Implementation and Evaluation of a gesture-based Input Method in Robotic Surgery

Christoph Staub, Salman Can, and Alois Knoll
Robotics and Embedded Systems
Technische Universität München
D-85748 Garching, Germany
{staub|cans|knoll}@in.tum.de

Verena Nitsch, Ines Karl, and Berthold Färber
Human Factors Institute
Universität der Bundeswehr München
D-85577 Neubiberg, Germany
{verena.nitsch|ines.karl}@unibw.de

***Abstract*— The introduction of robotic master-slave systems for minimally invasive surgery has created new opportunities in assisting surgeons with partial or fully autonomous functions. While autonomy is an ongoing field of research, the question of how the growing number of offered features can be triggered in a time-saving manner at the master console is not well investigated. We have implemented a gesture-based user interface, whereas the haptic input devices that are commonly used to control the surgical instruments, are used to trigger actions. Intuitive and customizable gestures are learned by the system once, linked to a certain command, and recalled during operation as the gesture is presented by the surgeon. Experimental user studies with 24 participants have been conducted to evaluate the efficiency, accuracy and user experience of this input method compared to a traditional menu. The results have shown the potential of gesture-based input, especially in terms of time savings and enhanced user experience.**

***Index Terms*— robot surgery, minimally invasive surgery, human-machine interaction, gesturing**

I. INTRODUCTION

Medical robots are well on their way to become as important to the surgical process as industrial robots have become to manufacturing within the last 30 years. Over the last years there has been a significant development in robot-assisted minimally invasive systems (RMIS). They enable the surgeon to overcome several barriers and limitations, compared to conventional interventions: scaling of movements, tremor filtering, improved dexterity and haptic feedback aid the surgeon just as well as teleoperative teamwork (possibly over long distances) does. Typically, all this functionality is provided by a master console, which is equipped with sophisticated input devices and an immersive real-time visualization of the situs. The instruments at the slave side of the system can be controlled remotely by a surgeon sitting at the console. A remarkable example of such achievements is the daVinci^{TM}machine [1].

Simultaneously, much effort is made in intelligent assistance functions that aim to improve both patient safety and operation time. Virtual constraints [2] and supervision of manipulation forces [3] minimize the risk of tissue damage. Augmentation of the patient with preoperative imaging modalities [4] and the associated information presentation, camera control and (partly) autonomous executed actions of error-prone and recurrent (sub-) tasks, such as knot-tying [5] or tissue piercing [6], have drawn the attention of researchers. This growing number of high level functionality that we might see in future (commercial) systems has to be executed, controlled and supervised by the human operator. However, we believe the number of existing functionality to be imbalanced with the currently available input options of master consoles. Most RMIS master consoles come with several foot pedals, e.g., for swapping between different surgical tools or repositioning the camera. Also a touchpad and buttons might be available for system settings, but the user is forced to remove his hand from the main control device. Using the devices for menu interaction is possible but time consuming, since the devices have to be relocated after menu interaction to resemble the posture of the end effector.

With regard to the growing number of autonomous functions it is desirable to keep the time used for system interaction at a minimum. If the call of an assistant function takes more time than its actual execution, the benefit of automation is questionable. Several groups investigate human friendly machine interfaces: A large body of literature is available with respect to automated camera guidance. The first commercial endoscope holder, the AEOSOP^{TM}, was released to the market in 1994 and controlled by the surgeon by means of either a foot or a hand controller [7]. To provide "hands free" control, camera assistants were quickly extended with new input modalities such as voice activated control [8], visual tracking of the surgeon's movements (e.g., head movements or "mouth gestures" [9]), which often comes with the burden of tracking markers, which have to be borne by the doctor.

While force reflecting devices are optimized with respect to the specific requirements of different types of interventions [10] and the overall integration of a multitude of human-machine interfaces into master consoles were studied [11], the problem of executing system functions is barely tackled. [12] investigated the direct execution of higher level functions by means of gaze contingent control. An eye tracker was integrated into the daVinci^{TM}stereoscopic console, which allows video capturing of the eyes at $50fps$ without obstructing the surgeon's view. The eyes are illuminated with a fixed infrared light source and the corneal reflection in relation to the position of the pupil is measured. The two centers of the eyes define a vector, which can be mapped to an unique gaze direction. The relationship between horizontal

978-1-4577-0499-4/11/$26.00 ©2011 IEEE

disparity and depth perception that varies with the viewing distance recovers depth information. They propose to use the surgeon's fixation point to recover tissue deformation for beating heart stabilization [12]. In [13] virtual fixtures are interactively prescribed and updated via eye tracking to guide the user to an incision point at the tissue surface. The same eye tracker setup is used to automate articulated instrument positioning [14] and 3D path planning in focused energy ablation [15].

Like most input modalities, gaze control is dedicated to a specific group of tasks and cannot cover all cases of human-machine interaction. Foot pedal or menu interaction interrupts the intervention, since the user needs to locate the pedals (for which s/he may have to take his eyes off the screen) or has to navigate through a (more or less) complex menu structure. The introduction of "menu shortcuts", which show a reduced and situation adapted number of items, is possible, but difficult to realize: On one hand, the current state of the medical workflow needs to be known precisely, on the other hand, only few foot pedals are available. If the haptic input devices are used as a mouse pointer for menu interaction, their posture needs to be readjusted to the current robot pose after use.

In general, we ask the following requirements for an input modality in medical robotics:

- shortest possible distraction of the surgeon from the operative situs
- little cognitive burden and mental stress
- fast and seamless integration into the surgical workflow
- correct interpretation and execution of the commands
- little training effort

Sign language and haptic gesturing promises to be an intuitive and easy understandable concept for human-machine interfaces, borrowed from everyday life. In [16], the authors propose that a human operator give hints to a teleoperated robot by natural sign language. The signs define a spatial-temporal context (e.g., the user points to an object) for subsequent robot behavior. The signs are interpreted from finger encoder readings of an exoskeleton. In a similar fashion, we propose the use of the haptic input devices of a surgical workstation to trigger commands with gestures. In comparison to traditional human-computer interfaces (e.g., menus), which are commonly used in robotic surgery, we hope to offer a customizable and intuitive input modality with reduced interaction time.

II. The EndoPar System

The Endoscopic Partial-Autonomous Robot (EndoPar) system is an experimental robotic surgical system, developed by the Robotics and Embedded Systems research group at the Technical University of Munich. The hardware and software components of the system have already been introduced to the research community [17]. Therefore, we limit the description to an extent necessary for understanding the subsequent sections. As depicted in Fig. 1 the slave part of the system is composed of four ceiling mounted

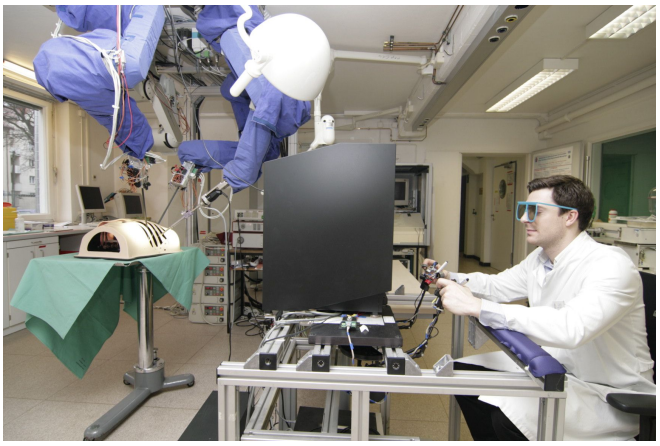


Fig. 1. **Hardware setup:** Ceiling mounted robots with surgical instruments

industrial robots that are either equipped with surgical instruments or with a stereoscopic camera. The utilized EndoWrist™instruments are originally deployed with the daVinci™system and are coupled by a magnetic clutch to the robots' flange. The instruments are powered by small servo motors, which are incorporated into the coupling mechanism. All instruments are augmented with strain gauge sensors in order to measure forces. The user is located in front of a master console that provides a stereoscopic view of the situs. Two Phantom™haptic displays are arranged upside down for less constricted flexibility of the stylus pen. The devices serve as input devices for the operator to control the surgical instruments and can simultaneously feed back forces measured at the instrument tip. As an additional input modality, four foot pedals can be taught with individual system functions.

III. Haptic Gesturing

Gesture recognition is widely studied as a computer input modality [18]. The identification and classification of human motions (e.g., facial gestures, pointing gestures, body poses) or movements executed by means of pointing device (e.g., a mouse, remote controllers, data gloves, or touch-sensitive displays) can be used to convey meaningful information or interactions with the environment.

Hand gestures are frequently used by humans, since they are very intuitive and expressive. Two categories can be distinguished: static finger configurations, also called *postures*, and dynamic *gestures*. Variations of the executed gestures usually occur between different instantiations as well as different performers. Hidden Markov Models (HMM) have been used successfully in the field of medical workflow modeling (e.g., [19]).

A. Gesture Recognition with HMMs

A Hidden Markov Model is a stochastic model, described by two random processes. An detailed description of Hidden Markov Modeling techniques can i.e., be found in [20]. The model is defined as a quintuple $\lambda = (N, M, A, B, \pi)$, where

Fig. 2. **Master console** with Phantom^{TM}devices and 3D screen

N is the number of states $S = \{s_1, \cdots, s_N\}$, M is the number of distinct observation symbols per state (the discrete alphabet size), $A = \{a_{ij}\}$ is the state transition probability distribution matrix and $B = \{b_j(k)\}$ is the observation symbol probability distribution in state j. Variable π denotes a probability distribution over the initial states. The first process of the model is a hidden Markov chain and describes the dependency of the state $q_t = s_j$, reached at time t, from the previous states $q_{t-1}, \cdots, q_1$. The second stochastic process defines the probability of the observation f_k in the state $q_k = s_i$ (a.k.a the *emission probability*).

In this work, the observations that allow to infer the state of the process are feature vectors f_k of time-series representing the surgical instruments. The chosen features are presented in Section III-C. The used type of HMM model follows the left-right topology: The states are arranged in a linear progression, whereas each state is entered at least once and no transitions to past states are allowed.

HMMs model the stochastic properties of a training set of time-series. To train a model λ and to optimize its set of parameters the Baum-Welch algorithm is a well-know tool. The Viterbi algorithm is then used to find the path with the highest likelihood $P(\Pi \mid \lambda)$ through the topology of λ that would generate the sequence Π). In order to prevent the underflow problem during longer time-series, a log-scaling is used.

B. Data Acquisition and Preprocessing

To obtain the data, we recorded 25 individual executions of each gesture. All demonstrations were performed by a person who was not involved in the later evaluation.

The trajectories of all grippers were sampled at a frequency of 10Hz and stored in a data base. The recorded raw data are represented by the vector $\vec{s}_{raw,i} = (x, y, z, f_x, f_y, f_z, g, t)^T$, where $(x, y, z)^T$ is the Cartesian position of the gripper with corresponding forces $(f_x, f_y, f_z)^T$. Variable g denotes the state of the gripper (open/closed) and t is a timestamp.

The raw trajectory data contain variations in execution speed that yield an inhomogeneous distribution of sampling points. In order to represent the data with regularly spaced points, the trajectory is resampled. Although the data is time-stamped, a time equidistant sampling is not preferable: the varying execution speed of trajectory parts yields to a different number of sampling points with unequal sampling distances between the points. For instance, segments that are executed with low velocity, such as tight turns, comprise more sampling points then fast movements, such as straight lines. Therefore, the data are resampled position equidistant that is with an uniform spatial spacing. The preprocessing removes demonstration dependent variations such as execution speed and small geometrical variations (i.e., jitter, caused by the human tremor). The Euclidean distance between two sampling points is used for a linear interpolation for each record (cf. Fig. 3). The resampled position data replace the raw data in $\vec{s}_{raw}$ and the vector used for feature extraction is then denoted as $\vec{s}$.

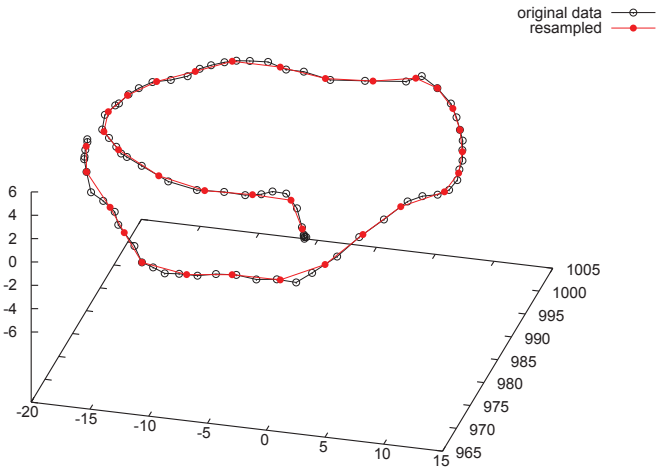


Fig. 3. Equidistant resampling of the original trajectory.

C. Features

After preprocessing, the trajectories are represented by sampling points with equal distance in combination with the recorded instrument forces and the gripper state. Variable $\vec{S}_i = (\vec{s}_{i,1}, \cdots, \vec{s}_{i,k})$ comprises all sampling points of a trajectory, where $i \in \{l, r\}$ indicates if the data belongs to the *left* or *right* instrument. The distance between two adjacent Cartesian points $p_{i,t}$ and $p_{i,t+1}$ is substituted by the vector $\Delta\vec{p}_{i,t} = \overline{p_{i,t}p_{i,t+1}}$. Note that none of the features is related to a global coordinate system, which is important to allow a location-independent execution of gestures. The features $F = \{f_1, \cdots, f_{10}\}$ are defined as follows:

1) **Directional change of the instrument trajectory:** Similar to [21], the change of the instrument direction, defined by two adjacent points $p_{i,t}$ and $p_{i,t+1}$ can be described by the angle ρ_t, enclosed by $\Delta\vec{p}_{i,t-1}$ and $\Delta\vec{p}_{i,t}$ (cf. Fig. 4). To ensure a smooth description, sin and cosine are calculated:

$$f_{1,t} = \cos\rho_t = \frac{\Delta\vec{p}_{i,t-1} \cdot \Delta\vec{p}_{i,t}}{\|\Delta\vec{p}_{i,t-1}\| \, \|\Delta\vec{p}_{i,t}\|} \quad (1)$$

$$f_{2,t} = \sin\rho_t = \frac{\Delta\vec{p}_{i,t-1} \times \Delta\vec{p}_{i,t}}{\|\Delta\vec{p}_{i,t-1}\| \, \|\Delta\vec{p}_{i,t}\|} \quad (2)$$

2) **Directional change of one instrument w.r.t. a second instrument:** The feature indicates if the first instrument tends to converge to the second instrument or veers away from it. Sin and cosine of the angle θ are calculated between the vectors $\Delta\vec{p}_{i,t}$ and $\vec{v} = \overline{p_{i,t}p_{j,t}}$. The corresponding features are $f_{3,t}$ and $f_{4,t}$ (cf. Fig. 4).
3) **Velocity of an instrument:** The velocity of the instrument tip can be derived from the recorded data. Each sampling point is labeled with a timestamp t. After the linear resampling of the trajectory, the velocity for each instrument is calculated by

$$f_{5/6,t} = \frac{l}{\vec{s}_{i,t} - \vec{s}_{i,t-1}} \tag{3}$$

where l is the sampling distance between two points after the preprocessing step. Note that from the position equidistant resampling follows a resampling of the timestamps that yields certain inaccuracies.
4) **Distance between the two instruments:** Depending on the gesture, the Euclidean distance between two instruments varies over time or keeps similar (e.g., in case of parallel moving instruments). The feature represents the current distance and is denoted as

$$f_{7,t} = \|p_{l,t} - p_{r,t}\| \tag{4}$$

5) **Temporal change of distance between two instruments :** In contrast to feature f_7, the distance change between two instruments over time indicates movement direction of one instrument w.r.t the second one:

$$f_{8,t} = \|p_{l,t} - p_{r,t}\| - \|p_{l,t-1} - p_{r,t-1}\| \tag{5}$$

6) **State of the gripper:** The state of the grippers (open/closed) can directly be taken from the data base and represented as features $f_{9,t}$ and $f_{10,t}$. For our experiments all subjects were told to keep the gripper closed during the gestures. On one hand this simplifies the gesture, on the other hand most novices assess the pose of the stylus more comfortable if the gripper is closed.

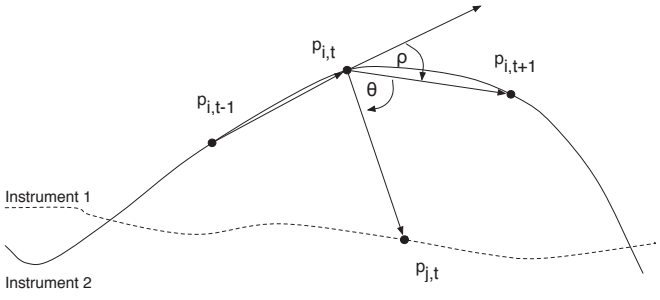


Fig. 4. Directional change of one instrument (angle ρ) and directional change of one instrument w.r.t to a second instrument (angle θ).

It is also possible to use a single instrument for gesturing, instead of two. In this case, we do not consider the change of the instrument w.r.t. the second one, but the directional change of the observed instrument. Instead of the distance between two instruments, the current distance between the instrument and the first point of the trajectory is measured, respectively the change of the distance.
The features are normalized over the range of all demonstrations and can be weighted afterwards to adjust their impact. To generate a discrete codebook from the data, the *k-means++* algorithm is used to quantize the feature vector.

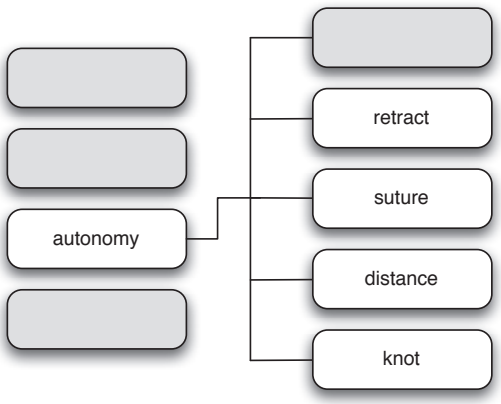


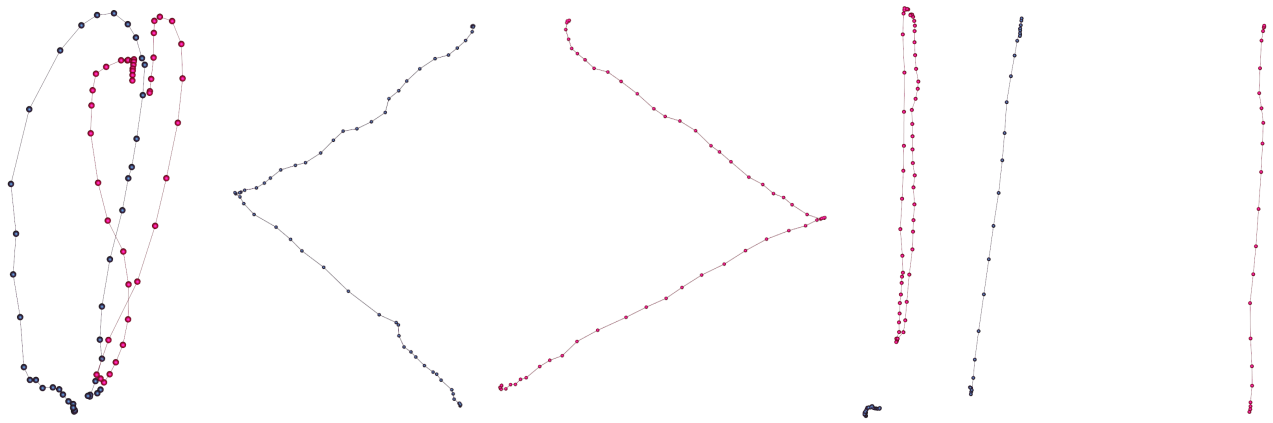
Fig. 5. Structure of the used menu.

IV. Experiments and Results

In order to evaluate the effectiveness of the proposed gesture recognition method within a realistic context, an experimental user study was conducted in which gesture-based input was compared to the traditional menu-based input. In addition to objective measures of performance with this system, another aspect evaluated in this study was the user experience. User experience is an extension of the concept of usability and has been defined as "all aspects of the user's experience when interacting with the product, service, environment or facility. [...] It includes all aspects of usability and desirability of a product, system or service from the user's perspective" [22]. User experience has been found to be a critical factor in the design of new products, as it was found to link to purchasing decisions and product use [23].

A. Identification of Intuitive Gestures

In principle, gesture-based input has the disadvantage that the gestures need to be remembered, whereas menu entries only need to be recognised. This requires greater cognitive effort and increases the risk of false input commands. This disadvantage is minimised, however, if the gestures that need to be remembered are intuitive. For a fair comparison of menu vs. gesture input, it was hence necessary, to first identify gestures for the system input, that would feel intuitive to the user and can therefore be easily remembered and executed. For this purpose, an exploratory pre-experiment was conducted with an opportunity sample of 22 participants ($M = 36\,yrs., SD = 14\,yrs.$), in which they were asked to spontaneously perform two alternative gestures



(a) (b) (c) (d)

Fig. 6. **Trajectories of gesture instances**: The blue lines indicate the left instrument, the red ones show the right instrument. Fig. 6(a) shows an instance of the "knot-tying" gesture, Fig. 6(b) shows the gesture for "suturing", Fig. 6(c) shows the gesture that initializes the "distance measuring". The picture is rotated 90° counterclockwise to save space. The gesture depicted in Fig. 6(d) would initialize the retraction of the 3rd robot arm to supply the surgeon with new material (e.g., threads).

that they would associate with each of nine pre-selected surgical assistance functions. After performing a particular gesture, they would also be asked to give a reason for their choice. The video recordings of the performed gestures were independently examined and rated by two trained raters. Based on the results of this study, four gestures which showed high consistency in ratings and high concordance amongst participants were selected for the evaluation study. The assistance functions associated with these gestures were "knot-tying", "suturing", "distance measuring" and "arm retraction". The trajectories corresponding to the selected gestures are depicted in Figure 6.

B. Gesturing vs. Menu Interaction

1) Method:

Participants: The main evaluation study was conducted with an opportunity sample of 24 participants ($M = 24\,yrs., SD = 3\,yrs.$), half of whom had surgical experience. Eleven participants were female and all but two were right-handed. None of the participants exhibited signs of motor impairment.

Experimental Design: A 2×4 ((input mode) $\times$ (gesture)) within-subject design was implemented, whereby gesture input was tested against menu input, with the four gestures described above. A plausible menu design was chosen for this experiment, with which participants had to select two options for each gesture: on the first screen of the menu, a general "surgical action" option had to be activated, which then led to the second screen on which the appropriate gesture had to be selected and confirmed (cf. Fig. 5). The time that it took people to activate a surgical action with the respective input mode was measured in each trial (input time), as well as the success rate in triggering the correct action (input success). The user experience of both input modes was assessed with the AttrakDiff2 [24], a well-tested questionnaire measuring four different aspects of user experience on a seven-point bipolar Likert-type scale. The four aspects of user experience measured are: pragmatic quality (PQ), attractiveness (ATT), hedonic quality-stimulation (HQ-S) and hedonic quality-identity (HQ-I). The construct of pragmatic quality refers to the perceived ability of a product to accomplish task goals by offering useful and usable functions and requires participants to rate the system on items such as complicated/simple and unpredictable/predictable. Attractiveness measures the users' global positive/negative evaluation of a product and contains items such as pretty/ugly and attractive/repulsive. Hedonic quality-stimulation refers to the ability of a product to satisfy the user's needs for the development of one's knowledge and skills and is rated with items such as unimaginative/creative and lame/mesmerizing. Finally, the construct of hedonic quality-identity measures the extent to which a product promotes one's self-worth and is comprised of items such as unstylish/stylish and cheap/valuable. The individual items comprising each scale were found to measure the respective constructs reliably, with Cronbach's $\alpha > 0.70$.

Procedure: Prior to the experiment, participants were trained in the use of both the menu and the gesture input modes in triggering the four gestures according to a standardized training procedure. Including familiarization with the system and filling in the questionnaire the procedure took about 1 hour per subject. On average, participants took 6.75 min. ($SD = 2.66\,min.$) to learn the four gestures, whereas it took on average 2.96 min. ($SD = 1.12\,min.$) to learn how to navigate the menu efficiently. Upon successful completion of the training phase, participants were then asked to either perform a certain gesture or select the appropriate menu items in order to trigger a particular action. The input modes were trained and tested in one block, meaning that participants would first be trained, then perform with one input mode, after which they would be trained and tested with the other input mode. The input mode and the tested gestures were systematically varied for each person in order to avoid learning or fatigue effects.

2) Results: Collected data were inspected for outliers and scores with z > 3.29 were removed for the statistical analysis. When the assumptions for parametric tests were violated, corrections were applied. A factorial ANOVA found a large and statistically significant effect of input mode on input time ($F(1, 22) = 38.44, p < .001, \eta^2 = .64$). The estimated marginal means indicate that, on average, it took significantly less time to trigger the surgical action via gesture input ($M = 4.45 sec., SD = 0.86\, sec.$) compared to activation via menu input ($M = 7.41 sec., SD = 2.06\, sec.$). The times needed to trigger a certain action are depicted in Fig. 7. There was also a significant main effect of gesture ($F(2.04, 44.79) = 23.79, p < .001, \eta^2 = .52$), 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. 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 slighlty 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). Table I shows the success rates for the individual actions. Finally, an ANOVA of the AttrakDiff2 scores indicates a significant main effect of input mode ($F(1, 23) = 23.74, p < .001, \eta^2 = .51$), whereby significantly higher mean user experience scores were given for gesture input ($M = 5.40, SD = 0.87$) than for menu input ($M = 4.21, SD = 0.85$). Bonferroni-adjusted post-hoc simple comparisons showed that only the scores to pragmatic quality did not differ significantly between the two input modes ($t(23) = 0.17, p = .87$), whereas gesture input received significantly higher ratings for hedonic quality-identity ($t(23) = 4.67, p < .001, r = .70$), hedonic-quality stimulation ($t(23) = 7.97, p < .001, r = .86$) and attractiveness ($t(23) = 4.39, p < .001, r = .68$). These findings indicate that, while the gesture input system was not necessarily considered to provide greater functionality than the menu, it was perceived to be more stimulating, creative and comfortable.

V. Discussion and Conclusion

In summary, a type of gesture-based input was implemented and evaluated as a time saving alternative to menu input for commanding frequently demanded, automated or semi-automated surgical actions in robot-assisted minimally invasive surgery. After identifying intuitive gestures associated with surgical assistance functions, a user study was conducted in order to compare the two input concepts

TABLE I

modality	knot	retraction	distance	suture	∅
gesture	95.83%	75.0%	100.0%	87.5%	89.58%
menu	95.83%	91.67%	95.83%	95.83%	94.79%

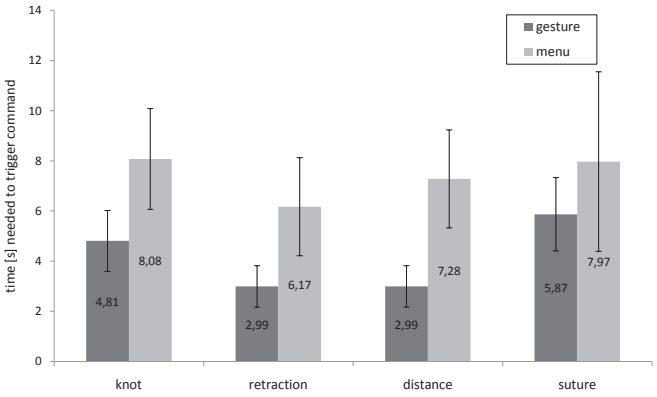


Fig. 7. Times needed to trigger an action (in sec): Gesture-based vs. menu.

in terms of performance and user experience. The results show that gesture-based input is faster and receives more favourable user experience ratings compared to the tested menu-mockup, even though this input method is still slightly more error prone. Furthermore, it must be noted that the subjects were in full control of the robots during menu-type interaction and the haptic devices were not decoupled during this time, which would be very uncommon during real interventions. Since resembling the posture of the end effectors usually takes the longest time, a command that was triggered with haptic gesturing would be significant faster. 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). In our case, a plausible and very simple menu structure of two layers was used, coupled with a foot pedal mechanism. Whether or not gesture-based input is equally effective for other set-up needs to be tested in future studies. Similarly, the user experience ratings should be interpreted with caution. Three of the four scales on the user experience questionnaire would favour 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. In particular, the possibility to intervene in the execution of (semi-) autonomous tasks in case of an emergency (or a misinterpreted command) needs to be investigated. Some actions may also require further user interaction, such as the selection of an appropriate piercing point. How this can be done, e.g., by task-specific visual servoing or by verbally describing the next step is, however, a complete open question. Some subjects

also had concerns about the feasibility of gesturing during a real intervention. Sweeping gestures might be a danger with regards to the small intra-operative dimensions in MIS. However, a decoupling of robots and input devices would be possible to perform the action in an virtual environment.

From a technical point of view, the recognition performance and robustness of the Hidden Markov Model can further be improved by fine-tuning. The linearity of a movement, as well as the curvature in a certain neighborhood of the trajectory could be considered in addition. If interaction with the environment or more complex primitive would be assumed during a gesture (e.g., pulling a thread), the measured force vector gives hints to environmental interaction. For the evaluation, all gestures were taught the system by an experienced user. This is not optimal with respect to two aspects: First, the demonstrations depend on the embodiment of the teacher and do not reflect the characteristics of individual users. This might influence the recognition rate of the HMM in a negative way. Second, the idea of intuitive and custom-made gestures is lost. The subjects had to mentally link an action with a gesture that might not be optimal in their understanding.

In conclusion, we see haptic gesturing as a potential, additional (but not replacing) input modality in robotic surgery. For short and easy movements (e.g. as in "measure distance"), where the surgical instruments can be controlled safe without being disconnected from the input devices, it offers a clear time-saving w.r.t menu interaction.

VI. ACKNOWLEDGMENTS

This work is supported by the German Research Foundation (DFG) within the Collaborative Research Center SFB 453 on "High-Fidelity Telepresence and Teleaction". The authors would also like to thank the German Heart Center (DHM) for the cooperation. Special thanks go to Jerome Haas and his medical colleagues for their support and advice.

REFERENCES

[1] G. Guthart and J. Salisbury, "The intuitive™telesurgery system: overview and application," *In proceedings of the IEEE International Conference on Robotics and Automation*, vol. 1, pp. 618–621, 2000.

[2] J. Abbott, P. Marayong, and A. Okamura, "Haptic virtual fixtures for robot-assisted manipulation," in *Robotics Research*. Springer Berlin / Heidelberg, 2007, vol. 28, pp. 49–64.

[3] H. Mayer, I. Nagy, A. Knoll, E. Braun, R. Bauernschmitt, and R. Lange, "Haptic feedback in a telepresence system for endoscopic heart surgery," *MIT PRESENCE: Teleoperators and Virtual Environments*, vol. 16, no. 5, pp. 459–470, 2007.

[4] P. Kazanzides, G. Fichtinger, G. Hager, A. Okamura, L. Whitcomb, and R. Taylor, "Surgical and interventional robotics - core concepts, technology, and design," *IEEE Robotics Automation Magazine*, vol. 15, no. 2, pp. 122 –130, jun. 2008.

[5] H. Mayer, D. Burschka, A. Knoll, E. Braun, R. Lange, and R. Bauernschmitt, "Human-machine skill transfer extended by a scaffolding framework," in *Proceedings of the IEEE International Conference on Robotics and Automation*, May 2008, pp. 2866–2871.

[6] C. Staub, T. Osa, A. Knoll, and R. Bauernschmitt, "Automation of tissue piercing using circular needles and vision guidance for computer aided laparoscopic surgery," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Anchorage, Alaska, USA, may 2010, pp. 4585–4590.

[7] J. Sackier and Y. Wang, "Robotically assisted laparoscopic surgery," *Surgical Endoscopy*, vol. 8, pp. 63–66, 1994.

[8] M. E. Allaf, S. V. Jackman, P. G. Schulam, J. A. Cadeddu, B. R. Lee, R. G. Moore, and L. R. Kavoussi, "Laparoscopic visual field," *Surgical Endoscopy*, vol. 12, pp. 1415–1418, 1998.

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

[10] H. Takahashi, T. Yonemura, N. Sugita, M. Mitsuishi, S. Sora, A. Morita, and R. Mochizuki, "Master manipulator with higher operability designed for micro neuro surgical system," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, may 2008, pp. 3902 –3907.

[11] A. Greer, P. Newhook, and G. Sutherland, "Human-machine interface for robotic surgery and stereotaxy," *IEEE/ASME Transactions on Mechatronics*, vol. 13, no. 3, pp. 355 –361, june 2008.

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

[13] G. Mylonas, K.-W. Kwok, A. Darzi, and G.-Z. Yang, "Gaze contingent motor channelling and haptic constraints for minimally invasive robotic surgery," in *Medical Image Computing and Computer-Assisted Intervention*. Springer Berlin / Heidelberg, 2008, vol. 5242, pp. 676–683.

[14] D. Noonan, G. Mylonas, A. Darzi, and G.-Z. Yang, "Gaze contingent articulated robot control for robot assisted minimally invasive surgery," in *In proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, sep. 2008, pp. 1186 –1191.

[15] D. Stoyanov, G. Mylonas, and G.-Z. Yang, "Gaze contingent 3d control for focused energy ablation in robotic assisted surgery," in *Medical Image Computing and Computer-Assisted Intervention*, D. Metaxas, L. Axel, G. Fichtinger, and G. Székely, Eds. Springer Berlin / Heidelberg, 2008, pp. 347–355.

[16] P. Pook and D. Ballard, "Deictic teleassistance," in *Intelligent Robots and Systems '94. 'Advanced Robotic Systems and the Real World', IROS '94. Proceedings of the IEEE/RSJ/GI International Conference on*, vol. 1, sep. 1994, pp. 245 –252.

[17] H. Mayer, I. Nagy, A. Knoll, E. Schirmbeck, and R. Bauernschmitt, "The Endo[PA]R system for minimally invasive robotic surgery," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sendai, Japan, September 2004, pp. 3637–3642.

[18] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 3, pp. 311 –324, may. 2007.

[19] J. Rosen, M. Solazzo, B. Hannaford, and M. Sinanan, "Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden markov model," *Computer Aided Surgery*, vol. 7, no. 1, pp. 49–61, 2002.

[20] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257 –286, feb. 1989.

[21] I. Guyon, P. Albrecht, Y. L. Cun, J. Denker, and W. Hubbard, "Design of a neural network character recognizer for a touch terminal," *Pattern Recognition*, vol. 24, no. 2, pp. 105 – 119, 1991.

[22] T. Stewart, "Usability oder user experience - what's the difference?" in *System Concepts*, 2008. [Online]. Available: http://www.system-concepts.com/articles/usability-articles/2008/usability-or-user-experience-whats-the-difference.html

[23] M. Kuniavsky, *Observing the user experience: A practitioner's guide to user resaerch*. Elsevier Science, 2003.

[24] M. Hassenzahl, M. Burmester, and F. Koller, "Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität," in *Mensch & Computer 2003: Interaktion in Bewegung*, G. Szwillus and J. Ziegler, Eds. Stuttgart: B. G. Teubner, 2003, pp. 187–196.