CrossMark

REVIEW ARTICLE

Touchless interaction with software in interventional radiology and surgery: a systematic literature review

André Mewes[1] · Bennet Hensen[2] · Frank Wacker[2] · Christian Hansen[1]

© CARS 2016

Abstract

Purpose In this article, we systematically examine the current state of research of systems that focus on touchless human–computer interaction in operating rooms and interventional radiology suites. We further discuss the drawbacks of current solutions and underline promising technologies for future development.

Methods A systematic literature search of scientific papers that deal with touchless control of medical software in the immediate environment of the operation room and interventional radiology suite was performed. This includes methods for touchless gesture interaction, voice control and eye tracking.

Results Fifty-five research papers were identified and analyzed in detail including 33 journal publications. Most of the identified literature (62 %) deals with the control of medical image viewers. The others present interaction techniques for laparoscopic assistance (13 %), telerobotic assistance and operating room control (9 % each) as well as for robotic operating room assistance and intraoperative registration (3.5 % each). Only 8 systems (14.5 %) were tested in a real clinical environment, and 7 (12.7 %) were not evaluated at all.

Conclusion In the last 10 years, many advancements have led to robust touchless interaction approaches. However, only a few have been systematically evaluated in real operating room settings. Further research is required to cope with current limitations of touchless software interfaces in clinical environments. The main challenges for future research are the improvement and evaluation of usability and intuitiveness of touchless human–computer interaction and the full integration into productive systems as well as the reduction of necessary interaction steps and further development of hands-free interaction.

Introduction

The key to successful medical interventions is immediate access to the patient's anatomical image data. Especially during minimally invasive procedures, in contrast to open surgery, physicians are not able to see their target, surrounding risk structures or their instruments inside the patient and therefore rely heavily on recent medical images and 3D models of the anatomy. Because of special working conditions in the operating room (OR) and interventional radiology suite, i.e., sterility, limited space and time pressure, physicians face challenging human–computer interaction tasks. These tasks include the control of medical image viewers, interactive registration of images and interaction with medical robots. In clinical routine, sterile covers enable the direct use of interaction devices, e.g., joysticks, touchscreens or control panels. In addition, foot pedals with little functionality are used to control software directly, but the interaction with software is still very often delegated to a nonsterile assistant using speech or gesture commands [25].

However, indirect interaction might be inefficient and error-prone. Technologies in the field of touchless human–computer interaction, e.g., range cameras, voice control or eye tracking, are promising. They present new ways of inter-

✉ André Mewes
andre.mewes@ovgu.de

[1] Faculty of Computer Science, University of Magdeburg, Magdeburg, Germany

[2] Institute for Diagnostic and Interventional Radiology, Medical School Hanover, Hanover, Germany

action with medical software under sterile conditions. Bauer et al. [5] already gave a first overview of touchless interaction in sterile environments. Nevertheless, they focused on body and hand gesture interaction and did not broaden the scope to other promising modalities, such as voice recognition. Another synopsis of touchless interaction has been given by O'Hara et al. [51], who also concentrate on body gestures, especially with the structured-light-based Microsoft Kinect 1 (Microsoft Corp., Redmond, WA, USA). They furthermore mention voice control as a useful addition to gesture input.

With respect to the growing interest in the area of touchless interaction in the OR we aim to give a broad and complete systematic overview of existing approaches that deal with the given interaction challenges in the classic and the hybrid OR. We additionally discuss the main problems and future trends in intraoperative touchless gesture interaction.

Methods

In the following, the literature search strategy and inclusion criteria are described.

Search strategy

A systematic literature search for scientific papers was conducted using the PubMed database. For this purpose the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines [42] were followed. Our PubMed search term consisted of 8 MeSH terms and 34 title/abstract search terms and is provided as supplementary material for this paper. Forward and backward search was performed using PubMed and Google Scholar. Therefore, similar, cited and citing papers of relevant literature from the PubMed database meeting the inclusion criteria (see below) were determined. Both literature search and review were performed by two reviewers independently.

Inclusion criteria

Relevant—in the sense of this literature review—are all scientific papers in English language that deal with prototypes as well as systems in productive use in the immediate environment of an OR which can be controlled partly or completely touchless and serve as aids to successfully complete the intervention. This includes voice recognition, body movement gestures and eye tracking.

Not included in this review is literature investigating the use of dictating software for radiological diagnosis as well as systems helping in rehabilitation and training or teaching (e.g., live streaming devices), since those approaches do not have a direct impact on the outcome of an intervention.
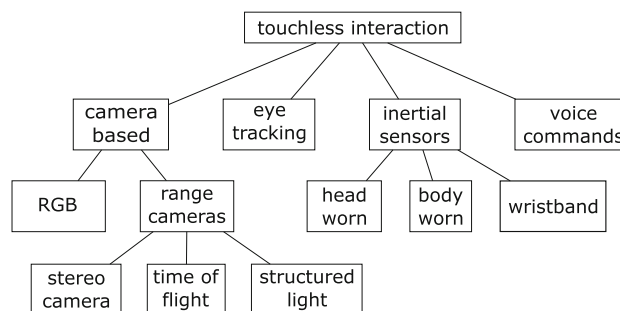

Fig. 1 Overview of used touchless interaction methods and devices

Results

The aforementioned search strategy yielded 403 references from the PubMed database of which 41 are relevant (see section "Inclusion criteria"). The Google Scholar search added 14 more papers of relevance for a total of 55 to be considered in this review. Thirty-three of the papers were published in peer-reviewed journals, and two are book chapters. Overviews of the implementations and evaluations of touchless gesture interaction systems for interventional use are presented in the tables in the corresponding sections containing a short description of the interaction type or device, the technical approach, and, if given, evaluation results of each paper.

Most of the authors describe methods for the touchless manipulation of medical image data (34 of 55). Other objectives are laparoscopic assistance (7), telerobotic assistance (5), OR control (5), robotic OR assistance (2) and intraoperative registration (2). The most popular device for touchless intraoperative gesture control is the Microsoft Kinect 1 structured-light-based range camera (21). Other relevant interaction devices or types in this review are stereo cameras (12)—9 of which are the Leap Motion Controller (LMC) (Leap Motion, Inc, San Francisco, CA, USA), body-worn inertial sensors (6), RGB camera or webcam (5), voice recognition (7), eye tracking (4), the Intel Realsense Creative (Intel Corporation, Santa Clara, CA, USA) structured-light camera (1) and a time-of-flight range camera (1). Figure 1 illustrates the connections between those devices and methods listed in this review. Most (40) of the systems described in the references have been evaluated under laboratory conditions or single experiments. Eight systems were tested in real interventions, and 7 research teams did not evaluate their work or did not provide information about it.

Two papers with the same content as Wachs et al. [72] and one paper similar to Jacob and Wachs [28] were excluded, because those did not make a contribution to the research. We decided to include the latest and most valuable of these publications in this literature review.

In the following, a synopsis of the relevant literature for each category is given. The papers are categorized by their objective in the OR, within which they are summarized according to the used interaction device in chronological order. Publications which provide new approaches or a significant contribution to the research area are described in greater detail than others.

Control of medical image viewers

A lot of fundamental research has been done on the interaction with the visualization of the patient's anatomy (see Table 1). A *camera-based* approach was followed by Wachs et al. [72], who developed a vision-based hand gesture and posture capture system to control a medical image viewer with 7 gestures. During calibration, the hand is segmented from a camera image by subtracting a detected moving blob from the image background. The hand color is saved in a histogram as a look-up table and used as reference in the gesture recognition process. The difference between two consecutive frames is computed and serves as motion cue. Within a defined interaction area the user can browse, zoom and rotate medical images. The usability was evaluated based on interviews with one surgeon and a questionnaire. According to this, the system is easy to use and has short training times at a recognition rate of 96 %. A similar system has been introduced by Achacon et al. [1]. Their hand gesture-controlled image viewer uses Haar-like features and the AdaBoost learning algorithm to train gestures as well as principal component analysis and distance matching to later recognize them in the camera image. The algorithm requires a clean background to work. Five unambiguous gestures were mapped onto the software functions. A 3-person experiment provided the findings that the gesture recognition works better in a well-lit (96–100 % recognition rate) than in a dark environment (16–96 %). The false-positive rate is high (64–75 % precision in well-lit environment, 35–53 % in dark environment).

A different technique to generate a depth map similar to the Microsoft Kinect 1 is the time-of-flight TOF method. Soutschek et al. [66] used such a *TOF camera* to define 5 hand gestures based on thresholds in the depth map as well as in the RGB image to interact with medical images. The user study with 15 subjects revealed a 94 % gesture classification rate and real-time capability (10 FPS). The users assessed the system to be intuitive, comfortable and with short response times.

Major issues of camera-based gesture control are line of sight and the fixed interaction area of the user. These problems can be avoided using *inertial sensors* worn on head, wrist or body, which enable a position-independent interaction. Schwarz et al. [64] introduced a technique which collects pose data from multiple body-worn inertial sensors and classifies them as low-dimensional body gestures. Those

are previously learned by the software and parameterized. This enables specialized and personalized gesture sets. A usability study with 10 subjects revealed a good wearability of the system and a 90 % recognition rate. This system was later extended by a voice-based and handheld switch unlock method by Bigdelou et al. [6]. Eight different gestures were defined and tested in a user study. The system does not inhibit usual movements of the users and is responsive and accurate. The handheld switch to unlock the interaction is preferred over the voice trigger, possibly due to a faster response time. Jalaliniya et al. [29] presented a single wristband sensor and SensFloor capacitive floor sensors (Future-Shape GmbH, Höhenkirchen, Germany) with 12 different universally defined hand and foot gestures. While foot gestures are for toggling and switching purpose, the hand gestures are used to interact with medical images. Single output artificial neural networks recognize the gestures in the sensor data stream. A user study with 5 subjects resulted in 93 % recognition precision and 98 % recall. The users described the system as precise, intuitive and responsive. An approach with a myo-electric armband (Myo armband, Thalmic Labs Inc., Kitchener, Ontario, Canada) was taken by Hettig et al. [22]. This armband has 8 surface electromyographic sensors that sense electrical signals from the muscle contractions of the forearm. Five gestures were mapped on four software functions, and haptic vibration feedback was implemented. Two user studies and one clinical test provided the knowledge that the device is not robust enough for clinical use (recognition rates 56–86 %) and has a high false-positive recognition rate.

In 2010, the introduction of the Microsoft Kinect 1, an inexpensive consumer market *structured-light-based depth sensor* and RGB camera with a source development kit (SDK) and user tracking and voice input capabilities, made it possible to easily build touchless gesture and voice-controlled interfaces. Kirmizibayrak [33] first presented a comparison of a two-handed Kinect 1 3D rotation and target localization (2D slicing) interface for medical images with mouse control. A user study with 15 participants revealed that the two-handed gesture control outperforms mouse interaction in rotation tasks in terms of accuracy and task completion time; mouse interaction is slower but more accurate when localizing targets. Ebert et al. [14] introduced a different medical image viewer control. Voice commands and range camera input were mapped onto keyboard and mouse events. The voice recognition is provided by the operating system. With vocal commands, the interaction modes can be switched. The images are manipulated (windowing, scrolling, moving) by arm gestures. A comparison with mouse interaction in a user study with 10 subjects resulted in mouse interaction being 1.4 times faster than gesture interaction and an overall usability rating of 3.4 out of 5, 3.4 for accuracy of the gesture control and 3 for accuracy of the voice control. A very similar concept was presented in Suelze et al.

Table 1 Touchless control of a *medical image viewer*, journal publications are marked with *

	Technical approach	Evaluation results
	RGB camera	
*Wachs et al. [72]	Color motion cues	96 % gesture accuracy, fast response, easy to use
Achacon et al. [1]	Vision-based hand gesture detection	64–75 % precision in well-lit environment
	Time-of-flight camera	
Soutschek et al. [66]	Threshold-based hand gesture recognition, 5 gestures	94 % recognition rate, 10 FPS
	Intertial sensors	
Schwarz et al. [64]	Trained arm gestures	Recognition rate >90 %, good wearability
Bigdelou et al. [6]	[64] + voice or switch unlock	Handheld switch preferred over voice unlock
Jalaliniya et al. [29]	Wristband + capacitive floor sensors	Precision >93 %, recall >98 %
Hettig et al. [22]	(Myo) gestures from myoelectric signals of the forearm	Recognition rates 71–86 %, high false-positive rate
	Structured-light sensor (Kinect 1)	
Kirmizibayrak et al. [33]	Comparison of two-handed Kinect 1 interface with mouse	Kinect 1 interface faster and more accurate in rotation
*Ebert et al. [14]	Gestures mapped onto mouse events	1.4 times faster than mouse interaction
*Ruppert et al. [60]	Self-defined arm gestures	–
Gallo [16]	Comparison of Kinect 1 and trackball interaction	Kinect 1 faster in ballistic phase, trackball more precise
*Hötker et al. [24]	6 gestures and 6 voice commands	97 % voice recognition, 88 % gesture recognition
Riduwan et al. [58]	Vision- and depth-map-based gesture detection	–
*Strickland et al. [68]	Hand gestures mapped on mouse events, user feedback on 2nd display	Robust, reliable, especially in darker environment
Suelze et al. [69]	5 threshold-based gestures and voice commands	–
*Tan et al. [70]	Arm gestures based on skeletal data	69 % of subjects would use the system
*Yusoff et al. [79]	4 arm and hand gestures based on skeletal data	–
*Kocev et al. [34]	Projector calibrated with Kinect 1, multi-touch gestures on projection	Algorithm is real-time capable
*Chao et al. [10]	Comparison of different gesture input devices	Comparable accuracy, Kinect 1 significantly slower, iPad has highest usability score
*Jacob and Wachs [28]	Intent recognition via body orientation	99 % accuracy
*Silva et al. [65]	Own skeletal tracking implementation	Visual feedback important, learning effect noted
*Nouei et al. [50]	Hand and finger gestures + RFID role tracking	Finger tracking preferred, direct, centralized access
*Wipfli et al. [77]	Comparison of Kinect 1 interface, task delegation and mouse	Mouse significantly faster than gesture control and delegation, no significant difference in effectiveness
	Stereo camera	
Kipshagen et al. [32]	2 webcams, computer vision hand recognition	Position error less than 2 cm in 96 % of cases
*Bizzotto et al. [7]	(LMC) 3 gestures mapped on keyboard shortcuts	No individual calibration, 5-min training time
*Ebert et al. [13]	(LMC) 5 gestures, one- and two-handed	Reliable with gloves, sensitive to smudges, short delay
*Mauser and Burgert [37]	(LMC) control of viewer and instruments, filtering of unintentional gestures	75 % recognition rate, high false-positive rate
*Rosa and Elizondo [59]	(LMC) hand and finger gestures	No technical errors, feasibility successfully tested
*Pauchot et al. [55]	(LMC) gestures defined with GameWave App	–
*Mewes et al. [41]	(LMC) CT radiation shield used as display, 5 gestures	Intuitive gestures, 3D rotation gesture ambiguous
Saalfeld et al. [61]	(LMC) comparison of improved gesture set of [41] and touchscreen	Freehand gestures significantly slower than touchscreen, except 3D rotation

Table 1 continued

	Technical approach	Evaluation results
*Opromolla et al. [52]	(LMC) custom gestures, focus on intuitiveness and memorability	Learning effect visible, natural interaction, too slow for real application, not robust enough
*Park et al. [53]	(LMC) universal gesture mapper	Recognition rate 77–100 %, significantly faster than mouse
	Voice commands and range camera (Kinect 1)	
Mentis et al. [39]	Gestures for continuous functions, voice commands as mode switch and trigger	Positive feedback from surgeon

[69]. Ruppert et al. [60] developed two solutions to interact with hand or arm gestures with the software. Hand recognition is realized with a depth threshold and by post-processing the noisy data. Then the center of gravity of the hand is calculated and used for cursor movement and mouse events. The OpenNI[1] framework and NiTE[2] (Kinect 1 and skeletal data tracking frameworks to work with the open source libfreenect driver) are the basis for arm gesture interaction. The user can move the mouse cursor with the right hand, while lifting the left arm triggers click events. The authors did not provide evaluation details.

The drawback of permanent user tracking and the resulting danger of triggering unintended actions was eliminated by Jacob and Wachs [28] by determining the user's intent with the torso orientation, previously executed commands and the time between subsequent commands. They created a Kinect 1-based gesture set, which is only active if the user is directed toward the display. A set of 10 gestures was chosen and trained with 10 surgeons. The interaction was evaluated in a user study with 20 subjects, which revealed a gesture recognition accuracy of 98 and 99 % intent recognition. Gallo [16] compared the Kinect 1 interaction with state-of-the-art trackball performance with a medical image viewer. A gesture set was developed, and the task completion time with 95 and 80 % pose accuracy was measured in two user studies. The author summarizes that the higher-degrees-of-freedom gesture set of the Kinect 1 performs better than trackball interaction method in the ballistic phase, i.e., 80 % pose accuracy and worse in the correction phase (90 %). Another comparison was drawn by Hötker et al. [24]. Six voice commands and six hand gestures with the same functions for medical image manipulation were implemented. A user study with 10 subjects indicated that voice commands (97 %) are better recognized than the body gestures (88 %) with an overall false-positive rate of 30 %. The Kinect 1, a gyroscopic mouse and a tablet PC were compared by Chao et al. [10] in a study with 29 users. Five tasks had to be executed with each device. The highest usability was measured for the

tablet (13.5 points), followed by the gyroscopic mouse (12.9 points) and the Kinect 1 (9.9 points). The task completion time was highest for the Kinect 1 (157s) and lowest for the tablet (41s). Only the measurement error did not differ significantly between those devices at about 1 cm each. Riduwan et al. [58] segment the user's hand from the depth image and use k-means clustering to find the hand pixels in the RGB image. After finding the hand's contours with a Graham scan and Moore-neighbor tracing, the fingertips are detected by finding the convex hull. Finger gestures were defined, but not evaluated. Strickland et al. [68] developed a mouse-emulating gesture control as well. It includes visual feedback on a second monitor and is tested in a study of six surgeries. The system is claimed to be robust and reliable, but no data were reported. Similar interaction concepts were introduced by Tan et al. [70] and Yusoff et al. [79]. Kocev et al. [34] combined the range camera with a projector and calibrated it as a spatial augmented reality system to interact with projected information on a deformable surface. A touchscreen-inspired multi-touch gesture set was implemented as well as contactless fingertip interaction. The algorithm execution time was evaluated and declared as real-time capable. Silva et al. [65] developed an own skeletal tracking method similar to OpenNI and NiTE. With this method, a gesture set was developed and evaluated in a user study with 16 subjects and 10 tasks. OR information software compatible with health level 7 (HL7) has been developed by Nouei et al. [50]. It provides all available data about a patient in one application, which can be controlled touchlessly by finger or hand gestures using the depth and pixel data of the hand. Additionally, radio-frequency identification (RFID) tags are used to determine the role of the current user. The evaluation was carried out during 30 surgeries. The surgeons and assistants pointed out the advantage of centralized, direct access to the patient data. Wipfli et al. [77] compared a Kinect 1 gesture interface with OR-typical interaction task delegation and mouse interaction with medical data. After a study with 30 participants they concluded that mouse interaction is significantly more efficient and has significant higher user satisfaction than gesture control and task delegation. However, there were no significant differences in error rates.

---

[1] http://structure.io/openni.

[2] http://www.openni.ru/files/nite/.

A depth map of the 3D space can also be obtained after calibrating two cameras as a *stereo camera* and calculating the disparity between the two images. Kipshagen et al. [32] introduced a medical image viewer that handles hand gestures with a stereo camera. The hand segmentation takes place after noise removal by segmenting the hand or glove colors, further removing noise and closing gaps. Low-frequency Fourier descriptors are used as unique feature vectors and compared to a previously trained feature database. By measuring distances and time between the frames, a hand's position offset and therefore velocity and direction can be detected. A user study with 15 subjects resulted in a position error of less than 2 cm in 96 % and less than 1 cm in 50 % of the cases. The processing is done in real time. A very similar functional principle underlies the LMC, which is basically a stereo camera with 3 infrared LEDs that illuminate the hand above to be segmented more easily from the images. Bizzotto et al. (2014) [7] first implemented an LMC-based gesture control plugin for a medical image viewer. They used the freely available GameWave App to define their gestures and map them to the software's functions. A study with 8 users was conducted, but no results or evaluation details were presented. The same approach was followed by Pauchot et al. (2015) [55] and compared the LMC with the Kinect 1 qualitatively. The authors assert that the LMC has a reduced work space, is less tiring, has greater precision and a much smaller casing. Ebert et al. [13] developed a two-handed gesture set for browsing and manipulating medical images, but did not evaluate their approach. However, the authors experienced a small delay before the gesture recognition and sensitivity to smudges on the device.

Mauser et al. (2014) [37] use the LMC to control medical instruments as well as a medical image viewer. A difference from other gesture sets is the lock and unlock gestures to avoid unintended gestures, as it was already suggested in [7]. No evaluation details were presented. Rosa et al. (2014) [59] tested the feasibility of medical image viewer control with the LMC during 11 dental surgeries. Hand gestures were developed as well as two-finger gestures for scaling, rotating, windowing, browsing images or measuring. The feasibility was proven without major technical errors. Mewes et al. (2015) [41] integrated a touchlessly controlled display into the radiation shield of a computed tomography (CT) angiography intervention room. The user interacts with 2D images and 3D planning models via hand gestures with the LMC. The gesture set was designed to be intuitive and metaphoric. The user study with 12 subjects showed robustness problems with the 3D rotation, although all gestures were rated intuitive and self-descriptive. Saalfeld et al. [61] improved the gesture set of Mewes et al. (2015) [41] (especially the 3D rotation) and compared it to state-of-the-art touchscreen interaction. In a study with 10 subjects the task duration and intuitiveness of the gesture set for medical image

manipulation were measured. The interaction with the LMC is significantly slower, except for 3D rotation, which leads to the conclusion that high-dimensional gestures are better for more complex interaction tasks. Additionally, the touchscreen interaction was described as more intuitive, which is partly ascribed to the more frequent use of touchscreens on smartphones and tablets. A two-handed gesture set with a similar focus was developed by Opromolla et al. [52]. The evaluation with 10 users led to the conclusion that the LMC is too slow, not robust and not flexible enough for use in the OR. An advantage is the natural interaction with the software. Park et al. [53] developed a universal LMC gesture mapper to work with arbitrary medical image viewers. Either two-handed gestures or one-handed gestures with a foot-pedal form the user interface. This is achieved by mapping hand gestures on mouse events. The system is modular and battery-powered to provide maximum flexibility. The evaluation with one surgeon resulted, unlike other publications, in the LMC being significantly faster than mouse interaction, which the authors explain with the possibility of concurrent zoom and rotation. However, the gesture recognition rate ranged from 77–100 %, with a false-positive rate of 52 % of the double click gesture.

Despite the high focus on body gestures in this subject, touchless interaction is not only possible through hand and arm movements, but also with *voice recognition systems*. Mentis et al. [39] partly integrated voice commands in their Kinect 1-based interaction and use them as function trigger or mode switch. Hand gestures can be used for continuous functions like browsing through a set of images. A surgeon examined the system qualitatively and found it useful for use in a clinical environment. However, no evaluation data were given.

Laparoscopic assistance

During laparoscopic interventions, the physician often needs an assistant to control the laparoscopic camera, the light or the insufflator. Interhuman communication in the OR lacks precision and requires much experience as a team. To eliminate the possible complications which an indirect control implies, El-Shallaly et al. [15] evaluated a commercial *voice recognition* interface. Via voice commands, the light can be activated, the camera is set up and white balanced and the insufflator can be controlled. After treating 100 patients with and without the system, the authors drew the conclusion that significantly less time needs to be spent for switching components on and off compared to manual control. Nevertheless, the authors underline that the absolute gain in efficiency is only about 1 min in total per operation. The same commercial system was evaluated by Salama and Schwaitzberg [63]. They investigated the availability of the system in comparison with an assistant. As a result, the nurse was not immediately ready

Table 2 Touchless control of *laparoscopic and endoscopic devices*, journal publications are marked with *

	Technical approach	Evaluation results
	Voice commands	
*El-Shallaly et al. [15]	Comparison of routine laparoscopic intervention with voice command-assisted light, camera and insufflator control	Significantly shorter equipment setup time
*Salama and Schwaitzberg [63]	Voice-activated control of light, camera and insufflator	Instant control of equipment, nurse is not present at 77 % of times commands were given
*Nathan et al. [46]	Voice control of endoscope holding and automatic positioning robot	Procedure not significantly longer, no second surgeon needed
	RGB camera	
*Nishikawa et al. [49]	Face motion control of endoscopic camera	Accurate, natural, more efficient than voice control
Wachs et al. [73]	Threshold-based head movement control of endoscopic camera	Slower than with keyboard, easier to learn
Yoshida et al. [78]	Switch multiple image streams with number of fingertips	5-s delay, reliable
	Inertial sensors	
Reilink et al. [57]	Head movements control gastroscope	Less space needed as with assistant, slight drift occurs

to execute commands in 77 % of the times voice commands were given to the voice control system. This implicates that voice commands can make a laparoscopic intervention more productive. Not only laparoscopic but also endoscopic procedures can benefit from voice recognition assistance. Nathan et al. [46] presented a robotic scope holder which is used to position an endoscope and controlled by the physician's spoken distinct commands. The system was evaluated with 10 cadaver heads. There is no significant increase in time to set up the endoscope or software. The major advantages are that the system does not misinterpret commands, like a second surgeon or assistant would, due to the direct interaction and that the robot is not affected by fatigue.

A laparoscopic camera can be controlled with head movements as well as with spoken commands. Nishikawa et al. [49] first developed a *camera*-based head movement control for this scenario. The user is monitored via an RGB camera, and head movements are interpreted as gestures and mapped on the laparoscopic camera actions tilt, pan, insert and retract. According to the laboratory experiments with three users the system is highly accurate and not misguiding. An in vivo experiment with a pig revealed signs of fatigue in the user's neck. Wachs et al. [73] implemented a similar solution. If the head angle is above a defined threshold, the camera turns to the desired direction. A simulated surgery with 4 users revealed that face orientation control is slower than keyboard control, but easier to learn. A drawback is the absence of a lock gesture or command, which makes unintended actions possible. Yoshida et al. [78] successfully tested their finger-based head-mounted display (HMD) view interaction during a laparoscopic intervention. The user is presented multiple views on the HMD, i.e., a video stream from the laparo-

scopic camera, medical image data and a video stream from the head-mounted camera. The number of the user's fingertips in front of the camera controls the viewports.

Reilink et al. [57] followed a similar approach as [49] and [73], but with *body-worn inertial sensors* on the head to track the physician's movements. A monitor can be used as display as well as a HMD. Three algorithms were implemented: position dependent, velocity dependent and hybrid movement control. The physician resets the initial position with a foot pedal. Fifteen subjects tested a two-directional gastroscope steering and preferred the velocity-dependent approach and the HMD. No delay between head and camera motion was noted. A disadvantage is that no information about the tip orientation is given, and thus, the users do not know which further movements are possible.

See Table 2 for a short summary.

Telerobotic assistance

Telerobotic surgery enables surgeons to conduct more precise and less invasive operations than conventional methods do. Nevertheless, the control consoles are complex and difficult to handle. To facilitate the telerobotic control, Mylonas et al. [45] developed an *eye-tracking*-assisted method to generate haptic constraints for the robot's movements based on the physician's gaze point (see Table 3). These constraints are experienced as haptic feedback with 6 degrees of freedom. The force opposed to the surgeon's movement is related to the distance between the eyes' fixation point and the surgical instrument and the underlying force profile (high, 1:1 scaling and linear spring force profile). This way, unwanted movements of the robot can be avoided. Ten subjects tested

Table 3 Touchless *telerobotic assistance*, journal publications are marked with *

	Technical approach	Evaluation results
	Eye tracking	
Mylonas et al. [45]	Gaze-point-generated haptic constraints using motor channeling	Improvement of task performance and accuracy, hands are not overpowered
Stoyanov et al. [67]	Gaze-based optimization of ablation path during robotic surgery	2.2-mm error, jitter-free 3D path
Visentini-Scarzanella et al. [71]	Gaze-contingent dense 3D reconstruction of structure surface in ROI	3.5-mm reconstruction error, decreased computational complexity
Clancy et al. [11]	Eye-tracker-based telerobot camera autofocus with liquid lens	Faster than manual mechanical focus, real-time capable
	Structured-light sensor (Kinect 1)	
*Wen et al. [75]	Hand gesture-controlled robotic surgery system with augmented reality visualization on patient	Less than 2-mm needle insertion error

the eye tracking and motor channeling and 6 additional subjects used it with a commercial telerobot. The linear spring force profile performed best. The hands of the users were not overpowered and no pre- or intraoperative registration was necessary. A similar system was used to optimize ablation paths on the surface of heart tissue with nonparametric clustering by Stoyanov et al. [67]. The surgeon's fixation points are determined by measuring the corneal reflection from a fixed infrared light source in relation to the center of the pupil. An user study with 8 subjects was conducted with a heart phantom model. The 3D path error was 2.2 mm; the path itself was jitter-free. Visentini-Scarzanella et al. [71] used this binocular eye tracking to localize the physician's region of interest (ROI) on deformable tissue, which enables a semidense stereo surface reconstruction with reduced computational complexity and better resolution of the desired area. A decent 3D reconstruction is mandatory for dynamic active constraints, motion stabilization and image guidance. The authors tested the method on a silicon heart phantom with 15 fiducials. CT-generated 2D images and the heart were temporarily and spatially aligned. The static reconstruction of smooth featureless areas showed a maximum error of 3.5 mm, and the real-time dynamic motion recovery error was 2.9 ± 2.3 mm. The difference between an eye-tracker-based autofocus and built-in foot-pedal-based mechanical focus was investigated by Clancy et al. [11]. A liquid lens at the end of the commercial telerobot endoscope was used to automatically focus the area the surgeon is fixating with the eyes. The evaluation with 17 subjects revealed that the eye-tracking autofocus method is not only faster but also feels more comfortable and natural to the users. The liquid lense response time was ~ 30 ms. One advantage over mechanical autofocus is that no moving parts are used, which implies a longer durability.

Another interaction method to control a robot is hand gesture-based with a *range camera*. Wen et al. [75] use a

Kinect 1 to recognize the physicians' hand gestures, which control a surgical robot to insert a needle into the operation field as well as a projected augmented reality needle guidance on the patient. Two modes of operation are possible: manual and semiautomatic generation of ablation paths. The automatically generated trajectory can be revised directly on the patient with the RFA planning models. The context can be selected with the palm; gestures are described by 90-dimensional feature descriptors. Twenty-two insertion tests were conducted to measure the accuracy of the whole system. The needle insertion error in a static scenario was less than 2 mm.

Robotic assistance

A different use of robots from actually operating on the patients is the assistance in the OR (Table 4). Li et al. [35] introduced a robotic scrub nurse, which hands medical instruments over to the physician after being instructed by hand gesture commands. A Microsoft Kinect 1 *range camera* is used to detect 5 different finger poses which represent a single instrument each, which will then be delivered to the surgeon. Usability tests with 4 subjects revealed a 97 % gesture recognition rate, 160-ms gesture recognition time and 5- to 6-s total interaction time including 2-s instrument delivery to a fixed spot. The delivery precision is 25 mm. Users rate the system moderately easy to use, remember and learn, moderately comfortable and safe. The robot delivery is 0.83s slower than human delivery.

Hartmann and Schlaefer [20] use a gesture-controlled robot to reposition the operating room light spot. An unlock gesture is used to activate the robot. After that, either the center of the palm or the center point between both hands is followed. Eighteen users were involved in the evaluation. A fixed track was to be followed by the light. Only one-handed interaction was tested for precision and speed. The system is

Table 4 Touchless control of *robotic OR assistance*, journal publications are marked with *

	Technical approach	Evaluation results
	Voice commands and range camera (Kinect 1)	
*Li et al. [35]	Robotic scrub nurse, controlled with 5 gestures and voice commands, hands over instruments	5- to 6-s total delivery time, 0.83 s slower than human, 25-mm delivery precision
	Structured-light sensor (Kinect 1)	
*Hartmann and Schlaefer [20]	Arm gesture control of OR lights	Direction-dependent error in hand tracking, effective, unlock gesture is robust

Table 5 Touchless *intraoperative registration*, journal publications are marked with *

	Technical approach	Evaluation results
	Inertial sensors	
Herniczek et al. [21]	Orientation sensor under sterile gloves, point placement for guidance in ultrasound (US)-guided nephrostomy	100 % recognition rate for back, reset and select gesture, 92 % for next gesture
	Structured-light sensor (Kinect 1)	
*Gong et al. [17]	Initial 2D–3D positioning for X-ray image and 3D volume registration, gestures based on skeletal and depth data of hands	Sufficiently accurate

robust, reliable, and the unlock gesture is suitable for clinical use.

Intraoperative registration

In addition to simple medical viewers, modern navigation systems also provide the possibility to plan needle insertion paths or register anatomical images from different imaging modalities. Herniczek et al. [21] investigated the use of *body-worn inertial sensors* on the hand under a sterile glove to place points for needle insertion guidance on ultrasound snapshots touchlessly. Four gestures were trained. No evaluation details were presented. The authors claim a gesture recognition rate of 100 % for all but one gesture (92 %).

Gong et al. [17] introduced an interactive 2D/3D registration method with a *depth-camera*-based hand gesture interaction. The user realizes an initial alignment of a 3D model to X-ray images with two gestures. The gestures are processed via the skeletal and depth data. A cursor is positioned directly via hand movements. Three users tested the system. The positioning error was 8.3 ± 5 mm in 140 ± 70 s positioning time.

Refer to Table 5 for a brief overview.

OR control

Sometimes the physician does want to control not only a specific navigation support or instrument touchlessly, but also other arbitrary software in the OR. To facilitate this, Graetzel et al. [18] developed a system based on a *stereo camera* to control any OR software contactlessly with hand gestures. The gestures are performed in a $50 \times 50 \times 50$ cm workspace and tracked and processed at 25 Hz. With this system, the user can move the mouse cursor and trigger mouse click events. The authors conducted a user study in the laboratory with 16 participants and a mock-up user interface. Click gestures were robust, and rapid gestures were not reliably detected. The pointer often jittered and hand-eye coordination problems occured. Another test during an intervention showed that the physicians preferred the hold-and-click gesture over pushing the hand forward. Grange et al. [19] extended this solution. The physician is permanently monitored by the stereo and an RGB camera to infer context information. This information can be used to automate the adaption of equipment settings. The whole process is workflow-step-aware. The authors did not evaluate their extension.

In many crisis situations the clinical staff in the OR have to use both hands to ascertain the patient's health. Therefore, Alapetite [2] developed a *voice recognition*-based anesthesia record system for the intraoperative use that allows anesthetists to control liquid flow and log events with spoken commands from a fixed dictionary. This was generated from medications that have been used in the hospital in the previous 2 years. However, the language is natural. The command recognition starts after an unlocking keyword and automatically stops after a fixed period. Free text input is also possible. Six anesthesia teams participated in the evaluation in two sessions each. Conventional keyboard and touchscreen input was compared to voice recognition interaction. The study provided the findings that mental workload is decreased by

Table 6 Touchless *OR control and logging*, journal publications are marked with *

	Technical approach	Evaluation results
	Stereo camera	
*Graetzel et al. [18]	Universal one-handed interaction with OR software; mouse mapper	Clicking reliable, scaling imprecise, rapid gestures not robust, pointer jitters
Grange et al. [19]	[18] extended by automatic intention detection; semantic description of gestures	–
	Voice commands	
*Alapetite [2]	Voice recognition-based anesthesia record in crisis situations	Less mental workload, good addition to conservative input
*Perrakis et al. [56]	Comparison of 2 voice recognition systems included in commercial integrated ORs	SIOS better performing than OR1, no significant recognition errors due to speaker's accent
	Structured-light sensor (intel realsense)	
*Meng et al. [38]	Multi-display control with head-worn RGBD sensor; finger tracking	Good usability, response must be improved, not suitable for 3D rotation of medical data

natural language interaction. In order to do that, the authors introduced the "average queue of events" metric as a mental workload indicator. Perrakis et al. [56] compared the voice recognition software of two integrated OR environments (Siemens Integrated OR System SIOS and Karl Storz OR1) in a user study with 74 subjects. The evaluation covered the adjustment of OR light, increasing gas pressure, switching on the video controller and the control of the endolight source. As a result, no significant difference in the number of repeated commands due to different accents has shown. However, the SIOS voice recognition performed significantly better than that of the OR1. The authors conclude that the SIOS is more reliable, but all actions are performed faster manually than with voice control.

Meng et al. [38] not only developed another gesture to mouse event mapper, but also present a system to connect the multi-user interaction to all software on the different displays in the OR. The user wears a *structured-light* sensor on the head and points with one finger to the direction where the mouse cursor shall be moved to. The RGBD sensor detects the finger and the corresponding display, which is calibrated with 2D markers. The sensor data are processed on a wearable computer-connected wireless to different mobile devices. Those are plugged into the different computers and control the cursor on each display. A user study with 7 participants was conducted in a simulated OR environment and revealed good usability. However, the system's response to the users' movements must be improved.

The aforementioned publications are summarized in Table 6.

Commercial state of the art

Some of the approaches that have been described before have already arrived in commercial products. The company Therapixel[3] developed a medical image viewer software called Fluid which can be controlled by a depth sensor that recognizes hand gestures. The GUI of the client software is particularly designed for touchless interaction. Different clients are connected to a central server which caches DICOM data, builds 3D reconstructions and serves as a PACS gateway. Another commercial solution is distributed by Gestsure.[4] The product emerged from the system presented by Strickland et al. [68]. A Kinect 1 is used to map arm gestures onto mouse actions, i.e., a USB mouse is emulated and can be used for every (OR) software. A different approach was followed by TedCas.[5] The company developed a connectivity box for medical applications called TedCube which takes a number of different gestural control sensors as input independent of the operating system and maps movements of hands, arms or eyes to keyboard or mouse commands. They also provide a LMC-based interface to control the mobile DICOM viewer TedSIGN. NZTech[6] introduced an projection-based augmented reality interface to interact touchlessly with medical image data. The system consists of a ceiling-mounted projector-camera unit and an optional self-developed hand gesture sensor integrated in a table. A user interface with control elements for a medical image viewer software is projected onto an arbitrary surface in the sterile area. By hovering the hand over the projected buttons or by executing circle and pointing gestures above the sensor the physician can trigger the desired actions directly and sterilely. The SCOPIS GmbH[7] provides touchless hand gesture control of a surgical navigation system based on a LMC. The user can interact

---

3 http://www.therapixel.com/.

4 http://www.gestsure.com/.

5 http://www.tedcas.com/.

6 http://www.nztech.ca/.

7 http://www.scopis.com/.

with medical images and 3D visualizations via a mouse emulation or gesture interface.

Discussion

This literature review presents the first broad overview of systems providing touchless software interaction for sterile and direct intraoperative or interventional use. The list of publications contains various technical approaches for a diverse set of objectives. While most of the papers deal with touchless control of medical image viewers and laparoscopic devices, there are also publications regarding robotic and telerobotic assistance or general OR control.

Before 2013 there was a great diversity in approaching a natural user interface in the medical domain with RGB cameras, stereo cameras, one time-of-flight camera, body-worn inertial sensors or voice recognition, but only 22 papers were published by this time (see Fig. 2). This is an average of 2.2 publications per year. The release of the consumer-grade console and PC gaming range camera Microsoft Kinect 1 in 2010 and the Kinect SDK in 2012 obviously had a big impact on the human–computer interaction community. With this depth sensor, it was made easy to create touchless interfaces at affordable price. This is reflected by the relatively high number of 21 publications and one commercial product introducing touchless interfaces with the Kinect 1 since 2013. The market introduction of the LMC had a similar effect, leading to 10 publications and two companies adopting the device since 2014. Between 2014 and 2015 10.3 papers were published in average per year. Compared to the years before, a rising trend in the number of publications and thus in interest in the research area can be seen.

The high number of publications introducing very similar approaches in touchless medical image viewer gesture control shows that in the future the research community does not have to produce yet another gesture interface for medical image interaction in a special application. Instead, it has
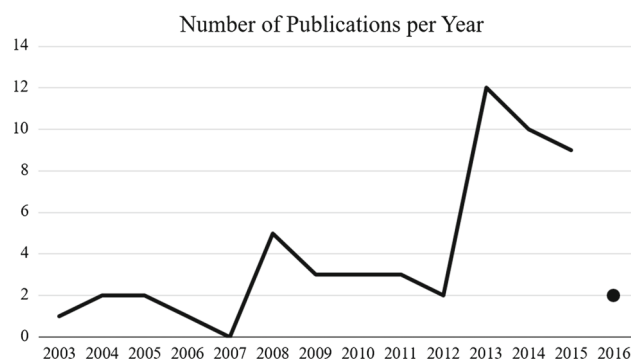

Fig. 2 Number of publications dealing with touchless interaction in the OR per year, number 2016 as of June

to be put more effort in improving and evaluating usability and intuitiveness. Especially in the medical domain, it is crucial to design user interfaces that help the physicians to be more effective and thus shorten treatment duration. Nevertheless, only a few touchless interaction systems have been evaluated properly in this respect: 8 groups tested their systems in a real clinical setting; 7 did not provide evaluation details at all. The remaining 40 systems have been examined in the laboratory, but often with too few participants (fewer than ten), which can skew the results. For example, Nathan et al. [46] and Hötker et al. [24] gained better voice recognition rates under laboratory conditions than Alapetite [2] or Perrakis et al. [56] in a real OR environment. Hence, the research community needs to rethink its user interface evaluation methods to eventually enable the integration of touchless interaction capabilities into the OR. Hettig et al. [22], Saalfeld et al. [61], Nouei et al. [50], Opromolla et al. [52], Meng et al. [38] or Wipfli et al. [77] are examples of appropriate usability testing, since they follow a well-defined study concept, use standardized usability questionnaires and provide qualitative and quantitative data as proof. However, those usability studies always have to be reviewed critically, since the tested implementation and study setup might not be the best solution possible. Therefore, it is not possible to draw a conclusion about the applicability of an interaction device based on a single implementation.

Further research also needs to be done in the aspect of hardware specialization and optimization. Most of the systems in the literature are built with commercial products like the Microsoft Kinect 1 or the LMC, which are general purpose hardware designed for games and home use with, in case of the Kinect 1, a large working space but very low resolution of the imprecise depth data. Custom-built, high-class, use-case-dependent devices could bring advantages for the robustness as well as for the usability. Additionally, it is promising to use multimodal interaction to fulfill the requirements of the clinical application. As proposed in Mentis et al. [39], voice commands could be used as trigger functions such as input unlocking or as functionality switch, and hand or body gestures may be used for continuous manipulation of parameters. If the user needs to interact with multiple applications on different displays, new gaze tracking approaches as in Meng et al. [38] can help to switch between them.

As proposed by Grange et al. [19] and Jacob and Wachs [28], clear software restrictions should be implemented to prevent unintended input. For the purpose of better understanding the intention of the interacting person, sensor fusion of all instruments and interaction modalities in the OR and interventional radiology suite will be of interest. Therefore, it must be possible to gather all the information in one place. The OR.NET[8] project aims at providing a signal bus and

---

[8] http://www.ornet.org/?lang=en.

software protocol for the safe interconnection of instruments in clinical environments. Together with standardized workflows for medical procedures [26,27], much of the manual physician–computer interaction can be replaced by automated presentation of relevant information in the respective workflow step. This way physicians can concentrate on the actual intervention by reducing the necessary interaction with software.

Another crucial condition for increasing the usability when using touchless input devices is feedback for the user. Compared to mouse, keyboard and touchscreen interaction there is no haptic feedback (except in Mylonas et al. [45]). Thus, as stated by Silva et al. [65], visual or even auditory feedback [30,54,74] is important to prevent confusion but often not provided.

Many of the technical solutions presented in the field of intraoperative gesture interaction are prototypes emulating mouse interaction which are set on top of existing medical software. Although there exist good approaches for integrating touchless interaction better into the existing software via mobile devices [38], one has to get past the 2D WIMP paradigm to enable real natural user interfaces [76]. The manufacturers of medical software will have to redesign their user interfaces, considering menus and action triggers, so that users can benefit from 3D interaction devices.

Although most of the groups present very sophisticated and robust methods to control a medical image viewer, telerobotic systems or other intraoperative assistance (with recognition rates mostly >90 %), it has to be noted that in the medical domain an almost always correct recognition of input commands is mandatory to decrease the risk for the patient and the workload of the clinical staff. Future research should therefore focus on further increasing robustness and accuracy of touchless control. In the past years, dozens of sophisticated and easy-to-use machine learning frameworks have emerged, e.g., Theano,[9] Tensorflow[10] or the Stanford CoreNLP Natural Language Processing Toolkit [36]. Those toolkits enable developers to train and classify voice commands or gestures [47,48], given that training data are available, and to reach much better recognition rates than with conventional methods even in the noisy or cluttered environment [12,23,43] of an OR or interventional radiology suite. Natural language processing is constantly improving. Chan et al. (2016) [9] presented a neural network that learns to transcribe speech utterances to characters at a word error rate of 14.1 % without a dictionary or language model. Before, as Cambria and White (2014) [8] pointed out, a large knowledge base was needed to match existing vocabulary and recorded voice. Speech recognition with large vocabulary knowledge works at a word error rate of 8 % [62]. With improved semantic

information acquisition [3], and with the promising advantages of neural network integrated circuits [40], it will be possible for computers to understand humans offline without the need for large data centers that are only available through internet connections.

Considering new interaction hardware approaches, a disruptive element in interactive medical software will be the augmented reality HMDs Meta 2 (Meta Company, Portola Valley, CA, USA) or Microsoft Hololens.[11] They provide lightweight augmented reality glasses, which can be used to display relevant information or images spatially aligned with the patient, and, in case of the Hololens, eye tracking, gesture and state-of-the-art speech recognition capabilities to interact with them. Such devices have the potential to bring together all information in one place, similar to presenting mobile augmented reality solutions [4,31,44], and additionally give the opportunity to interact with all relevant data naturally via sophisticated speech recognition (as mentioned before) and touchless gestures. The affordable price will make prototypes pop up in the medical domain, as the Kinect 1 and the LMC did.

With those technologies around the corner and if the research community adapts its efforts in increasing the usability, it will be possible for physicians to interact with medical software context-aware, workflow-step-dependent and finally hands-free.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Informed consent This article does not contain patient data.

Ethical standards This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Achacon DLM, Carlos DM, Puyaoan MK, Clarin CT, Naval Jr. PC (2009) Realism: real-time hand gesture interface for surgeons and medical experts. In: 9th Philippine computing science congress, Citeseer
2. Alapetite A (2008) Speech recognition for the anaesthesia record during crisis scenarios. Int J Medi Inform 77(7):448–460
3. Audhkhasi K, Sethy A, Ramabhadran B (2016) Semantic word embedding neural network language models for automatic speech recognition. IN: 2016 IEEE international conference on acoustics. Speech and signal processing (ICASSP), IEEE, pp 5995–5999

---

9 http://deeplearning.net/software/theano/.

10 https://www.tensorflow.org/.

11 https://blogs.windows.com/buildingapps/2016/01/21/hololens-interaction-model/.

4. Bane R, Höllerer T (2004) Interactive tools for virtual x-ray vision in mobile augmented reality. In: Third IEEE and ACM international symposium on mixed and augmented reality, 2004. ISMAR 2004, IEEE, pp 231–239

5. Bauer S, Seitel A, Hofmann H, Blum T, Wasza J, Balda M, Meinzer HP, Navab N, Hornegger J, Maier-Hein L (2013) Real-time range imaging in health care: a survey. In: Grzegorzek M, Theobalt C, Koch R, Kolb A (eds) Time-of-flight and depth imaging. Sensors, algorithms, and applications. Springer, Berlin, pp 228–254. doi:10.1007/978-3-642-44964-2_11

6. Bigdelou A, Schwarz L, Navab N (2012) An adaptive solution for intra-operative gesture-based human-machine interaction. In: Proceedings of the 2012 ACM international conference on intelligent user interfaces, ACM, pp 75–84

7. 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. Surg Innov 21(6):655–656

8. Cambria E, White B (2014) Jumping nlp curves: a review of natural language processing research [review article]. IEEE Comput Intell Mag 9(2):48–57

9. Chan W, Jaitly N, Le Q, Vinyals O (2016) Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. 2016 IEEE international conference on acoustics. Speech and signal processing (ICASSP), IEEE, pp 4960–4964

10. Chao C, Tan J, Castillo EM, Zawaideh M, Roberts AC, Kinney TB (2014) Comparative efficacy of new interfaces for intra-procedural imaging review: the microsoft kinect, hillcrest labs loop pointer, and the apple ipad. J Digit Imaging 27(4):463–469

11. Clancy NT, Mylonas GP, Yang GZ, Elson DS (2011) Gaze-contingent autofocus system for robotic-assisted minimally invasive surgery. In: Engineering in medicine and biology society, EMBC, 2011 annual international conference of the IEEE, IEEE, pp 5396–5399

12. Dahl GE, Yu D, Deng L, Acero A (2012) Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Trans Audio Speech Lang Process 20(1):30–42

13. Ebert L, Flach P, Thali M, Ross S (2014) Out of touch-a plugin for controlling osirix with gestures using the leap controller. J Forensic Radiol Imaging 2(3):126–128

14. Ebert LC, Hatch G, Ampanozi G, Thali MJ, Ross S (2012) You can't touch this touch-free navigation through radiological images. Surg Innov 19(3):301–307

15. El-Shallaly G, Mohammed B, Muhtaseb M, Hamouda A, Nassar A (2005) Voice recognition interfaces (vri) optimize the utilization of theatre staff and time during laparoscopic cholecystectomy. Minim Invasive Ther Allied Technol 14(6):369–371

16. Gallo L (2013) A study on the degrees of freedom in touchless interaction. In: SIGGRAPH Asia 2013 technical briefs, ACM, p 28

17. Gong RH, Güler Ö, Kürklüoglu M, Lovejoy J, Yaniv Z (2013) Interactive initialization of 2d/3d rigid registration. Med Phys 40(12):121,911

18. Graetzel C, Fong T, Grange S, Baur C (2004) A non-contact mouse for surgeon-computer interaction. Technol Health Care 12(3):245–257

19. Grange S, Fong T, Baur C (2004) M/oris: a medical/operating room interaction system. In: Proceedings of the 6th international conference on multimodal interfaces, ACM, pp 159–166

20. Hartmann F, Schlaefer A (2013) Feasibility of touch-less control of operating room lights. Int J Comput Assist Radiol Surg 8(2):259–268

21. Herniczek SK, Lasso A, Ungi T, Fichtinger G (2014) Feasibility of a touch-free user interface for ultrasound snapshot-guided nephrostomy. Proceedings of SPIE 9036, medical imaging 2014: image-guided procedures, robotic interventions, and modeling, 90362F. doi:10.1117/12.2043564

22. Hettig J, Mewes A, Riabikin O, Skalej M, Preim B, Hansen C (2015) Exploration of 3D medical image data for interventional radiology using myoelectric gesture control. In: Proceedings of Eurographics workshop on visual computing for biology and medicine, The Eurographics Association, pp 177–185

23. Hinton G, Deng L, Yu D, Dahl GE, Mohamed Ar, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN (2012) Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process Mag 29(6):82–97

24. Hötker AM, Pitton MB, Mildenberger P, Düber C (2013) Speech and motion control for interventional radiology: requirements and feasibility. Int J Comput Assist Radiol Surg 8(6):997–1002

25. Hübler A, Hansen C, Beuing O, Skalej M, Preim B (2014) Workflow analysis for interventional neuroradiology using frequent pattern mining. In: Proceedings of the annual meeting of the German Society of Computer- and Robot-Assisted Surgery, Munich, pp 165–168

26. Neumann J, Neumuth T (2015a) Standardized semantic workflow modeling in the surgical domain–proof-of-concept analysis and evaluation for a neurosurgical use-case. IEEE, Boston, pp 6–11

27. Neumann J, Neumuth T (2015b) Towards a framework for standardized semantic workflow modeling and management in the surgical domain. Curr Dir Biomed Eng 1(1):172–175

28. Jacob MG, Wachs JP (2014) Context-based hand gesture recognition for the operating room. Pattern Recognit Lett 36:196–203

29. Jalaliniya S, Smith J, Sousa M, Büthe L, Pederson T (2013) Touchless interaction with medical images using hand & foot gestures. In: Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication, ACM, pp 1265–1274

30. Kajastila R, Lokki T (2013) Eyes-free interaction with free-hand gestures and auditory menus. Int J Hum Comput Stud 71(5):627–640

31. Kilgus T, Bux R, Franz A, Johnen W, Heim E, Fangerau M, Müller M, Yen K, Maier-Hein L (2016) Structure sensor for mobile markerless augmented reality. Proceedings of SPIE 9786, medical imaging 2016: image-guided procedures, robotic interventions, and modeling, 97861L. doi:10.1117/12.2216057

32. Kipshagen T, Graw M, Tronnier V, Bonsanto M, Hofmann U (2009) Touch-and marker-free interaction with medical software. World congress on medical physics and biomedical engineering, September 7–12, 2009. Springer, Munich, pp 75–78

33. Kirmizibayrak C, Radeva N, Wakid M, Philbeck J, Sibert J, Hahn J (2011) 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, ACM, pp 69–74

34. Kocev B, Ritter F, Linsen L (2014) Projector-based surgeon-computer interaction on deformable surfaces. Int J Comput Assist Radiol Surg 9(2):301–312

35. Li YT, Jacob M, Akingba G, Wachs JP (2013) A cyber-physical management system for delivering and monitoring surgical instruments in the or. Surg Innov 20(4):377–384

36. Manning CD, Surdeanu M, Bauer J, Finkel JR, Bethard S, McClosky D (2014) The stanford coreNLP natural language processing toolkit. In: Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, pp 55–60

37. Mauser S, Burgert O (2014) Touch-free, gesture-based control of medical devices and software based on the leap motion controller. Stud Health Technol Inform 196:265–270

38. Meng M, Fallavollita P, Habert S, Weidert S, Navab N (2016) Device-and system-independent personal touchless user interface for operating rooms. Int J Comput Assist Radiol Surg 11(6):1–9

39. Mentis HM, O'Hara K, Gonzalez G, Sellen A, Corish R, Criminisi A, Trivedi R, Theodore P (2015) Voice or gesture in the operating room. In: Proceedings of the 33rd annual ACM conference extended abstracts on human factors in computing systems, ACM, pp 773–780

40. Merolla PA, Arthur JV, Alvarez-Icaza R, Cassidy AS, Sawada J, Akopyan F, Jackson BL, Imam N, Guo C, Nakamura Y, Brezzo B, Vo I, Esser SK, Rathinakumar A, Taba B, Amir A, Flickner MD, Risk WP, Monohar R, Modha DS (2014) A million spiking-neuron integrated circuit with a scalable communication network and interface. Science 345(6197):668–673

41. Mewes A, Saalfeld P, Riabikin O, Skalej M, Hansen C (2015) A gesture-controlled projection display for ct-guided interventions. Int J Comput Assist Radiol Surg 11(1):1–8

42. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. Ann Intern Med 151(4):264–269

43. Molchanov P, Gupta S, Kim K, Kautz J (2015) Hand gesture recognition with 3d convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 1–7

44. Müller M, Rassweiler MC, Klein J, Seitel A, Gondan M, Baumhauer M, Teber D, Rassweiler JJ, Meinzer HP, Maier-Hein L (2013) Mobile augmented reality for computer-assisted percutaneous nephrolithotomy. Int J Comput Assist Radiol Surg 8(4):663–675

45. Mylonas GP, Kwok KW, Darzi A, Yang GZ (2008) Gaze-contingent motor channelling and haptic constraints for minimally invasive robotic surgery. In: Medical image computing and computer-assisted intervention—MICCAI 2008, Springer, pp 676–683

46. Nathan CAO, Chakradeo V, Malhotra K, D'Agostino H, Patwardhan R (2006) The voice-controlled robotic assist scope holder aesop for the endoscopic approach to the sella. Skull Base 16(3):123

47. Neverova N, Wolf C, Taylor GW, Nebout F (2014) Multi-scale deep learning for gesture detection and localization. In: Computer vision-ECCV 2014 workshops, Springer, pp 474–490

48. Nishida N, Nakayama H (2015) Multimodal gesture recognition using multi-stream recurrent neural network. In: Pacific-rim symposium on image and video technology, Springer, pp 682–694

49. Nishikawa A, Hosoi T, Koara K, Negoro D, Hikita A, Asano S, Kakutani H, Miyazaki F, Sekimoto M, Yasui M, Miyake Y, Takiguchi S, Monden M (2003) Face mouse: a novel human-machine interface for controlling the position of a laparoscope. IEEE Trans Robot Automa 19(5):825–841

50. Nouei MT, Kamyad AV, Soroush AR, Ghazalbash S (2015) A comprehensive operating room information system using the kinect sensors and rfid. J Clin Monit Comput 29(2):251–261

51. O'Hara K, Gonzalez G, Sellen A, Penney G, Varnavas A, Mentis H, Criminisi A, Corish R, Rouncefield M, Dastur N (2014) Touchless interaction in surgery. Commun ACM 57(1):70–77

52. Opromolla A, Volpi V, Ingrosso A, Fabri S, Rapuano C, Passalacqua D, Medaglia CM (2015) A usability study of a gesture recognition system applied during the surgical procedures. In: Marcus A (ed) Design, user experience, and usability: interactive experience design. Springer, pp 682–692. doi:10.1007/978-3-319-20889-3_63

53. Park BJ, Jang T, Choi JW, Kim N (2016) Gesture-controlled interface for contactless control of various computer programs with a hooking-based keyboard and mouse-mapping technique in the operating room. Comput Math Methods Med 2016. doi:10.1155/2016/5170379

54. Park Y, Kim J, Lee K (2015) Effects of auditory feedback on menu selection in hand-gesture interfaces. IEEE MultiMed 22(1):32–40

55. Pauchot J, Di Tommaso L, Lounis A, Benassarou M, Mathieu P, Bernot D, Aubry S (2015) Leap motion gesture control with carestream software in the operating room to control imaging installation guide and discussion. Surg Innov 22:615–620

56. Perrakis A, Hohenberger W, Horbach T (2013) Integrated operation systems and voice recognition in minimally invasive surgery: comparison of two systems. Surg Endosc 27(2):575–579

57. Reilink R, De Bruin G, Franken M, Mariani M, Misra S, Stramigioli S (2010) Endoscopic camera control by head movements for thoracic surgery. In: 2010 3rd IEEE RAS and EMBS international conference on Biomedical robotics and biomechatronics (BioRob), IEEE, pp 510–515

58. Riduwan M, Basori AH, Mohamed F (2013) Finger-based gestural interaction for exploration of 3d heart visualization. Procedia Soc Behav Sci 97:684–690

59. Rosa GM, Elizondo ML (2014) Use of a gesture user interface as a touchless image navigation system in dental surgery: case series report. Imaging Sci Dent 44(2):155–160

60. Ruppert GCS, Reis LO, Amorim PHJ, de Moraes TF, da Silva JVL (2012) Touchless gesture user interface for interactive image visualization in urological surgery. World J Urol 30(5):687–691

61. Saalfeld P, Mewes A, Luz M, Preim B, Hansen C (2015) Comparative evaluation of gesture and touch input for medical software. In: Proceedings of Mensch und computer 2015

62. Sainath TN, Vinyals O, Senior A, Sak H (2015) Convolutional, long short-term memory, fully connected deep neural networks. In: 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, pp 4580–4584

63. Salama IA, Schwaitzberg SD (2005) Utility of a voice-activated system in minimally invasive surgery. J Laparoendosc Adv Surg Tech 15(5):443–446

64. Schwarz LA, Bigdelou A, Navab N (2011) Learning gestures for customizable human-computer interaction in the operating room. In: Medical image computing and computer-assisted intervention—MICCAI 2011, Springer, pp 129–136

65. Silva ÉS, Rodrigues MAF (2014) Design and evaluation of a gesture-controlled system for interactive manipulation of medical images and 3d models. SBC J Interact Syst 5(3):53–65

66. Soutschek S, Penne J, Hornegger J, Kornhuber J (2008) 3-d gesture-based scene navigation in medical imaging applications using time-of-flight cameras. In: IEEE computer society conference on computer vision and pattern recognition workshops, 2008. CVPRW'08, IEEE, pp 1–6

67. Stoyanov D, Mylonas GP, Yang GZ (2008) Gaze-contingent 3d control for focused energy ablation in robotic assisted surgery. In: Medical image computing and computer-assisted intervention–MICCAI 2008, Springer, pp 347–355

68. Strickland M, Tremaine J, Brigley G, Law C (2013) Using a depth-sensing infrared camera system to access and manipulate medical imaging from within the sterile operating field. Can J Surg 56(3):E1

69. Suelze B, Agten R, Bertrand PB, Vandenryt T, Thoelen R, Vandervoort P, Grieten L (2013) Waving at the heart: Implementation of a kinect-based real-time interactive control system for viewing cineangiogram loops during cardiac catheterization procedures. In: Computing in cardiology conference (CinC), 2013, IEEE, pp 229–232

70. Tan JH, Chao C, Zawaideh M, Roberts AC, Kinney TB (2013) Informatics in radiology: developing a touchless user interface for intraoperative image control during interventional radiology procedures. Radiographics 33(2):E61–E70

71. Visentini-Scarzanella M, Mylonas GP, Stoyanov D, Yang GZ (2009) i-brush: A gaze-contingent virtual paintbrush for dense 3d reconstruction in robotic assisted surgery. In: Medical image computing and computer-assisted intervention—MICCAI 2009, Springer, pp 353–360

72. Wachs JP, Stern HI, Edan Y, Gillam M, Handler J, Feied C, Smith M (2008) A gesture-based tool for sterile browsing of radiology images. J Am Med Inform Assoc 15(3):321–323

73. Wachs JP, Vujjeni K, Matson ET, Adams S (2010) a window on tissue-using facial orientation to control endoscopic views of tissue depth. In: 2010 annual international conference of the IEEE engineering in medicine and biology society (EMBC), IEEE, pp 935–938

74. Walker BN, Lindsay J, Nance A, Nakano Y, Palladino DK, Dingler T, Jeon M (2013) Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus. Hum Factors J Hum Factors Ergon Soc 55(1):157–182

75. Wen R, Tay WL, Nguyen BP, Chng CB, Chui CK (2014) Hand gesture guided robot-assisted surgery based on a direct augmented reality interface. Comput Methods Programs Biomed 116(2):68–80

76. Wigdor D, Wixon D (2011) Brave NUI world: designing natural user interfaces for touch and gesture. Elsevier, Amsterdam

77. Wipfli R, Dubois-Ferrière V, Budry S, Hoffmeyer P, Lovis C (2016) Gesture-controlled image management for operating room: a randomized crossover study to compare interaction using gestures, mouse, and third person relaying. PloS One 11(4):e0153,596

78. Yoshida S, Ito M, Tatokoro M, Yokoyama M, Ishioka J, Matsuoka Y, Numao N, Saito K, Fujii Y, Kihara K (2015) Multitask imaging monitor for surgical navigation: combination of touchless interface and head-mounted display. Urol Int. doi:10.1159/000381104

79. Yusoff YA, Basori AH, Mohamed F (2013) Interactive hand and arm gesture control for 2d medical image and 3d volumetric medical visualization. Procedia Soc Behav Sci 97:723–729