

Comparative Evaluation of Gesture and Touch Input for Medical Software

Patrick Saalfeld^{1,2}, André Mewes², Maria Luz³, Bernhard Preim¹, Christian Hansen²

Visualization Group, Otto-von-Guericke University Magdeburg, Germany¹

Computer-Assisted Surgery Group, Otto-von-Guericke University Magdeburg, Germany²

Department of Psychology and Ergonomics, Technical University Berlin, Germany³

Abstract

The interaction with medical software during interventions challenges physicians due to the limited space and the necessary sterility. Current input modalities such as touch screen control present a direct, natural interaction which addresses usability aspects but do not consider these challenges. A promising input modality is freehand gesture interaction, which allows sterile input and a possibly larger interaction space. This work compares gesture and touch input regarding *task duration* to perform typical intervention tasks and *intuitiveness*. A user study with ten medical students shows mostly significantly better results for touch screen interaction. Despite the advantages of freehand gestures, it is debatable whether these can compensate the better efficiency and usability results of touch screen interaction in the operating room.

1 Introduction

Many diseases are treated by interventions carried out by highly specialized radiologists. In contrast to open surgery, needles or catheters are moved through thin holes to the target anatomy, e.g., to drain an abscess or insert a stent to widen a narrow vascular structure. Due to the missing direct interaction with human tissue, constant imaging control is necessary to see the tip of the catheter or needle. Therefore, physicians need to interact with interventional imaging software. Patient-specific data visualized as volume renderings, 3D surface models and tomographic slice images is inspected and analyzed with different interaction techniques. For example, the physician needs to rotate the 3D model, select single structures, navigate through tomographic images and show additional information. Based on this, the physician, e.g., determines tumor locations and sizes, tracks surgery devices and decides which interventions are appropriate. Especially in critical situations, the challenging interaction with the

complex interventional imaging software needs to be efficient and usable. Currently, controlling the software is realized with different approaches:

- A technical assistant controls the software from a non-sterile room with voice commands from the surgeon (Hübler et al. 2014; O'Hara et al. 2014). This indirect interaction method is inefficient and error-prone, since misunderstandings easily occur.
- The physician moves to a separate room with a workstation and uses mouse and keyboard interaction (Hübler et al. 2014). Beneath the time for moving rooms, this approach is problematic due to the aspect of sterility. The physician needs resterilization, which not only leads to longer operation time but also increases the risk of possible infections. Additionally, the interaction with 3D visualizations by indirect mouse input is not ideal.
- The medical software in the operating room is controlled with a touch screen device. The touch screen is covered with a sterile transparent foil leading to reduced usability. Furthermore, dependent on the current position of the surgeon, the touch screen is out of reach. Therefore, the physician needs to move to it or lean over the operating table to interact with it. This is problematic due to the distraction of the workflow and the ergonomic disadvantage (Hanna et al. 1998; van Det et al. 2009; Mewes et al. 2015).

New interaction styles such as 3D User Interfaces (3DUIs) and Natural User Interfaces (NUIs) offer solutions to the described problems. This work focuses on freehand gestures which address the problems in the following ways. They allow direct interaction by the physician with the interventional imaging software without the necessity to delegate instructions to a technical assistant. The interaction allows more degrees of freedom (DOF) compared to, e.g., mouse input and therefore intuitive controlling of 3D models and navigation in 3D space. An important advantage is the touchless interaction which ensures sterility, and thus, lowers the infection risk of the patient. However, the constrained space due to the close proximity of the interventional team must be considered (O'Hara et al. 2014).

We employ and improved an existing gesture set presented in (Mewes et al. 2015) and compared it with touch screen interaction for interventional imaging software. Both input modalities are evaluated in a user study regarding quantitative and qualitative aspects: First, the duration to solve typical intervention tasks and secondly, the subjective consequences of intuitive use. Our work shows that the participants perform significantly worse with gesture interaction and rate the intuitiveness of touch screen interaction higher. To use the advantage of gesture interaction, longer training times and well selected gestures for different tasks are necessary.

2 Related Work

3DUIs are interfaces for realizing virtual 3D space, with a special set of input and output devices, interaction techniques, and metaphors (Bowman et al. 2004; Preim & Dachselt 2015). This work focuses on freehand gesture interaction and thus, on the input and metaphor aspects of 3DUIs. Since gesture-based input is also embedded in the field of NUIs,

3DUIs and NUIs overlap. This is supported by the fact that the user's behavior and feeling during interaction in NUIs should be close to real-world applications (Wigdor & Wixon 2011). Ritter et al. (2013) investigated the suitability of the WiiMote (Nintendo, Kyoto, Japan) to control a medical planning software during an intervention. The WiiMote was used to control a mouse cursor, hence, disadvantages such as indirect interaction and less DOF are adopted. Schwarz et al. (2011) tested gesture input in an operating room and point out that flexible and robust gesture detection helps to make the interaction and thus the interventional work more efficient. Mentis et al. (2012) present fieldwork observations in neurosurgery theatres which deal with touch and gesture interaction as a spatial concern, i.e. freehand gesture interaction supports a distal control of a medical device.

The tracking of hands is commonly realized with the motion sensing device Kinect (Microsoft, Redmond, USA). Alternatively, (Bizzotto et al. 2014) tested the Leap Motion Controller (LMC, Leap Motion Inc., San Francisco, USA) and point out better accuracy and shorter working distance compared to the Kinect. Therefore, we use the LMC in our work. There are several possibilities to obtain an appropriate gesture set. For example, Schwarz et al. (2011) individualized gestures for physicians with a gesture learning approach. This allows the integration of customized personal and workflow requirements. Alternatively, an existing gesture set can be used. We use gestures from (Mewes et al. 2015) for an intraoperative projection display prototype on the radiation shield of a multi-detector computed tomography scanner (MDCT). A user study demonstrated that this approach is useable by physicians. However, the robustness and intuitiveness need to be further improved, which is described in more detail in the next section.

3 Materials and Methods

This section describes the medical workflow and derives typical interaction tasks. For all tasks, gesture and touch-based interaction techniques are presented. After that, the experimental setup for the user study is explained, followed by the study design and the study procedure.

3.1 Medical Workflow and Interaction Tasks

Hübner et al. (2014) described and analyzed the workflow of interventional neuroradiology with frequent pattern mining. They revealed common tasks such as controlling operating room equipment, e.g., the operating table or the C-arm, a c-shaped computer tomography (CT) device. The C-arm is used to acquire computer tomographic images during the operation. The resulting data can be displayed in the operating room in different views: as *2D tomographic images* and a *3D model representation*. The surgeon needs to inspect this data to retrieve information about, e.g., the contrast agent and blood flow behavior in vessels or to determine current positions of operation devices such as a tracked ablation needle. For this, she interacts with the 2D and 3D representation of the acquired data. In the following, the

derived interaction tasks for the 2D tomographic images, the 3D model representation and both views are listed:

- 2D tomographic images:
 - Cycle through the stack of images
- 3D model representation:
 - Rotation around arbitrary axis,
 - Selection of structures
- Both views:
 - Trigger button selection, e.g., to show additional information or reset the scene,
 - Zoom in to interesting structures such as tumors,
 - Zoom out to get an overview,
 - Translation of the image position or the object position

These interaction tasks can be fulfilled with different devices and interaction techniques. In this work, state-of-the-art touch screen interaction is compared with gesture input.

Touch screen interaction. The touch-based control is modeled after interaction with modern interventional systems such as the *CAS-ONE Liver* (CAScination AG, Bern, Switzerland). The control is primarily based on pressing buttons. The cycling is realized with “up” and “down” buttons to change to the next or previous slice in the image stack. Also, the discrete zooming is realized with “+” and “-” buttons. There are three exceptions: the rotation of the 3D model and the translation of the 2D position are realized with drag or swipe interaction on the touch screen, and structures in the 3D view can be selected by touching on them.

Gesture interaction. The gesture-control is realized with an improved freehand gesture-set presented by Mewes et al. (2015). They introduced five gestures to control different interaction tasks, which are shown in Fig. 1. Their grab gestures to rotate the 3D model was modified due to robustness problems. Instead, the object can be continuously rotated through tilting a hand with all five fingers extended with 3DOF (*flying hand gesture*, Fig. 1(a)). A dead zone guarantees that no unwanted rotation is performed. Zooming and translation is available in both views and realized by virtually grabbing the objects on the screen and translating the hand forward/backward for zooming and left/right/up/down for translation (*fist gesture*, Fig. 1(b)). Cycling through the 2D image stack is provided through a *circle gesture* (Fig. 1(c)) with one extended finger. The user can influence the step size by varying the circle’s radius. A *click gesture* (Fig. 1(d)) is implemented for the selection of structures or buttons. To select an object, the user has to extend the index finger and thumb, point to the object and move the tip of the thumb to the knuckle of the middle finger. If no action is wanted by the user, a relaxed hand can be used as a *rest gesture* (Fig. 1(e)).

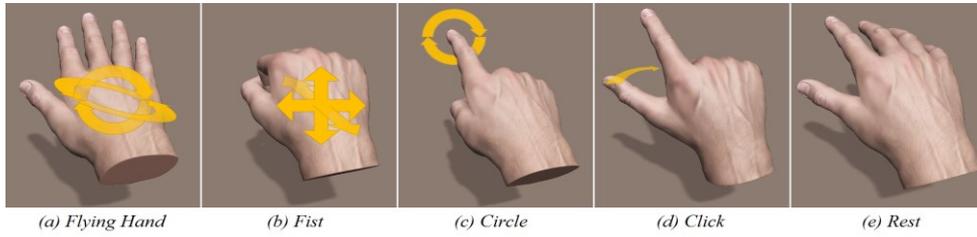


Fig. 1: The modified freehand gesture set presented from Mewes et al. (2015) with an improved rotation gesture for 3D rotation (a). Instead of rotating the object with a grab gesture, the hand can be tilted to perform 3DOF rotation.

3.2 Experimental Setup

We used the touch screen of the state-of-the-art commercial surgical navigation system *CAS-ONE Liver* for the study. The display is a resistive medical touch screen (ELO 2400 LM 24", Elo Touch Solutions, Inc.), see Fig. 2. To reconstruct the intraoperative setting, a surgical table with a body phantom is placed in front of the user. For gesture control, the Leap Motion Controller (LMC) is used, which is an optical device for observing the user's hands and providing position and orientation data for palm, fingers, bones and joints, which are used to define hand and finger gestures. The LMC is put on the edge of the table within the user's range. Our prototype, which has one mode for touch and one mode for gesture interaction is displayed on the touch screen, which is covered with a sterile transparent drape such as in an operating room. Tomographic image slices and a 3D model from a human liver with a hepatocellular carcinoma serve as test dataset within our study.

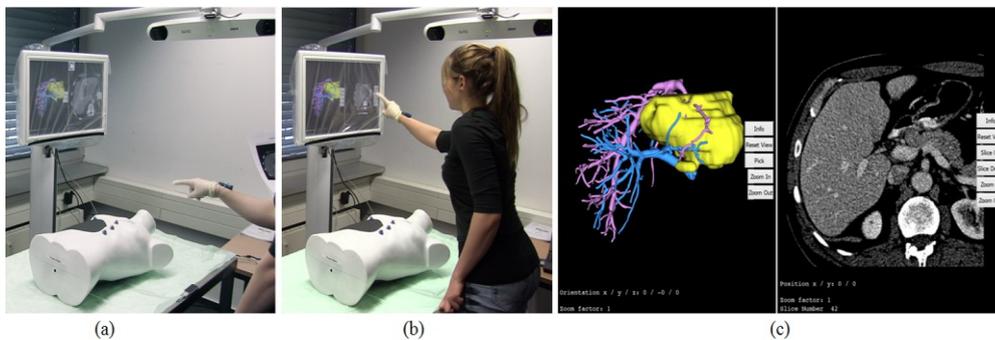


Fig. 2: Experimental setup with touch screen, plastic drape, surgical table and Leap Motion Controller (below hand, (a)). A user is performing tasks with freehand gestures (a) and touch screen interaction (b.) A screenshot of our prototype is shown in (c).

3.3 Study Design

The participants solved five tasks (see Table 1). These were selected based on observations in the operating room and on subsequent discussions with clinical partners. The first inde-

pendent variable is the interaction modality which has two levels: touch-based and gesture-based input. We consider the duration of the tasks as a first dependent variable and the intuitiveness of the two input modalities as a second dependent variable. For intuitiveness, we use the QUESI questionnaire (Questionnaire for the subjective consequences of intuitive use) (Hurtienne and Naumann 2010), which contains 14 items grouped into five sub-scales, such as *subjective mental workload* and *perceived achievement of goals* (see Fig. 5). The answer scale is a five-point Likert scale from 1 (fully disagree) to 5 (fully agree). The results of all items can be combined to a single score. Higher scores represents higher probability of intuitive use. The questionnaire is handed out for both input modalities which allows us to compare the two resulting scores.

The experiment is conducted as within-subject design, i.e., every participant fulfills the tasks with both input modalities. This prevents the influence of interpersonal differences. To avoid sequence effects, the experiment is performed as a crossover experiment. Thus, the order of input modalities changes. Here, we randomize the assignment of the order of input modalities. We ensure adaptive randomization (assignment depending on previous assignments) with *biased coin randomization* (Smith 2014). For every *odd* participant number (e.g., the first) a thrown coin decides the sequence of the input modalities: head means touch screen interaction first, then gesture interaction, tail means the opposite. For every *even* participant number, the coin is biased to favor the opposite result. Since we only have two order possibilities, the coin is biased to show tail or head with 100%.

Task	Description	Gesture	Touch interaction
1	Identify the range of slices in which a tumor is located in 2D	Circle gesture	Press buttons
2	Zoom the current 2D slice to factor 2.0 and center the tumor	Fist gesture	Press buttons
3	Rotate the 3D model identical to a given model	Flying hand gesture	Drag
4	Enlarge the 3D model to zoom factor 2.0	Fist gesture	Press buttons
5	Select the tumor in the 3D view	Click gesture	Press buttons

Table 1: Overview of the five tasks and the corresponding gesture or touch interaction to solve it.

3.4 Procedure

First, we handed out a pre-questionnaire with demographic questions and questions about the frequency of use (experience) with *interventional imaging software*, *gesture interaction*, *touch screen interaction on smartphones and tablets*, and *touch screen interaction in an operating room* on a scale from -2 (never) to 2 (very often). The experimental setup including the touchscreen and the leap motion controller were explained after that. Secondly, the medical viewer software was described including its different functionalities. Then, the participants were asked to put on rubber gloves and according to the result of the biased coin method, the participants started with one of the two interaction modalities. For both modali-

ties, the following sequence was the same: First, different functionalities were explained based on the input modality and corresponding gestures were shown, second, the participants could exercise the functions until they were confident in using them, third, the five tasks were stated subsequently and for each task the time was measured by analyzing video recordings and, fourth, the QUESI questionnaire was handed out. After the participants solved the tasks with the second modality and filled out the second questionnaire, they finished the study.

4 Results

The study was conducted with ten medical students (7 female, 3 male). Their age ranged from 20 to 27 years ($M = 22.7$ years) and one of them was left-handed. The experience with different modalities is shown in Fig. 3. The participants had little experience with interventional imaging software ($M = -1.3$, rarely; min: -2, max: 1), with gesture interaction ($M = -1.3$, rarely; min: -2, max: 2) and with touch interaction in the OR ($M = -1.5$, never; min: -2, max: 0). In contrast, they had more experience with touch interaction on smartphones or tablets ($M = 1.2$, often; min: -1, max: 2).

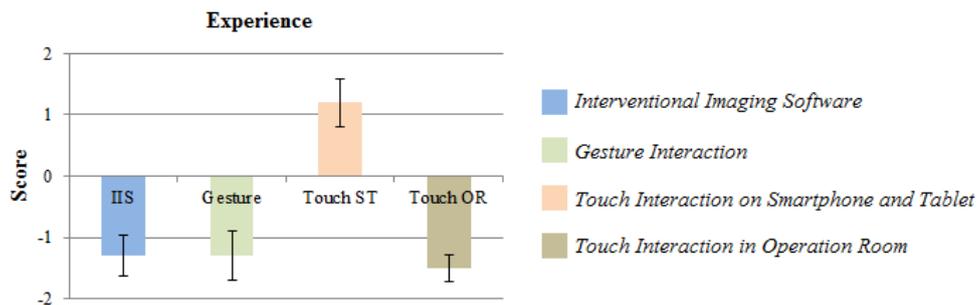


Fig. 3: Overview of the participants frequency of use with different systems and input modalities on a scale from -2 (never) to 2 (very often).

The training times for each modality were less than 10 min. The task duration was analyzed by a 2×5 (two conditions: gesture vs. touch \times five tasks) within-subjects ANOVA. The effects for task duration are shown in Fig. 4. Compared to touch interaction ($M = 25.4$ s, $SD = 35.3$ s), the participants needed almost twice as long to perform a task with freehand gestures ($M = 48.6$ s, $SD = 43.1$ s), reflected in a significant main effect of condition, $F(1,9) = 17.82$, $p < .01$, $\eta^2 = .66$. Further, the analysis revealed a significant main effect of task ($F(4,36) = 22.89$, $p < .01$, $\eta^2 = .72$), indicating the logical fact that different tasks require different times to be executed. The rotation of the 3D model (task 3: $M = 72.6$ s, $SD = 51.8$ s) takes the longest time, followed by the identification of the tumors' range of slices (task 1: $M = 62.0$ s, $SD = 44.4$ s), zooming of a 2D slice and centering the tumor (task 2: $M = 25.0$ s, $SD = 15.7$ s). In contrast, the tasks to zoom the 3D model (task 4: $M = 12.8$ s, $SD = 10.9$ s) and select

the tumor (task 5: $M = 12.6$, $SD = 17.2$) were performed very fast. There was no significant interaction effect ($F(4,36) = 2.32$, $p = .14$, $\eta^2 = .21$), although Fig. 4 implies this: while there seems to be no difference between the two conditions for rotation of the 3D model (task 3: $M = 68.5$ s, $SD = 57.9$ s vs. $M = 76.7$ s, $SD = 47.8$ s), participants need much longer to identify the range of slices with gestures than with touch interaction (e.g. task 1: $M = 90.9$ s, $SD = 45.2$ s vs. $M = 33.0$ s, $SD = 16.0$ s).

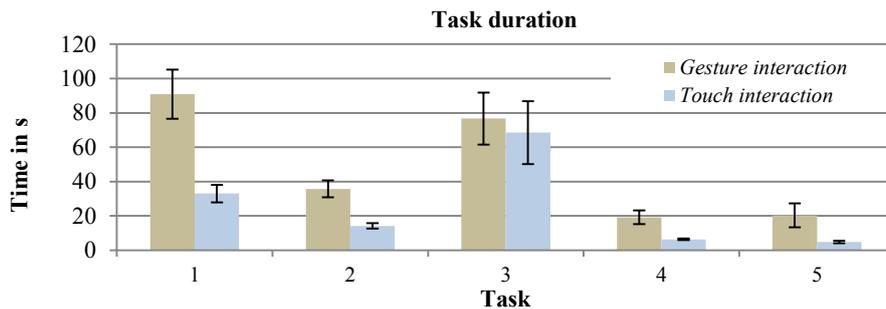


Fig. 4: Comparison of the duration of tasks with freehand gesture and touch screen interaction.

The intuitiveness measured by the QUESI questionnaire was analyzed with the Wilcoxon signed-rank test (see results in Fig. 5). Overall, users found touch interaction ($M = 4.2$, $SD = 0.5$) more intuitive than gesture interaction ($M = 3.5$, $SD = 0.7$), reflected in a significant effect (Wilcoxon- $U = -2.5$, $p < .01$). After the Bonferroni adjustment of the alpha level for the QUESI sub-scales, significantly higher scores emerged for touch interaction in comparison to gesture interaction only for two dimensions: *mental workload* (Wilcoxon- $U = -2.5$, $p < .01$; $M = 4.1$, $SD = 0.5$ vs. $M = 3.2$, $SD = 1.0$) and *familiarity* (Wilcoxon- $U = -2.7$, $p < .01$; $M = 4.3$, $SD = 0.6$ vs. $M = 3.4$, $SD = 0.9$). However, the data shows a trend for less *perceived effort of learning* for touch interaction (Wilcoxon- $U = -2.4$, $p = .01$; $M = 4.4$, $SD = 0.5$ vs. $M = 3.4$, $SD = 1.0$). There was no significant effect for the subscales *perceived achievement of goals* and *perceived error rate*.

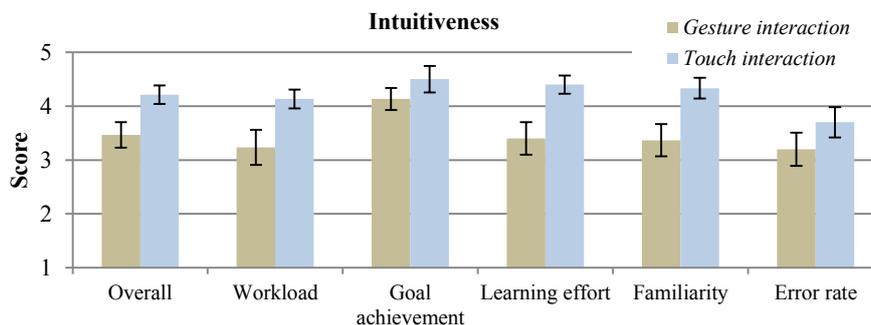


Fig. 5: Comparison of the intuitiveness of freehand gesture and touch screen interaction.

5 Discussion

Our participants were medical students and thus, had less experience in interventional settings compared to physicians. However, the medical knowledge necessary to fulfill the tasks is fairly basic. The participants showed significantly worse performance with gestures in almost all tasks. Only for 3D rotation there was no significant difference between the two conditions for the task duration. This indicates that for more complex interaction tasks, higher degrees of freedom of freehand gesture interaction can compete with touch interaction. Another fact during gesture interaction influenced the task duration, which is an important indicator for workflow efficiency: some users forgot about the correct execution of gestures, which lead to longer task durations. This issue could be avoided if they had more training time with the gesture interaction.

Although the effect size is relatively high (21 % of explained variance), an interaction effect missed to become significant. Due to the small sample size of ten participants, only very large effects can be identified. With a few more participants, the found interaction has a good chance to become significant.

The advantages of touch compared to gesture interaction was also found in terms of intuitiveness, i.e., the subconscious application of prior knowledge that leads to effective interaction. This explains the significant and marginally significant differences of interaction types on the dimensions *workload*, *learning effort* and *familiarity*. Indeed, if one considers the participants' experience with interaction types (Fig. 3), it stands out that the participants have strong experience with touch interaction and very little experience with gesture interaction, which may also have influenced the performance. However, no subjective differences emerged in terms of effectiveness (*goal achievement* and *error rate*).

Freehand gesture interaction ensures sterility, enables a larger working space, provides more degrees of freedom, and compensates disadvantages of touch screen interaction such as the need for plastic foil and a handicap due to interaction with rubber gloves. Still, touch screen interaction is superior regarding efficiency. To improve freehand gesture interaction, the gesture set needs to be improved regarding robustness and error tolerance and the participants need longer training times to equate lesser experience. Further studies could be performed with physicians. Here, it would be interesting to evaluate if more experience in interventional settings had an influence on the difference between the two input modalities.

Acknowledgements

This work was partly funded by the German Federal Ministry of Education and Research (BMBF) within the research campus STIMULATE under grant number '13GW0095A'

Contact information

saalfeld@isg.cs.uni-magdeburg.de

References

- Bizzotto, N., Costanzo, A., Bizzotto, L., Regis, D., Sandri, A., & Magnan, B. (2014). Leap Motion Gesture Control with OsiriX in the Operating Room to Control Imaging: First Experiences During Live Surgery. *Surgical Innovation* 21 (6), pp. 655–656.
- Bowman, D. A., Kruijff, E., LaViola J. & Poupyrev, I. (2004). *3D User Interfaces: Theory and Practice*. Redwood City, CA, USA: Addison-Wesley.
- Hanna, G. B., Shimi, S. M. & Cuschieri, A. (1998). Task performance in endoscopic surgery is influenced by location of the image display. *ANNALS OF SURGERY* 227, pp. 481–484.
- Holliman, N.S., Dodgson, N.A., Favalora, G.E. & Pockett, L. (2011). Three-Dimensional Displays: A Review and Application Analysis. *IEEE transactions on Broadcasting* 57 (2). pp. 362 - 371.
- Hübler, A., Hansen, C., Beuing, O., Skalej, M. & Preim, B. (2014). Workflow Analysis for Interventional Neuroradiology Using Frequent Pattern Mining. In Feußner, H. (Hrsg.): *Computer- und Robotergestützte Chirurgie. München (CURAC)*. pp. 165–168.
- Hurtienne, J. & Naumann, A. (2010). QUESI—A Questionnaire for Measuring the Subjective Consequences of Intuitive Use. In *Interdisciplinary College*, p. 536.
- O'Hara, K., Gonzalez, G., Sellen, A., Penney, G., Varnavas, A., Mentis, H., Criminisi, A., Corish, R., Rouncefield, M., Dastur, N., & Carrell, T., (2014). Touchless Interaction in Surgery. *In Commun. ACM*. 57(1). pp. 70-77.
- Mentis, H., O'Hara, K., Sellen, A., & Trivedi, R., (2012). Interaction proxemics and image use in neurosurgery. *In Proc.of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 927-936.
- Mewes, A. & Saalfeld, P., Riabikin, O., Skalej, M. & Hansen, C. (2015). A Gesture-Controlled Projection Display for CT-Guided Interventions. In Lemke H. U. (Hrsg.): *Computer Assisted Radiology and Surgery (CARS)* 29.
- Preim, B., & Dachselt, R., (2015). *Interaktive Systeme - Band 2: User Interface Engineering, 3D-Interaktion, Natural User Interfaces*. Berlin, Germany: Springer Vieweg Verlag.
- Ritter, F., Hansen, C., Wilkens, K., Köhn, A. & Peitgen, H.-O., (2013). User Interfaces for Direct Interaction with 3D Planning Data in the Operating Room. *i-com: Vol. 8, No. 1*. München: Oldenbourg Wissenschaftsverlag GmbH. (S. 24-31).
- Schwarz, L. A. & Bigdelou, A. & Navab, N. (2011). Learning Gestures for Customizable Human-Computer Interaction in the Operating Room. In Fichtinger, G. & Martel, A. & Peters, T. (Hrsg.): *In: Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 129–136.
- Smith, M. D. (2014): Biased Coin Randomization. In *Methods and Applications of Statistics in Clinical Trials*. John Wiley & Sons, Inc, pp. 90–105.
- van Det, M. J. & Meijerink, W. J. H. J., Hoff, C., Totté, E. R. & Pierie, J. P. E. N. (2009). Optimal Ergonomics for Laparoscopic Surgery in Minimally Invasive Surgery Suites: a Review and Guidelines. *Surgical Endoscopy* 23 (6), pp. 1279–1285.
- Wigdor, D. & Wixon, D. (2011). *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. San Francisco, CA, USA: Elsevier.