AHRQ Grant Final Progress Report

Title of Project: Context-Based Hand-Gesture Recognition for the Operating Room

Principal Investigator:

- Juan P. Wachs, Assistant Professor, School of Industrial Engineering, Purdue University

Co-Investigators:

- Rebecca A. Packer, Associate Professor, Purdue University College of Veterinary Medicine

Team Members:

- Mithun G. Jacob, PhD Candidate, School of Industrial Engineering, Purdue University

Organization: Purdue University

Inclusive Dates of Project: 09/30/2010 – 09/30/2013

Federal Project Officer: Denise Burgess

Acknowledgement of Agency Support: This project was supported by grant number R03HS019837 from the Agency for Healthcare Research and Quality (AHRQ). The content is solely the responsibility of the authors and does not necessarily represent the official views of the AHRQ.

Grant Award Number: R03HS019837

Contents

1 Structured Abstract

Purpose

The goal of this study is to investigate gaps in sterility and efficiency of traditional human-computer interaction (HCI) devices in the operation room (OR) by designing hand gesture technologies that utilize contextual cues in the OR for robust usage and validating them with a simulated surgical procedure.

Scope

HCI in the OR is ubiquitous due to the need of computers for surgical procedures. Currently, traditional HCI are used to interact with these machines, which can compromise sterility and spread infection. Additionally, conveyance of instructions to an assistant to operate the machine (or any imaging device) is known to be cumbersome and inefficient.

Methods

The first of two studies involved the collection of data of users utilizing a hand gesture system to navigate MRI images while performing an MRI image navigation task. The collected data was used to build a robust hand gesture recognition system that utilized surgical contextual cues to determine user intent. The efficiency of utilizing the contextual cues was also examined. The second study investigated usability, where the users performed a simulated neurobiopsy procedure with a prototype biopsy needle and the improved sterile, touchless hand gesture interface.

Results

The first study indicated that the intent recognition system was accurate and efficient to use. Additionally, it was shown that the gesture lexicon was easy to learn and remember. The second study showed that utilizing surgical context resulted in a faster, more efficient system, which was more natural than using an assistant to control the computer.

Keywords: User-computer interface, infection control, computer vision, operating room

2 Purpose

2.1 Objectives of Study

The goal of this study was to investigate methods to increase efficiency and reduce risk of infections in the OR due to the use of traditional HCI for the navigation of MRI images. The use of hand gesture technologies in the operating room were investigated, because gestures have been shown to be a natural and efficient way to manipulate images (1,2). Additionally, gesture-based control does not compromise sterility. Therefore, the use of this modality was investigated as an alternative to the keyboard and mouse, which has traditionally been the HCI of choice in the OR.

Preliminary research conducted by the PI (3) in the field of a touchless gesture-based MRI navigation system was to be used as a starting point. The system would then be improved by incorporating surgical contextual cues and a new set of dynamic two-handed gestures. Several experiments with human-computer interactions in a mock surgery were to be conducted to validate the new approach.

The specific aims of the study were as follows:

1. Improve the robustness of hand gesture recognition algorithms by incorporating contextual cues, such as the physical environment, the type of task, and user characteristics, into a gesture-based browsing and manipulation system for healthcare environments. The working hypothesis was that contextual information integrated with visual hand gesture information can significantly improve overall system recognition performance.
2. Validate the hand gesture recognition in the operating room using a simulated surgical procedure. Our working hypothesis was that integrating a hand gesture interface system for search and manipulation of medical images in the OR will result in high usability and effectiveness for the surgeon.

3 Scope

3.1 Background

Due to advances in computer-assisted surgery, human-computer interaction (HCI) in the operating room (OR) is gradually becoming commonplace. Several surgical procedures, such as tumor resections, mandate the use of computers (4) intraoperatively and during preoperative planning. Because HCI devices are possible sources of contamination due to the difficulty in sterilization, clinical protocols have been devised to delegate control of the terminal to a sterile human assistant (5–7). However, this mode of communication has been shown to be cumbersome (8) and prone to errors (5) and therefore increases the overall duration of the procedure. As a secondary effect, such indirect interaction could increase the surgeon's cognitive load (9–11) and highlights the need for a sterile method of HCI in the operating room. Computer systems used to navigate MRI images before and during the surgery (PACs) (12–14) conventionally requires the use of keyboard, mice, or touchscreens for MRI browsing. This project proposed a sterile method for the surgeon to naturally and efficiently manipulate MRI images through touchless, freehand gestures (2,15–19).

Image manipulation through gestural devices has been shown to be natural and intuitive (1,2) and does not compromise the sterility of the surgeon. An example of a touchless mouse (20) utilizes stereo vision to localize the hand in 3D, which allows the user to control the interface with hand gestures. A multimodal solution (21) for obtaining patient input using gestures has also been shown to be effective. Systems based on voice recognition have also been utilized in the OR, including AESOP, a voice-controlled robotic arm that handle a camera during surgery (22). The main drawback with voice recognition systems are the long reaction times, erratic responses, and user dependency (23). The uncontrolled and noisy environment characteristic to the OR has led to the development of gesture-based (24,25) interfaces for the operating room.

The need for sterile image manipulation motivated the development of touchless HCI based on the use of facial expressions (23), hand and body gestures (20,21,26,27), and gaze (23,28). It should be noted that none of this research has incorporated surgical contextual cues to disambiguate recognition of false gestures and improve gesture recognition performance. Thus, there exist gaps in sterility and efficiency that can be filled by an alternate modality to conventional HCI devices, such as the keyboard, mouse, and touchscreens traditionally used to navigate and manipulate a sequential set of MRI images.

3.2 Context

Based on the success of the initial prototype of Gestix developed by the PI (3), combined with the expertise in healthcare environments available through the Regenstrief Center for Healthcare Engineering and the Purdue University Veterinary School, our team was in a position to investigate the challenges involved with bringing this technology to the high-risk enterprise of the OR.

Experiments were conducted to determine the robustness and efficiency of a novel gesture recognition system that utilized several surgical contextual cues to improve recognition performance. The performance of users while using the system to perform simulated surgical procedures was studied to compare the use of traditional HCI with the novel system.

3.3 Setting and Participants

The setting of the research project was the hospitals' surgical suites.

The subjects in the study were affiliated with Purdue University and consisted of undergraduate and graduate students, visiting faculty, and staff. The only criteria required of the subject pool was the ability to perform gestures, unless impeded by any physical or cognitive disability.

A limitation of the study was that the sample used is possibly not representative of the target audience (the surgical community).

3.4 Incidence and Prevalence

Currently, imaging devices deployed in the OR are accessible through traditional interfaces, which can compromise sterility and spread infection (5,6).

For example, Schultz et al., 2004 (6) reported that 95% of the cultures from computer keyboards were positive for microorganisms; therefore, the hospital environment plays a critical role in the transmission of organisms associated with nosocomial infections. This has been demonstrated for several important pathogens, including *Clostridium difficile*, *Staphylococcus aureus*, and vancomycin-resistant *Enterococcus*.

Additional research has shown a stronger relation between keyboard use and nosocomial infections (37–39). Transmission of pathogens from contaminated keyboards to gloves and then to vials and syringes can initiate infection (35). Therefore we can conclude that there is evidence that supports keyboards serving as reservoirs of nosocomial pathogens and vectors for cross transmission in the ICU setting (36).

A touchless (40) interface would allow the surgeon to directly interact with images without compromising sterility (1,5,6). A clear need (17) for a sterile solution for surgeons to browse and manipulate medical images exists and must be addressed. Gestures are a natural and efficient way to manipulate images (3) and allow the surgeon to interact directly with the images without compromising the sterile environment (3, 20, 31–34).

The deployment of this technology has the capacity to increase efficiency during surgeries and also reduce infections. Prevention methods for healthcare-acquired infections have the capacity to save billions of dollars annually, according to the Centers for Disease Control (41).

4 Methods

Several quantitative and qualitative methods were utilized for the design and validation of the proposed system. Two studies were designed to investigate the working hypotheses defined in the specific aims. The data acquired from user interactions in the first study was used to improve the system and validate the accuracy and efficiency of the user intent recognition system. The data from the second study was used to validate the complete system, which utilized contextual cues from a simulated neurobiopsy procedure. The second study also compared the gesture-based system to the usual practice of asking an assistant to control the computer used for MRI navigation.

4.1 Study Design

4.1.1 Study Design 1

Two studies were conducted. The first study was designed to investigate how the incorporation of contextual cues would improve gesture recognition performance. This was completed by conducting the following experiments:

A. *Lexicon Design*

An ethnographic study was conducted with 10 surgeons from Purdue University's School of Veterinary Medicine to collect a set of gestures natural for the primary user of the system (clinicians and surgeons). First, surgeons were asked to specify functions they perform on MRI images in typical surgeries that would be useful in the OR. When asked about gestures that could be effective if the interface were only controlled via hand or body gestures, each surgeon provided a set of gestures corresponding to each aforementioned function. Each surgeon clearly showed the gesture assigned to the function (requiring one or both hands), which was recorded. The gestures were then assembled into lexicons and compared to find agreements.

B. *Intention Detection and Gesture Recognition*

Datasets were collected of users performing gestures (behavior showing intent to use the gesture-based system, henceforth referred to as *intentional behavior*). Additionally, observations of users not interacting with the system (i.e., talking, writing, etc.) were collected (henceforth referred to as *unintentional behavior*). The dataset consisted of color videos and 3D maps of the users (which included the 3D coordinates of key points on the user, such as the head, shoulders, and hands) participating in either behavior. Experiments were conducted to train a classifier to utilize anthropometric cues, such as torso orientation, hand position, and gaze (computed using aforementioned key 3D points on the user) to discriminate between intentional and unintentional behavior. Additionally, the 3D trajectories of the users' hands while performing the gestures were recorded and used to train a gesture recognition system. Experiments were conducted on the dataset to examine the efficiency of the gesture classifier as well.

C. *Usability Study*

A feasibility study was conducted with the aforementioned context-based gesture recognition system. Subjects performed a specific browsing and manipulation task using the MRI image browser in a laboratory environment. The task consisted of searching for a landmark image and performing image manipulation tasks on the image. At the end of each trial, each user was asked to assemble a surgical box. This activity served as a controlled way to force the user to shift the focus of attention from the image browser. Without contextual information, such activity could potentially trigger accidental gestures. Color videos and 3D maps of the users were collected while the task was performed. The collected data was manually annotated (e.g., from time 0:12 to 0:20, the user is performing gesture "up"; from time 0:20 to 0:50, the user is assembling the box). The annotated data was used to analyze the effect of incorporating context on the true- and false-positive rate of the system. Each subject was asked to complete a post-study questionnaire and rate the interface on a Likert-type scale.

4.1.2 Study Design 2

The second study design utilized the data collected from Study Design 1 to design a novel gesture recognition system. The intention classifier was also improved using the in-task observations of users exhibiting intentional and unintentional behavior. Another usability study was conducted to test the final version of the system. Several experiments were conducted to validate the system.

A. *Intention Detection and Gesture Recognition*

A novel gesture recognition system was designed. In addition to the experiments conducted in Study Design 1, further experiments were conducted to test the efficiency of the temporal segmentation component of the system (recognizing when a gesture begins and ends from a continuous stream of hand movements from the user).

B. *Usability Study*

A usability study was conducted to compare the performance of users who used the context-based gesture system and those who worked with an assistant using a keyboard and mouse interface to navigate and manipulate images based on verbal commands from the user. The two interaction paradigms are henceforth referred to as *Context* and *Assistant*, respectively. The interaction sequences and time taken to complete the task of finding and manipulating the landmark image of interest were recorded. Each participant worked on three biopsy sites on the mock model and performed the aforementioned task a total of 12 times for each interaction paradigm. After the mock operations, the participant completed a questionnaire on a Likert scale (maximum of 5) regarding their experience, which recorded their opinion of ease of use, naturalness, and precision when using paradigm *Context* compared to paradigm *Assistant*. The task completion times using both paradigms were analyzed and compared.

4.2 Data Sources/Collection

In Study Design 1, the data for analyzing intention detection and gesture recognition was collected from 2 and 10 subjects from Purdue University, respectively. The data for the usability study was collected from 20 subjects. The male-female ratio was 12:8; males were aged between 20 and 33 years (mean±SD, 25.67±4.48 years), and females were aged between 18 and 27 years (mean±SD, 22.50±3.42 years). In Study Design 2, the data for analyzing intention detection was obtained from the observations of the same subjects. The usability study was conducted with 19 subjects. The male-female ratio was 13:6; male subjects were aged between 19 and 29 years (mean±SD, 23.46±3.02 years) and female subjects were aged between 18 and 28 years (mean±SD, 21.66±4.32 years). All subjects are affiliated with Purdue University (undergraduate and graduate students, staff, and visiting faculty). The subjects who participated in the usability studies were compensated with pizza and soda. Questionnaires were filled out by each subject who participated in both usability studies.

4.3 Measures

The measures relevant to this study were:

1. True-positive rate (or hit rate) and false-positive rate (or false alarm rate). These objective measures are used extensively to study the efficiency of intent and gesture classification.
2. Receiver-operator characteristics (ROC) curve. A graph of the hit rate and false alarm rate for different operating parameters.
3. Confusion matrix. An element *(i,j)* at row *i* and column *j* on this matrix displays the mean hit rate of recognizing gesture i as gesture j. For example, element *(1,1)* is the hit rate of recognizing gesture 1 as gesture 1 and element *(1,2)* is the rate at which gesture 1 has been falsely recognized as gesture 2.
4. Detection and reliability rate of gestures. The detection rate corresponds to the efficiency of detecting a gesture from a continuous stream of hand movements, and the reliability rate corresponds to the accuracy of recognizing the detected gesture.
5. Learning rate. The rate at which proficiency is acquired with experience of the system. The learning rates in this study are conducted using the task completion times over successive trials.
6. Questionnaires on a Likert scale. These subjective measures were used to gauge the subjects' measure of characteristics, such as ease of use, naturalness, and precision of the system.

4.4 Limitations

The study possesses the following limitations. Only the study involving the hand gesture lexicon design involved actual surgeons (from the Purdue Veterinary School). For the remaining studies, the subjects who participated were not trained medical professionals. In a study of a novel interaction paradigm for the operating room, data collected from such professionals is invaluable. Similarly the study is also limited by the setting, because all experiments were conducted in a laboratory on simulators as opposed to a real surgery in an operating room. Future work involves addressing these issues by testing the use of the system during a surgical procedure.

5 Results

5.1 Principal Findings

5.1.1 Lexicon design

The lexicon (shown in Table I) includes gestures chosen by the surgeons. Within the lexicon, gestures (c & d) were most popular, with an agreement of six surgeons; the least popular were gestures (g & h), with only one surgeon choosing them. The size of the lexicon and the specific commands depend on the type of procedure; thus, each surgeon offered different choices. Overall, the surgeons suggested 21 commands in total (one gesture for each command), but most of the commands were agreed upon by one surgeon only. Only 10 commands were selected by at least two surgeons. This was the main criteria to set the lexicon size to 10 commands/gestures. Among the pool of selected gestures by the surgeons, 42% (14 of 33) used both hands simultaneously.

The 10 gestures selected for the MRI image browser are displayed in Table I; clockwise and counterclockwise rotate the image; browse-left and browse-right browse between images in a sequence; zoom-in and zoom-out toggle between magnified and normal view, respectively; browse-up and browse-down switch between sequences; increase-brightness and decrease-brightness alter brightness.

5.1.2 Study Design 1

In the first study, a dataset of 4750 observations (44% intentional and the rest unintentional) was collected from two subjects. The mean intent recognition hit rate and false positive rate were 97.9% and 1.36%, respectively, through two-fold cross-validation from the generated ROC curve (see Figure 1). Another dataset of 1000 gestures from 10 users (the users performed each gesture 10 times sequentially) was collected. The mean gesture recognition hit rate was found to be 97.23% through 10-fold cross-validation (see Figure 2). A significant reduction of 27.91s ($p<0.05$) of the mean task completion time between the first and last two trials was observed (see Table II). The usability study scores (on a Likert-type scale) indicated that the gesture lexicon was easy to learn (4.40 ± 0.68) and remember (4.05 ± 0.94).

In the usability study, in-task recognition performance was studied. A total of 4445 gestures were manually annotated from videos of the subjects interacting with the MRI browser. Figure 3 displays the isolated gesture recognition accuracy of the 4445 annotated gestures. Intent was correctly determined 98.7% of the time, and mean gesture recognition accuracy (ACC) of 92.58% and 93.6% was obtained for the system with and without context, respectively. ACC is the average of the recognition accuracies obtained (one for each gesture).

During the “nonintentional” phase of each trial, segmented gestures (false positives) were recognized. In the system without contextual cues, a false-positive rate (FPR) of 20.76% was obtained, whereas the false-positive rate was reduced to 2.33% when contextual cues were integrated. One-way ANOVA indicated that the mean task completion time (see Table II) of the first two trials (75.05s) was significantly ($p<0.05$) longer than the mean task completion time of the last two trials (47.14s).

5.1.3 Study Design 2

In the second study, a dataset of 900 gestures was collected from nine subjects. The mean detection and reliability rate of continuous gesture recognition was 92.26% and 89.97%, respectively (see Table III). Intent was recognized with mean hit rate of 99.55% and a false-positive rate of 1.3% from a dataset of 20 subjects obtained from the previous study (see Figure 4).

A usability study was also conducted, in which the task completion times were recorded. The learning rates for each subject were computed using the task completion times, and a representative learning curve was selected by choosing the subject with the median learning rate of 74.62%. The corresponding learning rate for the representative subject for the *Assistant* paradigm was 94.02%. The learning curves were fit using outlier rejection based on Cook’s distance, and it was observed that at least 10 of the 12 data points obtained from each trial were not rejected as outliers.

The usability study revealed that the system that utilized contextual cues from the simulated neurobiopsy procedure was easy to learn (learning rate of 74.62% from Figure 5), because usage was significantly faster than using an assistant ($p<0.05$ in Table IV) to control the computer after nine trials (see Figure 6).

It was revealed that the improvement in performance was due in part to the significant (p<0.05 in Table V) reduction in the number of commands when using the gesture system (mean±SD, 4.56±2.83) as opposed to verbally guiding an assistant (19.89±2.86) to navigate and manipulate MRI images. The usability study scores (on a Likert-type scale) indicated that the system was significantly more natural (p<0.05 in Table VI) than using an assistant (see Figure 7).

5.2 Discussion

In the first study, a gestural interface was designed for browsing MRI images in the OR. The results of this study evaluated the relative importance of hand gestures alone versus hand gestures combined with environmental context. The main finding was that it is possible to accurately recognize the user's (or a surgeon's, in an OR) intention to perform a gesture by observing environmental cues (context) with high accuracy.

Additionally, the hypothesis that contextual information can improve gesture recognition was validated by the decrease in the gesture recognition false-positive rate from 20.76% to 2.33%. The significant reduction (p<0.05) of 27.91s in the mean task completion time indicates that the user operates the interface more efficiently with experience.

Compared with prior work (20), which gauged intent by checking if gestures were performed in a pre-defined 3D workspace, our work uses environmental cues to determine intent, allowing the user to perform gestures anywhere in the field of view of the sensor.

Other relevant prior work (2) required the use of voice commands to switch between modes, allowing the tracked movement of the hands to manipulate images (similar to some studies (15,29,30), where the position of the hands was used like a mouse pointer). Alternatively, our system recognizes the movement of the hands to manipulate images. It was also observed that voice recognition was a problem due to the accents of the participants; this issue was cited as the main challenge in using the system.

The observations from the first study were used to improve the intention recognition system such that the mean hit rate is 99.5%, with a false-positive rate of 1.3%. Experimental results showed that the mean continuous gesture recognition detection ratio and reliability are 92.26% and 89.97%, respectively.

In the second study, a usability study was conducted to compare the *Context* and *Assistant* paradigms. The analysis of task completion times revealed a learning curve in the *Context* paradigm. The low coefficient of determination (R^2 value) and lower learning rate for the learning curve of the *Assistant* paradigm indicate that much learning does not occur. This is anticipated due to the assistant's proficiency with the keyboard and mouse.

It was also observed that the learning curves intersected close between the 5th and 6th trial, which indicates that a user can be expected to improve in the usage of the *Context* paradigm and be superior to the *Assistant* paradigm after the 6th trial. ANOVA was conducted on the task completion times, and it was observed that a significant (p<0.05) improvement in the task completion time was observed after the 9th trial.

5.3 Conclusions

It has been shown that contextual cues have been successfully integrated to improve task completion performance when navigating MRI images using a gesture-based interface. The system's ability to recognize user intent occurs at 99.55%, with a false-positive rate of 1.3%. It was shown that users learned to use the context-integrated interaction paradigm faster than the conventional assistant-based interaction paradigm. This is evinced by the significant reduction in task completion time over the last 25% of the trials when utilizing the context-based system. Also, the context-based system was shown to be significantly more natural than the assistant-based system upon analysis of the post-study questionnaire.

5.4 Significance

An obstacle in the adoption of touchless hand gesture technologies in the OR as a replacement for traditional interfaces, such as the keyboard and mouse, is the notion that gesture recognition is not accurate enough to replace these interfaces. In this study, it has been shown that this is not the case. Additionally, it has been shown that incorporation of contextual cues from a surgical task can significantly improve overall system performance compared with interfaces relying on gesture recognition alone. The results from this study indicate that, through the incorporation of contextual cues, it is possible for bodial interfaces, such as a gesture-based interface, to function more efficiently than traditional interfaces.

5.5 Implications

This study has utilized several contextual cues to improve the performance of an MRI navigation system. It has incorporated visual context, task knowledge, and anthropometric data to decrease ambiguity in communication. Therefore, this study has shown that incorporating surgical context can indeed improve the performance of such a system. Additionally, it has shown that bodial interfaces, which are known to be more natural and easy to use, can also be accurate, thus allowing for efficient usage in the operating room. It is believed that these results will lead to discovery of more discriminative contextual cues for various surgical procedures, which can be incorporated to improve overall system performance. Figure 8 shows the system developed while in use in order to browse MRI images with contextual information from the surgical needle.

6 List of Publications and Products

The current award resulted in two journal publications, one conference proceeding publication, two presentations (one awarded the best poster award), and national news coverage.

6.1 Journal Articles

1. Jacob MG, Wachs JP, Packer RA. Hand-gesture-based Sterile Interface for the Operating Room Using Contextual Cues for the Navigation of Radiological Images. Journal of American Medical Informatics Association. 2012 Dec 18;20(1):183–186.

2. Jacob MG, Wachs JP. Context-based Hand Gesture Recognition for the Operating Room. Pattern Recognition Letters, Accepted, Online First June 2013.

6.2 Conference Proceedings

3. Jacob M, Cange C, Packer R, Wachs JP. Intention, Context and Gesture Recognition for Sterile MRI Navigation in the Operating Room. In: Alvarez L, Mejail M, Gomez L, Jacobo J, editors. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications Lecture Notes in Computer Science Volume 7441, 2012, pp 220-227. Springer Berlin Heidelberg; 2012

6.3 Presentations

4. Jacob MG, Wachs JP. Context-based Hand Gesture Recognition for the Operating Room. Poster at the 2013 Industrial Engineering Graduate Student Organization Symposium Competition, Purdue University (***Best poster award***).

5. Jacob MG, Cange C, Wachs J, Packer, R. Intention, Context and Gesture Recognition for Sterile MRI Navigation in the Operating Room" in Proceedings of the 16th Iberoamerican Congress Conference on Progress in Pattern Recognition (CIARP), Image Analysis, Computer Vision, and Applications, Buenos Aires, Argentina, pp. 220-227.

6.4 Press

6. Surgeons may use hand gestures to manipulate MRI images in OR. *Purdue News*. January 10, 2013. http://www.purdue.edu/newsroom/releases/2013/Q1/surgeons-may-use-hand-gestures-to-manipulate-mri-images-in-or.html

7. Research helps surgeons use Xbox for better images. *The Exponent*. January 30, 2013. http://www.purdueexponent.org/campus/article_af1f5de5-c095-59a3-b9f1-0b10a5331e06.html

8. Surgeons May Use Hand Gestures to Manipulate MRI Images in OR. *Science Daily*. http://www.sciencedaily.com/releases/2013/01/130110111729.htm

9. Gesture-recognition system could reduce surgical delays. ***The Engineer***. 11 January 2013. http://www.theengineer.co.uk/medical-and-healthcare/news/gesture-recognition-system-could-reduce-surgical-delays/1015164.article

References

1. Hauptmann AG. Speech and Gestures for Graphic Image Manipulation. Acm Sigchi Bull. 1989. p. 241–5.

2. Ebert LC, Hatch G, Ampanozi G, Thali MJ, Ross S. You Can't Touch This: Touch-free Navigation Through Radiological Images. Surg. Innov. [Internet]. 2011 Nov 6 [cited 2012 Mar 29]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/22064490.

3. 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 May 1;15(3):321–3.

4. World Health Organization, others. WHO Guidelines for Safe Surgery 2009: Safe Surgery Saves Lives. World Heal. Organ. 2009.

5. Albu A. Vision-based User Interfaces for Health Applications: A Survey. Adv. Vis. Comput. 2006;771–82.

6. Schultz M, Gill J, Zubairi S, Huber R, Gordin F. Bacterial Contamination of Computer Keyboards in a Teaching Hospital. Infect. Control Hosp. Epidemiol. 2003;24(4):302–3.

7. Oliveira JF, Capote H, Pagador JB, Moyano JL, Margallo F. Perspectives on Computer Assisted Surgery (CAS) in Minimally Invasive Surgery and the Role of the CAS System Nurse. Int. J. Bioinforma. Biosci. 2012 Dec;2(4).

8. Maintz J, Viergever MA. A Survey of Medical Image Registration. Med. Image Anal. 1998;2(1):1–36.

9. Firth-Cozens J. Why Communication Fails in the Operating Room. Qual. Saf. Health Care. 2004;13(5):327–327.

10. Halverson AL, Casey JT, Andersson J, Anderson K, Park C, Rademaker AW, et al. Communication Failure in the Operating Room. Surgery. 2011;149(3):305–10.

11. Lingard L, Espin S, Whyte S, Regehr G, Baker GR, Reznick R, et al. Communication Failures in the Operating Room: An Observational Classification of Recurrent Types and Effects. Qual. Saf. Health Care. 2004;13(5):330–4.

12. Boochever SS. HIS/RIS/PACS Integration: Getting to the Gold Standard. Radiol. Manage. 2004 Jun;26(3):16–24; quiz 25–27.

13. Lemke HU, Berliner L. PACS for Surgery and Interventional Radiology: Features of a Therapy Imaging and Model Management System (TIMMS). Eur. J. Radiol. 2011 May;78(2):239–42.

14. Mulvaney J. The Case for RIS/PACS Integration. Radiol. Manage. 2002 Jun;24(3):24–9.

15. Gallo L, Placitelli AP, Ciampi M. Controller-free Exploration of Medical Image Data: Experiencing the Kinect. Comput.-Based Med. Syst. Cbms 2011 24th Int. Symp. 2011. p. 1–6.

16. Xbox Useful to Surgeons: Sunnybrook Hospital [Internet]. [cited 2013 Jan 14]. Available from: http://sunnybrook.ca/media/item.asp?i=616.

17. Johnson R, O'Hara K, Sellen A, Cousins C, Criminisi A. Exploring the Potential for Touchless Interaction in Image-guided Interventional Radiology. Proc. 2011 Annu. Conf. Hum. Factors Comput. Syst. [Internet]. New York, NY, USA: ACM; 2011 [cited 2012 Jun 11]. p. 3323–32. Available from: http://doi.acm.org/10.1145/1978942.1979436.

18. Mentis HM, O'Hara K, Sellen A, Trivedi R. Interaction Proxemics and Image Use in Neurosurgery. Proc. 2012 Acm Annu. Conf. Hum. Factors Comput. Syst. [Internet]. New York, NY, USA: ACM; 2012 [cited 2012 Jun 11]. p. 927–36. Available from: http://doi.acm.org/10.1145/2207676.2208536.

19. Micire M, Desai M, Courtemanche A, Tsui KM, Yanco HA. Analysis of Natural Gestures for Controlling Robot Teams on Multi-touch Tabletop Surfaces. Proc. Acm Int. Conf. Interact. Tabletops Surfaces [Internet]. New York, NY, USA: ACM; 2009 [cited 2013 Jan 14]. p. 41–8. Available from: http://doi.acm.org/10.1145/1731903.1731912.

20. Graetzel C, Fong T, Grange S, Baur C. A Non-contact Mouse for Surgeon-Computer Interaction. Technol. Heal. Care-Eur. Soc. Eng. Med. 2004;12(3):245–58.

21. Keskin C, Balci K, Aran O, Sankur B, Akarun L. A Multimodal 3D Healthcare Communication System. 3dtv Conf. 2007. p. 1–4.

22. Mettler L, Ibrahim M, Jonat W. One Year of Experience Working With the Aid of a Robotic Assistant (the voice-controlled optic holder AESOP) in Gynaecological Endoscopic Surgery. Hum. Reprod. 1998;13(10):2748–50.

23. 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. Robot. Autom. Ieee Trans. 2003;19(5):825–41.

24. Jacob MG, Yu-Ting Li, Wachs JP. A Gesture-driven Robotic Scrub Nurse. 2011 Ieee Int. Conf. Syst. Man Cybern. Smc. IEEE; 2011. p. 2039–44.

25. Jacob MG, Li Y-T, Wachs JP. Gestonurse: A Multimodal Robotic Scrub Nurse. Proc. Seventh Annu. Acmieee Int. Conf. Hum.-Robot Interact. [Internet]. New York, NY, USA: ACM; 2012 [cited 2012 Mar 29]. p. 153–4. Available from: http://doi.acm.org/10.1145/2157689.2157731.

26. Zeng J, Wang Y, Freedman M, Mun SK. Finger Tracking for Breast Palpation Quantification Using Color Image Features. Opt. Eng. 1997;36:3455.

27. Grange S, Fong T, Baur C. M/ORIS: A Medical/Operating Room Interaction System. Proc. 6th Int. Conf. Multimodal Interfaces. 2004. p. 159–66.

28. Yanagihara Y, Hama H. System for Selecting and Generating Images Controlled by Eye Movements Applicable to CT Image Display. Med. Imaging Technol. 2000;18(5):725–34.

29. Ruppert G, Reis L, Amorim P, de Moraes T, da Silva J. Touchless Gesture User Interface for Interactive Image Visualization in Urological Surgery. World J. Urol. 1–5.

30. Kirmizibayrak C, Radeva N, Wakid M, Philbeck J, Sibert J, Hahn J. Evaluation of Gesture-based Interfaces for Medical Volume Visualization T. Proc. 10th Int. Conf. Virtual Real. Contin. Its Appl. Ind. [Internet]. New York, NY, USA: ACM; 2011 [cited 2012 Jun 11]. p. 69–74. Available from: http://doi.acm.org/10.1145/2087756.2087764.

31. Wachs J, Kölsch M. Stern H. Edan Y. (2010). Vision-Based Hand Gesture Applications: Challenges and Innovations. Communications of the ACM.

32. Wachs J, Stern H, Edan Y, Gillam M, Feied C, Smith M, Handler J. (2008). A Real-time Hand Gesture Interface for a Medical Image Guided System. Int. J. Intell. Comput. Med. Sci. Image Proc. 1(3)1:175–85.

33. Zhao M, Nowatzyk AG, Lu T,Farkas DL. (2008). Intelligent Noncontact Surgeon-Computer Interface Using Hand Gesture Recognition. In Proc. SPIE, 6805, 68050U.

34. Garg P, Aggarwal N, Sofat S. (2009). Vision-Based Hand Gesture Recognition. World Academy of Science, Engineering and Technology, 49.

35. Fukada T, Iwakiri H, Ozaki M. Anaesthetists' Role in Computer Keyboard Contamination in an Operating Room. J Hosp Infect. 2008 Oct;70(2):148–53. Epub 2008 Aug 12.

36. Bures S, Fishbain JT, Uyehara CF, Parker JM, Berg BW. Computer Keyboards and Faucet Handles as Reservoirs of Nosocomial Pathogens in the Intensive Care Unit. Am J Infect Control. 2000 Dec;28(6):465–71.

37. Bellamy K, Laban KL, Barrett KE, Talbot DCS. Detection of Viruses and Body Fluids Which May Contain Viruses in the Domestic Environment. Epidemiol. Infect. 1998 Dec.;121(3):673–80.

38. Bebbington A, Parkin I, James PA, Chichester LJ, Kubiak EM. Patients' Case-notes: Look but Don't Touch. J Hosp Infect. 2003 Dec;55(4):299–301.

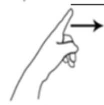
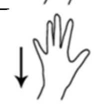
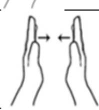
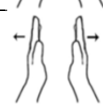
39. Hartmann B, Benson M, Junger A, Quinzio L, Röhrig R, Fengler B, Färber UW, Wille B, Hempelmann G. Computer Keyboard and Mouse as a Reservoir of Pathogens in an Intensive Care Unit. J Clin Monit Comput. 2004 Feb;18(1):7–12.

40. Detmer WM, Shiffman S, Wyatt JC, et al. A Continuous-speech Interface to a Decision Support System: II. An Evaluation Using a Wizard-of-Oz Experimental Paradigm. J Am Med Inform Assoc 1995;2:46–57.

41. Douglas SR, II. (2009). The Direct Medical Costs of Healthcare-associated Infections in U.S. Hospitals and the Benefits of Prevention. Division of Healthcare Quality Promotion National Center for Preparedness, Detection, and Control of Infectious Diseases Coordinating Center for Infectious Diseases, Centers for Disease Control and Prevention.

Appendix

Table I. Gesture Lexicon

MRI Image Viewer Command		Gesture
(a)	Rotate-clockwise	
(b)	Rotate-counterclockwise	
(c)	Browse-left	
(d)	Browse-right	
(e)	Browse-up	
(f)	Browse-down	
(g)	Increase-brightness	
(h)	Decrease-brightness	
(i)	Zoom-in	
(j)	Zoom-out	

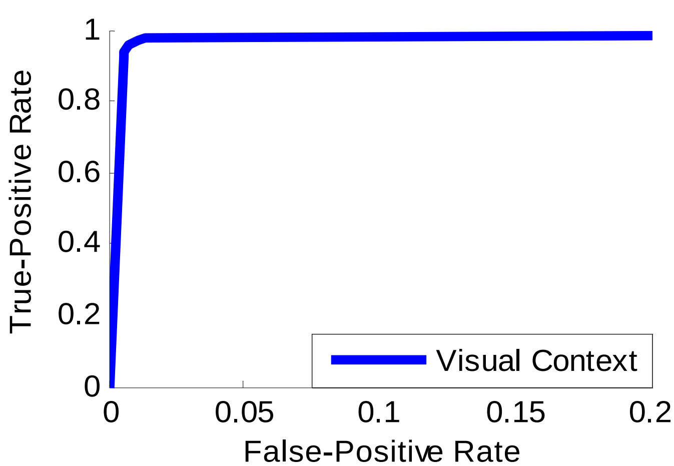


Figure 1. ROC curve for intention recognition

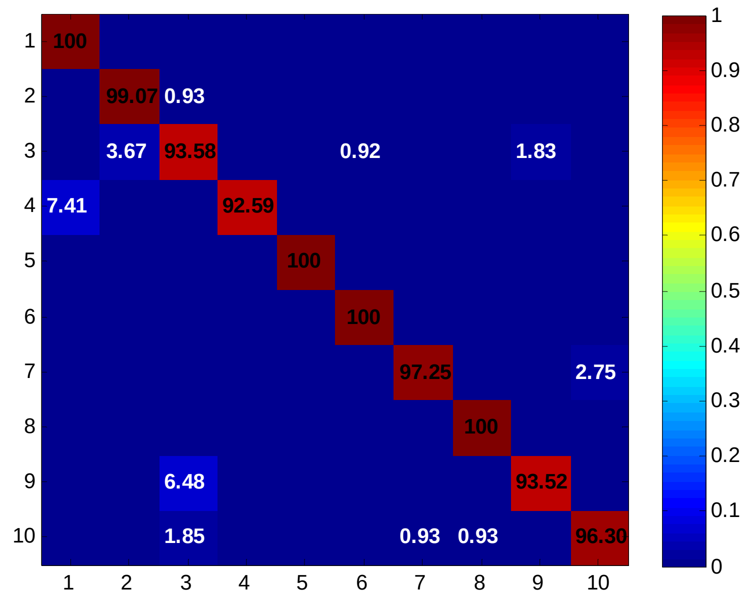


Figure 2. Confusion matrix. The rows represent the true class of the gestures labels, and the columns represent the class assigned by the algorithm. High values on the diagonal elements indicate high gesture recognition accuracy.

Table II. Mean task completion times of the first two and last two trials in the usability study of Study Design 1

	Trial #	Mean task completion time for trial (sec)	Mean task completion time (sec)
Initial	1	68.00	75.05
	2	82.11	
Final	9	49.91	47.14
	10	44.38	

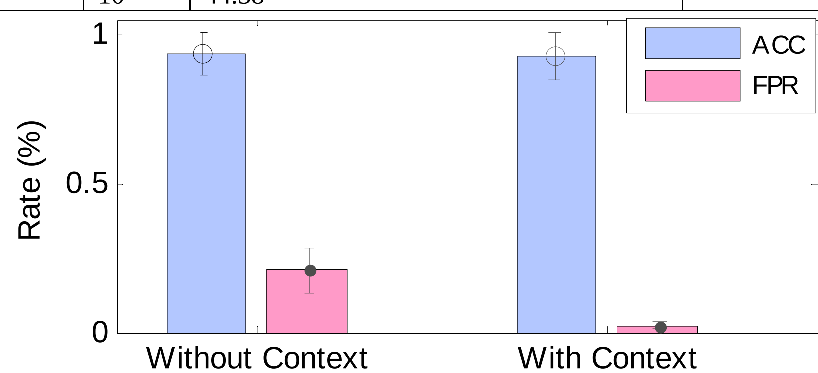


Figure 3. Comparison of gesture recognition with and without context

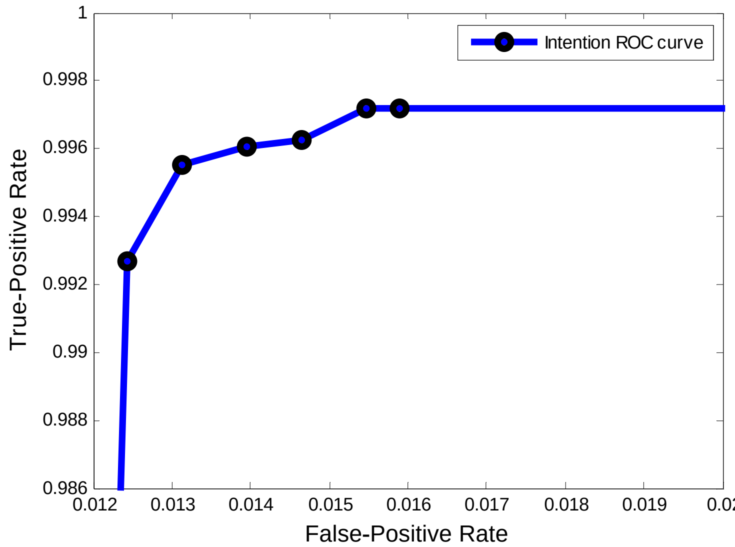


Figure 4. ROC curve for intention recognition

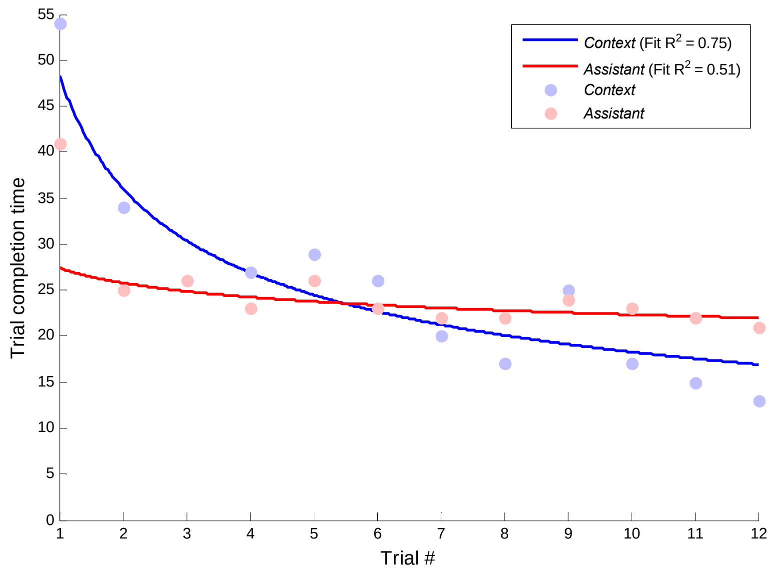


Figure 5. Learning curves for the representative participant for the *Context* and *Assistant* paradigms

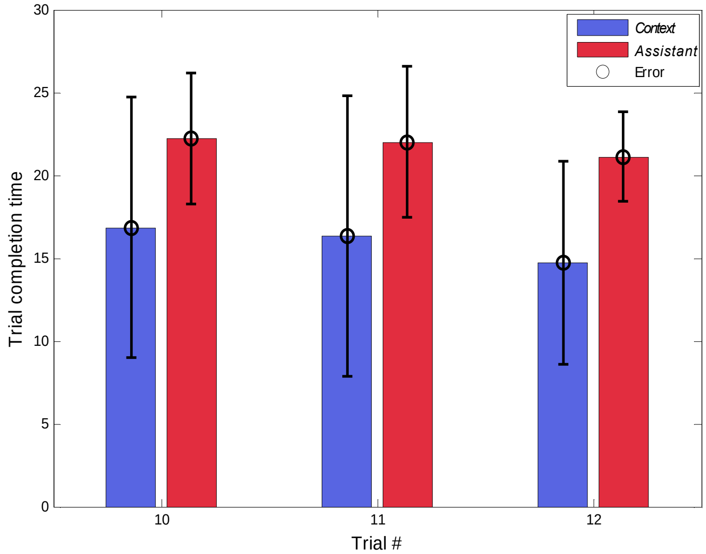


Figure 6. Mean task completion times for trials 10, 11, and 12 for the *Context* and *Assistant* paradigms

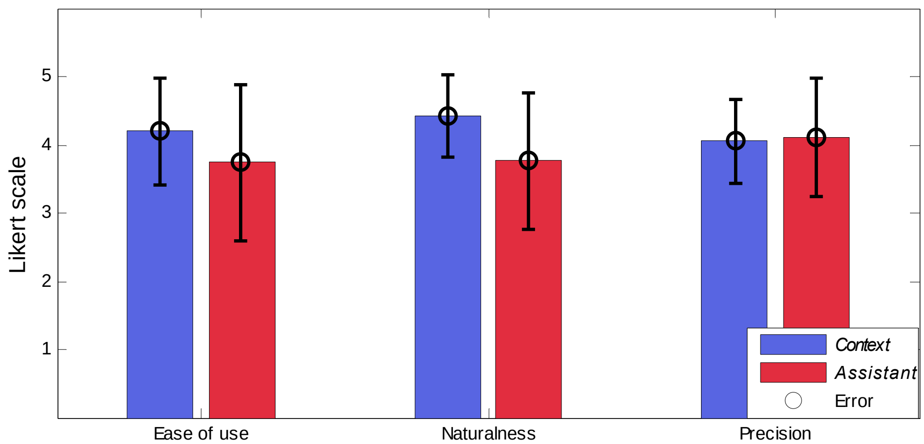


Figure 7. Mean questionnaire responses for the *Context* and *Assistant* paradigms

Table III. Detection and reliability of continuous gesture recognition

Gesture	**Detection (%)**	**Reliability (%)**
clockwise	0.9031987	0.9031987
counterclockwise	0.9480519	0.9480519
left	0.8591606	0.8306553
right	0.9713805	0.9417508
zoom-in	0.7981466	0.7397161
zoom-out	0.9225589	0.9154364
up	0.969697	0.9427609
down	0.9269619	0.8634907
brightness-down	0.969697	0.969697
brightness-up	0.9575758	0.9420357
Mean	**0.922643**	**0.899679**

Table IV. ANOVA on task completion times of both interaction paradigms for trials 10, 11, and 12

Trial #	Source	SS	df	MS	F	Prob > F
10	*Columns*	273.79	1	273.789	7.1	0.0114
	Error	1387.68	36	38.547		
	Total	1661.47	37			
11	*Columns*	306.95	1	306.947	6.64	0.0142
	Error	1664.11	36	46.225		
	Total	1971.05	37			
12	*Columns*	385.29	1	385.289	17.26	0.0002
	Error	803.47	36	22.319		
	Total	1188.76	37			

Table V. ANOVA on the number of commands required for task completion for both interaction paradigms across all 12 trials

Source	SS	df	MS	F	Prob > F
Columns	26787.3	1	26787.3	3307	0
Error	3677.5	454	8.1		
Total	30464.8	455			

Table VI. ANOVA on "Naturalness" questionnaire response for both paradigms

Source	SS	df	MS	F	Prob > F
Columns	4.1118	1	4.11184	5.97	0.0196
Error	24.8158	36	0.68933		
Total	28.9276	37			

Figure 8. Mock surgeon navigating radiological images with contextual information