

Hand-gesture-based sterile interface for the operating room using contextual cues for the navigation of radiological images

Mithun George Jacob,¹ Juan Pablo Wachs,¹ Rebecca A Packer²

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/amiajnl-2012-001212>).

¹School of Industrial Engineering, Purdue University, West Lafayette, Indiana, USA
²Departments of Basic Medical Sciences and Veterinary Clinical Sciences, College of Veterinary Medicine, Purdue University, West Lafayette, Indiana, USA

Correspondence to

Dr Juan Pablo Wachs, School of Industrial Engineering, Purdue University, 315 N Grant Street, West Lafayette, IN 47907, USA; jpwachs@purdue.edu

Received 13 July 2012

Revised 3 September 2012

Accepted 21 November 2012

Published Online First

18 December 2012

ABSTRACT

This paper presents a method to improve the navigation and manipulation of radiological images through a sterile hand gesture recognition interface based on attentional contextual cues. Computer vision algorithms were developed to extract intention and attention cues from the surgeon's behavior and combine them with sensory data from a commodity depth camera. The developed interface was tested in a usability experiment to assess the effectiveness of the new interface. An image navigation and manipulation task was performed, and the gesture recognition accuracy, false positives and task completion times were computed to evaluate system performance. Experimental results show that gesture interaction and surgeon behavior analysis can be used to accurately navigate, manipulate and access MRI images, and therefore this modality could replace the use of keyboard and mice-based interfaces.

BACKGROUND AND SIGNIFICANCE

One of the most ubiquitous pieces of equipment in US surgical units is the computer workstation, which allows access to medical images before and during surgery. While electronic medical records and images are widely used, efficiency and sterility are vital issues in the quality of their use in the operating room (OR), since computers and their peripherals are difficult to sterilize¹ and keyboards and mice have been found to be a source of contamination.¹ Currently, imaging devices deployed in the OR are accessible through traditional interfaces, which can compromise sterility and spread infection.^{1 2} In addition, when nurses or assistants operate the keyboard for the surgeon, the process of conveying information accurately has proven cumbersome and inefficient³ (since spoken dialog can be time-consuming⁴) and leads to frustration and prolonged time for performing the intervention.² This is a problem when the user (surgeon) is already performing a cognitively demanding task.

These gaps in sterility and efficiency can be addressed by adopting hand gesture technologies in the OR, which are already beginning to gain widespread acceptance in gaming.⁵ Gestures are a natural and efficient way to manipulate images⁶ and have been used in the OR to improve user performance for data entry.⁷ A touchless⁸ interface would allow the surgeon to directly interact with images without compromising sterility.^{1 6 9} A need¹⁰ for a sterile solution for surgeons to browse and manipulate medical images has led to the development of interfaces enabling non-contact gesture-controlled human-computer interfaces,

including facial expression,¹¹ hand^{3 5 9 12 13} and body gestures,^{2 5 14} and gaze.^{11 15} None of these works utilized surgical contextual information to disambiguate false recognition. Our working hypothesis (which extends our previous research³) is that contextual information such as the focus of attention, integrated with gestural information, can significantly improve overall system recognition performance compared with interfaces relying on gesture recognition alone.

MATERIALS AND METHODS











Twenty volunteers participated in the experiment conducted in February 2012. The male/female ratio was 12:8, with male subjects aged 20–33 years (mean±SD 25.67±4.48) and female subjects aged 18–27 years (mean 22.50±3.42). The study was approved by the university's ethics review committee. Written, informed consent was obtained from all participants. An MRI image browser application was developed to test the effectiveness of the gesture-based interface. The development and validation of the system was performed in three steps.

Step 1: lexicon generation—An ethnographic study was conducted with 10 surgeons from the 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 with find agreements.

The lexicon (shown in table 1) includes gestures chosen by the surgeons. Within the lexicon, gestures c and d were most popular, with an agreement of six surgeons; the least popular were gestures g and h, with only one surgeon choosing them. The size of the lexicon and the specific commands depend on the type of procedure, and 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 by one surgeon only. Only 10 commands were selected by at least two surgeons. This was the main criterion to set the lexicon size to 10 commands/gestures. Among the pool of

To cite: Jacob MG, Wachs JP, Packer RA. *J Am Med Inform Assoc* 2013;**20**:e183–e186.

Table 1 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	

gestures selected by the surgeons, 42% (14 of 33) used both hands simultaneously.

The 10 gestures selected for the MRI image browser are displayed in table 1: '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.

Step 2: development of gesture recognition software—Skeletal joints (figure 1) were tracked using a software library (OpenNI, V1.3.2.3) using the Kinect sensor, which fits a skeleton model to the user. To discriminate between intentional and non-intentional gestures,¹⁶ the continuous torso and head orientation, and angles of each arm to the torso were computed from the skeleton to train a pruned decision tree.¹⁷

If the gesture is found to be intentional, the three-dimensional hand trajectory corresponding to gestural motion is discretized into a sequence of symbols that encapsulates the velocity of each hand along each coordinate axis. The sequence of symbols is then recognized (and the corresponding command is sent to the MRI image browser) by combining the response obtained through the Hidden Markov Models (HMMs)¹⁸ and the

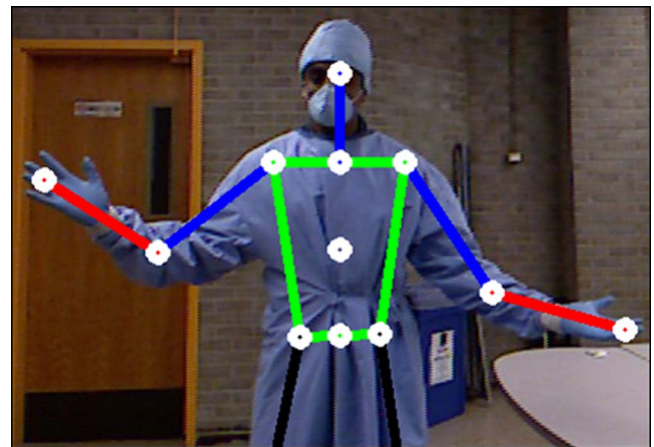


Figure 1 Skeleton model for the upper part of the body and tracked skeletal joints. This figure is only reproduced in colour in the online version.

probability of the evoked command given the previous command and time between commands.

Step 3: validation of the technology—Three experiments were conducted to validate the main hypothesis that contextual information can be used to accurately detect gestures evoked by the user.

First, a dataset of 4750 observations (44% 'intentional' and the rest 'unintentional' behavior) was collected from two users, and a decision tree classifier was trained using twofold cross-validation.

In the second experiment, gesture recognition was evaluated with a dataset of 1000 gestures performed by 10 users (the users performed each gesture 10 times sequentially). Each gesture segment in the video was manually annotated for training the HMMs through 10-fold cross-validation.

In the third experiment, users were asked to perform a specific browsing and manipulation task using the MRI image browser with hand gestures only. The task involved navigating images in different sequences while performing manipulation activities on landmarks over 10 trials. Each subject was briefly trained in the use of the gesture-based interface before the trials. Task completion time was recorded for each subject. Each subject was asked to complete a post-study questionnaire and rate the interface on a Likert-type scale.

Statistical analysis

Statistical analysis was performed with SPSS (V19.0). Continuous variables were expressed as mean±SD. Means of task completion times for the first and last two trials were compared with one-way analysis of variance. $p < 0.05$ was considered indicative of a statistically significant difference.

RESULTS

The results of the first study showed a mean recognition accuracy of 97.9% with 1.36% false positive rate (FPR) through twofold cross-validation at the peak operating point. True positives were obtained when users were facing the camera and gesturing with their hands. This supports our working hypothesis that environmental context can be accurately used to detect the surgeon's state. An analysis of the proportion of attributes used at different levels of the decision tree indicated that, initially, torso orientation was discriminative enough for intention recognition, but further down the tree, all contextual cues provided

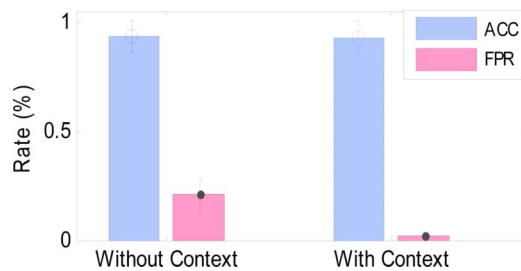


Figure 2 Comparison of gesture recognition with and without context. The mean gesture recognition accuracy (ACC) across all users and trials in the experiment testing in-task performance remains approximately the same with and without context (92.58% and 93.6%, respectively), and the false positive rate (FPR) drops from 20.76% to 2.33% upon integrating context. This figure is only reproduced in colour in the online version.

useful and necessary information (see figure 3 in the online supplementary appendix).

In the second study, several experiments were conducted on the collected dataset to determine the operating parameters, which resulted in a gesture recognition accuracy of 97.23% through 10-fold cross-validation.

In the third 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. At the end of each trial, each user was asked to assemble a surgical box. This served as a controlled way to force the user to shift the focus of attention from the image browser (by turning 90° away from the sensor). Without contextual information, such activity could potentially trigger accidental gestures. Figure 2 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). In the system with context, it was found that decrease brightness (table 1, gesture h) had the lowest mean accuracy (82.5%) and that browse right and browse down (table 1, gestures d and f) had the highest mean accuracy (100%).

During the 'non-intentional' phase of each trial, segmented gestures (false positives) were recognized. In the system without contextual cues, a FPR of 20.76% was obtained, whereas the FPR was reduced to 2.33% when contextual cues were integrated. One-way analysis of variance indicated that the mean task completion time (table 2) of the first two trials (75.05 s) was significantly ($p < 0.05$) longer than the mean task completion time of the last two trials (47.14 s).

The gestures performed by one typical user were analyzed to determine the maximum and minimum ranges of each gesture as functions of maximum grasp range (MGR) (defined as the

radius of the arcs from the shoulder pivots¹⁹). The increase brightness gesture (table 1, gesture g) required the most reach (100% MGR), and zoom in (table 1, gesture i) required the least reach (43.79% MGR).

Analysis of questionnaire data (maximum score of 5) showed that users rated the lexicon of gestures designed with surgeons as easy to learn (4.40 ± 0.68) and remember (4.05 ± 0.94) and moderately easy to perform (3.60 ± 0.88).

DISCUSSION

A gestural interface was designed for browsing MRI images in the OR. The results of this study evaluate the relative importance of hand gestures alone versus hand gestures combined with environmental context. The main finding is 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.

In comparison with prior work,⁹ which gauged intent by checking if gestures were performed in a predefined three-dimensional workspace, our work uses environmental cues to determine intent allowing the user to perform gestures anywhere in the field of view of the sensor.

In addition, our algorithm for recognizing dynamic hand gestures (which theoretically allows new gestures to be added) resulted in a mean gesture recognition accuracy of 97.23%. Compared with prior work,³ our framework can be extended to a large gesture vocabulary. These results outperform prior work²⁰ where Principle Component Analysis and Laplacian Eigenmaps were used to recognize a lexicon of four and 18 gestures with 95% and 85% accuracy, respectively.

Other relevant prior work¹³ required the use of voice commands to switch between modes allowing the *tracked* movement of the hands to manipulate images (similar to previous studies^{21–23} where the position of the hands is 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 because of the accents of the participants and was cited as the main challenge in using the system.

The hypothesis that contextual information can improve gesture recognition was validated by the decrease in the gesture recognition FPR from 20.76% to 2.33%. The significant reduction ($p < 0.05$) of 27.91 s in the mean task completion time (table 2) indicates that the user operates the interface more efficiently with experience.

For tracking, a commodity camera was used (Kinect) with a software development kit, which delivered satisfactory results for user detection. Nevertheless, the tracking algorithm occasionally failed in the presence of several people in the camera field of view. More research is required to improve the tracking algorithm.

Since people perform gestures differently and exhibit intent to use the system in different ways, it is imperative to collect high-variance training data (more users), which provides enough information to accurately train the classifiers. Future work involves adding more environmental cues such as the position of a surgical instrument within the patient's body requiring further data collection, training, and validation. Once the new contextual classifiers have been integrated into the system, a large-scale usability test must be conducted with a phantom model, similar to those used in surgical training.

Contributors The authors affirm that they have provided substantial contribution to (1) conception and design of the applied components of the study and (2)

Table 2 Mean task completion times of the first two and last two trials in experiment 3

	Trial No	Mean task completion time for trial (s)	Mean task completion time (s)
Initial	1	68.00	75.05
	2	82.11	
Final	9	49.91	47.14
	10	44.38	

drafting the applied components of the article as they pertain to the surgical needs and application, and revising it critically for important intellectual content, and (3) have provided final approval of the completed manuscript.

Funding This work was supported by the Agency for Healthcare Research and Quality (AHRQ) grant number R03HS019837.

Competing interests None.

Ethics approval Purdue University Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- Schultz M, Gill J, Zubairi S, et al. Bacterial contamination of computer keyboards in a teaching hospital. *Infect Control Hosp Epidemiol* 2003;24:302–3.
- Albu A. Vision-based user interfaces for health applications: a survey. *Adv Vis Comput* 2006;Part I:771–82.
- Wachs JP, Stern HI, Edan Y, et al. A gesture-based tool for sterile browsing of radiology images. *J Am Med Inform Assoc* 2008;15:321–3.
- Maintz J, Viergever MA. A survey of medical image registration. *Med Image Anal* 1998;2:1–36.
- Wachs JP, Kölsch M, Stern H, et al. Vision-based hand-gesture applications. *Commun ACM* 2011;54:60–71.
- Hauptmann AG. Speech and gestures for graphic image manipulation. *ACM SIGCHI Bulletin* 1989;20:241–5.
- Poon AD, Fagan LM, Shortliffe EH. The PEN-Ivory project: exploring user-interface design for the selection of items from large controlled vocabularies of medicine. *J Am Med Inform Assoc* 1996;3:168–83.
- 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.
- Gratzel C, Fong T, Grange S, et al. A non-contact mouse for surgeon-computer interaction. *Technol Health Care* 2004;12:245–58.
- Johnson R, O'Hara K, Sellen A, et al. Exploring the potential for touchless interaction in image-guided interventional radiology. Proceedings of the 2011 annual conference on Human factors in computing system. New York, NY, USA: ACM, 2011:3323–32.
- Nishikawa A, Hosoi T, Koara K, et al. FAcE MOUSE: A novel human-machine interface for controlling the position of a laparoscope. *Robot Automation, IEEE Trans on* 2003;19:825–41.
- Keskin C, Balci K, Aran O, et al. A multimodal 3D healthcare communication system. *3DTV Conference*, 2007:1–4.
- Ebert LC, Hatch G, Ampanozi G, et al. You can't touch this: touch-free navigation through radiological images. *Surg Innov* 2012;19:301–7.
- Grange S, Fong T, Baur C. M/ORIS: a medical/operating room interaction system. Proceedings of the 6th international conference on Multimodal interfaces. 2004:159–66.
- Yanagihara Y, Hama H. System for selecting and generating images controlled by eye movements applicable to CT image display. *Med Imaging Technol* 2000;18:725–34.
- Langton SRH. The mutual influence of gaze and head orientation in the analysis of social attention direction. *Q J Exp Psychol: Section A* 2000;53:825–45.
- Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
- Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 1989;77:257–86.
- Berry EC, Kohn ML. *Introduction to operating-room technique*. New York: McGraw-Hill, Blakiston Division, 1960.
- 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*. New York, NY, USA: ACM, 2012:75–84.
- Ruppert G, Reis L, Amorim P, et al. Touchless gesture user interface for interactive image visualization in urological surgery. *World J Urol* 2012;30:687–91.
- Gallo L, Placitelli AP, Ciampi M. Controller-free exploration of medical image data: Experiencing the Kinect. *Computer-Based Medical Systems (CBMS), 2011 24th International Symposium on*. 2011:1–6.
- Kirmizibayrak C, Radeva N, Wakid M, et al. Evaluation of gesture based interfaces for medical volume visualization tasks. *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry [Internet]*. New York, NY, USA: ACM, 2011: 69–74.