



Evaluating Usability of a Touchless Image Viewer in the Operating Room

Markus Bockhacker¹  Hannah Syrek²  Max Elstermann von Elster³  Sebastian Schmitt¹
Henning Roehl¹

¹Department of Orthopedic Surgery, Diakonissen Hospital, Mannheim, Germany

²mbits imaging GmbH, Heidelberg, Germany

³Department of Sports Surgery, ATOS Clinic, Heidelberg, Germany

Address for correspondence Markus Bockhacker, MD, Department of Orthopedic Surgery, Diakonissen Hospital Mannheim, Speyerer Str. 91-93, D-68163 Mannheim, Germany (e-mail: markus@bockhacker.com).

Appl Clin Inform 2020;11:88–94.

Abstract

Background Availability of patient-specific image data, gathered from preoperatively conducted studies, like computed tomography scans and magnetic resonance imaging studies, during a surgical procedure is a key factor for surgical success and patient safety. Several alternative input methods, including recognition of hand gestures, have been proposed for surgeons to interact with medical image viewers during an operation. Previous studies pointed out the need for usability evaluation of these systems.

Objectives We describe the accuracy and usability of a novel software system, which integrates gesture recognition via machine learning into an established image viewer.

Methods This pilot study is a prospective, observational trial, which asked surgeons to interact with software to perform two standardized tasks in a sterile environment, modeled closely to a real-life situation in an operating room. To assess usability, the validated “System Usability Scale” (SUS) was used. On a technical level, we also evaluated the accuracy of the underlying neural network.

Results The neural network reached 98.94% accuracy while predicting the gestures during validation. Eight surgeons with an average of 6.5 years of experience participated in the usability study. The system was rated on average with 80.25 points on the SUS.

Conclusion The system showed good overall usability; however, additional areas of potential improvement were identified and further usability studies are needed. Because the system uses standard PC hardware, it made for easy integration into the operating room.

Keywords

- ▶ orthopaedics
- ▶ machine learning
- ▶ neural networks
- ▶ gestures
- ▶ operating rooms

Background and Significance

Apart from surgical skills, preoperative planning and exact knowledge of patient-specific anatomy are needed to perform surgical procedures safely. In orthopaedic surgery, it is common to obtain images from diagnostic procedures like plain X-ray radiography, computed tomography (CT) scans and magnetic resonance imaging (MRI) during the preoperative planning stage. The availability of these images in the operating room has been identified as a factor for patient safety by the World

Health Organization¹ and is part of the preoperation checklist in our institution. During the procedure, these images serve as a reference to the surgeon to match the planned procedure to the surgical site at hand. In the field of spinal surgery, there is an inherent risk of “wrong level surgery” (falsely identifying the spinal region during the procedure),² so having CT and/or MRI images available during the procedure is a must. Traditionally, the patient’s X-ray films would be hung up in front of a lightbox in the operating room, but with the digitization of radiology departments, these lightboxes have been substituted by digital

received

September 30, 2019

accepted after revision

December 12, 2019

© 2020 Georg Thieme Verlag KG
Stuttgart · New York

DOI <https://doi.org/10.1055/s-0039-1701003>.
ISSN 1869-0327.

monitors connected to the hospital picture archiving and communication system (PACS). This offers the added benefit of being able to navigate (scroll) through a tomography data set while getting live-updating reference lines on a second, reconstructed plane. Also, the user is able to manipulate images by zooming into a region of interest (ROI) or changing contrast and brightness. All these features require the interaction of the user with the system, which leads to a unique set of challenges since a surgeon in the operating room is not able to use common input devices like keyboards and mice because of bacterial contamination.³ Apart from covering input devices with sterile plastic sheeting, researchers and medical technology companies have explored several alternative input methods, summarized in the concept of a “Contactless Operating Room.”^{4,5} The use of hand gestures is an established way of interacting with software in a sterile environment⁶ and their usage seems to be favorable compared to relaying verbal instructions to a non-sterile assistant.⁷ In the past, different methods of gesture recognition have been proposed and tested, ranging from the use of specialized hardware, like time of light cameras, down to regular off-the-shelf computer hardware. Both approaches have been demonstrated to be feasible and effective.^{8,9}

In 2017, Mewes et al conducted a systematic review of available evidence for the use of touchless human–computer interaction in sterile, medical environments.⁶ In their study, the majority of systems (34 studies or 62% of all examined studies) described systems to control medical image viewers. They concluded that further research should focus on evaluating usability since the general concepts have been established by now.

Methods

Hardware

The system uses off-the-shelf hardware, composed of a laptop with a dedicated graphics card, connected to an external 42-inch display and a 720p USB camera, both mounted at approximately eye-level height (→Fig. 1). To improve reliability, a wired Ethernet connection can be used.

Software

A prototype version of the mRay DICOM Viewer (mbits imaging GmbH, Heidelberg, Germany), running on the Windows operating system (Microsoft Corporation, Redmond, Washington, United States) was used. The prototype uses a deep learning algorithm, specifically the “Very Deep Convolutional Neural Network” called “VGG16” (Visual Geometry Group, University of Oxford¹⁰), to recognize hand gestures. This network was chosen because of its high performance on the “Imagenet” (Stanford Vision Lab, Stanford and Princeton University, United States¹¹) data set and its ranking in the “Large Scale Visual Recognition Challenge 2014.”¹¹ During the initial outset other machine learning algorithms were tested and showed less accuracy; however, we did not record these results.

The network was initially trained on the generic “Imagenet” data set and afterwards trained to recognize hand gestures with our train data set. This is a common approach in transfer learning.



Fig. 1 Surgeon in front of the hardware setup used in the usability study. The webcam is mounted above the screen.

Using the hardware described above, we took images of one surgeon performing hand gestures in various angles and distances from the camera (→Fig. 2). For each gesture, a series of images was recorded and each image was annotated with a label specific for the gesture. Afterwards the data set was split into a train and test data set. A total of 4,000 images were used to train the neural network and 1,000 were used to test the training process. An additional set of 1,300 images were reserved for later validation of the neural network.

Human–Computer Interaction

Using traditional mouse and keyboard input, the image data is loaded from the hospital PACS server ideally before surgery. Different grid layouts are available, wherein different imaging modalities like plain radiographs and tomography data sets can be displayed. To initiate interaction with the system, the user has to face the screen with his or her upper body (→Fig. 1). A set of five gestures is used to interact with the software (→Fig. 2). The user is able to cycle through four modes (scrolling through images, zooming, altering contrast, moving the ROI vertically, moving the ROI horizontally) and can modify the image in every mode. The system displays the current mode via onscreen text (→Fig. 3). This mode of operation including the gestures were selected in a joint effort by the software developers and M.B. during the study design and represents a compromise between recognizability and the range of motion a surgeon can perform in the operating room. To minimize the risk of accidental contamination, the surgeon’s hands should not leave chest level, so we chose our set of gestures accordingly (→Fig. 2), resulting in reduced movement

Gesture	Example 1	Example 2	Action
background			no action / stop
fist / fist			change mode
hand / fist			go left / go up
fist / hand			go right / go down
hand / hand			change active window

Fig. 2 Two sets of example images of hand gestures as fed to the neural network through the USB camera. A bounding box was used to exclude the surrounding environment.



Fig. 4 Spatial layout during the usability study. A surgeon performing a written test of attention.

of upper and lower arms compared with previous studies.^{12,13} Previous studies also described the use of single fingers to express gestures^{14,15} however, during initial testing we found that using single fingers for different actions were not detected reliably from different angles and distances.

Usability Study

We opted not to perform the usability study during regular hospital hours with patients present in the operating room. Instead, we used one of our operating rooms during the weekend and recreated the environmental conditions during a surgical procedure. Accordingly, surgery tables, lighting, sterile drapes, clothing, and gloves were used (→Fig. 4). We

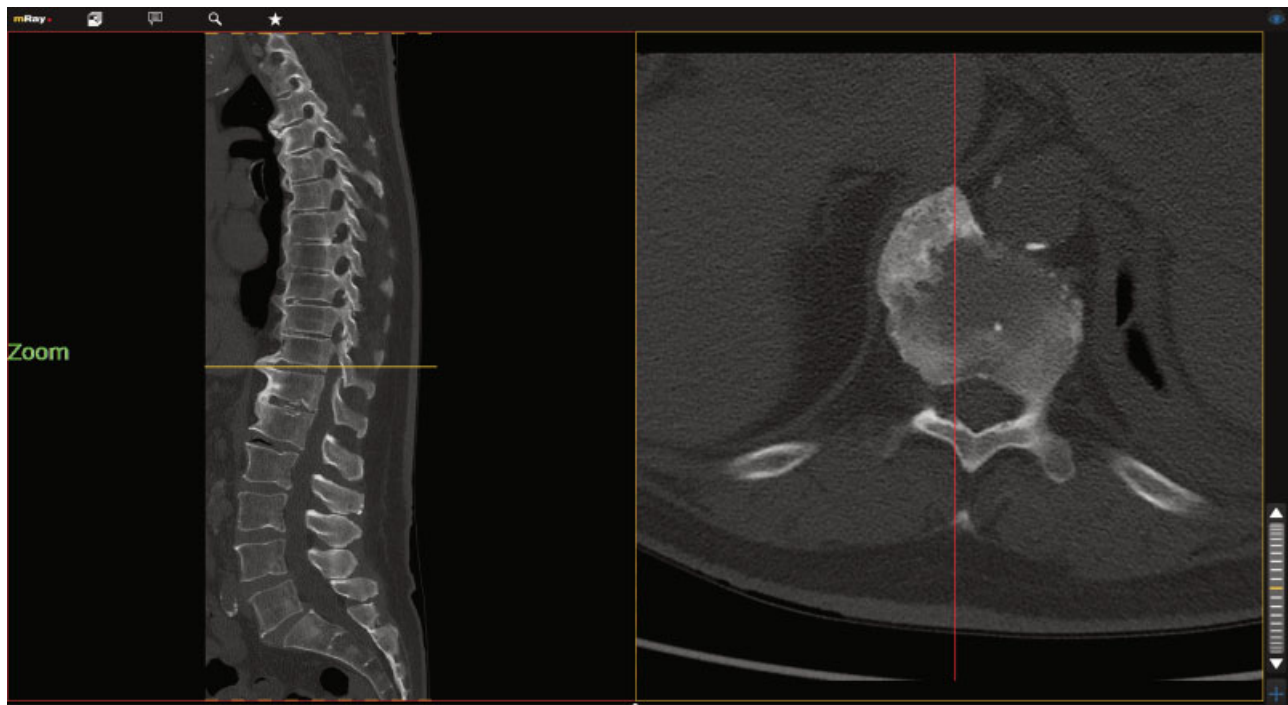


Fig. 3 Screenshot taken from the software system. We used a fully anonymized DICOM data set from a total spine computed tomography (CT) scan for every task in this study. The green text to the left indicates the mode currently selected: “Zoom.”

used a fully anonymized DICOM data set from a total spine CT scan with reconstructions in the transverse and sagittal plane (→Fig. 3). Participants were recruited from locally affiliated hospitals, provided they had experience in fields of orthopaedic surgery or neurosurgery and were available during the weekend. Participation was voluntary and without financial incentives. Every participant received a standardized introduction into the software and was then asked to perform a written test of selective and sustained attention and visual scanning speed (similar to a “D2-Test of Attention”) (→Fig. 4). We gave a time limit of 3 minutes, merely to provide an interlude between the introduction and the experiment. Hence, the results of these attention tests are irrelevant and will not be reported. After 3 minutes, the participant was asked to scroll the transversal plane of the CT images (→Fig. 3) to display both pedicles of the fourth lumbar vertebra. Upon completion of the first task, the participant was asked to scroll the transversal plane to show both pedicles of the second lumbar vertebra and to manipulate the sagittal plane such that the right pedicle of the second lumbar vertebra is clearly visible and the lumbar spine is zoomed in to exclude the thoracic spine. The time needed to complete each task was recorded and the simulation was reset for each participant. In summary, the first task consisted of one action (scrolling the transversal plane down toward the fourth lumbar vertebra) and the second task of five actions (scrolling the transversal plane, switching windows, scrolling the sagittal plane, moving the sagittal image upwards, and zooming in). Upon completing both tasks the participant was asked to fill out the System Usability Scale (SUS) questionnaire and general comments of the participants were noted in the study protocol.

Assessing Usability with the System Usability Scale

To ensure comparability with other studies, we chose the SUS¹⁶ to assess the overall usability of the system because it is widely used in the industry.¹⁷ The SUS is a 10-item questionnaire with Likert scales ranging from 1 to 5 for each item and is quick to administer.¹⁸ A German translation is freely available and was used in this study.¹⁹ The SUS has been empirically validated.²⁰

Results

Validation of the Neural Network

After 20 epochs the network can detect all trained gestures with an accuracy of 0.9894 on the given validation set, which included 1,300 images. To infer the correct classification of given images, the network requires 0.01 seconds per image on reasonably fast off-the-shelf PC hardware with six cores running at up to 3.6 GHz. →Fig. 5 shows the confusion matrix of classification during validation.

Usability Study

In total, we were able to recruit 8 physicians (2 female, 6 male), the mean age was 34.88 (range 30–44 years, median 33 years). The group was formed by five residents, one fellow, one attending, and the chief of medicine, with a

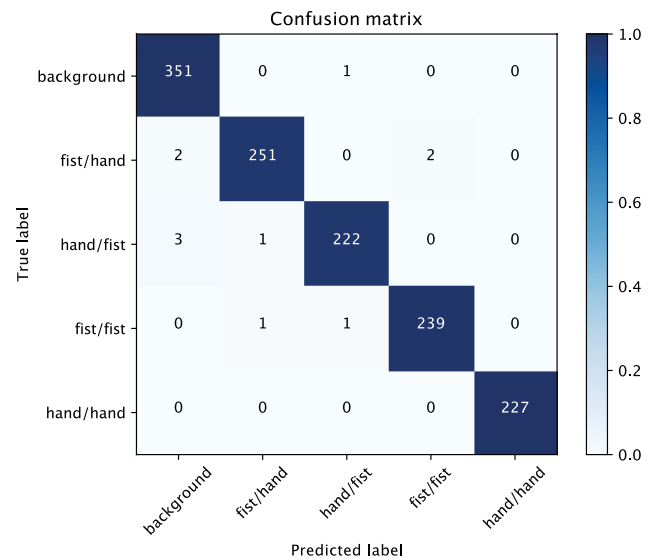


Fig. 5 Confusion matrix during validation of the neural network. The x-axis shows the true labels of gestures as recorded during image acquisition, the y-axis the labels predicted by the neural network during validation.

mean of 6.5 years of experience as physicians (range 1–17 years, median 5 years).

The first task took the group on average 114.4 seconds (range 62–202 seconds, median 101.5 seconds), while the second task took them on average 109 seconds (range 65–160 seconds, median 104.5 seconds) (→Fig. 6). The difference in means was not statistically significant ($p = 0.7301$, t -test for paired samples [two-sided] after testing for normality with Shapiro–Wilk’s test). We did not find any significant correlation between age or surgical experience and the time needed to complete the tasks.

On average, the group rated the system with 80.25 points on the SUS (range 70–93 points, median 81.5 points).

Four users noted, that the delay caused by switching modes was unfavorable, one user remarked, that it was unclear which window was active, and two users added that the reference lines were too small to be easily distinguishable from a distance.

Discussion

This pilot study describes the usability evaluation of a prototype touchless medical image viewer by surgeons in the operating room. Since the general concept of the “Touchless Operating Room” is not novel and several studies examined the use of hand gestures as a mode of touchless interaction before,^{6,8} we sought to focus our study on evaluating the system in a modeled close-to-real scenario by end users, because the need for usability testing has been pointed out previously.^{6,17,21,22}

Compared with previous studies we chose a different mode of human–computer interaction with a different set of gestures. These were informed by surgeons and the constraints in their range-of-motion in the operating room to reduce the risk of accidental contamination. Therefore, the

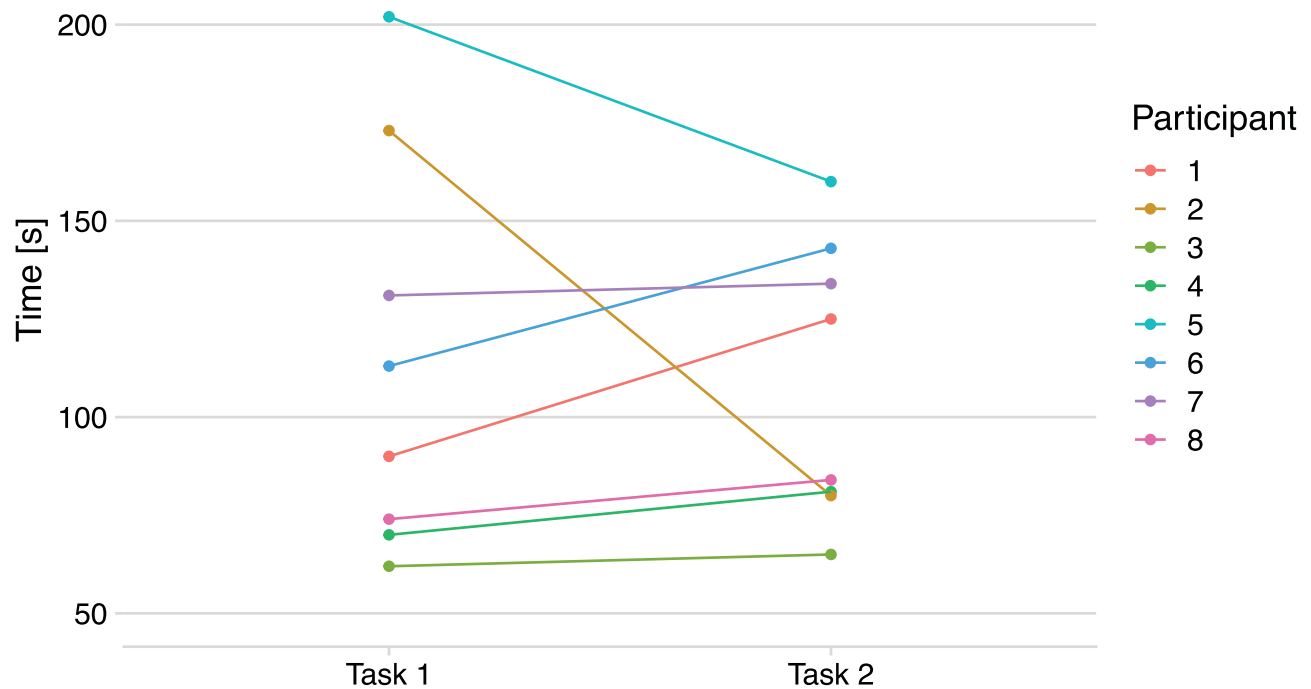


Fig. 6 Times taken for each participant to complete both tasks of the usability study. The difference in means between both tasks was not statistically significant ($p = 0.7301$).

surgeon's hands do not leave chest level when interacting with the system, a process that initially seemed unnatural to the developers but was accepted well by the users. To reduce complexity, the number of gestures was also reduced during the development process and we opted for a single gesture to cycle through modes of operation. This point was however specifically criticized by the users during usability testing because of the delay it caused. Previous systems used time-of-flight (TOF) cameras and segmentation to emulate a pointing device which allowed for different user interfaces.^{5,8,9} Our system was specifically designed to use standard PC hardware without the need for TOF cameras and is able to run on a laptop with a front-facing webcam, thus reducing investment cost. However, using a bigger screen and a camera mounted at eye level improved accuracy at greater distances from the screen during initial testing. During validation, the convolutional neural network reached 98.65% accuracy while classifying hand gestures and we feel comfortable in using this system outside of a laboratory environment. During the study, we were initially surprised by how long it took for the individual participants to complete each task, and the difference between the means of the times needed to complete task 1 and task 2 did not differ significantly (→ Fig. 6). Considering the fact, that task 2 requires five distinct actions and task 1 only one, this indifference could be a sign of a steep learning curve while using the system during the study. While similar effects have been demonstrated by other novel human-computer interfaces in surgery,²³ six of our eight participants noted that they disagreed or strongly disagreed with the statement that they did have to learn a lot before they could get going with the system (item 10 on the SUS). This needs to be investigated further with a bigger group of participants. Apart from the

small number of participants, this study has additional limitations. We used a specialized test scenario using a single-image data set, which makes it difficult to apply these results to different surgical specialties like vascular surgery. Also, the usage of additional usability questionnaires could potentially have allowed us to further investigate additional usability attributes like efficiency and errors.¹⁷ Although the results gained from the SUS (mean 80.25 points, median 81.5 points) suggest good overall usability, we gained valuable insights in how the system can be improved further and plan to conduct additional usability studies in the future using more complex scenarios (like switching between different imaging modalities) and different surgical specialties.

Conclusion

In summary, we were satisfied with the performance of the system and gained insight on how to further improve usability. Because the system uses standard PC hardware, it was very easy to integrate into our operating room. Additional usability studies will be needed using different scenarios and different medical specialties. We also want to encourage our fellow physicians to conduct or participate in usability studies under close-to-realistic conditions to help inform the development and improvement of novel systems.

Clinical Relevance Statement

Several studies in the past have focused on establishing new concepts of human-computer interaction for medical professionals, enabling the use of complex software systems in sterile environments. There is, however, evidence suggesting that not enough effort has been put into evaluating the

usability of these systems outside laboratory environments. This study demonstrates a usability evaluation of a prototype software system in a controlled, close-to-real environment by real end users (in this case surgeons).

Multiple Choice Questions

1. The problem of “wrong-level-surgery” in spinal surgery is defined as:
 - a. Using the incorrect height of the operating table during the procedure.
 - b. Using the incorrect tilt (leveling) of the operating table during the procedure.
 - c. Incorrectly identifying the spinal level (the vertebra) during the procedure.
 - d. Playing loud music during the procedure, thus limiting communication.

Correct Answer: The correct answer is option c. Falsely identifying the level of the spinal column during the operation is as bad as operating on the left knee instead of the right one. The surgeon needs to match the available images of the patient to the surgical site in the operating room to reduce the risk of wrongly identifying the spinal level. Apart from using intraoperative fluoroscopy, pre-operatively taken tomography images are used for reference by the surgical team.

2. The prototype software system used in this study is a neural network to classify the gestures performed by surgeons. The process of training this network can be classified as:
 - a. Supervised learning.
 - b. Semisupervised learning.
 - c. Unsupervised learning.
 - d. Meta-learning.

Correct Answer: The correct answer is option a. As described in “Methods,” we used 4,000 annotated images to train the network. Annotated means, every image is assigned its true label before training. This is a classic example of supervised learning.

3. The training process of the neural network used in the software prototype is a transfer learning (TL) approach. TL is a research problem in machine learning that:
 - a. Focuses on storing knowledge gained while solving one problem and applying it to a different but related problem.
 - b. Aims the implementation of frameworks that belong to the area of supervised learning.
 - c. Specializes the reduction of the number of extracted features to avoid overfitting throughout training.
 - d. Ensures that the model that results from learning is easily transferable to novel input variables.

Correct Answer: The correct answer is option a. In the context of this study, the neural network was initially trained using the Imagenet data set, which is a general-

purpose image data set. The knowledge gained from learning to recognize any kind of object in images is then applied to recognize specific hand gestures during further training.

Protection of Human and Animal Subjects

All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors. Informed consent was obtained from all individual participants included in the study.

Funding

None.

Conflict of Interest

M.B., M.E.v.E., S.S., and H.R. declare that they have no conflict of interest. H.S. is employed by mbits imaging GmbH as data scientist and contributed the technical evaluation of the software system.

Acknowledgments

We would like to thank our dear colleagues for their participation in this study. We would also like to thank mbits imaging GmbH for providing the system (hardware and software) for the duration of this study.

References

- 1 WHO Guidelines for Safe Surgery 2009: Safe Surgery Saves Lives. Geneva: World Health Organization; 2009. Available at: <http://www.ncbi.nlm.nih.gov/books/NBK143243/>. Accessed August 6, 2019
- 2 Grimm BD, Laxer EB, Blessinger BJ, Rhyne AL, Darden BV. Wrong-level spine surgery. *JBJS Rev* 2014;2(03):1
- 3 Schultz M, Gill J, Zubairi S, Huber R, Gordin F. Bacterial contamination of computer keyboards in a teaching hospital. *Infect Control Hosp Epidemiol* 2003;24(04):302–303
- 4 Cho Y, Lee A, Park J, Ko B, Kim N. Enhancement of gesture recognition for contactless interface using a personalized classifier in the operating room. *Comput Methods Programs Biomed* 2018;161:39–44
- 5 O'Hara K, Dastur N, Carrell T, et al. Touchless interaction in surgery. *Commun ACM* 2014;57(01):70–77
- 6 Mewes A, Hensen B, Wacker F, Hansen C. Touchless interaction with software in interventional radiology and surgery: a systematic literature review. *Int J CARS* 2017;12(02):291–305
- 7 Wipfli R, Dubois-Ferrière V, Budry S, Hoffmeyer P, Lovis C. Gesture-controlled image management for operating room: a randomized crossover study to compare interaction using gestures, mouse, and third person relaying. *PLoS One* 2016;11(04):e0153596
- 8 Alvarez-Lopez F, Maina MF, Saigí-Rubió F. Use of commercial off-the-shelf devices for the detection of manual gestures in surgery: systematic literature review. *J Med Internet Res* 2019;21(05):e11925
- 9 Strickland M, Tremaine J, Brigley G, Law C. Using a depth-sensing infrared camera system to access and manipulate medical imaging from within the sterile operating field. *Can J Surg* 2013;56(03):E1–E6

- 10 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *ArXiv14091556* Cs. September 2014. Available at: <http://arxiv.org/abs/1409.1556>. Accessed August 15, 2019
- 11 Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *ArXiv14090575* Cs. September 2014. Available at: <http://arxiv.org/abs/1409.0575>. Accessed August 17, 2019
- 12 Tan JH, Chao C, Zawaideh M, Roberts AC, Kinney TB. Informatics in radiology: developing a touchless user interface for intra-operative image control during interventional radiology procedures. *Radiographics* 2013;33(02):E61–E70
- 13 Jacob MG, Wachs JP, Packer RA. Hand-gesture-based sterile interface for the operating room using contextual cues for the navigation of radiological images. *J Am Med Inform Assoc* 2013;20(e1):e183–e186
- 14 Ebert LC, Hatch G, Thali MJ, Ross S. Invisible touch—control of a DICOM viewer with finger gestures using the Kinect depth camera. *J Forensic Radiol Imaging* 2013;1(01):10–14
- 15 Rossol N, Cheng I, Shen S, Basu A. Touchfree medical interfaces. *Conf Proc IEEE Eng Med Biol Soc* 2014;2014:6597–6600
- 16 Brooke J. SUS—a quick and dirty usability scale. In: *Usability Evaluation in Industry*. London, Bristol, PA: Taylor & Francis; 1996:189–194
- 17 Sousa VEC, Dunn Lopez K. Towards usable E-Health. A systematic review of usability questionnaires. *Appl Clin Inform* 2017;8(02):470–490
- 18 Brooke J. SUS: a retrospective. *J Usability Stud* 2013;8(02):29–40
- 19 Rummel B. System Usability Scale – jetzt auch auf Deutsch. SAP User Experience Community. Published January 13, 2015. Available at: <https://experience.sap.com/skillup/system-usability-scale-jetzt-auch-auf-deutsch/>. Accessed March 8, 2019
- 20 Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. *Int J Hum Comput Interact* 2008;24(06):574–594
- 21 Hultman G, Marquard J, Arsoniadis E, et al. Usability testing of two ambulatory EHR navigators. *Appl Clin Inform* 2016;7(02):502–515
- 22 Staggers N, Xiao Y, Chapman L. Debunking health IT usability myths. *Appl Clin Inform* 2013;4(02):241–250
- 23 Chaudhry A, Sutton C, Wood J, Stone R, McCloy R. Learning rate for laparoscopic surgical skills on MIST VR, a virtual reality simulator: quality of human-computer interface. *Ann R Coll Surg Engl* 1999;81(04):281–286