Haptic Gesturing as Human-Machine Interface in Minimally Invasive Robotic Surgery

C. Staub, S. Can, A. Knoll

Robotics and Embedded Systems, Technische Universität München,Germany

Corresponding author: staub@in.tum.de

INTRODUCTION

Tele-operated master-slave systems have created new opportunities in assisting surgeons during complex interventions. In addition to sophisticated input devices, tremor filtering, force feedback, and a stereoscopic view, researchers have drawn their attention to higher level functionality. Intelligent and automated camera tracking, knot-tying and partially autonomous executed tasks, as well as augmented reality applications are prominent examples. In contrast, the available number of input options to control these functions is limited, not last more because of the traditional design of master consoles. For instance, the high immersiveness of master consoles makes it hard to distinguish more than 4 to 5 pedals with the feet. Since autonomous functions are to be primarily actions to shorten operation time, new input modalities have to be time-saving alternatives to convectional human-machine interaction. If triggering a system function takes more time than its actual execution, the benefit of autonomy in medical robotics is diminished significantly. "Hands free" camera control has been introduced quickly after the AESOP camera holder was introduced to the market. Voice activated control and visual tracking of the surgeon's head movement replaced foot pedals and hand controller [1]. Vision-based control often comes with the burdon of additional tracking markers, which have to be borne by the doctor. A promising alternative might be to execute suitable system functions by means of gaze contingent control. Mylonas et al. investigated the integration of an eye tracking system into the daVinci master console [2], with application to e.g., depth recovery, virtual fixtures, and instrument control.

Fig. 1 Master console with haptic input devices, stereoscopic screen, and slave system in the background

We have implemented a gesture-based user interface, whereas the haptic input devices of the master console are used to trigger actions. Intuitive and customizable gestures are learned by the system once, linked by the surgeon to the preferred command, and recalled during operation as the gesture is presented to the system. In particular, the aspect of time consumption is interesting, since, in contrast to many conventional input modalities, full control of the surgical instrument can be kept during system interaction and the time-consuming resembling of the original robot posture is avoided.

MATERIALS AND METHODS

The setup of our robotic system for minimally invasive surgery consists of a master console, where the surgeon is located at, and a robotic slave part. The surgical workstation is equipped with two PHANToM™ haptic in-/output devices, a stereoscopic monitor and several foot pedals that can be programmed individually for system interaction (cp. Fig. 1). The slave part of the system is composed of four ceiling mounted robots, carrying EndoWrist™ surgical instruments from Intuitive, Inc. The instruments have been modified to measure occurring forces at the instrument tip via strain-gauge sensors. In order to recognize a set of four pre-defined surgical gestures, a discrete Hidden Markov Model (HMM) with left-right topology is used. The gestures have been chosen after an initial interview and reflect movements that subjects found to be intuitive for the autonomous actions "knot-tying", "suturing", "retract 3rd arm", and "measure distance". The HMM was trained on a basis of 15 individual executions of each gesture. Observations for building the HMM are derived from the vectors of the time-series representing the instrument trajectory. The trajectory data was sampled with a frequency of 10Hz and stored in a data base, comprising the Cartesian position of the end effector, corresponding forces, the state of the forceps (open/closed), and a time stamp. In a preprocessing step, the raw trajectory data is resampled position equidistant. This smoothes variation in the execution speeds and removes outliers, which arise due to the human tremor. The used features either deal with a single instrument (left l and right r), or refer to the interaction of both. The following features have been used: (1) change of the instrument trajectory (l/r) over time, (2) directional change of one instrument w.r.t. the second instrument (3), velocity of an instrument (*l/r*), (4) distance between the two instruments, (5) temporal change of the distance between two instruments. To simplify the evaluation for untrained subjects, the state of the gripper was neglected at this time. If environmental interaction is assumed during a gesture, the readings of the force sensors can also be included.

RESULTS

The evaluation study was conducted with 24 medical students (*M=24yrs., SD=3yrs.*), half of whom had surgical experience. The gesture input described above, was tested against menu input. A plausible two-tiered menu design was chosen: on the first screen of the menu, a general "surgical action" option had to be selected out of 4 possibilities. The second screen then offered the 4 surgical actions plus one arbitrary dummy menu entry, whereas the appropriate action had to be selected and confirmed with the foot pedal. Since none of the subjects were familiar with a robotic surgery system, participants had training time to become acquainted with the system and where asked to trigger all actions with both gesture and menu input mode. The time that it took people to activate the asked action as well as the success rate in triggering the correct action was measured.

The usability of both input modes was assessed with the AttrakDiff2 [3], a well-tested questionnaire measuring four different aspects of usability on Likert-type scales. Each of these scales was found to show high reliability with Cronbach's $\alpha > 0.70$. A factorial ANOVA found a large and statistically significant effect of input mode on input time ($F(1,22)=38.44$; $p<.001$; $\eta=.64$). The estimated marginal means indicate that, on average, it took significantly less time to trigger the surgical action via gesture input (*M=4.45sec., SD=0.86sec.*) compared to activation via menu input (*M=7.41sec., SD= 2.06sec.*). There was also a significant main effect of gesture ($F(2.04,44.79)=23.79$; $p<.001$, $\eta=.52$), but no significant interaction effect ($F(3,66)=2.18$, $p=.10$). Together, these results suggest that while some surgical actions (e.g. arm retraction) took longer to activate than others (e.g. suturing), input times were consistently shorter with gesture input than with menu input (cp. Fig. 2). A look at the input errors suggest that, while it took less time to input a command for a surgical action via gesture, this mode is slightly more error prone with 10.42% of gesture inputs classified as false compared to 5.21% of false inputs via the menu (out of 96 commands). Finally, an ANOVA of the usability scores indicates a significant main effect of input mode ($F(1,23)=23.74$, $p<.001$; $\eta=.51$), whereby significantly higher mean usability ratings were given for gesture input (*M=5.40, SD=0.87*) than for menu input (*M=4.21; SD=0.85*).

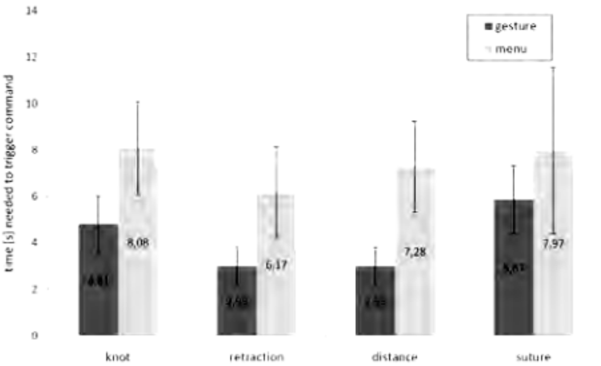

Fig. 2 Times needed to trigger a surgical action. The menu interaction time does not comprise the time needed to resemble the posture of the robots afterwards.

DISCUSSION

The results show that gesture-based input is faster and receives higher usability ratings compared to the tested menu-mockup, even though this input method is still slightly more error prone. Although the effect of learning on input success has not been explicitly investigated in this study, it seems likely that, despite the rigorous training protocol implemented in this experiment, participants were more practiced in menu-based input than in gesture-based input. Hence, one might assume that the likelihood to commit an error with gesture input would decrease with further practice. Nevertheless, further studies are required to determine the factors that mitigate the effectiveness of gesture-based input. For example, obviously, the superior effectiveness of gesture-based input over the traditional menu input strongly depends on the complexity of the menu, as well as the input mechanisms (e.g. foot pedals vs. mouse-type interaction). Whether or not gesture-based input is equally effective for other setup needs to be tested in future studies. Similarly, the usability ratings should be interpreted with caution. Three of the four scales on the usability questionnaire would favor technology that would be considered novel and exciting. The pragmatic qualities of gesture-based input, such as its ability to integrate into the surgical workflow, need to be tested in long-term user studies.

REFERENCES

[1] J.-B. Gómez, A. Ceballos, F. Prieto, and T. Redarce, "Mouth gesture and voice command based robot command interface", Proceedings of the IEEE International Conference on Robotics and Automation, *2009, pp. 333–338.*

[2] G. Mylonas, A. Darzi, and G.-Z. Yang, "Gaze contingent depth recovery and motion stabilisation for minimally invasive robotic surgery", Proceedings of the International Workshop on Medical Imaging and Augmented Reality. *Springer Berlin / Heidelberg, 2004, pp. 311–319.*

[3] M. Hassenzahl, M. Burmester, and F. Koller. "Attrakdiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität", Mensch & Computer 2003: Interaktion in Bewegung, 2003, pp. 187–196.