

REVIEW

Open Access



Advances in the development and application of non-contact intraoperative image access systems

Zhengnan Liu¹, Chengrun Li², Jixing Lin³, Hengliang Xu³, Yan Xu¹, Haoning Nan¹, Wei Cheng¹, Jie Li^{2*} and Bin Wang^{2*}

*Correspondence:
13801010576@163.com;
banxy0119@163.com

¹ Chinese PLA Medical School,
Beijing, China

² Department of Thoracic
Surgery, The First Medical Center,
Chinese PLA General Hospital,
Beijing, China

³ Department of Thoracic
Surgery, Hainan Hospital
of the General Hospital
of the People's Liberation Army
of China, Sanya, China

Abstract

This article provides an overview of recent progress in the achievement of non-contact intraoperative image control through the use of vision and sensor technologies in operating room (OR) environments. A discussion of approaches to improving and optimizing associated technologies is also provided, together with a survey of important challenges and directions for future development aimed at improving the use of non-contact intraoperative image access systems.

Keywords: Non-contact interaction, Gesture control, Human–computer interaction

Introduction

Unlike traditional open surgery, minimally invasive surgical procedures require that surgeons use an endoscopic system equipped with a camera to display the surgical operation area on a screen, owing to the lack of direct visibility of the operative field. In order to ensure that surgical procedures can be performed with precision, however, surgeons must still repeatedly intraoperatively refer to medical records or imaging data for the patient. Ensuring that surgeons can access this information in an efficient manner can thus reduce the overall operative duration while increasing patient safety. At present, access to medical electronic imaging systems in operating room (OR) settings primarily consists of surgeons asking for assistance from others outside the operating table (including circulating nurses and anesthesiologists). These personnel, however, are more likely to make operational errors owing to their limited experience with the imaging system, potentially requiring the surgeon or their assistant to leave the sterile operating area so that they can operate the computerized imaging systems directly, after which re-sterilization is required before continuing the procedure, thus interfering with the overall operating workflow.

There have been ongoing efforts since the early 1990s to apply advanced technologies as a means of establishing contact-free computerized manipulation systems that can be used when performing surgical procedures. Nishikawa et al. [1], for instance, designed a



© The Author(s) 2024, corrected publication 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

human–computer interaction (HCI) system that they designated “Face Mouse”, allowing surgeons to conduct simple contact-free operations by making appropriate facial movements. However, these typical facial movements can only be used to perform certain discrete commands including “zoom/tilt/pan”, whereas they are poorly suited for continuous control, such as the viewing of computed tomography (CT) images in a layer-by-layer manner. However, there have been significant technological advances over time. In contrast with facial recognition technologies, gesture recognition systems can function independently of the facial emotional state of the operator while also necessitating the use of less data to learn and exhibiting lower hardware requirements. The two mainstream classes of gesture recognition systems consist of camera vision-based and inertial sensor-based methods [2].

Vision-based approaches rely on acquiring videos or images of hand gestures using a video system, and can be broadly classified into four main categories: (1) monocular cameras: these cameras are generally composed of a lens, a sensor, and a processor responsible for capturing images via a single lens and internally processing those images (e.g., regular cameras, video cameras, smartphone cameras); (2) multi-ocular cameras: these camera systems are generally composed of at least two monocular cameras, with each being responsible for capturing the same scene from different angles to generate multiple images that can be compared, providing information regarding 3D coordinates and depth for objects in the scene; (3) active techniques: these technologies take advantage of structured light projection, which entails the projection of patterns of light from a structured source onto object surfaces and capturing the reflected light with a camera or sensor to acquire textural and geometric data related to the surface of the object in question (e.g., Kinect, Leap Motion, etc.); (4) invasive techniques: these technologies rely on the use of body markers including wrist bands, LED lights, and/or colored gloves, but relatively few associated studies are currently available owing to technical constraints.

Sensor-based approaches rely on capturing hand position, motion, and velocity data with motion sensors, and include the following: (1) inertial measurement unit (IMU) systems: these approaches make use of accelerometers and gyroscopes to assess finger position, degree of freedom, and acceleration; (2) electromyography (EMG): these approaches detect finger movements by harnessing the electrical bio-signals associated with human muscles; (3) WiFi and radar: these approaches detect changes in the in-air signal strength through the use of radio waves, broad-beam radar, or spectrograms; (4) other approaches: these strategies can include the use of electromagnetic, ultrasonic, and/or haptic technologies.

Vision-based non-contact intraoperative image access systems

Gestix system

Early uses of gesture recognition systems can be traced to the “Gestix” system developed in 2006 by Wachs et al. [3]. This system relies on the recognition of gestures using a 2D camera, converting these gestures into trends according to gesture temporal trajectory. Using this system, surgeons can enable surgeons to use hand gestures made in the air to select, rotate, scale, and move 3D images, achieving 96% gesture recognition accuracy. However, this system requires prolonged installation and setup time (~20 min for full setup), and it is only capable of recognizing gesture commands made with a single hand

at a time. More recently, Oshiro et al. [4] devised the contactless “Dr.aeroTAP” system, which can function with regular cameras while also providing support for infrared (IR) and stereo cameras, and can be readily set up by connecting a USB-based web camera and launching the aeroTAP system. This offers a clear advantage over the longer setup time associated with the Gestix system, and these innovative non-contact systems have been demonstrated to be effective surgical imaging aids in two reports to date [4].

Kinect system

Initially released in November 2010, the Microsoft Kinect sensor was designed with a primary focus on the gaming sector. The system consists of an IR emitter, color camera, and array microphone capable of sensing the position, movement, and voice of the operator. The Kinect device achieved depth sensing based on the structured light principle using data acquired from projected IR dot patterns and the IR camera [5]. Rupert et al. [6] were the first to describe a Kinect-based OpenNI/NITE component that was able to detect and segment multiple users in a scene while tracking 15 parts of the body in real-time (including hands, elbows, shoulders, hips, knees, feet, shoulders, head, neck, and torso), allowing for effective skeletal tracking that was used to resect four tumors in 3 male patients. Subsequently, Tan et al. [7] designed the customized “TRICS” software program that can function with the Kinect device to enable the tracking of 3D coordinates associated with skeletal and human body movement and the translation of these movements into gestures that can be used for medical image manipulation. TRICS provides a high level of control and flexibility through its support for specific gesture types, including circular gestures. Yoshimitsu et al. [8] developed a novel Kinect-based medical device designated “OPECT” and evaluated its performance across 30 neurosurgical procedures. They found that their OPECT system was capable of displaying images with excellent quality intraoperatively while accurately recognizing individual operator characteristics. Gobhiran et al. [9] explore the development of a new technique based on hand movement patterns and a square guide grid, utilizing the Kinect 3D sensor to capture these hand movements and encoding their movement paths with a chain code technique, ultimately employing a KNN algorithm for feature vector classification. In the context of gesture recognition, the screen displays a grid direction guide for the hand movements to minimize the potential for inter-operator error. The authors implemented 7 total commands for image browsing including movement, zoom, contrast adjustment, and image retrieval, and they ultimately achieved an average recognition accuracy of 95.72%. Liu et al. [10] designed a Kinect-based real-time gesture interaction system to allow for the intraoperative visualization of hepatic structures in a contact-free manner by tracking hand movements and combining three hand states to enable the effective control of hepatic structure visualization through zooming, rotation, transparency adjustment, fusion, and the selection of blood vessels. Glinkowski et al. [11] also established a Kinect-based application (Ortho_Kinect_OR) capable of controlling intraoperative medical image access while also providing support for telemedicine applications including intraoperative telementoring and teleconsultation.

While it does offer certain innovative features in the context of gesture recognition, there are some limitations associated with the Kinect system. For one, the performance of the Kinect depth camera is impaired under low-light conditions or in highly reflective

settings, contributing to suboptimal recognition accuracy. To correctly capture movements, the system also requires sufficient space (at least 6 m² of floor space), which may not be feasible in confined OR settings. Additionally, this system lacks the necessary sensitivity to capture motion at longer distances, particularly against complex backgrounds, potentially resulting in missed or misjudged movements. The Kinect also has the potential to experience delays when processing high-speed movements, and it lacks the accuracy needed to recognize small or rapid gestures [12]. Further optimization of the Kinect hardware and associated algorithms is thus necessary to achieve better gesture recognition performance and response speeds under a range of conditions and in various environments.

Leap motion controller

The LMC sensor, which was first made available in 2013, consists of two IR stereo cameras and three IR LEDs, allowing for the tracking of 27 distinct hand elements including joints and bones with an accuracy of up to 1/100th of a millimeter [13], making it better suited than the Kinect system to use in an OR setting. Feng et al. [14] evaluated the accuracy, efficacy, and operator satisfaction levels associated with these two devices when used for intraoperative imaging. When 10 surgeons used Kinect and LMC devices as well as a conventional mouse to perform five image interaction tasks (zooming, panning, step-by-step navigation, circle measurements, and line measurements), the Kinect and LMC yielded comparable accuracy for most tasks, with the Kinect exhibiting higher error rates during the step-by-step navigation task. The LMC system yielded shorter completion times relative to the Kinect system, and was preferred by these surgeons when performing measurement tasks, particularly those associated with a high degree of precision. Rosa et al. [15] were the first to implement a contactless natural user interface (NUI) LMC system when performing dental surgery, allowing for the zooming or rotation of images by surgeons with one or two finger movements. Chiang et al. [16] employed an LMC-based 3D stereo image observation system capable of providing two medical image observation tools usable by surgeons to quickly observe image cross-sections while incorporating view-through functionality to allow for the dissection of hidden information in 3D stereo images in a layer-by-layer manner. This LMC system can be used to intraoperatively access both CT and MRI images [17]. Zhang et al. [18] deployed an LMC-based approach to 3D spatial perception with Aruco visual tags. In their system, initial gesture recognition was achieved with an LMC device, whereas Aruco tags were used to improve overall operative accuracy based on their unique recognition and localization features. Further efforts to employ Aruco tags in the context of intraoperative contactless image access have the potential to improve operative efficiency and accuracy for surgeons performing gesture operations. Sa-nguannarm et al. [19] further employed an LMC system for the classifications of 10 gestures matched to 6 commands on the program screen (waiting, selecting, adjusting brightness, etc.) as well as four button commands (zooming in, zooming out, clockwise rotation, and counter-clockwise rotation). Using this classification system, the authors achieved an average accuracy rate of 95.83%, ensuring accurate and efficient contactless access to medical images.

Although the LMC has a high degree of potential utility in surgical settings, further improvements will be necessary for it to more effectively meet the specific needs of surgeons. For one, the system must be capable of accurately recognizing complex gestures and ensuring intuitive and responsive operations with high levels of customization and personalization suited to the operating habits of particular surgeons and the specific considerations associated with specific surgical procedures. Secondly, this device needs to be capable of effectively integrating the functions of the imaging and control software in real time through the development of appropriate tools and plugins capable of ensuring accurate, smooth operation while taking the compatibility of different software programs into account. Cho et al. [20] suggested the use of a personalized automated classifier through the use of the LMC for gesture acquisition in combination with support vector machine (SVM) and Naïve Bayes classifier-based training and testing strategies. This classifier is capable of being trained in accordance with the gestures of individual operators, thus enabling gesture recognition with greater accuracy and superior motion sensitivity together with a lower rate of inaccurate results. Ameer et al. [21] successfully recognized 11 gestures employed for contact-free medical image manipulation by acquiring data with an LMC device and combining this with a range of classification methods (e.g., SVM and multilayer perceptron) to achieve up to 91.73% accuracy. Using this optimal recognition rate, they were able to further simplify contactless interactions with medical images in OR settings.

Excessive numbers of gestures or overly cluttered gestures can markedly increase the difficulty of starting and the associated operation error rate. Selecting appropriate sets of gestures associated with particular tasks is thus vital when designing novel systems [22]. Research has demonstrated that people are generally only capable of remembering up to 6 gestures [23]. Tommaso et al. [24] proposed new approaches to the configuration and ergonomic optimization of LMC systems, simplifying the associated gesture configuration through the adoption of a steering mode such that operators need only utilize three gestures for five functions, thereby reducing the overall number of gestures and simplifying the utilization of the system as a whole.

Sensor-based contact-free intraoperative image access systems

Inertial sensors and MYO

Vision-based gesture recognition systems are subject to limitations associated with ambient light levels, and may be disagreeable to surgeons who would prefer not to be monitored constantly by a camera. These issues can be overcome by enabling position-independent interaction through the use of wearable inertial sensors on the head, wrists, and/or body [25, 26]. These sensors eliminate any need for the direct gaze of the operator, and ensure that only the individual wearing the sensors can interact with the system [27]. Bigdelou et al. [28] examined potential hardware issues associated with such inertial sensor-based systems including drift and noise, which have the potential to impair sensor output accuracy or to propagate gradual increases in measurement error levels. While these technologies generally necessitate lower amounts of OR space and are free of any issues related to line-of-sight, they typically necessitate training based on training based on pre-acquired datasets, in addition to being limited by the intuitive nature of the gestures [29].

Hettig et al. [30] proposed an innovative approach to completely contactless medical image viewer control through the implementation of an input device consisting of a myoelectric gesture-controlled (MYO) armband. This technique can analyze and interpret EMG signals corresponding to muscle activity captured using surface EMG (sEMG) sensors. By using specific gestures to activate particular muscle groups (such as fist clenching or the spreading of the fingers), surgeons can generate appropriate EMG signals that are captured by the sensor and interpreted as specific actions or commands. This MYO armband consists of eight sEMG sensors capable of mapping five gestures to four different software functions while also providing haptic vibration feedback. In clinical testing, however, this device achieved recognition rates that were too low for reliable clinical use (56–86%), including high false-positive recognition rates. Margallo et al. [31] first employed the MYO armband to control preoperative images during laparoscopic surgical procedures, comparing this system to Kinect and Leap Motion-based systems. Of the three, the Kinect system was regarded as being the most labor-intensive, whereas the MYO armband and associated voice commands were regarded as being the most precise and intuitive. Even so, many individual-specific factors can affect sEMG signals including electrode positioning, skin impedance, and the amount of subcutaneous fat. As a result, there tend to be marked differences in sEMG signals generated among operators even when making the same movement. There thus remains a pressing need to develop efficient strategies capable of recognizing movements across individuals in order to enable the more practical implementation of EMG-based systems [32, 33].

Radar sensors

Relative to other approaches, radar-based technologies offer advantages including low costs, high levels of accuracy, environmental resilience, and the ability to ensure privacy [34]. To overcome the limitations associated with wearable devices and computer-based vision, devices, Miller et al. [35] devised a directional radar gesture recognition system that they designated “RadSense”. This system leverages the Doppler effect to capture gestures with a continuous wave (CW) radar sensor while transmitting the associated gesture signals to a computer through a Bluetooth Low Energy (BLE) network, thereby enabling the classification and control of associated images. The system can be worn on the human body or affixed using Velcro to a range of objects including shadowless lamps on the operating table, allowing for gesture classification with 94.5% accuracy. However, the CW radar sensors have difficulty obtaining distance-related information, and variations in received power to the radar sensors as a consequence of changes in motion distance can interfere with the recognition of gestures in this system. As a result, prior studies have relied on a fixed sensing position for hand gestures. Separately, Yang et al. [36] proposed the use of a frequency-shift keying (FSK) radar-based system for gesture recognition that is capable of functioning at a range of distances between the radar sensor and the operator, yielding a similar rate of gesture recognition irrespective of the motion distance associated with the hand gestures in question.

Electromagnetic induction technology

Unlike optical or vision-based systems, capacitive sensors are better able to tolerate mechanical impacts or dirt and are not adversely affected by poor lighting or occlusion

[37]. The contact-free “HoverTap-MD” technology takes advantage of these systems to enable the detection of the position of a surgeon’s finger in 3D space using capacitive sensors mounted in a frame around the screen, eliminating the need for any camera. This allows for the smooth, contact-free operation of touchscreens, physical buttons, and other surfaces via finger taps and aerial sliding even through sterile gloves, sterile sheets, and thick glass. Importantly, the performance of this technology remains robust even in cases where the screen is covered with a range of liquids [38]. In a similar strategy, a wireless portable tablet designated an “AirPad” has been developed utilizing capacitive sensors within the tablet to allow for effective navigation and manipulation of real-time and historical medical images by surgeons without the need for contact. This system obviates the need to learn complex gestures or to employ unusual body movements by instead sensing finger movements, including rotation and swipe movements, on its surface and enabling a range of appropriate imaging controls including rotating, panning, zooming, and scrolling [39].

Other technologies

There have also been studies exploring the utility of a brain–computer interface (BCI) as a novel approach to enabling the control of medical images without the need for voice or gesture recognition. BCI technologies entail the use of small sensors capable of monitoring the brain activity of the surgeon through the real-time capture of steady-state visually evoked potentials, which are brain signals detected by dry electrodes in contact with the operator’s skull that monitor electrical activity in the visual cortex. When an operator is wearing the sensor and directs appropriate situational attention to specific buttons exhibiting flickering visual patterns mounted on the user interface, these buttons are capable of sending appropriate signals to the software. The associated lightweight sensor can be comfortably worn under a surgical cap. In a simulated surgical setting, Esfandari et al. [29] evaluated this technique by inviting 10 orthopedic surgeons to use it as a means of navigating and localizing preselected positions in CT images. They found that these surgeons exhibited good technique receptivity, reporting average Likert scores of 4.07 with a corresponding overall impression score of 3.77. In many cases, however, the participants reported that the BCI device had a slow response time, which was noted as the greatest drawback associated with this technology. Yang et al. [40] further suggested the implementation of a contact-free 3D virtual keyboard capable of combining gesture recognition and holographic display technologies to enable more convenient, intuitive, and accurate navigation during surgical procedures.

Multimodal fusion

Modality is a term referring to the channels via which organisms receive information through their perceptual organs and experiences. Humans, for instance, exhibit visual, auditory, tactile, gustatory, and olfactory perceptual channels. Multimodal systems take advantage of multiple such sensory channels [23], with the aim of overcoming the inherent limitations associated with any single sensory modality that can compromise the utility and naturalness of any mid-air gesture interactions based on that modality. Multimodal gesture recognition strategies have thus emerged as an important area of research interest. Hui et al. [41] designed a contactless multimodal medical image

interaction system equipped with synergistic 2D laser localization, 3D gesture, and voice recognition tools to allow for multimodal interaction that is seamless and capable of overcoming the limitations associated with other systems, such as issues associated with the need for large movements to achieve gesture recognition, the need for frequent trips to the display by the surgeon, and difficulties associated with the selection and comparison of target images.

Voice recognition

Voice recognition is an important component of minimally invasive surgical systems, as it can afford accuracy superior to that associated with gesture control [42, 43]. Ebert et al. [44] developed a gesture recognition-based voice command system through the combination of a Kinect device and the “OsiriX” medical image access software. Their established approach entails the use of a blob detection algorithm for the recognition of hand positioning, with voice recognition software translating data from the camera and voice commands into mouse and keyboard commands that can then be transmitted to the imaging system. Through the use of 14 voice commands, operators are thus able to switch between a range of system operating modes. To assess the usability of this system and associated response times, it was evaluated by 10 medical professionals who assigned an overall rating of 3.4/5. The main problems noted to be associated with this system included issues with ambient noise and operator accent, as the conversational style of a given surgeon or their discussions with others when using this system can interfere with overall system performance. A survey focused on speech recognition in OR settings found that noise in the OR can result in incorrect symptom responses [45]. The method of blob recognition is also relatively primitive, and the movement of the hand into and out of the viewing area can result in involuntary gesture changes, while prolonged periods of gesture-based recognition can contribute to operator fatigue. Other researchers including Nishihori et al. [46] have further combined the Kinect device and voice recognition programs to facilitate contactless 3D image manipulation during vascular neurosurgery. Their system captures voice commands through a headset microphone worn by the operator, with these commands then being transmitted to a nearby computer via Bluetooth, utilizing the Julius software to convert these voice commands into text messages. This system, when successfully implemented, can enable surgeons to focus more directly on the surgical procedure being performed while increasing the overall accuracy and smoothness of gesture control through the integration of voice commands.

Hand and foot linkage

Total reliance on gesture inputs can make achieving accurate interactions difficult, as distinguishing between hand movements as commands or interaction modes can be challenging, often contributing to accidental system activation and command misrecognition. To address this issue, Lopes et al. [47] devised the novel multimodal FEETICHE interaction system, which captures the hand and foot movements of the operator with a depth-sensing camera such that the operator can use gestures for selection and control, while utilizing foot movements such as heel rotations or toe taps to switch modes or make fine adjustments, improving the intuitive nature of the system. This multimodal

interaction can enable the operator to achieve constant physical stability while lessening the burden on their hands and reducing the levels of fatigue associated with extended sessions of gesture manipulation. Paulo et al. [48] developed an alternative contactless medical image access system suitable for use while both sitting and standing, allowing dentists to use both hands as 3D gesture cursors while enabling image navigation and manipulation through simple one-foot movements. These authors interviewed 18 dental specialists and found that in most cases the doctors agreed that this system would offer substantial value in clinical practice.

Optimization and improvement

Gesture overlap and jitter

The LMC system exhibits instability in its detection of hand movements, contributing to poorer accuracy when moving the hand to a position that results in the obstruction of the controller, such as when the fingers overlap or when the hand is rotated perpendicular to the controller [49]. In these instances, the controller is not capable of tracking or reading hand movements. To address this issue, Meta lab has developed new gesture algorithms capable of recognizing the hands even when occluded [23]. Jinshu et al. [50] proposed the development of a spring-based model for overcoming issues associated with gesture jittering. Their model selects the first gesture within a given time period as a template, with subsequent gestures then being compared to that template gesture. Gestures that are sufficiently close to the template gesture will be locked to the template gesture such that the hand position no longer changes. When the gesture is no longer sufficiently close to the template gesture in question, the system instead selects the gesture as a new template gesture. Using this spring-based model can help bridge issues that arise in the context of contactless gesture interaction, storing and analyzing the most recent sequence of gestures as a means of determining the gesture state of an operator (i.e., whether they are preparing, executing, or completing a gesture). This model can also overcome difficulties in directly accessing the gesture state for traditional interactions while providing methodological support that can enable reliable contact-free gesture-based interactions in an OR setting.

Multi-user environments and camera occlusion

OR environments generally contain multiple medical staff any given time such that any camera-based systems will likely capture several people working simultaneously around the operating table when compiling a 3D dataset. To permit cameras to segment multiple people and to individually recognize hand types such that the system is only able to interact with the attending surgeon on a one-to-one basis, there are several potential approaches that can be employed to improve the system. For one, the camera can be mounted at a high position or on the ceiling such that it can clearly capture the actions of all users. Moreover, an algorithm can be used to select the frontal view of the target user of interest, such as a facial landmark detection algorithm [51]. In addition, multiple cameras can enable the capture of data from multiple angles of view that can then be used for the reconstruction of 3D objects, replacing or reconstructing occluded parts via an appropriate reconstruction technique [52].

The Midas touch problem

The so-called Midas Touch problem is one which refers to the accidental activation of contactless gesture interaction systems or the generation of undesirable commands as a result of the sensor system having mistakenly recognized an unintentional gesture such as a hand wave or pointing at a particular object as being an intentional input [53]. This problem has the potential to impair the practical real-world adoption of these sorts of contactless systems. Cronin et al. [54] employed four clutching techniques including voice, gaze, gesture, and active zone to regulate the activation and deactivation of control over a system, thus enabling contactless systems to more accurately interpret whether operators are intentionally providing commands. The authors then assessed this technique and its interactive performance. Additionally, Schreiter et al. [55] demonstrated that the combination of gestures and voice commands in a multimodal interface could help achieve interaction that was perceived as more natural and intuitive. By using voice commands for clutching control and gesture actions for continuous control, operators can select the most appropriate mode of interaction as appropriate, leading to better overall operative accuracy and efficiency.

Pinch gestures

Contactless gesture systems often employ virtual cursor-based interactions. This requires the operator to target the part of the screen on which clicking is desired and to maintain the gesture in the air, activating this area using a pushing gesture by moving the hand towards the display in place of a “click”. Owing to cursor instability, ambiguities between whether or not the cursor is in the clicked or unclicked state, and unintentional cursor displacement when pushing the hand forward can interfere with the ability of the operator to effectively control these continuous interactions. When viewing CT images, sustained zooming and scrolling is often necessary for each image. Pinch gestures can enable continuous interactions without being dependent on the depth of the hand’s forward thrust when determining an action in contrast to other gestures, providing two clearly defined states that can be readily distinguished (i.e., fingers together vs. spread apart for an appropriate interval). Introducing pinch gestures can thus provide operators with the ability to control continuous interactions, reducing inappropriate outcomes and improving the overall interactive experience [56, 57].

Discussion

This review offers an overview of recent advances in research focused on the development of contactless image access systems with a focus on their feasibility and potential utility in future clinical practice. To date, the majority of the developed techniques appear to have yet to be rigorously tested [58], and most extant studies of these systems have largely focused on qualitative metrics or task completion times [29], with no focused quantitative analyses of the accuracy of image control. Despite marked advances in recent decades, most trials have been performed in surgical environments that are simulated, whereas relatively few have been performed in real-world surgical settings in the clinic, and of those most have only been tested in a single hospital. This lack of large-scale studies and ecological validation efforts represents a major barrier

to the adoption of the technology. Future efforts should focus on the total integration of machine intelligence and artificial intelligence systems. For instance, adaptive gesture control systems can provide the system with a means of learning the gesture habits of a given surgeon and automatically adjusting the recognition algorithm in a personalized manner. The further maturation of augmented reality technologies and the resultant use of gestures to facilitate interaction and navigation in augmented reality settings matures, operators will be able to achieve more natural experiences. Additional developments of interaction systems that do not necessitate the use of both hands will also have important future implications.

Conclusion

The studies discussed in this review highlight a gradual shift from specific technical challenges to more pressing fundamental images. While extant technologies hold great promise as a means of enabling contactless image system operation in OR settings, at present the literature suggests that no one technology is sufficiently mature to achieve widespread acceptance or adoption. These contactless HCI systems nonetheless represent an important step towards overcoming persistent issues of sterility that have long plagued OR settings, and these revolutionary techniques thus offer great potential to transform surgical practice. Importantly, these techniques will have profound benefits for the surgical treatment of patients with many diseases, achieving the paramount goal of improving surgical outcomes for patients.

Abbreviations

OR	Operating room
HCI	Human–computer interaction
CT	Computed tomography
IMU	Inertial measurement unit
EMG	Electromyography
IR	Infrared
LMC	Leap motion controller
NUI	Natural user interface
SVM	Support vector machine
MYO	Myoelectric gesture-controlled
CW	Continuous wave
BLE	Bluetooth low energy
FSK	Frequency-shift keying
BCI	Brain–computer interface

Author contributions

(i) Funding acquisition: JL; (ii) conception and design: JL, BW; (iii) provision of study materials: ZNL, CLL; YX; (iv) investigation: BW, ZNL, HLX; (v) visualization: JXL, WC, HNN; (vi) manuscript writing: ZNL, BW; (vii) final approval of manuscript: JL, BW.

Funding

This work was supported by Hainan Provincial Department of Science and Technology (ZDYF2021GXJS004).

Availability of data and materials

No datasets were generated or analyzed during the current study.

Declarations

Competing interests

The authors declare no competing interests.

Received: 14 July 2024 Accepted: 25 October 2024
Published: 30 October 2024

References

- Nishikawa A, Hosoi T, Koara K, Negoro D, Hikita A, Asano S, et al. FACE MOUSE: a novel human-machine interface for controlling the position of a laparoscope. *IEEE Trans Robot Automat*. 2003;19:825–41. <https://doi.org/10.1109/TRA.2003.817093>.
- Wang H, Ru B, Miao X, Gao Q, Habib M, Liu L, et al. MEMS devices-based hand gesture recognition via wearable computing. *Micromachines*. 2023;14:947. <https://doi.org/10.3390/mi14050947>.
- Wachs JP, Stern HI, Edan Y, Gillam M, Handler J, Feied C, et al. A gesture-based tool for sterile browsing of radiology images. *J Am Med Inform Assoc*. 2008;15:321–3. <https://doi.org/10.1197/jamia.M2410>.
- Oshiro Y, Ohuchida K, Okada T, Hashizume M, Ohkohchi N. Novel imaging using a touchless display for computer-assisted hepato-biliary surgery. *Surg Today*. 2017;47:1512–8. <https://doi.org/10.1007/s00595-017-1541-7>.
- Kurillo G, Hemingway E, Cheng M-L, Cheng L. Evaluating the accuracy of the azure kinect and kinect v2. *Sensors*. 2022;22:2469. <https://doi.org/10.3390/s22072469>.
- Ruppert GCS, Reis LO, Amorim PHJ, De Moraes TF, Da Silva JVL. Touchless gesture user interface for interactive image visualization in urological surgery. *World J Urol*. 2012;30:687–91. <https://doi.org/10.1007/s00345-012-0879-0>.
- Tan JH, Chao C, Zawaideh M, Roberts AC, Kinney TB. Informatics in radiology: developing a touchless user interface for intraoperative image control during interventional radiology procedures. *Radiographics*. 2013;33:E61–70. <https://doi.org/10.1148/rg.332125101>.
- Yoshimitsu K, Muragaki Y, Maruyama T, Yamato M, Iseki H. Development and initial clinical testing of “opect”: an innovative device for fully intangible control of the intraoperative image-displaying monitor by the surgeon. *Oper Neurosurg*. 2014;10:46–50. <https://doi.org/10.1227/NEU.0000000000000214>.
- Gobhiran A, Wongjunda D, Kiatsoontorn K, Charoenpong T. Hand movement-controlled image viewer in an operating room by using hand movement pattern code. *Wirel Pers Commun*. 2022;123:103–21. <https://doi.org/10.1007/s11277-021-09121-8>.
- Liu J, Tateyama T, IWAMOTO Y, Chen Y-W. A preliminary study of kinect-based real-time hand gesture interaction systems for touchless visualizations of hepatic structures in surgery. *Med Imag Infor Sci*. 2019;36:128–35. <https://doi.org/10.11318/mii.36.128>.
- Glinkowski WM, Miściór T, Sitnik R. Remote, touchless interaction with medical images and telementoring in the operating room using a kinect-based application—a usability study. *Appl Sci*. 2023;13:11982. <https://doi.org/10.3390/app132111982>.
- Weichert F, Bachmann D, Rudak B, Fisseler D. Analysis of the accuracy and robustness of the leap motion controller. *Sensors*. 2013;13:6380–93. <https://doi.org/10.3390/s130506380>.
- Vysocký A, Grushko S, Oščádal P, Kot T, Babjak J, Jánoš R, et al. Analysis of precision and stability of hand tracking with leap motion sensor. *Sensors*. 2020;20:4088. <https://doi.org/10.3390/s20154088>.
- Feng Y, Uchidiuno UA, Zahir HR, George I, Park AE, Mentis H. Comparison of kinect and leap motion for intraoperative image interaction. *Surg Innov*. 2021;28:33–40. <https://doi.org/10.1177/1553350620947206>.
- Rosa GM, Elizondo ML. Use of a gesture user interface as a touchless image navigation system in dental surgery: case series report. *Imaging Sci Dent*. 2014;44:155. <https://doi.org/10.5624/isd.2014.44.2.155>.
- Chiang P-Y, Chen C-C, Hsia C-H. A touchless interaction interface for observing medical imaging. *J Vis Commun Image Represent*. 2019;58:363–73. <https://doi.org/10.1016/j.jvcir.2018.12.004>.
- Hatscher B, Mewes A, Pannicke E, Kägebein U, Wacker F, Hansen C, et al. Touchless scanner control to support MRI-guided interventions. *Int J CARS*. 2020;15:545–53. <https://doi.org/10.1007/s11548-019-02058-1>.
- Zhang X, Wang J, Dai X, Shen S, Chen X. A non-contact interactive system for multimodal surgical robots based on LeapMotion and visual tags. *Front Neurosci*. 2023;17:1287053. <https://doi.org/10.3389/fnins.2023.1287053>.
- Sa-nguannarm P, Charoenpong T, Chianrabutra C, Kiatsoontorn K. A Method of 3D Hand Movement Recognition by a Leap Motion Sensor for Controlling Medical Image in an Operating Room. 2019 First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP), Bangkok, Thailand: IEEE; 2019; 17–20. <https://doi.org/10.1109/ICA-SYMP.2019.8645985>.
- 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. <https://doi.org/10.1016/j.cmpb.2018.04.003>.
- Ameur S, Ben Khalifa A, Bouhlel MS. Hand-Gesture-Based Touchless Exploration of Medical Images with Leap Motion Controller. 2020 17th International Multi-Conference on Systems, Signals & Devices (SSD), Monastir, Tunisia: IEEE; 2020; 6–11. <https://doi.org/10.1109/SSD49366.2020.9364244>.
- Cronin S, Doherty G. Touchless computer interfaces in hospitals: a review. *Health Informatics J*. 2019;25:1325–42. <https://doi.org/10.1177/1460458217748342>.
- Xue Z. Foresight interaction: from voice and gesture design to multimode convergence. Beijing: Publishing House of Electronics Industry; 2022.
- di Tommaso L, Aubry S, Godard J, Katranji H, Pauchot J. Un nouvel interface homme machine en neurochirurgie: le Leap Motion®. Note technique sur une nouvelle interface homme machine sans contact. *Neurochirurgie*. 2016;62:178–81. <https://doi.org/10.1016/j.neuchi.2016.01.006>.
- 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;12:291–305. <https://doi.org/10.1007/s11548-016-1480-6>.
- Ahmed S, Kallu KD, Ahmed S, Cho SH. Hand gestures recognition using radar sensors for human-computer-interaction: a review. *Remote Sens*. 2021;13:527. <https://doi.org/10.3390/rs13030527>.

27. Jalaliniya S, Smith J, Sousa M, Bütthe L, Pederson T. Touch-less interaction with medical images using hand & foot gestures. Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication, Zurich Switzerland: ACM. 2013; 1265–74. <https://doi.org/10.1145/2494091.2497332>.
28. Bigdelou A, Schwarz L, Navab N. An adaptive solution for intra-operative gesture-based human-machine interaction. Proceedings of the 2012 ACM international conference on Intelligent User Interfaces, Lisbon Portugal: ACM. 2012; 75–84. <https://doi.org/10.1145/2166966.2166981>.
29. Esfandiari H, Troxler P, Hodel S, Suter D, Farshad M, Collaboration Group, et al. Introducing a brain-computer interface to facilitate intraoperative medical imaging control—a feasibility study. *BMC Musculoskelet Disord*. 2022;23:701. <https://doi.org/10.1186/s12891-022-05384-9>.
30. Hettig J, Mewes A, Riabikin O, Skalej M, Preim B, Hansen C. Exploration of 3D Medical Image Data for Interventional Radiology using Myoelectric Gesture Control. *Eurographics Workshop on Visual Computing for Biomedicine* 2015.
31. Sánchez-Margallo FM, Sánchez-Margallo JA, Moyano-Cuevas JL, Pérez EM, Maestre J. Use of natural user interfaces for image navigation during laparoscopic surgery: initial experience. *Minim Invasive Ther Allied Technol*. 2017;26:253–61. <https://doi.org/10.1080/13645706.2017.1304964>.
32. Liu Y, Peng X, Tan Y, Oyemakinde TT, Wang M, Li G, et al. A novel unsupervised dynamic feature domain adaptation strategy for cross-individual myoelectric gesture recognition. *J Neural Eng*. 2023;20:066044. <https://doi.org/10.1088/1741-2552/ad184f>.
33. Xu H, Xiong A. Advances and disturbances in sEMG-based intentions and movements recognition: a review. *IEEE Sens J*. 2021;21:13019–28. <https://doi.org/10.1109/JSEN.2021.3068521>.
34. Ha M-K, Phan T-L, Nguyen D, Quan N, Ha-Phan N-Q, Ching C, et al. Comparative analysis of audio processing techniques on doppler radar signature of human walking motion using CNN models. *Sensors*. 2023;23:8743. <https://doi.org/10.3390/s23218743>.
35. Miller E, Li Z, Mentis H, Park A, Zhu T, Banerjee N. RadSense: enabling one hand and no hands interaction for sterile manipulation of medical images using Doppler radar. *Smart Health*. 2020;15:100089. <https://doi.org/10.1016/j.smhl.2019.100089>.
36. Yang K, Kim M, Jung Y, Lee S. Hand gesture recognition using FSK radar sensors. *Sensors*. 2024;24:349. <https://doi.org/10.3390/s24020349>.
37. Stetco C, Muhlbacher-Karrer S, Lucchi M, Weyrer M, Faller L-M, Zangl H. Gesture-based Contactless Control of Mobile Manipulators using Capacitive Sensing. 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Dubrovnik, Croatia: IEEE. 2020; 1–6. <https://doi.org/10.1109/I2MTC43012.2020.9128751>.
38. NZ technologies inc. HoverTap MD™ - NZ Technologies Inc. - Touchless Technology n.d. <https://www.nztech.ca/hovertap-md/>. Accessed 4 Jan 2024.
39. NZ technologies inc. TIPSO AirPad™ - NZ Technologies Inc. - Touchless Technology n.d. <https://www.nztech.ca/airpad/>. Accessed 4 Jan 2024.
40. Yang Y, Gao Y, Liu K, He Z, Cao L. Contactless human–computer interaction system based on three-dimensional holographic display and gesture recognition. *Appl Phys B*. 2023;129:192. <https://doi.org/10.1007/s00340-023-08128-2>.
41. Hui WS, Huang W, Hu J, Tao K, Peng Y. A new precise contactless medical image multimodal interaction system for surgical practice. *IEEE Access*. 2020;8:121811–20. <https://doi.org/10.1109/ACCESS.2019.2946404>.
42. Perrakis A, Hohenberger W, Horbach T. Integrated operation systems and voice recognition in minimally invasive surgery: comparison of two systems. *Surg Endosc*. 2013;27:575–9. <https://doi.org/10.1007/s00464-012-2488-9>.
43. Argoty JA, Figueroa P. Design and development of a prototype of an interactive hospital room with Kinect. Proceedings of the XV International Conference on Human Computer Interaction, Puerto de La Cruz Tenerife Spain: ACM. 2014; 1–4. <https://doi.org/10.1145/2662253.2662290>.
44. Ebert LC, Hatch G, Ampanozi G, Thali MJ, Ross S. You can't touch this: touch-free navigation through radiological images. *Surg Innov*. 2012;19:301–7. <https://doi.org/10.1177/1553350611425508>.
45. Schulte A, Suarez-Ibarrola R, Wegen D, Pohlmann P-F, Petersen E, Miernik A. Automatic speech recognition in the operating room—an essential contemporary tool or a redundant gadget? A survey evaluation among physicians in form of a qualitative study. *Ann Med Surg*. 2020;59:81–5. <https://doi.org/10.1016/j.jamsu.2020.09.015>.
46. Nishihori M, Izumi T, Nagano Y, Sato M, Tsukada T, Kropp AE, et al. Development and clinical evaluation of a contactless operating interface for three-dimensional image-guided navigation for endovascular neurosurgery. *Int J CARS*. 2021;16:663–71. <https://doi.org/10.1007/s11548-021-02330-3>.
47. Lopes D, Relvas F, Paulo S, Rekik Y, Grisoni L, Jorge J. FEETICHE: FEET Input for Contactless Hand gEsture Interaction. Proceedings of the 17th International Conference on Virtual-Reality Continuum and its Applications in Industry, Brisbane QLD Australia: ACM. 2019; 1–10. <https://doi.org/10.1145/3359997.3365704>.
48. Paulo SF, Relvas F, Nicolau H, Rekik Y, Machado V, Botelho J, et al. Touchless interaction with medical images based on 3D hand cursors supported by single-foot input: a case study in dentistry. *J Biomed Inform*. 2019;100:103316. <https://doi.org/10.1016/j.jbi.2019.103316>.
49. Potter LE, Araullo J, Carter L. The Leap Motion controller: a view on sign language. Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration, Adelaide Australia: ACM. 2013; 175–8. <https://doi.org/10.1145/2541016.2541072>.
50. Lei J, Wang S, Zhu D, Wu Y. Non-contact gesture interaction method based on cursor model in immersive medical visualization. *J Comput Aided Des Comput Gr*. 2019;31:208–17. <https://doi.org/10.3724/SPJ.1089.2019.17593>.
51. Wu B-F, Chen B-R, Hsu C-F. Design of a facial landmark detection system using a dynamic optical flow approach. *IEEE Access*. 2021;9:68737–45. <https://doi.org/10.1109/ACCESS.2021.3077479>.
52. Siratanita S, Chamnongthai K, Muneyasu M. A method of football-offside detection using multiple cameras for an automatic linesman assistance system. *Wirel Pers Commun*. 2021;118:1883–905. <https://doi.org/10.1007/s11277-019-06635-0>.
53. Freitas A, Santos D, Lima R, Santos CG, Meiguins B. Pactolo bar: an approach to mitigate the Midas touch problem in non-conventional interaction. *Sensors*. 2023;23:2110. <https://doi.org/10.3390/s23042110>.

54. Cronin S, Freeman E, Doherty G. Investigating Clutching Interactions for Touchless Medical Imaging Systems. CHI Conference on Human Factors in Computing Systems, New Orleans LA USA: ACM. 2022; 1–14. <https://doi.org/10.1145/3491102.3517512>.
55. Schreiter J, Mielke T, Schott D, Thormann M, Omari J, Pech M, et al. A multimodal user interface for touchless control of robotic ultrasound. *Int J CARS*. 2022;18:1429–36. <https://doi.org/10.1007/s11548-022-02810-0>.
56. Waugh K, McGill M, Freeman E. Push or Pinch? Exploring Slider Control Gestures for Touchless User Interfaces. Nordic Human-Computer Interaction Conference, Aarhus Denmark: ACM. 2022; 1–10. <https://doi.org/10.1145/3546155.3546702>.
57. Waugh K, McGill M, Freeman E. Proxemic Cursor Interactions for Touchless Widget Control. Proceedings of the 2023 ACM Symposium on Spatial User Interaction, Sydney NSW Australia: ACM. 2023; 1–12. <https://doi.org/10.1145/3607822.3614525>.
58. Chung J, Liu DM. Experimental assessment of a novel touchless interface for intraprocedural imaging review. *Cardiovasc Intervent Radiol*. 2019;42:1192–8. <https://doi.org/10.1007/s00270-019-02207-8>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.