Gestonurse: A Multimodal Robotic Scrub Nurse

Mithun G. Jacob, Yu-Ting Li, Juan P. Wachs*
School of Industrial Engineering
Purdue University
315 N. Grant St, West Lafayette, IN 47907
{mithunjacob, yutingli, jpwachs}@purdue.edu

ABSTRACT

A novel multimodal robotic scrub nurse (RSN) system for the operating room (OR) is presented. The RSN assists the main surgeon by passing surgical instruments. Experiments were conducted to test the system with speech and gesture modalities and average instrument acquisition times were compared. Experimental results showed that 97% of the gestures were recognized correctly under changes in scale and rotation and that the multimodal system responded faster than the unimodal systems. A relationship similar in form to Fitts's law for instrument picking accuracy is also presented.

Categories and Subject Descriptors

I.2.9 [**Robotics**]: Operator Interfaces

General Terms

Algorithms, Design, Experimentation

1. INTRODUCTION

An RSN has the potential to reduce errors in the OR by automating the passing of surgical instruments which allows personnel to focus on more complicated tasks such as maintaining a sterile environment, preparing required surgical supplies and monitoring the state of the patient. This allows a possible reduction of errors due to communication failures. These failures can lead to wastage of resources, procedural errors, delays, distraction and inefficiency. A real-time RSN (see Figure 1(a)) dedicated to passing surgical instruments to the surgeon by speech and/or gesture is presented. An advantage of gesture interaction is that it is not affected by ambient noise and does not require surgeon re-training since hand signals are used by surgeons to request surgical instruments as part of standard OR procedure [1].

2. GESTURE RECOGNITION MODULE

Color and depth-based algorithms for hand segmentation are compared. The color-based algorithm was discussed in previous work [1]. The depth-based method used depth information from the Kinect sensor [2] to obtain a hand mask by thresholding the depth of objects within a bounding box in front of the camera. A fingertip detection method was used to recognize gestures [1]. Analysis shows that the gesture recognition algorithm runs in $O(n \log n)$ time in the number of points on the hand contour. Further classification was performed from hand masks obtained from both depth and color-based segmentation methods and fingertip detection and gesture recognition performances are compared in section 4.1.

3. ROBOT CONTROL SCHEME

The RSN uses a FANUC LR Mate 200iC robotic arm (see Figure 1(b)) and the CMU Sphinx [3] system for speech recognition. A state machine model (see Figure 2) controls the states of the robotic system. The system includes the following system states: ACTIVE, wait for the user commands; PASS, delivering an instrument; SLEEP on hold; MUTE speech recognition module is switched off (only gesture interaction is used). A sequence of two gestures is used to avoid accidental changes between states. The SPEECH (SP) ON/OFF gesture sequence is used to switch the system between ACTIVE and MUTE modes. Similarly the SLEEP/WAKE gesture sequence switches the system between SLEEP and the ACTIVE or MUTE states. If speech and gesture commands are performed simultaneously by the user, the first of the two events triggers the system.

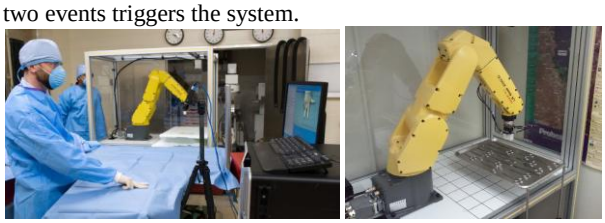

Figure 1. (a) A prototype of the real-time RSN tested at an OR (b) FANUC LR Mate 200iC and Mayo stand with instruments

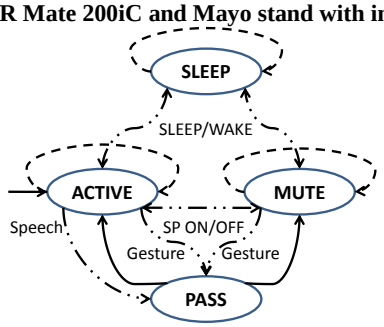

Figure 2. State Machine

4. EXPERIMENTS

4.1 Gesture Analysis Performance

A gesture analysis database consisting of 600 images of hand postures was collected from 4 users with different hand sizes and skin color (see Figure 3), and was used to assess the performance of the system. It included 30 images of postures per user per gesture exhibiting varying rotation, scale and translation of the fingertips. ROC curves for five gestures corresponding to five surgical instruments for depth and color-based hand segmentation were obtained.

4.2 Comparison of Modalities

The performance of the multimodal (speech and gesture combined) and unimodal speech/gesture systems were studied. Experiments were performed on 8 users between 20-30 years who performed 6 consecutive trials on each system. A trial is defined as an arbitrarily ordered sequence of 5 instrument types. The name of each instrument from the trial sequence was displayed to the subject who requested the instrument. The instrument acquisition time (the dependent variable) is defined as the time

*Corresponding author

HRI'12, March 5–8, 2012, Boston, Massachusetts, USA.
ACM 978-1-4503-1063-9/12/03.

Figure 3. Samples from the use-case with detected fingertips marked with red circles

elapsed between the displayed instrument name and the subject receiving the instrument. Each user performed a task with an instrument before requesting the next instrument and instrument acquisition times was recorded.

5. RESULTS & DISCUSSION

5.1 Gesture Analysis Performance

The fingertip detection algorithm achieved high average recognition rates for both color and depth-based segmentation algorithms (97% and 99% respectively). Conversely, the color-based segmentation algorithm exhibited a higher average false-positive rate of 10.8% compared to 0.5% for the depth-based algorithm (see Figure 4). This is attributed to the sensitivity of the color-based method to illumination. The average gesture recognition accuracy of the depth-based system was 97% (see Table 1) but in practice, users learned to avoid error-inducing poses such as out-of-plane rotation. Note that ϕ in Table 1 refers to instances where gestures were not recognized.

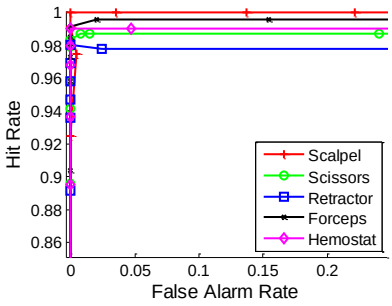

Figure 4. ROC Curve for Depth-based segmentation

5.2 Instrument Picking Performance

The robot can reliably pick an instrument from the Mayo stand and hand it to the user when the instruments (of normalized area A) are separated by at least $\lambda = 25$mm [1]. A linear relationship (see Figure 5) similar to Fitts's law [4] was observed ($\alpha = a + b\,ID$) between index of difficulty $ID = \log_2(\lambda/A + 1)$ and accuracy α. This was validated by high R^2 values of 0.89, 0.94 and 0.86 for the forceps, scissors and hemostat instrument types respectively. This result validates an intuitive tradeoff between λ and the accuracy of the picking and delivery task.

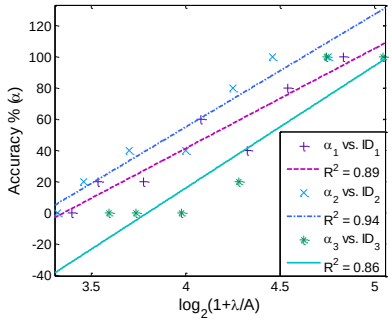

Figure 5. Accuracy vs. ID for forceps scissors, and hemostat

5.3 Comparison of Modalities

The instrument acquisition time for the six trials of each individual is presented in Figure 6. It was observed during trials that the gesture recognition system was faster (see Figure 6), and superseded the speech recognition system in triggering the state machine (shown by lower average instrument acquisition times). In events where the gesture was performed after the speech command or made incorrectly delaying recognition, the speech module compensated resulting in lower average instrument acquisition time for the multimodal system.

Table 1: Depth-based Segmentation: Confusion Matrix

	Scalpel	Scissors	Retractor	Forceps	Hemostat	∅
Scalpel	**100**	0.00	0.00	0.00	0.00	0.00
Scissors	0.84	**97.50**	0.83	0.00	0.00	0.83
Retractor	0.83	4.17	**95.00**	0.00	0.00	0.00
Forceps	0.00	0.00	1.67	**96.67**	1.66	0.00
Hemostat	0.00	0.00	0.84	3.33	**95.83**	0.00

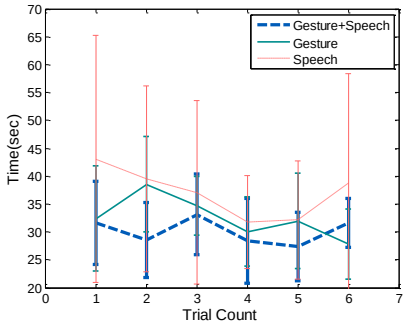

Figure 6. Mean instrument acquisition times for each trial and 95% confidence intervals for unimodal & multimodal systems

6. CONCLUSION & FUTURE WORK

A multimodal robotic scrub nurse was presented capable of reliably passing surgical instruments. Fingertips were detected and gestures recognized with 99% and 97% accuracy on average. A relationship similar to Fitts's law has been shown between picking accuracy, inter-instrument distance and instrument area. It was noted that the multimodal system performed faster than either unimodal system and that it is more suitable to uncontrolled environments, such as the OR.

Future work includes fusing the speech and gesture recognition data in a probabilistic fashion and testing the system in a live OR.

7. REFERENCES

[1] Jacob, M.G., Li, Y.-T, Wachs, J. P., "A gesture driven robotic scrub nurse," Systems, Man, and Cybernetics, IEEE International Conference on, pp. 2039-2044, 9-12 Oct. 2011

[2] Microsoft Corp. The Kinect sensor, www.xbox.com/kinect

[3] CMU Sphinx http://cmusphinx.sourceforge.net/

[4] Fitts, P., "The information capacity of the human motor system in controlling the amplitude of movement," Journal of experimental psychology, vol. 47, no. 6, p. 381, 1954.