

Received August 9, 2019, accepted September 1, 2019, date of publication October 10, 2019, date of current version July 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2946404

# A New Precise Contactless Medical Image Multimodal Interaction System for Surgical Practice

WANG SHI HUI<sup>1,2</sup>, WEIHONG HUANG<sup>1</sup>, JIANZHONG HU<sup>1</sup>,  
KUN TAO<sup>3</sup>, AND YONGHONG PENG<sup>4</sup>

<sup>1</sup>Mobile Health Ministry of Education–China Mobile Joint Laboratory, Xiangya Hospital Central South University, Changsha 410008, China

<sup>2</sup>School of Architecture and Art, Central South University, Changsha 410008, China

<sup>3</sup>AIHealthX Company Ltd., Beijing 100085, China

<sup>4</sup>Faculty of Computer Science, University of Sunderland, Sunderland SR6 0DD, U.K.

Corresponding authors: Weihong Huang (whuangcn@qq.com) and Yonghong Peng (yonghong.peng@sunderland.ac.uk)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0909900 and Grant 2018YFC2002400, and in part by the Key Research and Development Program of the Hunan Province of China under Grant 2017SK2013.

**ABSTRACT** Blind spot confirmation as a major challenge in surgical practice has led to booming of 3D modeling and printing technology and applications, but existing technologies could not offer readily available service of medical images in the operation room (OR) for surgeons. Contacting with keyboards and mice or asking assistants to help in viewing medical images during the operation could increase operation risk, infection and prolong of the operation time. This paper presents a new precise contactless medical image multimodal interaction system for surgeons in the OR. By supporting seamless multimodal interaction via 2D laser positioning, 3D gesture and voice recognition, it overcomes the design flaws in other similar solutions such as big motion for gesture recognition, round trip towards the screen of surgeons, difficulty in selection and comparison of images, and more. Following the system engineering and evidence-based approaches, the design and development of this new precise contactless multimodal medical image interaction system meet surgeons' real world requirements for efficient and effective 2D medical images and 3D model operation under both preoperative and intraoperative scenarios. The system evaluation study is carried out in conjunction with a group of 10 neurosurgeons with 3 to 10 years' of surgical experience from different hospitals. Evaluation results successfully demonstrate the system's feasibility, efficiency and usability of in surgical contexts.

**INDEX TERMS** Contactless multimodal medical image system, human-computer interaction, usability, operation context.

## I. INTRODUCTION

Healthcare-associated infections (HAIs) are a major problem. In the United States, HAIs cause 99,000 attributable deaths and cost \$6,500,000,000 US every year. In Europe, they result in 16,000,000 extra days of hospital stays, 37,000 attributable deaths and cost €7,000,000,000 every year [1]. In China, over 4 million HAIs occur annually, for a total economic loss of 112.29–331.87 million RMB [2].

The causes of HAIs are complex and diverse. The best way to prevent HAIs is to improve hand hygiene and

The associate editor coordinating the review of this manuscript and approving it for publication was Ying Song<sup>id</sup>.

contactless operation in pathogen-sensitive hospital scenarios, such as operating rooms [3].

Under the sterile operation restriction in the OR, surgeons confronted with blind spots under variable operating situations cannot make a solo confirmation immediately; instead, they must ask a nurse for assistance to acquire 2D medical images and 3D models [4]–[6].

Given the limited size and number of printed CT and MRI films, imprecise 3D print models, and the delays and inaccuracies of operating assistants in OR situations, clarifying blind spots becomes even more difficult and affects the operating procedure, patient safety, takes time and notably raises costs [7]–[9].

To clarify blind spots immediately, some surgeons pull their gowns over their hands to manipulate a mouse through the gown; however, this is not entirely free from contamination risk [10].

In contrast to the numerous recent rapid developments in medical image software, 3D modeling and contactless human computer interaction technology, an overall usable solution in operating scenarios to support surgeons for effective and efficient contactless operation of medical imagery equipment and 3D models is lacking.

Although surgeons expect contactless medical image systems to be available in surgical scenarios, providing proper solutions is not an easy task because it involves in-depth observation and understanding of complex operating scenarios, domain knowledge of the medical images used in surgical practice, and expertise with suitable contactless technologies—including human computer interaction methods. This variety of requirements makes contactless medical image operation a challenging interdisciplinary research, design and development topic in the fields of clinical practice and human computer interaction.

To develop a contactless medical image operation system suitable for surgical scenarios, we initially conducted interviews with neurosurgeons to gather the requirements for contactless medical image operation. Then, contextual inquiries about operating processes were conducted to clarify the operative scenarios in which contactless medical image operation should be applied. Later, a novel contactless multimodal medical image operation system for surgical scenarios was proposed with both a suitable industrial design (ID) and a corresponding UI (User Interface) design. Finally, system usability was evaluated and demonstrated through usability experiments with 10 neurosurgeons.

The main contributions of this paper are as follows:

- 1) It summarizes surgeons' general requirements in operating scenarios as follows: contactless, efficient, effective and easy to use operation of 2D medical image and 3D models solution;
- 2) It clarifies the layout of surgeons, nurses and equipment positions in both small and large Ors during operating procedures to acquire contactless multimodal medical image system ID references;
- 3) It reveals the surgeon's operation habits and cognitive experiences concerning operating medical image software with a mouse, providing corresponding UI (User Interface) design references for contactless multimodal medical image systems;
- 4) It proposes a novel contactless multimodal medical image operation system, demonstrated through usability experiments with 10 neurosurgeons with between 3 and 10 years of surgical experience that has the required characteristics: ease of use, learnability, efficiency, effectiveness, low error rate during task completion and high recognition rates in simulated surgical scenarios.

The remainder of this paper is organized as follows: Section II provides an extensive review of related works, revealing that the current systems fail to demonstrate usability in terms of ease of use, learnability, efficiency, effectiveness and accuracy rate. Section III reports the results of a requirements study: it presents the existing problems and requirements from interviews and a field study in both preoperative and intraoperative scenarios. Section IV presents suitable HW (hardware) interactions, HW components, ID proposal, and SW (software) interactions with their corresponding UI designs for the new contactless multimodal medical image operation system. Section V reports the design and execution of usability experiments, including their implementation and testing processes, and the evaluation results. Finally, Section VI concludes this work and discusses possibilities for future research and design directions.

## II. LITERATURE REVIEW

To date, 60% of the contactless human-computer interaction technologies in OR have been applied to medical image operations. These efforts can be divided into the following 4 major research directions: (1) gesture and body posture capture based on RGB or webcam, structured-light-based ranging (Kinect), or stereo cameras (Leap Motion) [10]–[12]; (2) foot posture, arm movement and finger pointing direction capture based on ground or wearable sensors [13]–[15]; (3) eye movement tracking [16]; and (4) voice recognition [17].

Studies investigating applications of these technologies to medical image operation in OR have revealed the following major usability problems that affect application acceptance in surgical scenarios [18]–[20]:

- 1) Single-mode operation has low efficiency;
- 2) Mis-recognition of users' unconscious commands;
- 3) High user cognitive load and low learnability;

The definition of usability from the HCI doctor and usability expert Jakob Nielsen provides evaluation quality properties for human computer interactions and includes 5 indicators: ease of use, learnability, effectiveness, efficiency, errors and satisfaction [21].

In terms of efficiency issues, because medical image operations have both discrete and continuous characteristics, applying single-mode contactless human-computer interaction limits the ability to improve efficiency for both discrete and continuous tasks [22]. Recent efficiency experiments for medical image operation by human computer interaction modes of voice and gesture [23], showed that gesture interactions are more suitable for continuous tasks, while speech interaction is more efficient for discrete tasks. Thus, to improve the efficiency of medical image operations, multimodal contactless human-computer interaction must be applied.

From the mis-operation and mis-recognition of user unconscious commands point of view, previous research has shown that compared to traditional mouse-user interaction mode, applications using new contactless human computer

interactions in surgical scenarios exhibit high error rates. These system can easily misinterpret a surgeon's unconscious gestures or speech during the operating process as intended commands or recognize irrelevant commands in an OR environment [24]–[28].

To solve this problem, Rossol, Tan, Jacob, etc. have proposed solutions in system settings, multidimensional gesture design, and AI intention recognition [29]–[31]. However, these studies provided no test or validation results to demonstrate the effectiveness of their proposals; thus, these results must be addressed through further verification.

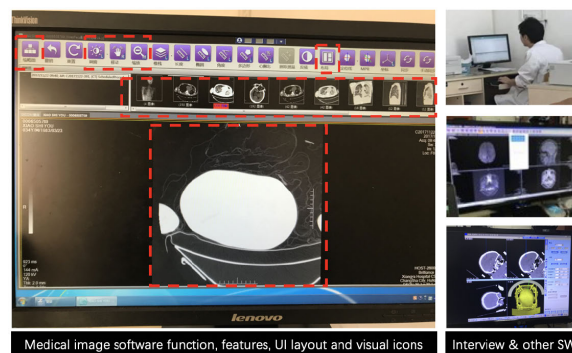
To reduce user cognitive load and increase the learnability of the new contactless interactive mode, several approaches have been explored. Yongki Park proposed a gesture-selected menu with sound and visual feedback to improve learnability for task completion [32]. David added sound feedback for selection and confirmation of an eye movement tracking system and provided a subsequent experimental demonstration showing that sound feedback can reduce a user's reaction time and cognitive load compared to visual feedback [33].

However, little attention has been paid to operating room scenarios. Excessive and improper sound feedback may interfere with the operation, especially when mixed with the noise of medical equipment and communications between surgeons and nurses.

From the above studies, it is clear that a good and usable contactless medical image operation system requires a comprehensive consideration of the efficiency, effectiveness, and learnability of its human computer interaction components in a surgical context. It is important to gain an in-depth understanding of medical image applications, operating procedures, and surgeons' previous cognitive experiences in surgical practices. A highly usable system should be built based on practical needs and be able to be applied seamlessly in modern surgical environments.

To reach the goal, this research is divided into 5 phases:

- 1) First, current problems in surgical practice and expectations for contactless medical image systems were collected through in-depth interviews with surgeons;
- 2) Then, further design requirements in surgical scenarios were revealed through a field study of operating practices and OR layouts;
- 3) Based on the results from the above studies, a new contactless human computer interaction system for medical image operation was designed and developed, including hardware and ID design, and software with corresponding UI design;
- 4) Furthermore, tests were performed, and design improvements were made to improve the performance of the contactless multimodal system;
- 5) Finally, by carrying out usability experiments with surgeons, each of whom had between 3 and 10 years' of operating experience, the new contactless multimodal human computer interaction system for medical image operation was shown to be applicable and effective for clinical practice.



**FIGURE 1. In-depth interviews with neurosurgeons and investigations of frequently used medical image software.**

**TABLE 1. Problems and operational demands.**

Requirements	Solution & Problems	Expectation
1. Confirm lesions	Review limited number and size of CT films at film viewer	Access to film reviews with unlimited numbers and large sizes
2. Confirm lesion is fully removed	Compare CT films at film viewer with assistant from nurse	Quick review and surgeon-controlled operations
3. Confirm blind spot	View 3D printed model (sometimes the model is printed incorrectly)	Quickly review a correct 3D model
4. Confirm blurred area	View PAC software on PC with nurse's assistance	Quickly enlarge images through surgeon-controlled reviewing windows

### III. FIELD INVESTIGATION AND DEMAND ANALYSIS

#### A. REQUIREMENTS INTERVIEWS IN DOCTORS' OFFICES

In-depth interviews with neurosurgeons to discover the requirements for a contactless medical image operation system were conducted in doctors' offices at the Neurosurgery Department of Xiangya Hospital of Central South University, as shown in Figure 1.

Based on the in-depth interviews with neurosurgeons and investigations of frequently used medical image software, we describe the operational problems, current solutions and surgeons' expectations for a new contactless medical image system in Table 1.

#### B. SURGERY SCENARIO FIELD RESEARCH

Based on the user requirements from in-depth interviews, we conducted contextual inquires and participatory observations of operating procedures at the OR of the Neurosurgery Department of Xiangya Hospital of Central South University.

The surgical equipment positions and the distributions of surgeons and nurses in both small and large operating rooms were clarified as shown in Figure 2.

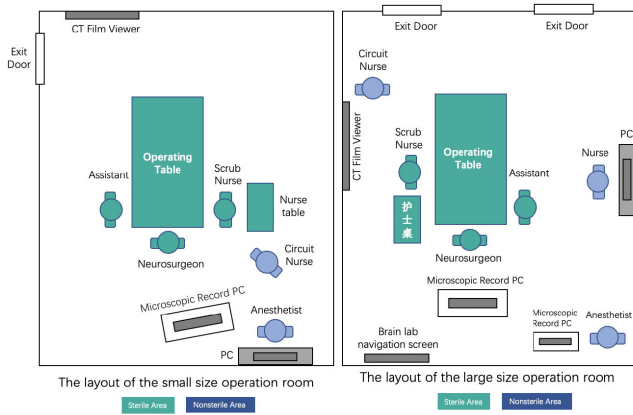


FIGURE 2. Surgical equipment and personnel positioning in ORs.



FIGURE 3. A doctor confirms medical images in front of a CT film viewer or next to the operation table with nurse assistance.

We identified three major design challenges for integrating contactless medical imaging systems into the current operating procedures and ORs:

1) Limited space exists for contactless medical image interaction;

In OR layouts both large and small, the CT film viewer is installed against the wall and located obliquely in front of the neurosurgeon, who is surrounded by the operating table, microscopic equipment, a first assistant, a scrub nurse and the nurse table, which leaves limited space for contactless medical image interaction.

By considering the current layout, the medical imaging display, together with all contactless interaction hardware, must be properly placed or installed against the wall in the same position as the current CT film viewer. To minimize interference, the contactless interaction hardware ID design should occupy minimal space when not in use.

2) The equipment must be suitable for both preoperative and intraoperative scenarios.

Before the operation starts, neurosurgeons usually review printed CT/MRI films at the CT film viewer.

During the operation, if there is a need for lesion or blind spot confirmation, the neurosurgeon usually walks to the CR/MRI film viewer or checks the viewer beside the operation table with the assistance of a nurse, as shown in Figure 3.

To integrate into both preoperative and intraoperative scenarios, the contactless multimodal medical image operations system must support operation both in front of the medical screen and adjacent to the operating table.

3) Noise disturbances in the OR

TABLE 2. Uses cases for contactless medical image systems.

USE CASE TYPE	DESCRIPTION
1. Wake up	Wake the system via speech or gesture
2. Identity validation	Access to medical image database via voice recognition of responsible surgeons
3. Patient & CT / MRI images retrieval	Effective and efficient retrieval of target patient info and CT/MRI images via the speech/gesture interactions
4. Patient & CT / MRI image set selection	Precise selection of target patients and CT / MRI image set via speech/gesture interactions
5. 2D image operation	Windowing, zooming, quick browsing and comparison, measuring via multimodal interaction both beside the operating table and in front of the medical screen
6. 3D model operation	Rotating, zooming, measuring, selection and hiding via multimodal interaction both beside the operating table and in front of the medical screen

Before the operation starts and during the operating process, noise from medical equipment and from communications between surgeons and nurses may affect the performance of voice recognition systems to some extent. Thus, there is a need for systems that can capture and identify true commands in a noisy environment.

#### IV. MULTIMODAL INTERACTIVE DESIGN

##### A. USE CASES FOR INTERACTIVE SOLUTIONS

By considering neurosurgeons' requirements and the potential problems and environments of contactless medical imaging equipment in the OR, we proposed use cases for a new solution, as shown in Table 2.

##### B. HARDWARE INTERACTION AND ID DESIGN

To develop the use cases and meet the design requirements of the OR layout, the new contactless multimodal medical image system provides 2 modes of interaction:

In a preoperative scenario, surgeons can interact through voice and gestures captured via an infrared laser, camera and microphone in front of a wall-mounted/ground-mounted medical screen.

In an intraoperative scenario, surgeons can additionally interact next to the operating table through a Leap Motion controller.

The ID design for preoperative scenario HW interaction consists of the camera (No. 1) and infrared laser (No. 2) to capture a surgeon's gestures in front of the screen, sound entry apertures and microphone (No. 3) for capturing speech commands, rotating arm (No. 4) and bracket (No. 5) for hardware installation fixed behind the screen, as shown in Figure 4.

To save limited space in OR, the rotating arm can be folded and aligned with the screen when the system is inactive.



FIGURE 4. HW & ID design in the preoperative scenario.

When the surgeon walks into the operating room and is about to review target patient information and medical images prior to an operation, he could say “hello, Xiaoya” to wake the system. In response to the wake-up command, the machine arm extends to capture gestures and speech commands from surgeons in front of the screen.

Because the interaction area is focused on the limited space in front of the screen, few misconstrued operations would occur.

The ID design for HW interaction in the intraoperative scenario supports gestures near the operation table and is composed of a Leap Motion controller that captures surgeons’ gestures (No. 1), a gooseneck arm with freely adjustable angle and height (No. 2), and a clamp fixed on the operation handstand near the operation table (No. 3), as shown in Figure 5.

**C. SOFTWARE INTERACTION AND UI DESIGN**

The learnability of the new contactless medical image interactive system poses a challenge. To solve this problem, we investigated the cognitive experiences of surgeons with traditional interaction mechanism of medical image software operation.

When a user moves the mouse by hand, the on-screen cursor moves accordingly. When the mouse cursor slides over a functional icon, the icon highlights with a light red color, indicating that the function is available for activation. When the user clicks the mouse over an icon, the active icon highlights in dark red, indicating that the user can operate a related function in the image area.

To apply surgeons’ previous cognitive experiences to the new contactless interaction, the gestural interactions are designed with inactive, highlighted and active status, as shown in Figure 6.

When a surgeon’s gesture moves the cursor over a button, the button activates by turning blue and a clockwise loading animation begins. If the surgeon’s gesture hovers over this button for 2 s, the loading animation ends and the relevant function command activates.

With timely and obvious visual feedback mechanisms that function similarly to their previous cognitive experiences,



FIGURE 5. HW & ID design in intraoperative scenario.

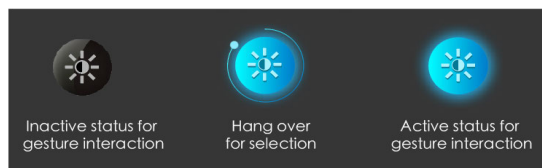


FIGURE 6. Contactless gesture interaction design.

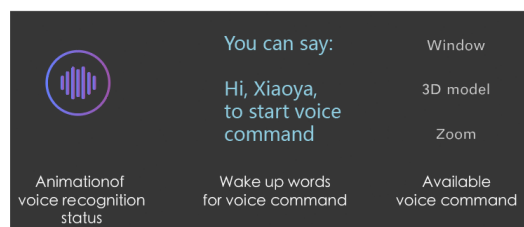


FIGURE 7. Voice interaction design.

surgeons can easily master browsing and button selection via contactless gesture interactions.

In terms of learnability and mis-operations with voice commands, the main problems are that users are unfamiliar with the available voice commands and that the system can easily capture unintended voice commands. To solve these problems, we design an animation to reveal the active voice command capture status, propose a voice command trigger phrase “Hi, Xiaoya,” and make the voice command list available on each page, as shown in Figure 7.

To reduce the cognitive load for surgeons during their initial interactions with the contactless multimodal system, we designed visual function icons by referring to the medical image software PACS, which is frequently used by surgeons, as shown in Figure 8.

**D. QUICK USER TESTS**

From 2017 to 2018, we conducted quick user tests of the new contactless multimodal medical image system with 6 surgeons.

Based on the quick test feedback, the ID design was optimized.

One ID design optimization associated with the medical image screen was to add sound apertures to the top and sides of the machine arm to improve the remote speech recognition rate beside the operating table.

In addition, a distance adjustment gear was added to the installation position of the infrared laser of the machine arm

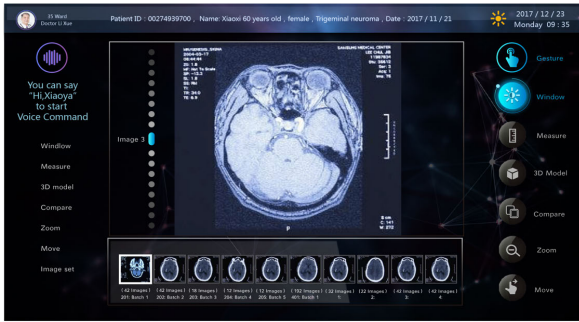


FIGURE 8. UI layout and visual design of contactless multimodal medical image systems.

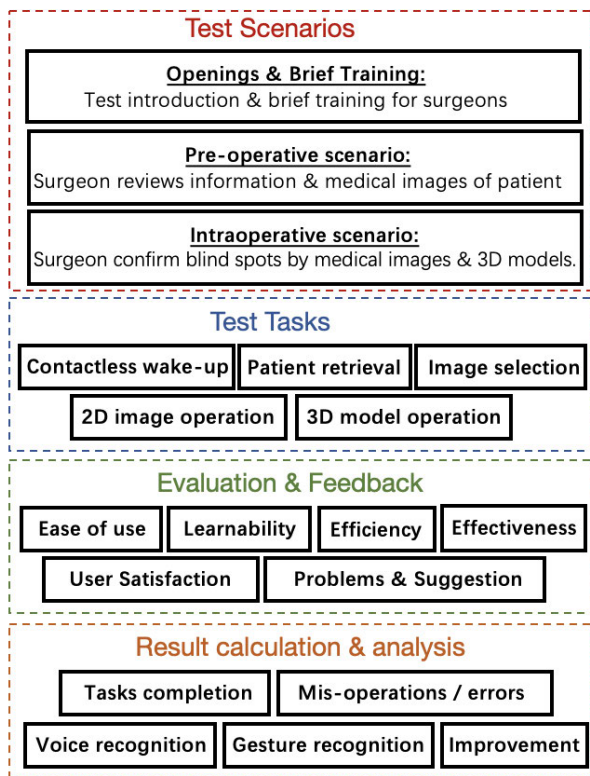


FIGURE 9. Usability experiment process design.

to allow the gesture interaction distance in front of the medical screen to be adjusted.

Another ID design optimization of the LMC besides the operating table was to add heat dissipation holes (see Figure 12, right side), because the device temperature can increase substantially during extended operations.

## V. USABILITY EXPERIMENT DESIGN

### A. EXPERIMENTAL SUBJECTS AND SETTINGS

Ten neurosurgeons with 3–10 years of surgical experience and active surgical assignments were selected as experimental subjects. Their average age was 38.

The experiment was set in the Mobile Health Ministry of Education-China Mobile Joint Laboratory, Xiangya Hospital Central South University. The experimental equipment

included the prototype multimodal interaction medical image hardware, which was placed in front of the screen and near the operating table, the medical image software system, and a camera used to record the experiment.

### B. EXPERIMENT PROCESS DESIGN

Based on the medical image application scenario in the surgery room, three experiment test scenarios were defined as follows:

- 1) Introduction and brief training; the researcher introduced the equipment to participants and encouraged them to try the new interactions freely to become accustomed to them;
- 2) Preoperative scenario: in this scenario, neurosurgeons review key information and relevant patient medical images;
- 3) Intraoperative scenario: in this scenario, neurosurgeons confirm blind and ambiguous spots using medical images and 3D models.

Under these 3 experimental test scenarios, the participants were asked to complete basic tasks such as awakening the contactless system, retrieving data for a patient, selecting CT/MRI images, performing medical 2D image operations (zooming, adjusting the window, comparing hung films, moving images, selecting an image sequence, quickly browsing, arranging multiple images, etc.) and 3D model operations (rotating, zooming, moving, hiding, etc.).

After task completion, the subjects were asked to evaluate the ease of use, learnability/memorability, efficiency, effectiveness and satisfaction of the completed operation tasks using a 5-level evaluation scale (5 = very good, 4 = good, 3 = general, 2 = slightly poor 1 = very poor) and to provide feedback for improvement in a subsequent interview.

Finally, using the video recordings of the experiments and reviewing the experiment record, the task completion rate, mis-operation/error rate, voice recognition accuracy rate, and gesture recognition accuracy rate were calculated to perform further design analysis to improve the system's performance, as shown in Figure 9.

## VI. USABILITY TEST

### A. EXPERIMENTAL PROCESS

The usability test was performed in March of 2019 in the operation simulation scenario in the Mobile Health Ministry of Education-China Mobile Joint Laboratory, Xiangya Hospital Central South University, as shown in Figure 10.

Ten neurosurgeons completed medical image operation tasks with the Xiaoya contactless multimodal human-computer interaction medical image system both in front of the screen (Figure 10, top) and at an operating table (Figure 10, bottom) via speech and gesture interaction. (Note: As the test is carried out in lab, surgeons did not wear surgeon clothes as they are in the operating room.)

### B. EXPERIMENTAL RESULTS

After the usability tests, the task completion rate, mis-operation/error rate, speech recognition rate and gesture

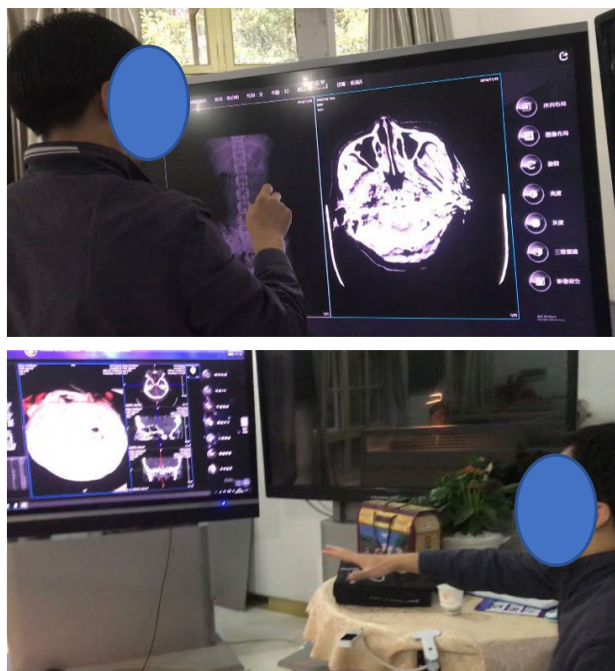


FIGURE 10. Usability test process.

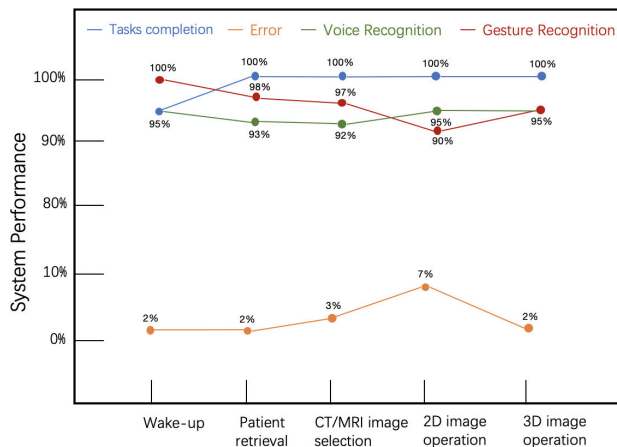


FIGURE 12. Intraoperative scenario test result.

centimeters. To solve this problem, the subjects were told after the test that they should point to the icons at the upper edge using a vertical motion; this reduced the mis-operations significantly.

In the intraoperative scenario, due to the limitations of the Leap Motion hardware, a gesture captured at certain heights can be misinterpreted as other gestures. To solve this problem, we modified the gesture recognition parameter settings to reduce misinterpretation.

Based on the comparisons with the reports of prior studies, the gesture recognition rate of the medical image software in front of a screen based on an RGB camera is approximately 64–75% [34] in well-lit environments, the gesture recognition rate based on an antebrachial muscle electrical sensor is 71%–86% [35]. Gesture and foot posture recognition rates based on wrist and floor sensors are 93% [36], and the gesture recognition rate based on a Kinect is 88% [37]. Close distance gesture recognition rate based on a Leap Motion ranges from 77% to 100% [38], and the speech recognition rate is approximately 90% [39].

Using the Xiaoya contactless multimodal human-computer interaction medical image system, the gesture interaction recognition rate by 2D laser positioning in front of the medical screen and by Leap Motion beside an operating table range from 90% to 100%. The speech recognition rate based on matched and customized interface gesture design and speech interaction design can reach 95–98%. These results validate the system’s viability and effectiveness.

The usability evaluation of the system in the preoperative and intraoperative scenarios regarding its ease of use, learnability/memorability, efficiency, effectiveness and satisfaction are shown in Figures 13 and 14.

According to the evaluation results, in the preoperative scenario, the ease of use is approximately 4.0, and learnability/memorability is approximately 4.2–4.5. The ease of system wake-up is below 4.0 because no visual indication or out-of-the-box guidance for operating the system was available, resulting in some confusion, particularly from the inexperienced users.

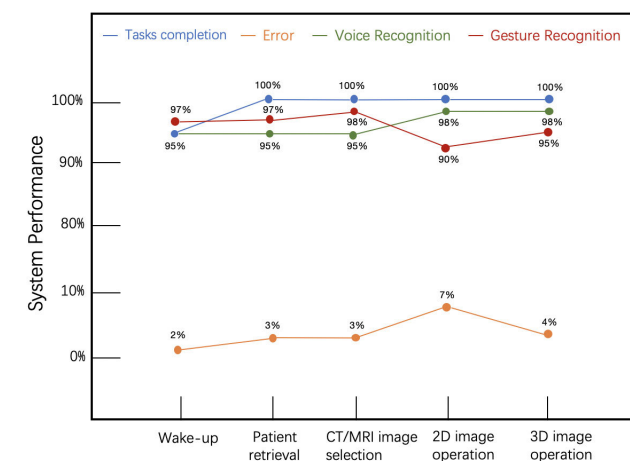


FIGURE 11. Preoperative scenario test result.

recognition rate under the preoperative and intraoperative scenarios were calculated, as shown in Figures 11 and 12.

According to the test results shown above, the task completion rate is high, the mis-operation/error rate is low, and the recognition rates for voice and gestures are above 95% in the preoperative scenario and 90% in the intraoperative scenario.

The voice recognition rate decreases slightly because distance affects the voice recognition rate. Moreover, due to the limitations of the Leap Motion hardware, the gesture recognition accuracy rate also decreases slightly.

In the preoperative scenario, some mis-operations occur when the user points a finger at the icons near the upper edge of the screen. The gesture mapping at the edge captured by the 2D laser positioning results in deviations of a few

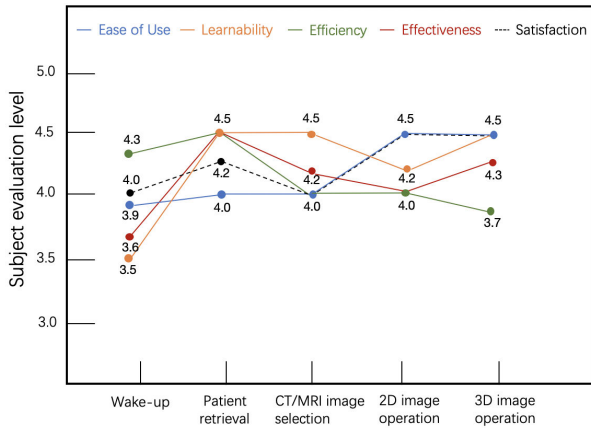


FIGURE 13. Preoperative scenario evaluation result.

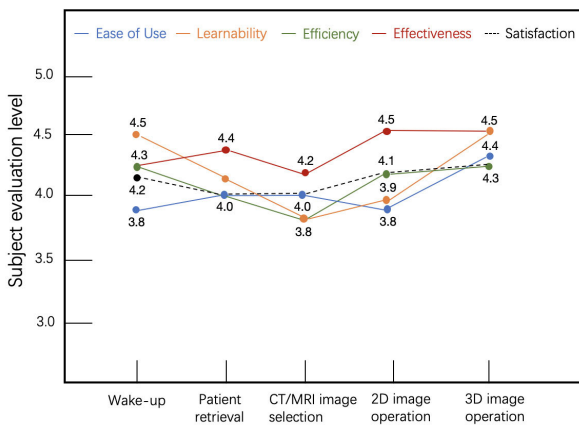


FIGURE 14. Intraoperative scenario evaluation result.

The efficiency ranges from approximately 4.0 to 4.3. Due to limitations in server performance, the speed of 3D model reconstruction and operation had a slight delay. The effectiveness approximately 4.0 to 4.3. Although infrared laser- and camera-based hardware are effective in neurosurgical scenarios, some application limitations exist during operation interventions because the medical screen hangs under the ceiling, necessitating smaller multimodal interaction hardware. The overall satisfaction evaluation ranged from approximately 4.0 to 4.5, verifying the usability of the system for clinical applications.

As shown in Figure 14, in the intraoperative scenario, the ease of use is approximately 4.0, learnability/memorability is approximately 4.0–4.5, efficiency is approximately 4.0–4.3, effectiveness is approximately 4.2 to 4.5, and the overall satisfaction level rating ranged from approximately 4.0 to 4.3, indicating the system’s usability for clinical applications.

Although the results of the usability tests and the user feedback results were positive, the subsequent user interviews indicated that further system improvements were needed.

- 1) System response time: Compared to mouse operation, the system exhibits a short response delay during

- gesture and voice interactions. This aspect must be further prioritized to improve the system’s efficiency;
- 2) Recognition accuracy: The gesture recognition rates for 1–2 icons located near the screen boundary were low due to software settings and hardware limitations; this aspect should also be improved;
- 3) Additional requirements in the intraoperative scenario: Some surgeons suggested that additional features were needed in intraoperative scenarios, such as the ability to record voice memos, shortcuts to access 3D models from different perspectives, hide and cut operations with 3D models, measurements in 3D models, and the ability to control the video connection during interventional operation.

VII. CONCLUSION AND FUTURE WORK

To design and develop a highly usable contactless medical image multimodal interaction system for surgeons in the OR, this paper provided a systematic review of the requirements of contactless medical images under surgical contexts in clinical practice. Through in-depth interviews with surgeons and investigating the distributions of surgeons, nurses and medical equipment in ORs during surgery, we clarified the existing challenges and needs for a new precise multimodal interactive medical image systems for surgeons in the OR. We designed and developed a contactless medical image multimodal interaction system prototype, ID and corresponding UI to address the major challenges in surgical practice. The system’s features include effective blind spot confirmation, immediate 2D medical image retrieval, accurate 3D model manipulation, and sterile contactless operation.

The system’s effectiveness was validated through usability experiments. The performance results and subjective evaluation feedback from 10 neurosurgeons at the Neurosurgery Department of Xiangya Hospital Central South University China were analyzed.

In addition to the multimodal interactive medical image functions provided by the current prototype system, in the future, we plan to create new features and make further technical improvements to smarter and improved system performance as listed below.

- 1) System response time: the algorithm will be prioritized to reduce delays in voice and gesture recognition;
- 2) Recognition accuracy: For large medical screens, the interactive icons should be placed appropriately to avoid low recognition rates near screen boundaries;
- 3) Other functional requirements: For different types of neurosurgeries, further functions are required, such as the ability to record voice memos, shortcuts to enable 3D model access from different perspectives, hide and cut operations with 3D models, measurement operations with 3D models, and the ability to control the video connection in interventional operation.
- 4) Effectiveness in real clinical practice: Despite our validation in simulated surgical experimental settings, system usability should be further validated in real-world



clinical practices conditions to reveal which design details still require improvements.

## ACKNOWLEDGMENT

(Shihui Wang and Weihong Huang are co-first authors.)

## REFERENCES

- [1] *Report on the Burden of Endemic Health Care-Associated Infection Worldwide*, WHO, Geneva, Switzerland, 2011.
- [2] J. Wang, F. Liu, E. Tartari, J. Huang, S. Harbarth, D. Pittet, and W. Zingg, "The prevalence of healthcare-associated infections in mainland China: A systematic review and meta-analysis," *Infection Control Hospital Epidemiol.*, vol. 39, no. 6, pp. 701–709, 2018.
- [3] S. Cronin and G. Doherty, "Touchless computer interfaces in hospitals: A review," *Health Inf. J.*, vol. 1, p. 18, 2018.
- [4] K. O'Hara, G. Gonzalez, A. Sellen, G. P. Penney, A. Varnavas, H. M. Mentis, A. Criminisi, R. Corish, M. Rouncefield, N. Dastur, and T. Carrell, "Touchless interaction in surgery," *Commun ACM*, vol. 57, no. 1, pp. 70–77, 2014.
- [5] K. O'Hara, G. Gonzalez, A. Sellen, R. Corish, A. Criminisi, M. Rouncefield, N. Dastur, T. Carrell, A. Varnavas, H. Mentis, and G. Penney, "Interactional order and constructed ways of seeing with touchless imaging systems in surgery," *Comput. Supported Cooperat. Work*, vol. 23, no. 3, pp. 299–337, 2014.
- [6] P.-Y. Chiang, C.-C. Chen, and C.-H. Hsia, "A touchless interaction interface for observing medical imaging," *J. Vis. Commun. Image Represent.*, vol. 58, pp. 363–373, Jan. 2019.
- [7] T. M. Bücking, E. R. Hill, J. L. Robertson, E. Maneas, A. A. Plumb, and D. I. Nikitichev, "From medical imaging data to 3D printed anatomical models," *PLoS ONE*, vol. 12, no. 5, 2017, Art. no. e0178540.
- [8] T. M. Bücking, E. R. Hill, J. L. Robertson, E. Maneas, A. A. Plumb, and D. I. Nikitichev, "From medical imaging data to 3D printed anatomical models," *PLoS ONE*, vol. 12, no. 5, 2017, Art. no. e0178540.
- [9] G. B. Kim, S. Lee, H. Kim, D. H. Yang, Y.-H. Kim, Y. S. Kyung, C.-S. Kim, S. H. Choi, B. J. Kim, H. Ha, S. U. Kwon, and N. Kim, "Three-dimensional printing: Basic principles and applications in medicine and radiology," *Korean J. Radiol.*, vol. 17, no. 2, pp. 182–197, 2016.
- [10] N. Bizzotto, A. Costanzo, and L. Bizzotto, "Leap motion gesture control with OsiriX in the operating room to control imaging: First experiences during live surgery," *Surgical Innov.*, vol. 21, no. 6, pp. 655–656, 2014.
- [11] S. Bauer et al., "Real-time range imaging in health care: A survey," in *Proc. Dagstuhl Seminar Time-Flight Imag., Sensors, Algorithms, Appl. Workshop Imag. New Modalities, Held German Conf. Pattern Recognit. (GCPR)*. Springer-Verlag, 2013.
- [12] S. K. Herniczek, A. Lasso, T. Ungi, and G. Fichtinger, "Feasibility of a touch-free user interface for ultrasound snapshot-guided nephrostomy," *Proc. SPIE*, vol. 9036, Mar. 2014, Art. no. 90362F.
- [13] B. Hatscher, M. Luz, and C. Hansen, "Foot interaction concepts to support radiological interventions," *I-Com*, vol. 17, no. 1, pp. 3–13, 2018.
- [14] M. Meng, P. Fallavollita, S. Habert, S. Weidert, and N. Navab, "Device- and system-independent personal touchless user interface for operating rooms," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 6, pp. 853–861, 2016.
- [15] J. Hettig, A. Mewes, O. Riabikin, M. Skalej, B. Preim, and C. Hansen, "Exploration of 3D medical image data for interventional radiology using myoelectric gesture control," in *Proc. Eurograph. Workshop Vis. Comput. Biol. Med.*, Sep. 2015, pp. 177–185.
- [16] S. Eivazi, A. Hafez, H. Afkari, E. Kasneci, M. Lehecka, R. Bednarik, and W. Fuhl, "Optimal eye movement strategies: A comparison of neurosurgeons gaze patterns when using a surgical microscope," *Acta neurochirurgica*, vol. 159, no. 6, pp. 959–966, 2017.
- [17] R. H. Taylor, A. Mencias, P. Fiorini, P. Dario, and G. Fichtinger, "Medical robotics and computer-integrated surgery," in *Springer Handbook of Robotics*. Cham, Switzerland: Springer, 2016, pp. 1657–1684.
- [18] B. Hatscher and C. Hansen, "Hand, foot or voice: Alternative input modalities for touchless interaction in the medical domain," in *Proc. Int. Conf. Multimodal Interact.*, Oct. 2018, pp. 145–153.
- [19] A. Mewes, B. Hensen, C. Hansen, and F. Wacker, "Touchless interaction with software in interventional radiology and surgery: A systematic literature review," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 2, pp. 291–305, 2017.
- [20] J. P. Wachs, M. Kölsch, Y. Edan, and H. Stern, "Vision-based hand-gesture applications," *Commun. ACM*, vol. 54, no. 2, pp. 60–71, 2011.
- [21] J. Nielsen. Usability 101: Introduction to usability. Nielsen Norman Group. Accessed: Oct. 27, 2014. [Online]. Available: <http://www.ppl/www.nngroup.com/articles/usability-101-introduction-to-usability/>
- [22] P. Suetens, *Fundamentals of Medical Imaging*. Cambridge, U.K.: Cambridge Univ. Press, 2017.
- [23] H. M. Mentis, K. O'Hara, G. Gonzalez, A. Sellen, R. Corish, A. Criminisi, R. Trivedi, and P. Theodore, "Voice or gesture in the operating room," in *Proc. 33rd Annu. ACM Conf. Extended Abstr. Hum. Factors Comput. Syst.*, Apr. 2015, pp. 773–780.
- [24] D. Stevenson, H. Gardner, E. Beenen, S. Gananadha, J. Fergusson, and W. Neilson, "Evidence from the surgeons: Gesture control of image data displayed during surgery," *Behav. Inf. Technol.*, vol. 35, no. 12, pp. 1063–1079, 2016.
- [25] L. C. Ebert, G. Hatch, M. J. Thali, and S. Ross, "Invisible touch—Control of a DICOM viewer with finger gestures using the Kinect depth camera," *J. Forensic Radiol. Imag.*, vol. 1, no. 1, Jan. 2013, pp. 10–14.
- [26] K. O'Hara, R. Harper, H. Mentis, A. Sellen, and A. Taylor, "On the naturalness of touchless: Putting the 'interaction' back into NUI," *Trans. Computer-Human Interaction*, vol. 20, no. 1, pp. 1–25, Mar. 2013.
- [27] G. S. Meltzner, J. T. Heaton, and Y. Deng, "Silent speech recognition as an alternative communication device for persons with laryngectomy," *ACM Trans. Comput.-Hum. Interact.*, vol. 20, no. 1, Mar. 2013, Art. no. 5.
- [28] A. Dietz, S. Schröder, K. Frank, E. Reithmeier, and A. Pösch, "Contactless Surgery Light Control based on 3D Gesture Recognition," in *Proc. 2nd Global Conf. Artif. Intell.*, 2016, pp. 138–146.
- [29] J. H. Tan, C. Chao, A. C. Roberts, T. B. Kinney, and M. Zawaideh, "Informatics in radiology: Developing a touchless user interface for intraoperative image control during interventional radiology procedures," *Radiographics*, vol. 33, no. 2, pp. E61–E70, 2013.
- [30] Y. A. Yusoff, A. H. Basori, and F. Mohamed, "Interactive hand and arm gesture control for 2D medical image and 3D volumetric medical visualization," *Procedia—Social Behav. Sci.* vol. 97, pp. 723–729, Nov. 2013.
- [31] N. Rossol, I. Cheng, and A. Basu, "A multisensor technique for gesture recognition through intelligent skeletal pose analysis," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 3, pp. 350–359, Jun. 2016.
- [32] Y. Park, J. Kim, and K. Lee, "Effects of auditory feedback on menu selection in hand-gesture interfaces," *IEEE MultiMedia*, vol. 22, no. 1, pp. 32–40, Jan./Mar. 2015.
- [33] D. Black, M. Unger, N. Fischer, R. Kikinis, H. Hahn, T. Neumuth, and B. Glaser, "Auditory display as feedback for a novel eye-tracking system for sterile operating room interaction," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 1, pp. 37–45, Jan. 2018.



**WANG SHI HUI** received the bachelor's degree in industrial design and the master's degree in design management from Hunan University, China, in 2007 and 2010, respectively, and the Ph.D. degree in design science from the Engineering Department, Chiba University, Japan, in 2014.

From 2014 to 2017, she was a UI Designer and Usability Analyst with the UI Team, Local Application Software Development Department, Sony, Tokyo. Since 2017, she has been a Postdoctoral Research Fellow with the Mobile Health Ministry of Education—China Mobile Joint Laboratory, Xiangya Hospital Central South University, China. Her research interests include human–computer interaction, usability, voice/gesture interaction, and artificial intelligence in healthcare.



**WEIHONG HUANG** received the B.Eng. degree in automation and the M.Eng. degree in pattern recognition and smart control from Southeast University, China, in 1995 and 1998, respectively, and the Ph.D. degree in computer science from Nanjing University, China, in 2001.

From 2001 to 2002, he was a Postdoctoral Research Fellow with CNRS University Lyon 1, France. From 2002 to 2005, he was a Lecturer with the Department of Computer Science, University of Hull, U.K. From 2005 to 2014, he was a Senior Lecturer with the School of Computer and Information Systems, Kingston University London, U.K. Since 2016, he has been a Professor and the Deputy Director of the Mobile Health Ministry of Education–China Mobile Joint Laboratory, Xiangya Hospital Central South University, China. His research interests include mobile health, artificial intelligence in healthcare, cognitive computing for healthcare, semantic multimedia computing, and knowledge graph applications.

Dr. Huang is a Committee Member of the China Hospital Information Management Association, a Standing Committee Member of the Medical and Health Big Data Evaluation and Assurance Board of the Chinese Health Information and Big Data Association, and the Chairman of the Specialized Committee of Information Management of the Hunan Health Management Association.



**JIANZHONG HU** received the M.D. degree. He is currently a Professor of spine surgery, a Level 1 Chief Physician, a Doctoral Supervisor, and the Vice-President of Xiangya Hospital Central South University, China, the Director of the Mobile Health Ministry of Education–China Mobile Joint Laboratory, and the Director of the Digital Medical Research Center, Central South University. His main research interests include acute spinal column injury, osteoarthritis, digital health technologies, smart healthcare, and hospital management.

Prof. Hu is a Committee Member of the Spine Surgery Group of the Orthopedic Branch of the Chinese Medical Association, the Chairman of the Specialized Committee of Orthopedics of the Hunan Medical Association, a Committee Member of the Chinese Health Information and Big Data Association, and a Standing Committee Member of the China Hospital Information Management Association. He serves as the Editor-in-Chief for the *Journal of Xiangya Medicine*, an open access international journal covering medicine and surgery.



**KUN TAO** is the Chief Technical Officer with AIHealthX Company Ltd., and a Senior Artificial Intelligence Expert. His main research interests include multimodal human–computer interaction, computer vision, natural language processing, and AI-aided diagnosis.

Dr. Tao has participated in many national scientific research projects as a technical consultant for video and image content analysis, complex acoustic scene recognition, a combination of medical care and pensions, and others.



**YONGHONG PENG** is currently a Professor of data science and the Leader for Data Science Research with the University of Sunderland, U.K. His research interests include data science, machine learning, data mining, and artificial intelligence. He is the Chair of the Big Data Task Force (BDTF) and a member of the Data Mining and Big Data Analytics Technical Committee of the IEEE Computational Intelligence Society (CIS). He is also a founding member of the Technical Committee on Big Data (TCBD) for the *IEEE Communications Magazine* and an Advisory Board Member of the IEEE Special Interest Group (SIG) on Big Data for Cyber Security and Privacy. He is an Associate Editor of the *IEEE TRANSACTIONS ON BIG DATA* and an Academic Editor of *PeerJ Computer Science*.

• • •