

A Gesture Driven Robotic Scrub Nurse

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

Abstract—A gesture driven robotic scrub nurse (GRSN) for the operating room (OR) is presented. The GRSN passes surgical instruments to the surgeon during surgery which reduces the workload of a human scrub nurse. This system offers several advantages such as freeing human nurses to perform concurrent tasks, and reducing errors in the OR due to miscommunication or absence of surgical staff. Hand gestures are recognized from a video stream, converted to instructions, and sent to a robotic arm which passes the required surgical instruments to the surgeon. Experimental results show that 95% of the gestures were recognized correctly. The gesture recognition algorithm presented is robust to changes in scale and rotation of the hand gestures. The system was compared to human task performance and was found to be only 0.83 seconds slower on average.

Keywords: *surgical robotics; healthcare robotics; human robot interaction; robotic scrub nurse*

I. INTRODUCTION

Communication failures in the operating room (OR) has been shown to create delays, inefficiency and waste resources [1]. Studies have shown that 30% [2] of procedurally relevant exchanges between members in a surgical team result in a failure in communication. Such mistakes increase the risk of complications and morbidity [3] and have been shown to affect surgical performance 36.4% of the time.

A gesture driven robotic scrub nurse (GRSN) automates part of the responsibilities of a human nurse such as passing surgical instruments to the surgeon. This frees the nurse to concurrently perform more complicated tasks such as maintaining a sterile environment, preparation of surgical supplies required for the surgery [4] and has the potential to reduce errors due to long chains of verbal exchanges.

Previous surgical robots include an assistive robot for object retrieval [5], robots with haptic feedback such as SOFIE [6], and the da Vinci® surgical system [7] which is designed to simplify and create minimally invasive procedures for planned endoscopic and laparoscopic surgeries.

Previous work on robotic scrub nurses (RSN) includes “Penelope” [8] developed by Kochan *et al.* which localizes, recognizes and returns a used instrument and is voice-controlled. A similar voice-controlled robot [9] was developed by Carpintero *et al.* also uses computer vision techniques to recognize, deliver and retrieve surgical instruments. A problem with voice-only systems lies in the notable performance degradation [10] of such systems in noisy environments such as an OR. Drills, anesthesia machines, surgical staff side conversations, and other equipment result in an environment which can compromise the patient safety. In such an

environment, the surgeon may say “50,000 units” and the anesthetist may hear “15,000 units.” [11]. These errors can have dramatic consequences in the patient’s well-being.

Another voice-controlled RSN for laparoscopic surgeries developed by Yoshimitsu *et al.* [12] uses depth-based action recognition for instrument prediction. The 3D point estimation method requires the surgeon to wear markers for action recognition which can compromise sterility.

A real-time GRSN (see Figure 1) is presented which passes surgical instruments when requested with a static hand gesture. The gestures are recognized from the video stream acquired by a network camera and then a FANUC robotic arm delivers the requested instrument to the surgeon. A significant advantage of gesture-based communication is that it does not require surgeon training. Additionally, hand signs are currently used to request surgical instruments according to standard OR procedure [13–16]. Also, gestures are not affected by ambient noise in the OR.



Figure 1. The real-time RSN in operation at an OR

The rest of the paper is divided as follows. Section II provides an overview of the complete system with brief description of various processing modules. Section III describes the static hand pose gesture recognition system and Section IV discusses the operation of the robotic arm. The experiments on recognition accuracy, instrument picking accuracy and speed are described in Section V. The results are discussed in Section VI and we conclude in Section VII with a discussion on future work.

II. SYSTEM ARCHITECTURE

The system architecture of the GRSN is illustrated in Figure 2. The streaming video serves as input for the gesture analysis module which is composed of the hand segmentation, fingertip localization and gesture recognition modules. The recognized gesture is interpreted as a command which is passed to the

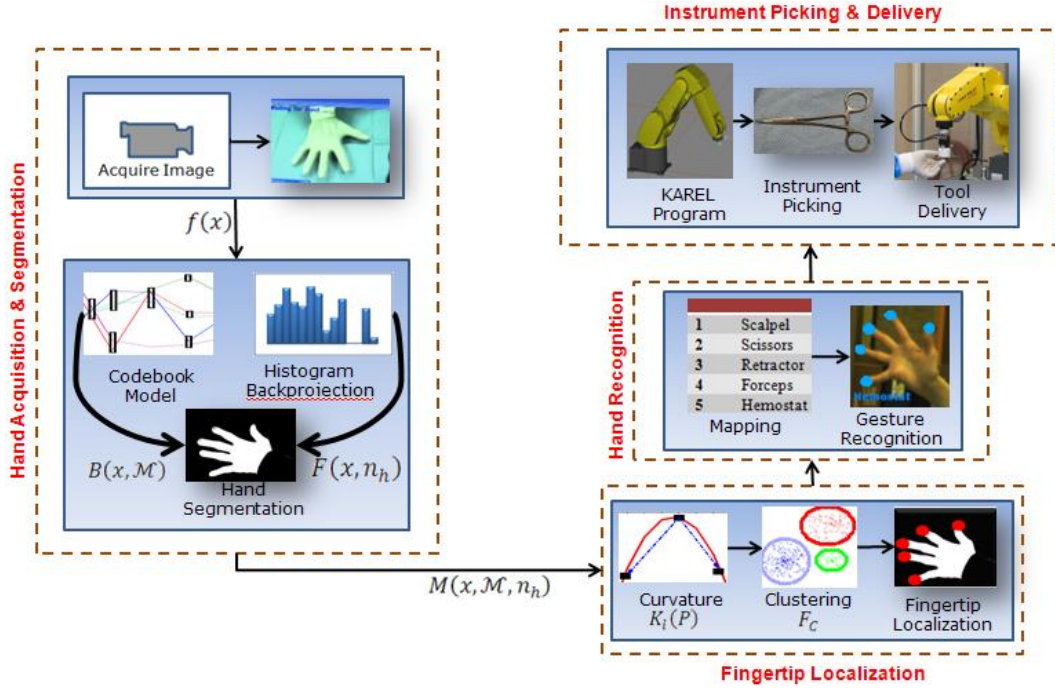


Figure 2. System Architecture

robot. An application module controls the FANUC robotic arm across the network through the Telnet interface with the tool delivery module. Finally, the GRSN hands the required surgical instrument to the surgeon and awaits the next command.

The processing modules in the system architecture (see Figure 2) are briefly described in the following subsections.

A. Hand Segmentation Module:

A background model of the scene and a hand color model are first created. The segmentation masks from these models are combined and morphologically processed. The mask of the hand is obtained by choosing the largest blob in the scene and thresholding its area to filter out noise. The contour and convex hull is extracted from the hand mask and passed to the fingertip detection module.

B. Fingertip Localization Module:

The curvature of the contour is computed at different scales and a point on the contour is accepted as a candidate for a fingertip if it is a local minima and is less than an empirically determined threshold. Due to contour discretization and noise, several candidates satisfying aforementioned criteria can exist around a fingertip. The candidates close to the convex hull are retained and are clustered together. The centroid of each cluster is then designated as the fingertip.

C. Gesture Recognition Module:

The number of fingertips is bijectively mapped to a set of surgical instruments. Therefore, once the fingertips have been localized and counted, we can recognize the static hand posture performed by the surgeon.

D. Tool Delivery Module:

The FANUC robot is programmed using KAREL, a scripted language used to control FANUC robots. The

corresponding KAREL program is selected for the recognized gesture and is executed on the FANUC controller over the Telnet interface.

The robot proceeds to pick the instrument from its predefined position and pass it to the surgeon and then returns to the starting position.

III. GESTURE ANALYSIS

A. Hand Segmentation

The codebook algorithm [17] is used to model a background scene based on multiple samples of the background. The algorithm observes the YUV values of a pixel x during the training phase and creates/expands existing sets in YUV space to cover the values observed over time. Since the size of a set is limited, it cannot expand to cover all the possible values of x and thus, a new set is created.

The sets or codebook entries constitute the learned codebook model \mathcal{M} which we use to create a mask to segment the hand from the scene, i.e. a foreground pixel or a pixel which cannot be explained by \mathcal{M} . Let the YUV values of a pixel x be $f(x)$ such that $f^Y(x)$, $f^U(x)$, and $f^V(x)$ represent the Y, U, and V components of x respectively.

Let a codebook entry be $R \in \mathcal{M}$. A codebook entry is associated by 6 values, the upper and lower bounds for each component in $f(x)$ i.e. R_U^k and R_L^k respectively for the Y component. Therefore, R can be defined as the set of points in YUV space which lie within aforementioned bounds for each component of $f(x)$.

$$R = \{f(x): R_L^k \leq f^k(x) \leq R_U^k \text{ for } k \in \{Y, U, V\}\} \quad (1)$$

Hence, we can generate a mask B for the background pixels using the codebook model and the YUV values of a pixel x as follows:

$$B(x, \mathcal{M}) = \begin{cases} 0 & : f(x) \notin R \quad \forall R \in \mathcal{M} \\ 1 & : \text{otherwise} \end{cases} \quad (2)$$

A foreground mask (see Figure 3(a)) based on the hand color is also generated which is stored as a histogram n_h . Histogram back-projection [18] is used to determine the probability $p(x, n_h)$ that a pixel belongs to the hand color histogram n_h . A foreground mask $F(x, n_h)$ (see Figure 3(b)) is generated by thresholding this probability with a constant γ :

$$F(x, n_h) = \begin{cases} 1 & : p(x, n_h) > \gamma \\ 0 & : \text{otherwise} \end{cases} \quad (3)$$



Figure 3. (a) Background mask B (b) Foreground mask F (c) M

The combined hand mask M (see Figure 3(c)) is obtained as follows. Note that $\overline{B(x, \mathcal{M})}$ denotes the logical negation of the output of B .

$$M(x, \mathcal{M}, n_h) = \overline{B(x, \mathcal{M})} \cap F(x, n_h) \quad (4)$$

A morphological closing operation is used to clean M and remove any spurious mask elements. Additionally, the largest blob in the scene is selected after its area is thresholded and the contour C and convex hull H is computed for use in the fingertip detection module.

B. Fingertip Detection

We build on the finger detection method used by Argyros *et al.* [19]. Their curvature measure $K_l(P)$, is modified so that it lies in the range $[0, 1]$. Let P_1, P , and P_2 denote successive points on the contour and let θ be angle between vectors $\overrightarrow{P_1P}$ and $\overrightarrow{PP_2}$. Also, let P_1, P , and P_2 be separated by l points. Then, the curvature measure is defined as:

$$K_l(P) = \frac{1}{2} (1 + \cos \theta) \quad (5)$$

The parameter l is used to detect fingertip candidates and valleys at several scales by constructing a set L of detected local minima on the contour in $K_l(P) : P \in C$.

Then, the candidate set S of finger tips is created from local minima, thresholded by the curvature with κ :

$$S = \{P \in L : K_l(P) \leq \kappa\} \quad (6)$$

Furthermore, the local minima can be effectively detected by distinguishing valleys between fingers from the fingertips using the convex hull of the whole hand (see Figure 4(a)).

We empirically found that the fingertips lie close to the convex hull H . Let $NN(P, H)$ return the nearest-neighbor of P in H , and d is the threshold on the Euclidean distance from the convex hull. We refine the candidate set S based on the proximity to the convex hull H and define F_C as:

$$F_C = \{P \in S : \|P - NN(P, H)\|_2 \leq d\} \quad (7)$$

Discretization of the contour or imperfect hand segmentation can cause F_C to contain several points around the

true fingertip since several local minima satisfy the criteria in Equations (6) and (7). Therefore, points in F_C which are close to each other or whose separation distance is thresholded by a constant δ are clustered together and a representative point is chosen as the finger tip.

The ordering of points known from the contour C is used to improve clustering performance. Let P_a and P_{a+1} denote two adjacent points on the contour C and let a set of contiguous points on C from P_a to P_b inclusive be defined as:

$$P_{ab} = \{P_a, P_{a+1}, P_{a+2}, \dots, P_{b-1}, P_b\} : a \leq b \quad (8)$$

Then, the candidate points are grouped into the set of clusters T as follows:

$$T = \{P_{ab} : \|P_i - P_{i+1}\|_2 \leq \delta \text{ for } a \leq i < b\} \quad (9)$$

The size of each cluster in T is thresholded and the representative point for each cluster is its centroid (see Figure 4(b)).

C. Gesture Recognition

Figure 5 displays the localized fingertips on a hand contour. The aforementioned algorithm is capable of localizing fingertips and determining local minima which satisfy the constraints in Equations (6)-(9). Five poses are determined using the fingertip detection method and are mapped to surgical instruments (see Table 1).

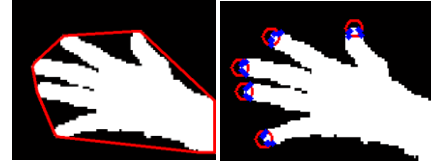


Figure 4. (a) Convex Hull (b) Clustering of candidate fingertips

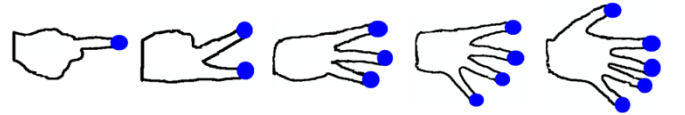


Figure 5. Fingertips on (a) Hemostat (b) Scissors (c) Scalpel gestures

IV. ROBOTIC TOOL DELIVERY

The FANUC LR Mate 200iC (see Figure 6(a)) robotic arm is used to pass surgical instruments to the surgeon. Teach-pendant (TP) programs were recorded for the delivery of each surgical instrument. The KAREL program interacts with the gesture analysis module over the network through aTelnet interface.

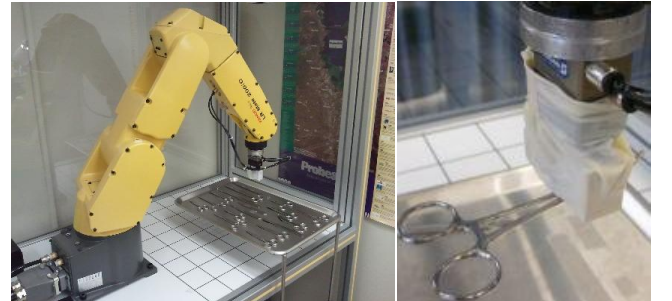












Figure 6. (a) FANUC LR Mate 200iC (b) Gripper with instrument

TABLE 1
GESTURE MAPPING

Name	Instrument	Gesture
Scalpel		
Scissors		
Retractor		
Forceps		
Hemostat		

The complete system is illustrated in a flowchart (see Figure 7). A router is used to connect the PC, FANUC robot and network camera. Additionally, a latex-encased magnetic gripper is used in order to maintain instrument sterility (see Figure 6(b)) when the robot hands the instrument to the surgeon.

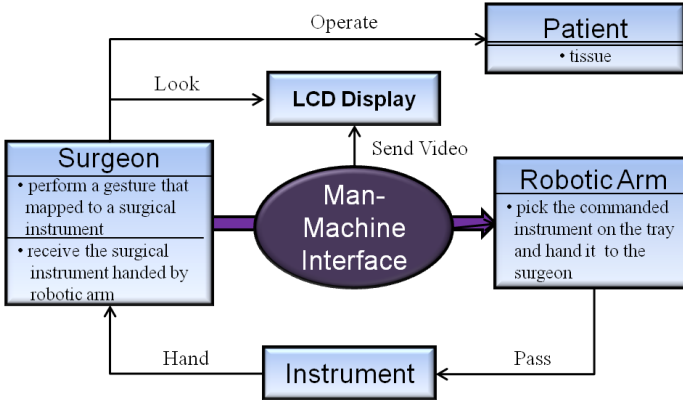


Figure 7. System Flowchart

V. EXPERIMENTS

A. Experiment 1: Instrument Picking Performance

Forceps, hemostats and scissors-type instruments may be required in large quantities (300-400) and are usually placed close together (see Figure 8). High precision and reliability is required to accurately deliver instruments packed close together in small clusters.



Figure 8. Instrument picking accuracy vs. inter-instrument distance (λ)

This scenario is tested in the following experiment. The distance between the centerlines of two instruments in a cluster is defined as λ . The performance of the system in picking different types of surgical instruments clustered in a small area of the Mayo stand is studied by recording the picking accuracy of the GRNS with respect to λ per instrument type.

The results are displayed in Figure 9. A picking and delivering task is evaluated. The number of trials which resulted in successful delivery was recorded from ten trials per instrument. It is counted as an error when the gripper of the GRNS either does not pick the instrument, drops the instrument before reaching the surgeon's hand, picks more than one instrument or picks the wrong instrument.

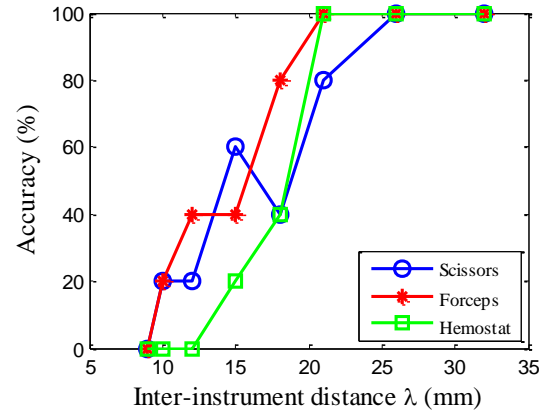


Figure 9. Instrument picking accuracy vs. inter-instrument distance (λ)

B. Experiment 2: Gesture Analysis Performance

A gesture analysis database consists of 300 cluttered background images (see Figure 10 for samples) for training the codebook model, and 1000 RGB images of size 720x480 pixels per gesture performed by a single user. The dataset used for testing consists of 2 databases captured from users of different hand color, shape and size.

Each user was instructed to keep their hands parallel to the image plane and move their hands in the image plane. This resulted in images of static pose gestures at different scales, rotations and positions.

The curvature κ was varied to obtain the ROC curves (see Figure 11) for each gesture and for fingertip detection across all gestures. The confusion matrix in Table 2 was generated with $\kappa = 0.30$ over our database of static hand pose gestures. We use the ϕ class to represent cases where zero or more than 5 fingertips were detected. On average, the recognition accuracy was found to be 94.65%.

TABLE 2
CONFUSION MATRIX (%) FOR $\kappa = 0.30$

	Scalpel	Scissors	Retractor	Forceps	Hemostat	ϕ
Scalpel	97.15	2.30	0.25	0	0	0.3
Scissors	3.20	96.75	0.05	0	0	0
Retractor	0	7.30	92.70	0	0	0
Forceps	0	0	8.55	91.40	0.05	0
Hemostat	0.1	0.7	1.65	2.20	95.25	0.1

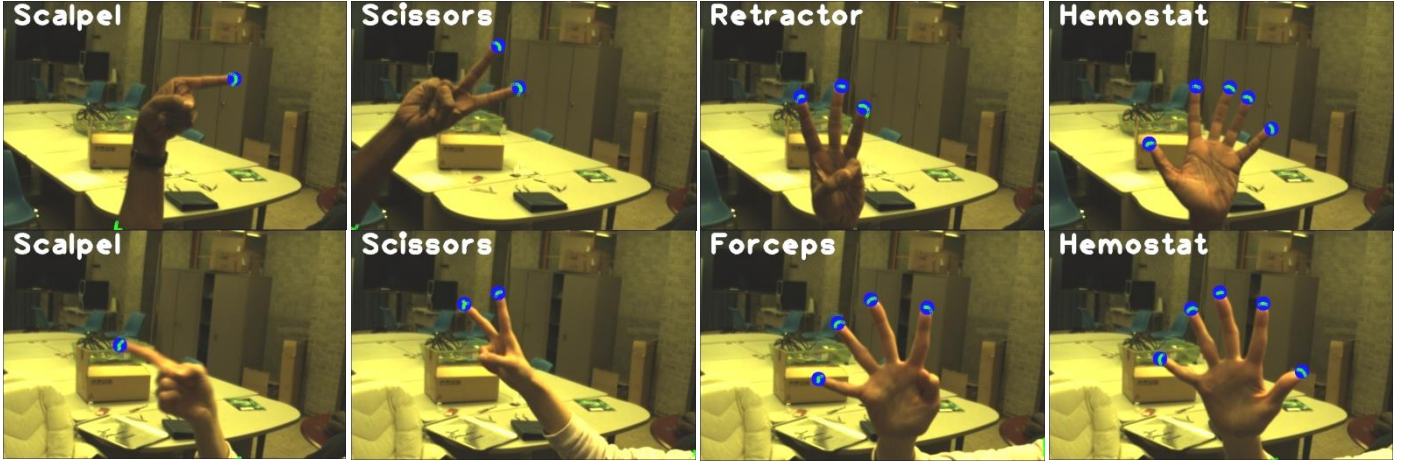


Figure 10. Samples from the database with correct fingertip detection

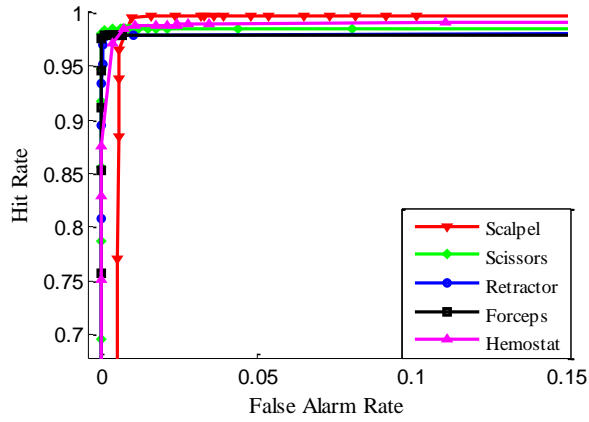


Figure 11. ROC curves for different fingertip detection algorithm

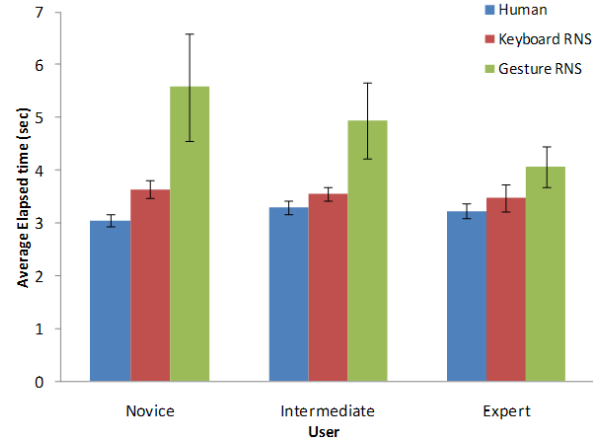


Figure 12. Comparison of different systems with the mean and 95% CI

C. Experiment 3: System Speed Comparison

The speed of the complete system (GRNS) was compared to that of a person passing the instrument (Human) and when a keyboard is used to request the required instrument from the robot (KRNS). Figure 12 displays the mean system time for of each class for the aforementioned systems as well as the 95% confidence intervals (CI).

The experiment was conducted as follows. An arbitrary sequence of instruments was randomly generated and used as the test sequence for all the subjects and all the systems being compared. The name of each instrument from the test sequence was displayed to the subject who requested the instrument.

The system time is defined as the time elapsed between the instrument name and the subject receiving the instrument and was recorded for each system which was compared (GRNS, KRNS and Human).

The human system was simulated with the subject saying the name of the instrument out loud and a person handing over the instrument from the Mayo stand (the person does not see the displayed name from the test sequence).

The KRNS is the same as the GRNS except a keyboard interface replaces the gesture interface. Each instrument is represented by an alphabet (most of the time, the first letter of

the instrument name) and the subject sees the instrument name and presses the appropriate key on the keyboard.

The test sequence has 25 instruments and three classes of three users per class were studied. Each class corresponds to the amount of experience with the GRNS. A novice has little experience with the system and is allowed to test a few gestures before starting the trials. The intermediate and expert users are allowed to “warm up” with a few gestures as well and have already been tested as a novice and intermediate user respectively.

VI. RESULTS & DISCUSSION

A. Experiment 1: Instrument Picking Performance

In Experiment 1, it is observed that the robot can reliably pick the instrument from the Mayo stand and hand it to the user when instruments are separated by at least 25mm (see Figure 11). Additionally, the magnetic gripper has been shown to be effective at picking instruments even if they are stacked close together.

It is apparent that there exists an intuitive relation between the degree of packing instruments in a small area of the Mayo stand is (or equivalently the inter-instrument distance λ) and the accuracy of the picking and delivery task. In addition, there is a link between the shape of the instrument and the overall

performance of the task. It is seen that picking instruments with smaller areas, like forceps have higher accuracy but the picking performance of the hemostat type of instruments is worse due to its larger area.

Since larger portions of instruments overlap, when instruments are packed close together, the accuracy of task delivery is impacted because events such as a falling or an unpicked instrument occur thus resulting in interference between adjacent instruments.

B. Experiment 2: Gesture Analysis Performance

Experiment 2 showed that the fingertip detection algorithm achieved a very high average hit rate of 98.17% with a false positive rate of 0.63% at $\kappa = 0.30$. The lower average gesture recognition accuracy of 94.65% is explained by out-of-plane rotations (see Figure 13) in the database. Since the curvature changes dramatically during these rotations, the fingertip detection algorithm fails to detect the fingertips.

In practice, users learn to keep dramatic out-of-plane rotations to a minimum and thus achieve high gesture recognition accuracy. This is apparent from the decreasing average system time as users gain more experience (see Figure 12).



Figure 13. Examples of out of plane rotation

C. Experiment 3: System Speed Comparison

Experiment 3 shows that the time elapsed between users observing the instrument name and receiving the instrument from the GRNS was 4.06s on average with expert users which is only 0.83s slower than the human system. This result shows that further incremental improvements on the GRNS implementation can result in a fully operational system in the OR.

VII. CONCLUSION & FUTURE WORK

A gesture driven robotic scrub nurse is presented capable of reliably passing surgical instruments to a user through gesture recognition. The gesture recognition module has a fingertip detection accuracy of 98.17% and gesture recognition accuracy of 94.65% on average.

It has been shown that the system is capable of accurately picking up stacked instruments when separated at least by 25mm. Also, experienced users can obtain instruments from the GRNS only 0.83s slower than a human-based system on average.

Future work includes using a more robust method to obtain the hand segmentation. Additionally, it would be advantageous to add a speech recognition module to work in confluence with gesture recognition since it has the potential to improve overall recognition.

Instrument localization and recognition is another useful addition since it will ease the constraint on maintaining a fixed position of the surgical tray relative to the robot.

ACKNOWLEDGMENT

This project was partially supported by the Intelligent Robotics Systems Lab. The authors would like to thank Dr. Steve Adams for his consultancy and the use of the OR and Dr. Rebecca Packer for the surgical supplies used in the experiment.

REFERENCES

- [1] P. McCulloch, A. Mishra, A. Handa, T. Dale, G. Hirst, and K. Catchpole, "The effects of aviation-style non-technical skills training on technical performance and outcome in the operating theatre," *British Medical Journal*, vol. 18, no. 2, p. 109, 2009.
- [2] L. Lingard, S. Espin, S. Whyte, G. Regehr, G. Baker, R. Reznick, J. Bohnen, B. Orser, D. Doran, and E. Grober, "Communication failures in the operating room: an observational classification of recurrent types and effects," *Quality and Safety in Health Care*, vol. 13, no. 5, p. 330, 2004.
- [3] L. Kohn, J. Corrigan, M. Donaldson et al., "To err is human: building a safer health system," Washington, DC, 1999.
- [4] L. Mitchell and R. Flin, "Non-technical skills of the operating theatre scrub nurse: literature review," *Journal of advanced nursing*, vol. 63, no. 1, pp. 15–24, 2008.
- [5] J. Borenstein and Y. Koren, "A mobile platform for nursing robots," *Industrial Electronics, IEEE Transactions on*, no. 2, pp. 158–165, 2007.
- [6] L. van den Bedem, "Realization of a demonstrator slave for robotic minimally invasive surgery," Ph.D. dissertation, Technische Universiteit Eindhoven, 2010, 2010.
- [7] I. Surgical, "Da Vinci Surgical System," 2004.
- [8] A. Kochan, "Scalpel please, robot: Penelope's debut in the operating room," *Industrial Robot: An International Journal*, vol. 32, no. 6, pp. 449–451, 2005.
- [9] E. Carpintero, C. Perez, R. Morales, N. Garcia, A. Candela, and J. Azorin, "Development of a robotic scrub nurse for the operating theatre," in *Biomedical Robotics and Biomechanics (BioRob) 2010*, pp. 504–509.
- [10] Y. Takahashi, T. Takatani, K. Osako, H. Saruwatari, and K. Shikano, "Blind spatial subtraction array for speech enhancement in noisy environment," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 17, no. 4, pp. 650–664, May 2009.
- [11] SC Beyea, "Noise: a distraction, interruption, and safety hazard," *AORN J*, 2007;86(2):281–285.
- [12] K. Yoshimitsu, F. Miyawaki, T. Sadahiro, K. Ohnuma, Y. Fukui, D. Hashimoto, and K. Masamune, "Development and evaluation of the second version of scrub nurse robot (SNR) for endoscopic and laparoscopic surgery," in *IROS 2007*, pp. 2288–2294.
- [13] Laws HL, "Intraoperative communication," *Surg Gynecol Obstet*. 1985 Mar;160(3):268–9..
- [14] Fulchiero GJ, Vujevich JJ., Goldberg LH, "Nonverbal hand signals: a tool for increasing patient comfort during dermatologic surgery," *Dermatol Surg* 2009 May; 35(5):856–7.
- [15] N. Phillips, E. Berry, and M. Kohn, Berry & Kohn's operating room technique. Mosby Inc, 2004.
- [16] A. Pezzella, "Hand Signals in Surgery", *AORN Journal* vol. 63, no. 4, p.769, 1996
- [17] K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-time imaging*, vol. 11, no. 3, pp. 172–185, 2005.
- [18] G. Bradski, "Computer vision face tracking for use in a perceptual user interface," 1998.
- [19] A. Argyros and M.Lourakis, "Vision-based interpretation of hand gestures for remote control of a computer mouse", in *Computer Vision in Human-Computer Interaction, series Lecture Notes in Computer Science*, T. Huang, N. Sebe, M. Lew, V. Pavlovic