Florian Heinrich, Kai Bornemann,

Kai Lawonn, Christian Hansen

Interacting with Medical Volume Data in Projective Augmented Reality

Pre-print version

Florian Heinrich, Kai Bornemann,

Christian Hansen

Department of Simulation & Graphics,
Research Campus *STIMULATE*
Otto-von-Guericke University Magdeburg, Germany
hansen@isg.cs.uni-magdeburg.de

Kai Lawonn

Department of Computer Science,
University of Jena, Germany
kai.lawonn@uni-jena.de

This is a pre-print of an article presented at the 23rd international conference on Medical Image Computing & Computer Assisted Intervention (MICCAI), 4-8 October 2020, Lima, Peru

Interacting with Medical Volume Data in Projective Augmented Reality

Florian Heinrich[1,2], Kai Bornemann[1,2], Kai Lawonn[3], and Christian Hansen[1,2][0000−0002−5734−7529]

[1] University of Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany
[2] Research Campus *STIMULATE*, Sandtorstrasse 23, 39106 Magdeburg, Germany
[3] University of Jena, Fürstengraben 1, 07743 Jena, Germany

Abstract. Medical volume data is usually explored on monoscopic monitors. Displaying this data in three-dimensional space facilitates the development of mental maps and the identification of anatomical structures and their spatial relations. Using augmented reality (AR) may further enhance these effects by spatially aligning the volume data with the patient. However, conventional interaction methods, e.g. mouse and keyboard, may not be applicable in this environment. Appropriate interaction techniques are needed to naturally and intuitively manipulate the image data. To this end, a user study comparing four gestural interaction techniques with respect to both clipping and windowing tasks was conducted. Image data was directly displayed on a phantom using stereoscopic projective AR and direct volume visualization. Participants were able to complete both tasks with all interaction techniques with respectively similar clipping accuracy and windowing efficiency. However, results suggest advantages of gestures based on motion-sensitive devices in terms of reduced task completion time and less subjective workload. This work presents an important first step towards a surgical AR visualization system enabling intuitive exploration of volume data. Yet, more research is required to assess the interaction techniques' applicability for intraoperative use.

Keywords: Interaction techniques · Medical volume data · Projective augmented reality.

1 Introduction

Exploring medical image data is essential for many modern surgical procedures in terms of diagnosis, planning and image guided surgery. This volume data is usually visualized as two-dimensional (2D) slices on conventional PC monitors, thus requiring surgeons to build mental maps of the actual patient anatomy. Three-dimensional (3D) data representations can facilitate this process [12, 19]. Moreover, such visualization enables a better understanding of spatial relations and an easier identification of anatomical or pathological structures [1].

These effects can potentially be enhanced by the concept of augmented reality (AR). By showing relevant anatomy directly on the patient, surgeons no longer need to split their focus between spatially separated monitors and the operation site. Additionally, mental effort for the understanding of anatomical spatial relations can be further reduced, because of less distance between these information sources [16, 21]. In the past, different technical solutions were developed to fuse both views on the patient and on the image data, e.g. augmented camera views displayed on monitors [22], optical see-through head mounted displays [17] or projector-camera-systems projecting volumetric image data directly onto the patient [21]. In comparison, projective AR does not require additional monitors in already crowded operating rooms or the user to wear head sets, which are often uncomfortable to wear for longer periods of time and could potentially compromise sterility.

To effectively explore and manipulate 3D image data, appropriate interaction methods are needed. Typical interaction tasks often require the manipulation of more degrees of freedom (DoF), than provided by conventional methods, e.g. mouse and keyboard [3]. Moreover, the need for sterility and limited space in the operating room further restrict needed interaction paradigms [15]. Therefore, more efficient techniques have been developed, that allow users to control more than two DoF simultaneously [11] and can be executed touchlessly [14]. Such intuitive methods are often based on natural human interaction in the form of gestures [6].

Related work identified advantages of using hand gestures [12, 20, 24] and evaluated different eligible input devices [9]. Foot gestures were identified as viable alternatives to manipulate image data even hands-freely [7, 10, 23]. In contrast, also interaction techniques using hand-held devices were explored [4]. Similar approaches were followed for immersive virtual reality applications using motion-sensitive controllers [2, 18]. Most of these techniques were developed for the interaction with image data displayed on monitors. However, methods developed for desktop environments may not be intuitive in AR systems because of the more complex dimensionality. Additionally, in AR, position, orientation and scale of image data are fixed in space while these parameters can be modified in desktop applications. Therefore, previous research findings may not be applicable for AR environments. Hence, this work evaluates four gestural interaction techniques with respect to their applicability for exploring medical volume data in projective AR. This type of AR was chosen because of its advantages in terms of sterility and space requirements. Moreover, interactive AR system projecting image data onto the operation site have not been investigated before.

2 Materials and Methods

A user study was conducted to evaluate different gestural interaction methods for the exploration of projected medical volume data. For this study, 26 medical students (16 female, median age: 24, median year of university: 4) with basic knowledge about medical image data were recruited. In the following, details and rationales of the study are presented. A supplementary video demonstrates implemented interaction concepts and tasks of the study.

Fig. 1. Apparatus of implemented interaction techniques. a) Hand gestures. b) Foot gestures. c) Surgical instrument. d) Controller.

2.1 Interaction Techniques

Related work suggests advantages of gestural interaction techniques over conventional input modalities. To ensure, that all methods can be executed during interventions, we restricted this work's techniques to single-handed use, leaving one hand still available. Implementations should provide methods to manipulate linear parameters, as well as means to activate or deactivate these modes.

First, an interaction concept based on *hand gestures* was implemented using the Myo gesture control armband (Thalmic Labs, Canada) for gesture recognition. A positively evaluated fist gesture [8] and an easy to distinguish finger double tap gesture were used for mode activation. Two inertial measurement units (IMU; Xsens, Netherlands) were attached to the users wrist and upper arm (see Fig. 1a). Using direct kinematics, the relative hand position of the user could be determined. Changes of that position were mapped to linear parameters, resulting in an overall grabbing and moving interaction concept.

We also implemented an interaction technique based on *foot gestures*. A toe tap, i.e. lifting and lowering the forefoot [23], was included as activation gesture. Heel rotations, i.e. rotating the forefoot around the heel, were implemented to change linear parameters. Degrees of rotation were mapped to the speed at which parameters were changed. Only one DoF could be manipulated at the same time. Therefore, toe taps were used to switch between parameters. An IMU attached to the forefoot measured relative foot movements, thus recognizing gestures (see Fig. 1b).

Surgical navigation systems are commonly used in image-guided surgery [17]. Device-based gestures could by adapted by using related tracked *surgical instruments*. We attached an HTC Vive tracker (HTC Corp., Taiwan) to a 3D-printed surgical pointer (see Fig. 1c). Movements of that instrument were mapped to changes of linear parameters. Mode activations were performed using toe taps because direct means of interaction, e.g. buttons, are not available on surgical instruments.

Finally, a *controller-based interaction* method using an HTC Vive controller was implemented for our prototype (see Fig. 1d). Manipulation modes were activated by pressing and holding the trigger button at the index finger or by touching and holding the trackpad at the thumb's position. Movement of the controller in space was than mapped to changes of linear parameters.

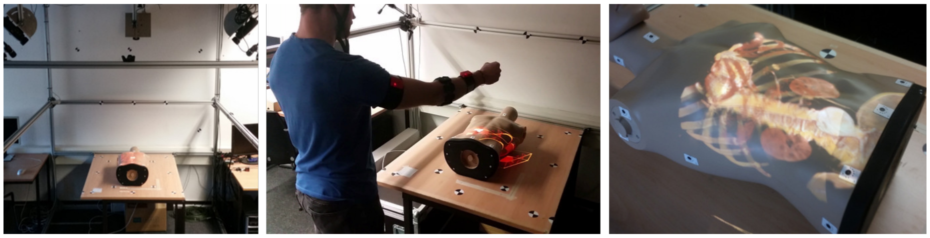

Fig. 2. The experimental setup. Projectors mounted above a table superimpose a torso phantom with a medical volume data set.

2.2 Projection of Medical Volume Data

For this work, the Panoramix DICOM example data set provided by the software 3D Slicer (Kitware, USA) was scaled to match the dimensions of a human torso phantom, on which two stereoscopic projectors with an active shutter 3D system (Barco F50 WQXGA, Barco GmbH, Germany) displayed virtual contents. Both projectors were calibrated using a photogrammetric measurement system. Calibration results included a surface model of the projection surface, as well as extrinsic and intrinsic projector parameters. The user's head position was tracked with an HTC Vive tracker attached to a helmet worn by the user. That way, both binocular and kinematic depth cues were supported. Images based on the user's spatial position were rendered for each eye using the game engine Unity (Unity Technologies, USA). Then, these images were mapped onto the projection surface scan. Rendering that textured surface from the view of individual projectors resulted in the projection of undistorted, perspectivly correct images (see Fig. 2). GPU accelerated volume ray casting was performed to determine, which parts of the data were currently visible. Thereby, current windowing width and windowing level parameters mapped the intensity values of the used DICOM data set to the range 0 to 1. Afterwards these values were used to apply color and transparency via a transfer function. Furthermore, the visualized volume could be reduced using a clipping box.

2.3 Tasks

This work's exploration tasks were limited to the manipulation of windowing parameters and the position and size of a clipping box, because in spatial projective AR systems, tasks manipulating transformation or camera parameters no longer apply. To this end, a search task was implemented that had to be completed with either windowing or clipping techniques. For that, a foreign body, i.e. a screw, was inserted into the volume data. Position, size, orientation and intensity value of the object were varied between trials to avoid bias. Only one target was present for each trial. For the windowing task, window level and window width needed to be manipulated, so that the hidden search object became visible (see Fig. 3a-b). An object was considered visible, when the windowing

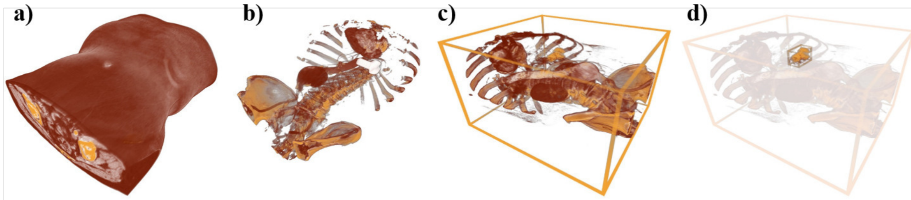

Fig. 3. Exploration task procedures. a) - b) Windowing task. c) - d) Clipping task.

parameters mapped its intensity value to a specific range of the used transfer function. After the target was found, windowing parameters were changed and a new target was inserted, until a total of four objects were detected. The clipping tasks required participants to change the position and size of a clipping box so that it encapsulated the search object as closely as possible without clipping the object itself (see Fig. 3c-d). Scaling and translation of the clipping box could be performed individually for each main axis. Both tasks are demonstrated in the supplementary video. Interaction techniques' mode activation methods were used to enable changing task-specific parameters. For the windowing task, only one method was needed to start changing both windowing parameters by vertical and horizontal movements, analogous to mouse input for 2D image slices. For the clipping task, individual activation gestures were used to enable either a change in clipping box size or position. Movements were then mapped to changes in respective main axis directions. These changes were always limited to the direction of greatest change, to avoid unwilling manipulation of more than one axis. As for foot-based interaction methods, the toe tap activation gesture needed to be performed consecutively to rotate between individual parameters.

2.4 Variables

Task completion time, total number of mode activations, i.e. how often the mode activation gestures were performed, and subjective workload were measured for both windowing and clipping tasks. Time measurement started after participants signalized their readiness and stopped when the individual task was considered finished. Subjective workload was estimated using the standardized raw TLX questionnaire. For clipping tasks, accuracy was measured by calculating the Jaccard index between the user-set and a perfect clipping box. This resulted in accuracy values between 0 and 100 with higher percentages representing higher congruence. Participant performance during windowing tasks was further evaluated by a degree of efficiency that was defined as the relation between the smallest possible amount both windowing parameters needed to be changed and their total accumulated sum of changes. Results were interpreted as percentages with higher values representing higher efficiency.

Table 1. Summary of the ANOVAs' results ($\alpha < .05$).

Variable	df	F	p	Sig.	η^2	Effect	Figure
Windowing Task							
Efficiency	3	2.24	0.088		0.063	medium effect	Fig. 4a
Task completion time	3	12.40	<0.001	*	0.271	large effect	Fig. 4b
Mode switches	3	6.80	<0.001	*	0.169	large effect	Fig. 4c
Subjective workload	3	5.65	0.001	*	0.145	large effect	Fig. 4d
Clipping Task							
Accuracy	3	0.47	0.705		0.014	small effect	Fig. 5a
Task completion time	3	9.48	<0.001	*	0.221	large effect	Fig. 5b
Mode switches	3	6.28	<0.001	*	0.159	large effect	Fig. 5c
Subjective workload	3	8.68	<0.001	*	0.207	large effect	Fig. 5d

2.5 Procedure

After initial instructions and collecting demographic participant data, the head tracking system was set up and calibrated according to the subject's eye position. Then the first windowing and clipping task trials began for the first interaction technique. For each participant, the order of interaction techniques were randomized to avoid bias caused by learning effects. Moreover, the order of windowing and clipping tasks was alternated between participants. Before each task, participants were given the chance to train under the current experimental conditions until they felt comfortable to start. After finishing a task, participants were asked to fill out the raw TLX questionnaire. The next interaction technique was evaluated after both tasks were fulfilled. The experiment concluded with a brief informal questionnaire about participant feedback.

3 Results

Data for the windowing and for the clipping tasks were analyzed separately. One-way ANOVAs were conducted for each task-related variable to investigate effects of the interaction technique factor. Their results are summarized by Table 1. Afterwards, post-hoc tests were conducted to analyze individual differences between factor levels. To this end, pairwise t-tests with Bonferroni correction were performed. Fig. 4 and Fig. 5 visualize identified effects, as well as statistically significant post-hoc test results.

4 Discussion and Conclusion

Participants were able to fulfill both exploration tasks using all interaction techniques, as supported by similar accuracy and efficiency results. However, differences between concepts are shown for the other variables. Foot gestures generally performed worst, as indicated by higher subjective workload and longer task completion times. This may have been because the concept only allowed for the

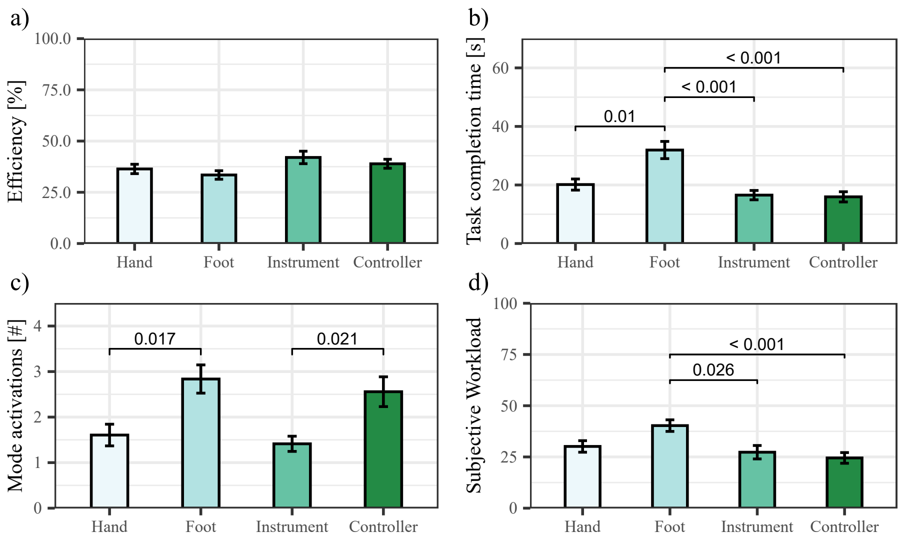

Fig. 4. Windowing task results. Effects of interaction techniques on a) efficiency, b) task completion time*, c) mode activations* and d) subjective workload*. Error bars respresent standard errors. Horizontal lines indicate statistically significant post-hoc test results.

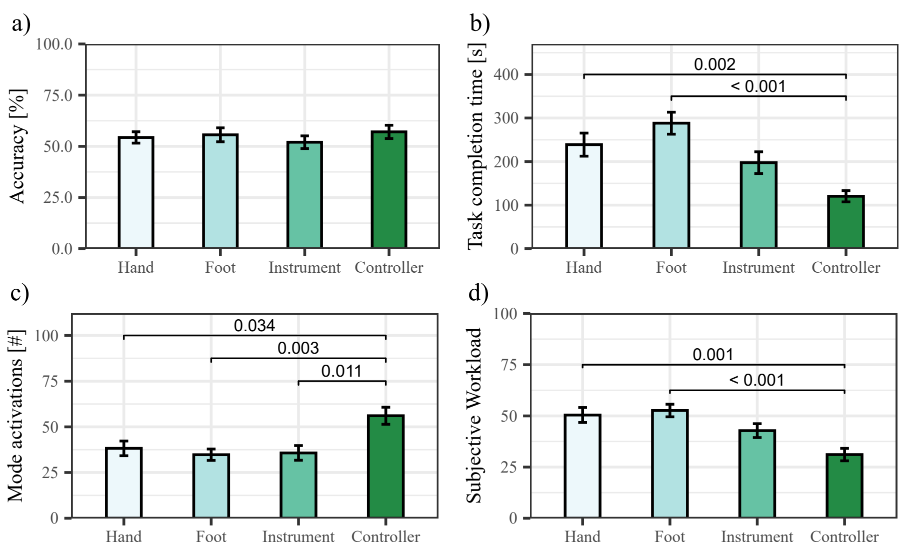

Fig. 5. Clipping task results. Effects of interaction techniques on a) accuracy, b) task completion time*, c) mode activations* and d) subjective workload*. Error bars respresent standard errors. Horizontal lines indicate statistically significant post-hoc test results.

manipulation of one DoF at the same time. As a result, the concept required more mode activations for the windowing task than other methods. Moreover, some participants reported problems in successfully performing the toe tap activation gesture and found the heel rotation gesture to be exhausting, which is partially reflected by a higher subjective workload. Hand gestures performed better than foot gestures, but not as well as controller-based interaction. These differences may have been caused by the concept only mapping relative hand movement to parameters, as opposed by global hand movement implemented by the instrument and controller concepts. As a result, movement axes and and changed parameter dimensions not always coincided. Implemented mode activation gestures may have also influenced the results, as participants reported gesture recognition and latency problems. Additionally, performing the finger double tap gesture sometimes coincided with inadvertent manipulation of parameters. Interaction using a surgical instrument resulted in similar but slightly better results compared to the hand gesture concept. The concept combined aspects of the controller concept, i.e. global tool position used for parameter manipulation, and the foot gesture concept, i.e. the activation gesture. This may have contributed to the results because of reported issues regarding the toe tap recognition. With button presses, the controller concept provided for simpler mode activation methods. This may have influenced the concepts overall advantages in terms of task completion time and subjective workload and is potentially reflected by the method's higher number of total mode activations. The easier to perform gesture may have led to participants switching between modes more frequently, resulting in an overall faster and iterative work flow.

The choice of implemented mode activation gestures seems crucial for user performance during investigated tasks. Foot gestures and hand gestures were both affected by recognition problems. More training could have solved these issues. However, more research should be conducted in improving recognition robustness. Wen et al. [24] also proposed using hand gestures for manipulating projected radiological 3d images. They used a different set of gestures and tracking hardware, that may have performed differently in this work's evaluation. Future research could, therefore, investigate effects of different gestures and recognition systems. Since the foot gesture concept performed worst in the user study, linear parameters should be rather modified using hand gestures or hand-controlled devices because of more controllable DoF and less exhaustion. In these cases, foot gestures seem to more suited than hand gestures for mode activations because of possible inadvertent modification of linear parameters controlled by the same hand. Yet, future work should evaluate different alternatives to the evaluated toe tap. Voice commands were not included in this experiment because of low expected speech recognition rates in noisy operating rooms. However, they still may be viable mode activation alternatives [13]. Thus, voice control may complement identified limitations of the current setup and should, therefore, be investigated in a future iteration of this work.

A meaningful continuation of this work would also be a comparison of investigated methods and conventionally used approaches to interact with medical image data, i.e. mouse input and visualizations on a monitor. Compared to conventional mouse interaction, using hand gestures was shown to be advantageous for the rotation of 3D models in terms of speed and precision [12]. Hettig et al. [9] compared different hand gestures detected by the Leap Motion controller (Ultraleap, UK & USA) and the Myo armband with the conventional methods of joystick input and verbal task delegation for rotation and navigation tasks. They found that both evaluated touchless interaction methods were viable alternatives for the conventional approaches. Similar comparisons could be conducted regarding this work's methods.

The present study was conducted under laboratory conditions only. The integration of used hardware systems in clinical environments is dependent on spatial and sterility conditions. The proposed controller-based method entails some sterility issues, similar to those of mice and keyboards in the operating room. Initial findings suggest, that these problems may be solved by wrapping the controller in a sterile plastic bag. However, potential resulting performance issues would need to be examined in future research. The instrument-based method is dependent on external tracking systems. The currently used, cumbersome HTC Vive tracker would need to be replaced by different techniques. Surgical navigation systems usually involve infrared-based optical tracking cameras, which could be used for this method, as well [17]. Additionally, the implemented head tracking method using a head-worn helmet would need to be replaced for clinical environments, as well. Gierwialo et al. [5] describe an alternative solution requiring only a small marker attached to a headband, that could be more easily integrated into surgical workflow. Also, markerless methods using depth cameras, like the Microsoft Kinect sensor (Microsoft Corporation, USA) could be implemented [12, 24]. Finally, dynamic patient registration methods need to be integrated to the system to correctly superimpose projected images on moving patients. Tissue deformation and respiratory movement models would also be meaningful additions. However, this topic was not focused on in this work, because first research objectives only included interaction and visualization aspects. Therefore, more research is needed to evaluate the applicability of developed techniques for the intraoperative use.

In conclusion, this work represents an important step towards the development of interaction techniques for surgical AR and has the potential to foster the clinical integration and acceptance of advanced AR visualization techniques in the operating room.

5 Acknowledgments

This work was funded by the German Research Foundation (HA 7819/1-2 & LA 3855/1-2).

References

1. D'Agostino, J., Diana, M., Vix, M., Soler, L., Marescaux, J.: Three-dimensional virtual neck exploration before parathyroidectomy. New England Journal of Medicine

367(11), 1072–1073 (2012)
2. Egger, J., Gall, M., Wallner, J., Boechat, P., Hann, A., Li, X., Chen, X., Schmalstieg, D.: Htc vive mevislab integration via openvr for medical applications. PloS one **12**(3), e0173972 (2017)
3. Gallo, L.: A study on the degrees of freedom in touchless interaction. In: SIGGRAPH Asia 2013 Technical Briefs. p. 28. ACM (2013)
4. Gallo, L., De Pietro, G., Marra, I.: 3d interaction with volumetric medical data: Experiencing the wiimote. In: Proceedings of the 1st International Conference on Ambient Media and Systems. Ambi-Sys '08, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL (2008)
5. Gierwiało, R., Witkowski, M., Kosieradzki, M., Lisik, W., Groszkowski, Ł., Sitnik, R.: Medical augmented-reality visualizer for surgical training and education in medicine. Applied Sciences **9**(13), 2732 (2019)
6. Goth, G.: Brave nui world. Communications of the ACM **54**(12), 14–16 (2011)
7. Hatscher, B., Luz, M., Hansen, C.: Foot interaction concepts to support radiological interventions. i-com **17**(1), 3–13 (2018)
8. Hettig, J., Mewes, A., Riabikin, O., Skalej, M., Preim, B., Hansen, C.: Exploration of 3d medical image data for interventional radiology using myoelectric gesture control. In: Proceedings of the Eurographics Workshop on Visual Computing for Biology and Medicine. pp. 177–185. Eurographics Association (2015)
9. Hettig, J., Saalfeld, P., Luz, M., Becker, M., Skalej, M., Hansen, C.: Comparison of gesture and conventional interaction techniques for interventional neuroradiology. International journal of computer assisted radiology and surgery **12**(9), 1643–1653 (2017)
10. Jalaliniya, S., Smith, J., Sousa, M., Büthe, L., Pederson, T.: Touch-less interaction with medical images using hand & foot gestures. In: Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication. pp. 1265–1274 (2013)
11. Jankowski, J., Hachet, M.: A survey of interaction techniques for interactive 3d environments. In: Eurographics (2013)
12. Kirmizibayrak, C., Radeva, N., Wakid, M., Philbeck, J., Sibert, J., Hahn, J.: Evaluation of gesture based interfaces for medical volume visualization tasks. In: Proceedings of the 10th international conference on Virtual reality continuum and its applications in industry. pp. 69–74. ACM (2011)
13. Mentis, H.M., O'Hara, K., Gonzalez, G., Sellen, A., Corish, R., Criminisi, A., Trivedi, R., Theodore, P.: Voice or gesture in the operating room. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. pp. 773–780. ACM (2015)
14. Mewes, A., Hensen, B., Wacker, F., Hansen, C.: Touchless interaction with software in interventional radiology and surgery: a systematic literature review. International journal of computer assisted radiology and surgery **12**(2), 291–305 (2017)
15. Mewes, A., Saalfeld, P., Riabikin, O., Skalej, M., Hansen, C.: A gesture-controlled projection display for ct-guided interventions. International journal of computer assisted radiology and surgery **11**(1), 157–164 (2016)
16. Nicolau, S., Soler, L., Mutter, D., Marescaux, J.: Augmented reality in laparoscopic surgical oncology. Surgical oncology **20**(3), 189–201 (2011)
17. de Oliveira, M.E., Debarba, H.G., Lädermann, A., Chagué, S., Charbonnier, C.: A hand-eye calibration method for augmented reality applied to computer-assisted orthopedic surgery. The international journal of medical robotics and computer assisted surgery **15**, e1969 (2019). https://doi.org/10.1002/rcs.1969

18. Pfeiffer, M., Kenngott, H., Preukschas, A., Huber, M., Bettscheider, L., Müller-Stich, B., Speidel, S.: Imhotep: virtual reality framework for surgical applications. International journal of computer assisted radiology and surgery **13**(5), 741–748 (2018)
19. Silén, C., Wirell, S., Kvist, J., Nylander, E., Smedby, Ö.: Advanced 3d visualization in student-centred medical education. Medical teacher **30**(5), e115–e124 (2008)
20. Silva, É.S., Rodrigues, M.A.F.: Design and evaluation of a gesture-controlled system for interactive manipulation of medical images and 3d models. SBC Journal on Interactive Systems **5**(3), 53–65 (2014)
21. Sugimoto, M., Yasuda, H., Koda, K., Suzuki, M., Yamazaki, M., Tezuka, T., Kosugi, C., Higuchi, R., Watayo, Y., Yagawa, Y., et al.: Image overlay navigation by markerless surface registration in gastrointestinal, hepatobiliary and pancreatic surgery. Journal of Hepato-Biliary-Pancreatic Sciences **17**(5), 629–636 (2010)
22. Thomas, R.G., William John, N., Delieu, J.M.: Augmented reality for anatomical education. Journal of visual communication in medicine **33**(1), 6–15 (2010)
23. Velloso, E., Schmidt, D., Alexander, J., Gellersen, H., Bulling, A.: The feet in human–computer interaction: A survey of foot-based interaction. ACM Computing Surveys (CSUR) **48**(2), 21 (2015)
24. Wen, R., Nguyen, B.P., Chng, C.B., Chui, C.K.: In situ spatial ar surgical planning using projector-kinect system. In: Proceedings of the Fourth Symposium on Information and Communication Technology. pp. 164–171 (2013)