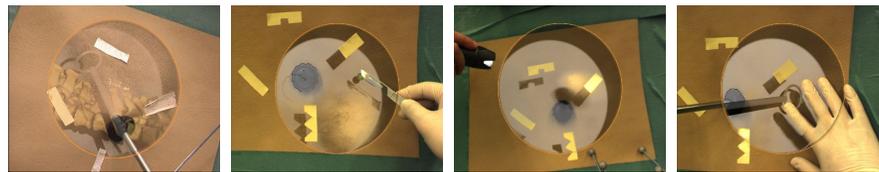
Here, we present the status quo of an iteratively tested and improved evaluation environment that has been designed to assess the quality of different methods for image composition for instance our proposed *Contextual In-Situ Visualization* (see Fig. 2). The experiment requires subjects to insert a tracked endoscopic instrument through a trocar to reach target positions behind a synthetic skin layer (see Fig. 1(h)). The instrument is superimposed with its virtual counterpart that has been augmented with a virtual ring at the instrument tip. The subject's task is to guide the ring of a set of virtual, unicolored, randomly sized and positioned balls that are physically located behind the skin layer. The ball changes its color in case of collision. In addition, to the number of collisions, we measure time of performance, time of collision, number of tries (according to the protocol the ring has to be guided straight forward over the ball), path length of the instrument tip, time it took to first reach the area near the ball as measured from a start position, and path length of the instrument tip along the instrument axes (motion into depth).

3 Results

Figure 2(a) shows an endoscopic instrument penetrating a previously inserted trocar above a simulated operation site for key hole surgery. The skin around the trocar shows in this case only homogeneous color distribution providing almost no information for MATERIAL1 and MATERIAL2. However, guidance of instruments causes deformation of the skin and influences the color brightness of the skin. This reactively changing light reflection can be detected by MATERIAL2 to generate reactive transparency changes. In order to feed MATERIAL1, we attached small glue strips having a slightly different color than the skin (see Fig. 2(a)). For intra operative surgical planning in some cases sterile markers are used to draw rough sketches of the hidden anatomy onto the patient's skin. MATERIAL1 and MATERIAL2 can be configured to detect also the color of such sterile markers to enable additional context information. Alternatively light patterns can be projected onto the skin. Figure 2(c) shows a hand held flashlight causing a bright spot on the skin that is detected by MATERIAL2. This sterile light spot again causes partial opacity of the vision channel. In addition, its shadow is thrown onto

the virtual anatomy and a virtual plane in the background of the scene. SHADOW1 can produce shadow from any object that is shown in the video image. This includes the surgeon’s hand (see Fig. 2(d)) and surgical instruments such as a scalpel (see Fig. 2(b)) or the endoscopic instrument with its attached marker tree for optical tracking. The effect of masking the scene with the user’s hand has been previously shown with distinguishable blueish gloves having an exclusive color in the AR scene [3]. The current configuration of MATERIAL1 can also deal with skin colored gloves that are mainly used in today’s ORs (see Fig. 1(c)).

Although, we got positive feedback from the majority of 26 subjects that tested the



(a) Deformation of the skin (b) Extracted instruments cast shadow (c) Flashlight causes context information. (d) Hand causes shadow feedback.

Fig. 2. Different features of the proposed image composition.

proposed method with our experimental setup described in section 2.3 our measuring data does not show significant differences when comparing our Contextual In-Situ Visualization with simple superimposed opaque or transparent balls onto an opaque skin. The positive feedback is derived from a questionnaire that had to be completed after the experiment. To the question *How supportive do you consider the shown visualization modes for your task*, the number answering either supportive or very supportive is 6 for the opaque balls, 4 for the transparent balls and 18 for our new method. With the question *How realistic do you find each visualization mode*, the number answering realistic or very realistic is 4 for the opaque balls mode, 4 for the transparent balls and 20 for our new method.

4 Discussion

Fig. 2(c) shows the generation of depth cues using a light source changing the illumination of the context layer, i.e. the skin. Instead of light being visible for the human eye, infrared light can be used to illuminate the skin for MATERIAL2. Infrared light can be reflected by the skin and is visible for cameras. In this case the homogeneous illumination conditions in ORs are not disturbed. Currently parametrization of MATERIAL1, MATERIAL2, SHADOW1 and SHADOW2 is configured manually by the user. However, future work will address the dynamic parametrization of the proposed functions. This can involve the analysis of the color histogram of the video over time to decide on the thresholding and weighting of MATERIAL1 and MATERIAL2. Also the blurriness and contrast of shadow effects can be adjusted according to analysis of the current lighting conditions. A major part of future work will be the interdisciplinary improvement of

the evaluation environment together with psychologist having a strong background in methodology to guarantee reliability, validity and objectivity of the study. The positive feedback from subjects strongly motivates us to study intensively whether the visualization method or the evaluation design causes the lack of significant measuring data.

5 Conclusion

This paper proposes a novel, flexible and practical approach of creating a context preserving view inside the patient during an operation. It exclusively employs information from video data from video see-through AR systems to compute context information, which allows for the processing of deformable context layers. The generated graphical effects from MATERIAL1 and MATERIAL2 that are part of the proposed visualization pipeline provide the depth cues of occlusion and motion parallax even though the illumination, color and geometry conditions completely change during the surgical procedure, e.g. the skin is cut or deformed, minimally invasive trocars are installed or an open operation is being performed. By adjusting parameters the masking of the virtual part of the scene with conventional surgical hand gloves can be achieved. The effects SHADOW1 and SHADOW2 can improve the 3D layout of the AR scene. In addition, they provide visual feedback when the user starts to interact, e.g. guidance of a surgical instrument, with the scene. The proposed method can be applied for any video see-through technology being used for AR applications. In the medical field this includes for instance laparoscope or arthroscope cameras. We strongly believe that also video see through HMDs used for medical procedures would benefit from our approach once they are integrated into the operating theater.

6 Acknowledgements

We would like to thank Frank Sauer, Ali Khamene, and Sebastian Vogt from Siemens Corporate Research (SCR) for the design, setup, and implementation of the RAMP system they provided us. Thanks also to Tobias Sielhorst having constructed the new HMD. We would also like to thank A.R.T. GmbH for providing cameras and software for the outside-in tracking system. We want to express our gratitude to the radiologists and surgeons of Klinikum Innenstadt in Munich for their precious contribution in obtaining medical data and evaluating our systems. Many thanks also to the crew at NARVIS lab.

References

1. Kalkofen, D., Mendez, E., Schmalstieg, D.: Comprehensible visualization for augmented reality. *IEEE Transactions on Visualization and Computer Graphics* **15**(2) (2009) 193–204
2. Stoyanov, D., Mylonas, G.P., Lerotic, M., Chung, A.J., Yang, G.Z.: Intra-operative visualizations: Perceptual fidelity and human factors. *J. Display Technol.* **4**(4) (2008) 491–501
3. Kutter, O., Aichert, A., Bichlmeier, C., Heining, S.M., Ockert, B., Euler, E., Navab, N.: Real-time Volume Rendering for High Quality Visualization in Augmented Reality. In: *International Workshop on Augmented environments for Medical Imaging including Augmented Reality in Computer-aided Surgery (AMI-ARCS 2008)*, New York, USA, MICCAI Society (Sept. 2008)

4. Bichlmeier, C., Wimmer, F., Heining, S., Navab, N.: Contextual Anatomic Mimesis: Hybrid In-Situ Visualization Method for Improving Multi-Sensory Depth Perception in Medical Augmented Reality. In: Proceedings of the 6th International Symposium on Mixed and Augmented Reality (ISMAR). (November 2007) 129–138
5. Swan, J.E.I., Livingston, M.A., Smallman, H.S., Brown, D., Baillot, Y., Gabbard, J.L., Hix, D.: A perceptual matching technique for depth judgments in optical, see-through augmented reality. In: VR '06: Proceedings of the IEEE Conference on Virtual Reality, Washington, DC, USA, IEEE Computer Society (2006) 19–26
6. Avery, B., S.C., Thomas, B.: Improving spatial perception for augmented reality x-ray vision. In: VR '09: Proceedings of the IEEE Conference on Virtual Reality, Louisiana, USA, IEEE Computer Society (2009) 79–82
7. Nicolaou, M., James, A., Lo, B., Darzi, A., Guang-Zhong, Y.: Invisible shadow for navigation and planning in minimal invasive surgery. (2005) 25–32
8. 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
9. Feuerstein, M., Mussack, T., Heining, S.M., Navab, N.: Intraoperative laparoscope augmentation for port placement and resection planning in minimally invasive liver resection. *IEEE Trans. Med. Imag.* **27**(3) (March 2008) 355–369
10. 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
11. Fuchs, H., A., S., Pisano, E., Garrett, W., Hirota, G., Livingston, M., Whitton, M., Pizer, S.: Towards performing ultrasound-guided needle biopsies from within a head-mounted display. (1996) 591–600
12. Wanschitz, F., Birkfellner, W., Figl, M., Patruta, S., Wagner, A., Watzinger, F., Yerit, K., Schicho, K., Hanel, R., Kainberger, F., Imhof, H., Bergmann, H., Ewers, R.: Computer-enhanced stereoscopic vision in a head-mounted display for oral implant surgery. *Clinical Oral Implants Research* **13**(6) (2002)
13. Bruckner, S., Grimm, S., Kanitsar, A.: Illustrative context-preserving exploration of volume data. *IEEE Transactions on Visualization and Computer Graphics* **12**(6) (2006) 1559–1569
Member-M. Eduard Groller.
14. Burns, M., Haidacher, M., Wein, W., Viola, I., Groeller, E.: Feature emphasis and contextual cutaways for multimodal medical visualization. In: EuroVis 2007 Proceedings. (May 2007)
15. Krueger, J., Schneider, J., Westermann, R.: Clearview: An interactive context preserving hotspot visualization technique. *IEEE Transactions on Visualization and Computer Graphics* **12**(5) (2006) 941–948
16. Hoff, W., Vincent, T.: Analysis of head pose accuracy in augmented reality. *IEEE Transactions on Visualization and Computer Graphics* **6**(4) (2000) 319–334
17. Pinna, B., Tanca, M.: Perceptual organization reconsidered in the light of the watercolor illusion: The problem of perception of holes and the object-hole effect. *J. Vis.* **8**(7) (5 2008) 1–15
18. Williams, L.: Casting curved shadows on curved surfaces. In: In Computer Graphics (SIGGRAPH '78 Proceedings). (1978) 270–274
19. Reeves, W.T., Salesin, D.H., Cook, R.L.: Rendering antialiased shadows with depth maps. *SIGGRAPH Comput. Graph.* **21**(4) (1987) 283–291