

Human-Computer Interfaces for Interaction with Surgical Tools in Robotic Surgery

C. Staub, S. Can, B. Jensen, and A. Knoll
Robotics and Embedded Systems
Technische Universität München
{staub|cans|jensen|knoll}@in.tum.de

S. Kohlbecher
Department of Neurology
University of Munich
skohlbecher@nefo.med.uni-muenchen.de

Abstract—The current trend toward robot assistants and the execution of autonomous tasks in minimally invasive surgery increases the operation complexity of telepresence systems. The available input channels are currently limited to traditional human-computer interfaces. We introduce two human-robot interfacing modalities that aim to make robotic surgery more intuitive. To reduce the surgeon’s mental load, gaze-contingent camera control is implemented. Eye tracking is performed by means of head worn tracking goggles. The tracking goggles are tightly integrated with a stereoscopic visualization system, based on the polarization method. The second technique supports scrub nurses during surgical tool interaction, e.g. tool exchange, via haptic gestures executed on the robot. Strain-gauges sensors installed at the instrument are used to detect hand tapping sequences, which trigger activation of specified commands.

I. INTRODUCTION

The technique of minimally invasive surgery (MIS) has pioneered its way into various surgical disciplines at the beginning of the 1980s. Especially patients benefited massively from the new method at the beginning, suffering less pain, less scarring and shorter hospitalization time. The introduction of telemanipulation systems, which were designed to overcome the fulcrum effect of endoscopic instruments and allowed Cartesian control, facilitated the work of surgeons. This intuitive user interface enabled the assistance of surgeons with the scaling of motions, tremor filtering, and a stereo vision interface at the master console. A remarkable example of this progress is the DaVinci™ robot [1]. More recently, researchers investigate in methods that aim to make robot-assisted systems more intelligent, in particular targeting the improvement of both patient safety and operating time. Both factors are interrelated with the system’s degree of automation: the (partial) autonomous execution of frequently performed and error-prone tasks, e.g. surgical knot-tying or piercing of tissue, have shown enormous potential to speed up procedures and to mentally relieve the surgeon during exhausting tasks [2]. Combined with virtual constraints [3], which might either provide haptic or audio-visual feedback, the surgeon still has full control over the system and can intervene if necessary.

Obviously, this growing amount of functionality reflects the demand for shared- and cooperative control, whether it is *human-* or *machine-*assisted. At the same time, new input modalities must facilitate the handling of these functionality to pave their way into clinical practice. Addressing this deficiency, a prototypic implementation of gesture-based

fine manipulation of the daVinci’s surgical tools using the Microsoft Kinect™ depth camera has been demonstrated in [4]. While this kind of interface allows two surgeons to control the robot simultaneously, the number of input options at the master console is still limited to the traditional haptic devices. This fact forces surgeons to interrupt the surgical workflow and to decouple the haptic input device from the surgical tool in order to gain control over the endoscopic camera. Gaze-contingent camera control has been proposed to replace this time-consuming procedure. However, existing systems are either integrated directly into the master console, anticipating head movements, or rely on remote trackers. The later configuration can be dicey in combination with stereo screens, which require the user to wear polarization glasses. The glasses used in our setup obscure the eye and introduce specular reflections. Passive tracking systems might not be able to detect the eye sufficiently through the polarization foil. Due to the emission of IR light, active trackers promote even more reflections on the glasses. Shutter systems black out the eyes alternately, making it impossible to perform remote tracking. The introduced tracking system combines both polarization foil and the eye tracker in one pair of glasses. The polarization foil sits between the viewer and the monitor. Therefore it does not disturb the view of the laterally mounted cameras onto the eye.

The topic of including OR personnel other than the surgeon performing the intervention into the aspects of human machine interaction was often underestimated in the past. For instance, scrub nurses frequently interact with the system to change surgical instrument and to supply medical consumables. To our knowledge no efforts have been made in developing new input methods for this purpose. We investigate a method that utilizes haptic gestures executed at the surgical tool. Scrub nurses can trigger system commands by performing unambiguous tapping sequences at the instrument tip. Therewith the robot itself becomes an input channel and no further devices are necessary.

II. ENDOSCOPE CONTROL STRATEGIES

In this section we will review methods, previously proposed in literature, to control endoscopes in minimally invasive surgery. The methods can be divided roughly into two groups. While some of the methods simply execute human input commands, given by the surgeon, the second group acts in a more intelligent or (semi-)

autonomous way, based on contextual information of the current surgical conditions. This information can be of very different nature.

Manual endoscope control by an endoscopic specialist is still a widespread approach. Often frequent repositioning of the endoscope is required to guarantee an optimal field of view. The surgeon gives verbal instructions to his assistant, which can possibly lead to confusion and debate.

In the case of robotic surgery, the physician typically has two input devices available, controlling the left and right surgical instrument respectively. In order to gain control over the endoscopic camera, the operator can actuate a clutch to disconnect the input devices from the manipulator and connect it to the camera. However, the control device needs to be relocated beforehand, in order to synchronize the posture with the end effector intended for action. This is necessary to maintain a reasonable hand-eye coordination. During the repositioning of the camera and the recovery of control over the instruments the procedure needs to be performed twice. An alternative, more time-saving option would be an additional endoscopic specialist who shares the real-time video image with the actual surgeon, but can use his own input device to control the camera.

Voice-activated control and head tracking, also known as “hands-free” methods, entered the operating room with the emergence of so-called *robot assistants*. Voice-activated control allows the surgeon to move the camera based on a limited and predefined number of voice commands (e.g., “left, right, start, stop”), while still handling the surgical tools. Already the famous *AESOP* manipulator [5], which was the first FDA approved system, implemented this technique, and was later adopted by the *ZEUS* system [6]. The surgeon must often wear a dedicated microphone to ensure an adequate voice quality for speech recognition systems. The *Freehand* system [7] works with head movements: head worn IR-emitting tracking markers are detected and their movement is interpreted as input commands. As the detector is placed above the monitor, which displays the endoscopic video stream, the surgeon must not avert his gaze from the carried out task.

The systems often come with additional foot pedals to activate and deactivate the control of the camera. Some studies even confirm foot pedals a faster and error-free operation over the other methods mentioned [8]. Another drawback is that only one movement can be triggered at a time (either horizontally, vertically, or zoom), prolonging and complicating the alignment.

Positioning based on instrument tracking allows endoscope control in a fully automated manner. The position of the instruments can be seen as a strong cue for a region of interest. A main difficulty of the method is the fault-tolerant and reliable detection of the instrument pose. The methods proposed in literature range from instruments, which are labeled with artificial color markers to markerless methods, which e.g., maximize the difference between foreground and background color alongside a geometrical shape [9], and

machine learning techniques that aim to learn and segment the instrument’s appearance from the background. A problem shared by most methods is the handling of occlusion (by tissue, body-liquids, or other surgical tools) and changing appearance of the instruments and background, e.g., caused by non-uniform, varying lighting conditions, smoke caused by electro-dissection, and organ movement.

Positioning based on workflow analysis: medical workflow analysis drew the attention of researchers in order to model and analyze medical procedures. Besides the application to documentation and benchmarking surgical interventions [10], autonomous and shared control methods rely on the knowledge gained, e.g. for assisted knot-tying. Regarding automated camera guidance, orientation data of the two surgical end-effectors as well as the camera, recorded during former interventions, can be used to predict the current phase of the intervention and to derive an optimal camera pose accordingly [11]. With respect to the extent of autonomy, this is probably the most “intelligent” of the suggested methods, but usually needs to be combined with the tracking of tools. However, the complexity of surgical trajectories represent a significant obstacle for a robust prediction [11]. To cope with varying surgical situations, the underlying knowledge base must be adequate in size, which is difficult to obtain.

Gaze contingent camera control: Noonon *et al.* introduced the gaze contingent control of an articulated robotic endoscope with a regular monitor and evaluated control techniques for joints selection [12]. As mentioned in the introduction, the combination of remote eye trackers with polarization screens might be troublesome. They also integrated a 3D eye tracker into the master console of the daVinci [13]. This implementation is, however, quite device specific and requires a fixed head position of the surgeon. Therewith it is not transferable to other setups.

III. SYSTEM DESIGN AND EYE-TRACKING INTEGRATION

We will now introduce the underlying hard- and software testbed.

A. Robotic System

The ARAMIS research platform (cf. Fig. 1) is a robotic system for minimally invasive surgery with a focus on autonomous functionality. Typical for telepresence, the system consists of a master console and a robotic slave part. The slave part comprises four Mitsubishi RV-6SL manipulators that are ceiling mounted on an aluminum gantry. A magnetic clutch mechanism allows to couple the DaVinci’s EndoWrist™ surgical instruments to the robot’s flange, while establishing all necessary mechanical and electrical connections. The master console, often referred to as the medical workstation, offers doctors the ability to operate the remote slave via two PHANTOM™ Premium 1.5 devices (Sensable Inc.). A customized handle snap-on, which is similar to handling a pair of tweezers, replaces the original switch in order to substitute the digital behavior of the micro-grippers at the distal end of the surgical tools with a continuous

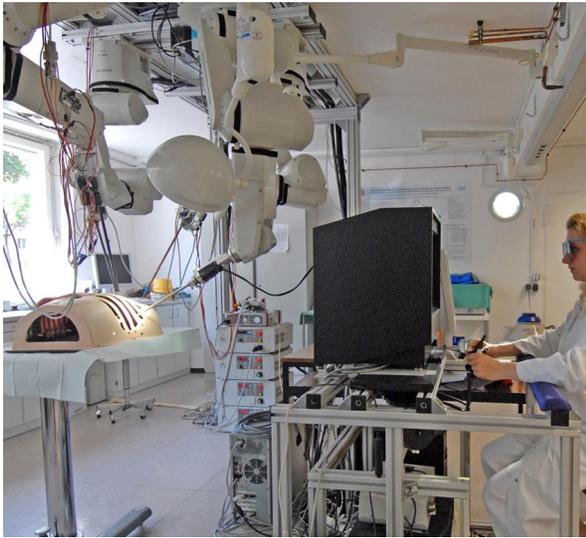


Fig. 1. **The ARAMIS system.** Four ceiling mounted manipulators are tele-operated via a master console. The robots can either be equipped with a surgical tool or an endoscopic stereo camera. For a high immersion, the surgeon is located in front of a stereo display and operates the system with haptic devices.

input option. The forefinger has to be fixed at a rocker, which is connected to a small DC motor with an integrated position sensor. Forces acting on the surgical instruments are measured by means of sensitive strain-gauges sensors and displayed to the operator via the force-feedback capabilities of the haptic devices. Usually, three of the manipulators can be equipped with surgical instruments, while the fourth arm carries an endoscopic stereo camera.

For efficient telemanipulation, it is crucial to get a good depth impression of the operating area. In order to allow for such a feature, our surgical workstation is equipped with a 3D screen. The optical system operates with a semi transparent mirror that displays for each eye the corresponding camera view. To get a depth impression, the user has to wear polarized goggles.

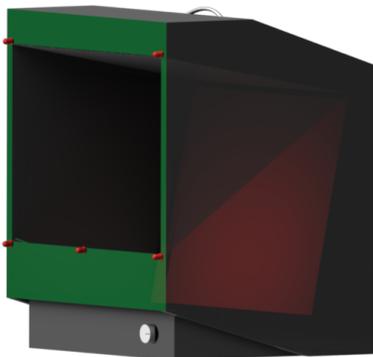


Fig. 2. **Stereo Display.** The five attached infrared LED's (red dots) form a plane (green) that needs to be calibrated to the actual view plane of the monitor (red). The stereo image is generated by fusing two images via a semi-transparent mirror.

B. Gaze Point Detection

We now describe how to determine the user's point of gaze with respect to the display. In order to maximize the signal to noise ratio, we chose to implement a head-mounted eye tracking solution. Such kind of eye trackers yield the gaze direction in head coordinates, hence head tracking is required to determine the intersection of the line of sight with a given display plane. We were able to adapt the solution presented in [14] to fit our needs in this application. In particular, our solution has the advantage of giving the surgeon a certain degree of mobility in front of the screen, instead of requiring a fixed head position at the master console. We directly integrate the polarization foil, necessary for the stereo display, into the tracking goggles, with which the amount of equipment remains the same for the surgeon. The detection of the gaze point involves three steps: (1) the detection of the pupil position, (2) head tracking, and (3) the projection of the gaze point onto the screen plane. Each single step is explained below.

1) *Eye Tracking:* The eye tracker is a monocular version of the one developed in the EyeSeeCam project [15] and consists of one infrared camera which is mounted laterally to a goggle frame. A so called hot mirror, which reflects infrared light, but allows visible light to pass through is installed such that the camera is able to see the eye. Due to its special characteristics, the mirror does not constrain the wearer's field of vision, as it appears to humans like a piece of glass. The eye is illuminated by infrared LEDs positioned in the goggle frame which allows for a clean signal independent of the ambient lighting conditions. The camera image is transmitted at a frame rate of 220 Hz via FireWire to a Mac, where image processing algorithms are used to detect the exact dark pupil position. To determine the relation between this pupil position and the gaze vector, the system has to be calibrated after having been put on. During the calibration procedure a diffraction grating is used to split up a laser beam mounted on the goggle between the eyes, thereby generating five reference points 8.5° left and right of, as well as above and below the primary position. The projected points need to be aligned with their counterparts on the screen. The procedure takes less than a minute. The parallax error emerging from the translational displacement between the eye ball and laser origins can be minimized by using a sufficiently large distance to the projection surface.

2) *Head Tracking:* A wide angle scene camera is used to determine the wearer's current head position. It is mounted on the goggles on the forehead facing away from the user and records markers in order to localise the display plane. To facilitate the detection of these markers, infrared LEDs have been used, which can be easily detected by attaching an IR filter to the scene camera. Because of the infrared mirror in front of the eyes, the marker LEDs do not produce corneal glints which could affect the pupil detection. A plane in 3D space is defined by three points, but since we are only able to observe a two-dimensional projection by the camera, a fourth point is needed to calculate the original position and

orientation [14]. Therefore we placed the four markers in each corner of the display. This allows us to calculate the display position with respect to the scene camera. As the LEDs (cf. Fig. 2, LEDs are colored red) are mounted on the front of the housing (green plane) of the stereo screen, but the actual image plane is located on the semi-transparent mirror (between the green and the red planes), we have to consider this transformation in addition. Since we know the goggle geometry we can use a transformation to convert these coordinates into a system defined by the calibration laser, which allows us to specify both screen position and gaze vector in the same coordinate system.

3) *Gaze On Plane*: Since the range of operation is typically within arm’s reach, the above-mentioned parallax error cannot be ignored and has to be compensated for. Hence, the eye ball position must be determined with respect to the calibration laser coordinate system. This can be achieved by having the user fixate two arbitrary calibration points on the display and measuring the corresponding gaze vectors from the eye tracker. Then, the two lines originating in each calibration point in the direction of the calculated gaze vectors intersect in the eye ball center. For this to work, the head must not be moved during the calibration procedure. The exact fixation point on the display can be calculated by intersecting the gaze vector originating at the eye ball with the display plane. These two-dimensional coordinates are then transmitted over network via a simple UDP protocol. On one hand our distributed system software uses the data to show an overlay of the gaze position on the screen, on the other hand, it is used to generate the endoscope trajectory.

C. Distributed Data Processing

The distributed architecture of our system’s software comes with the advantage of a highly flexible structure that quickly adapts to new scenarios. The framework organizes the software into *modules*. Each module is a software component that provides a certain system functionality, either on a software level or it integrates hardware respectively, and can be deployed and run as a separate process on any machine. Data as well as functionality can be shared with other modules over network. For instance, in the case of gaze contingent camera control, one module realizes the eye tracker integration, forwarding the data to modules for trajectory generation and visualization.

The eye tracker samples 220 fixation positions each second. To obtain a stable, jitterless fixation point, the values are smoothed by means of a recursive implementation of an exponential filter [16]. The observation period may therefore be chosen arbitrary long without the need of storing previous data. A smoothed value x_{n+1} of observation $n + 1$ can be written as a linear combination of the smoothed value of observation n and the data value z_{n+1} obtained in the new observation $n + 1$, whereas the observed data value is labeled with a time stamp t .

$$x_{n+1} = x_n + K_{n+1} (z_{n+1} - x_n) \quad (1)$$

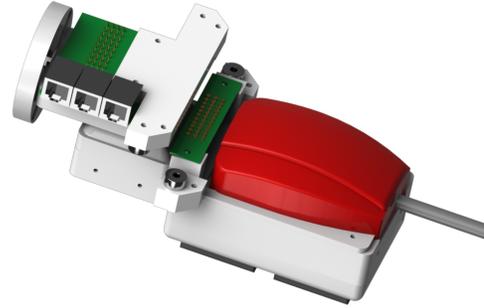


Fig. 3. **Hot-plug instrument** with drive box (red). All necessary electrical connections are established with spring contacts (golden pins on green circuit board). An integrated potentiometer allows a unique tool identification.

where K is a gain factor

$$K_{n+1} = \frac{K_n}{K_n + \exp(-(t_{n+1} - t_n)/\tau)} \quad (2)$$

and τ is the filter’s time scale. While a shorter time scale attributes to high-frequency noise, a long time scale does not reflect the current gaze point very accurate anymore, but rather gives a weighted gaze direction over time. The choice of τ is quite crucial with respect to the desired behavior of the endoscope control. When τ is chosen just large enough to smooth outliers, the endoscope control quickly adapts to new gaze positions and constantly guides the endoscope. A large value of τ , however, is robust against short-term changes in the viewing direction and the endoscope gradually adapts to the area of attention.

The processed data is consumed by a control algorithm, generating the robot’s trajectory. The algorithm orientates the camera so that the current fixation point gets aligned with the center of the endoscopic image. Insertion or retraction movements into the patient’s body are neglected for safety reasons. The velocity of the alignment process is calculated based on the distance between image center and fixation point. The velocity is reduced with decreasing distance, while keeping the trocar constraint at all times. The execution of movement commands is coupled with a foot pedal. Two modes can be chosen: Mode 1 interprets the gaze position as input command as long as the foot pedal is pressed. The movement is executed at a fairly high velocity. The second mode is activated after tapping the pedal once and deactivated after a second tap. During mode 2 the gaze position is constantly tracked, while τ is set to a large value, resulting in a slow trajectory adaption.

IV. HR-INTERFACE FOR TOOL EXCHANGE

Apart from an intuitive control of the surgical instruments via the master console, OR personnel is permanently in contact with the robotic system in order to change surgical tools and to supply consumables, such as suture material. Robotic scrub nurses, e.g., [17], [18], capable to identify and distinguish different surgical tools via machine vision have been developed to facilitate tool exchange in MIS.

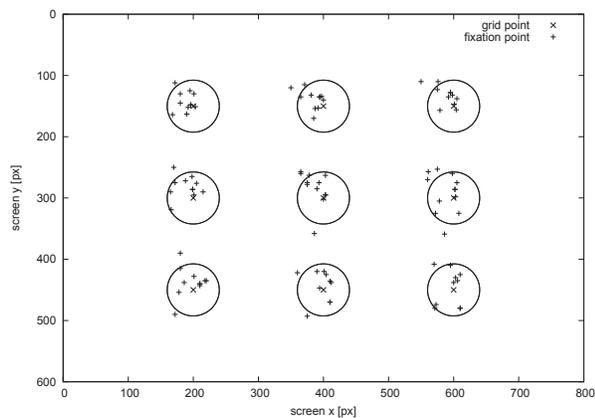


Fig. 4. Subject fixation validation grid on a screen with a resolution of $800 \times 600px$. The circles have a radius of $40px$. Subjects were looking straight at the monitor, at a distance of $30cm$ to the LED markers, and an additional $25cm$, caused by the stereo monitors geometry, to actual image plane. The overall distance between subject and image plane is thus approx. $45cm$.

Interaction with the staff is often realized by means of voice recognition. The tools are picked up and delivered to the surgeon using an electro-magnetic wrist appendage. With respect to fully automated robotic surgery Friedman *et al.* introduced the *TraumaPod* tool changer. The system handles 14 EndoWristTM tools of the daVinciTM robot. The instruments are arranged in a sterilizable carousel and can be identified using RFID transponders.

These tool changers put the focus on a fast and safe tool exchange, but preclude, or at least complicate, the integration of humans into the process. However, the responsibility of a scrub nurse is beyond the simple exchange of the instruments, e.g., nurses do also supply suture material. This task often requires both hands to hold the loose ends of the thread, as the thread needs to be tightened to properly pass it to the micro-gripper at the instrument tip. Therefore, input options which are operated by hand will not be appropriate. Foot pedals or voice control might complicate the procedure with additionally required hardware or restrict the mobility of the nurses.

We propose an approach that utilizes the force measurement capabilities of our surgical instruments. Already slight contacts can be recognized by the sensitive strain-gauges sensors and unique contact sequences, similar to a Morse code, can be identified. Predefined codes are used to trigger a certain system command. For instance, a scrub nurse can provide suture material, holding the material between thumb and index finger with both hands to tighten it, while triggering the closing of the micro gripper by the input of a “tapping sequence” at the instrument shaft using her middle finger. Fig. 6 illustrates an example sequence, where 3 peaks are detected within a certain time interval, uniquely representing this command. After passing the thread to the instrument, another sequence triggers the incision process of the instrument into the patient. If the instrument has been changed by the nurse, the system is automatically notified about the connected instrument type. All instruments are

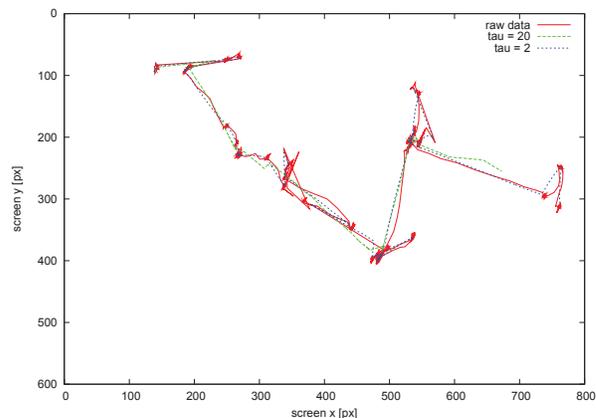


Fig. 5. Evolution of the fixation point trajectory on the screen over time for different values of τ . Variable τ governs the influence of outliers and the inclusion of only temporarily considered screen regions. Higher values result in more smoothing of the trajectory.

equipped with a potentiometer, set to a unique resistance value that uniquely identifies the tool. Calibration data, which compensates for the tool’s eccentricity, is loaded accordingly. All electrical connections are established with a hot-plug shoe, which is integrated into the magnetic coupling mechanism (cp. Fig. 3).

V. EXPERIMENTS

This section gives details on the accuracy of the eye tracking system in our particular setup and evaluates the generation of the trajectories for the endoscope control. 10 different subjects were instructed to fixate 9 points on the stereo screen at an overall distance of approx. $40cm$ to the image plane. The distance between subject and the frame of the screen, where the LED markers are mounted, is $30cm$. The distance between the monitors frame and the actual image plane is approx. $25cm$ (cf. Fig. 2). Please note that the distance varies from the upper to the lower edge of the screen, as the image plane is not parallel to the monitor frame. The stimulus points (cf. Fig. 4, (\times)) were arranged in image space with an image resolution of $800 \times 600px$. The intersections between the image plane and the right eye’s line of sight were plotted ($+$). The subjects looked straight at the monitor. The achieved averaged absolute accuracy of $22.5px$ is suitable for our application. The error range is indicated by the circles with a radius of $40px$. The better performance of the tracker on regular screens [14] can be attributed to the introduced transformation between LED markers and actual image plane. In addition our image plane is tilted.

The processing of the fixation points stream is crucial for the movement characteristics of the endoscope. The choice of the filter’s time scale, governed by variable τ (cf. Eqn. 2), influences the smoothing of the filter. Without any smoothing the control algorithm would directly interpret the fixation points as input command, yielding to a continuous endoscope movement and jitter (cf. Fig. 5, red trajectory). The larger τ is chosen, the more short-term changes of the fixation point are smoothed from the raw data. With a very large time scale chosen, the endoscope is commanded only to

regions permanently focused on by the surgeon. This allows us to easily adjust the control of the endoscope to a desired behaviour.

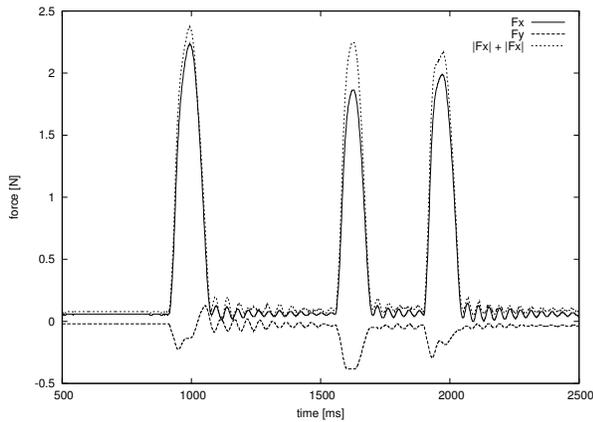


Fig. 6. **Tapping sequence:** A predefined sequence is interpreted as a certain command, e.g., closing the micro-gripper at the distal end of an instrument. As input, the sum of the absolute values of the two axis, which are perpendicular to the instrument shaft, is used.

In order to evaluate the haptic interface during a tool exchange we defined a sequence, consisting of 3 subsequent instrument contacts, triggered within a time interval of 2s and a maximum gap of 0.5s between each of the contacts. The contact has been considered as an intended input command, if the absolute force magnitude was greater than 1.6N. Other contacts were classified as accidental and not considered further. The sequence itself was classified by means of a state machine. Both force sensors, measuring the forces perpendicular to the instrument shaft, were used. Fig. 6 illustrates the sensor readings during the sequence.

VI. CONCLUSION AND OUTLOOK

We have presented a gaze contingent camera control that can be combined with polarized screens. Due to the head worn eye tracker the system can be integrated with both active and passive polarization glasses. The achieved precision is sufficient for the closed loop endoscope positioning and could probably be improved by precisely measuring the transformation between markers and image plane. To keep the system simple at this stage, we decided not to use the corneal glints from the illumination LEDs, as they are heavily dependent on the shape of the cornea. Taking these reflexions into account could make the system more robust to goggle slippage. An extension of the tracking system to two cameras, one for each eye, would improve the fixation accuracy. As a stereo tracker can provide 3D gaze information, the direct alignment of the camera with a certain point in space becomes possible, instead of minimizing the distance between image center and fixation point. If gaze control could also be an alternative to traditional menu-based navigation needs to be investigated in future. Haptic interaction between surgical tools and the OR staff was proposed as an alternative input method. Hand tapping sequences are used to trigger activation of system commands. This allows a seamless integration of tool change and the

delivery of surgical material into the surgical workflow. The application of a more sophisticated method to detect and learn a sequence, e.g., Hidden Markov Models, seems to be worthwhile.

REFERENCES

- [1] G. Guthart and J. Salisbury, "The intuitiveTMtelesurgery system: overview and application," *In proceedings of the IEEE International Conference on Robotics and Automation*, vol. 1, pp. 618–621, 2000.
- [2] 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*, 2008, pp. 2866–2871.
- [3] J. Abbott, P. Marayong, and A. Okamura, "Haptic virtual fixtures for robot-assisted manipulation," in *Robotics Research*, ser. Springer Tracts in Advanced Robotics, S. Thrun, R. Brooks, and H. Durrant-Whyte, Eds. Springer Berlin / Heidelberg, 2007, vol. 28, pp. 49–64.
- [4] K. Guerin, B. Vagvolgyi, A. Deguet, C. Chen, D. Yuh, and R. Kumar, "Reachin: A modular vision based interface for teleoperation," *Workshop of the International Conference of Medical Image Computing and Computer Assisted Intervention*, 08 2010.
- [5] J. Sackier and Y. Wang, "Robotically assisted laparoscopic surgery," *Surgical Endoscopy*, vol. 8, pp. 63–66, 1994.
- [6] H. Reichensperner, R. J. Damiano, M. Mack, D. H. Boehm, H. Gulbins, C. Detter, B. Meiser, R. Ellgass, and B. Reichart, "Use of the voice-controlled and computer-assisted surgical system zeus for endoscopic coronary artery bypass grafting," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 118, no. 1, pp. 11 – 16, 1999.
- [7] D. Sharma, C. Brown, C. Kouriefs, H. Sood, P. Grange, and H. Patel, "Initial experience with the freehand robotic camera holder in laparoscopic urology," *Journal of Endourology*, vol. 23, no. 1, p. A249, 2009.
- [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] C. Staub, C. Lenz, G. Panin, A. Knoll, and R. Bauernschmitt, "Contour-based surgical instrument tracking supported by kinematic prediction," in *Proceedings of the IEEE/RAS International Conference on Biomedical Robotics and Biomechatronics*, Tokyo, Japan, 2010, pp. 746–752.
- [10] H. C. Lin, I. Shafran, D. Yuh, and G. D. Hager, "Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions," September 2006.
- [11] O. Weede, H. Monnich, B. Muller, and H. Worn, "An intelligent and autonomous endoscopic guidance system for minimally invasive surgery," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2011, pp. 5762 –5768.
- [12] 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/RAS International Conference on Intelligent Robots and Systems*, 2008, pp. 1186 –1191.
- [13] D. Noonan, G. Mylonas, J. Shang, C. Payne, A. Darzi, and Y. G., "Gaze contingent control for an articulated mechatronic laparoscope," in *Proceedings of the IEEE/RAS International Conference on Biomedical Robotics and Biomechatronics*, 2010, pp. 759–764.
- [14] S. Kohlbecher, K. Bartl, S. Bardins, and E. Schneider, "Low-latency combined eye and head tracking system for teleoperating a robotic head in real-time," in *ETRA*. ACM, 2010, pp. 117–120.
- [15] E. Schneider, T. Villgratner, J. Vockeroth, K. Bartl, S. Kohlbecher, S. Bardins, H. Ulbrich, and T. Brandt, "Eyeseecam: An eye movement-driven head camera for the examination of natural visual exploration," *Annals of the New York Academy of Sciences*, vol. 1164, no. 1, pp. 461–467, 2009.
- [16] P. D. Stroud, "A recursive exponential filter for time-sensitive data," Los Alamos National Laboratory, Tech. Rep. LAUR-99-5573, 1999.
- [17] E. Carpintero, C. Pérez, R. Morales, N. García, A. Candela, and J. Azorí, "Development of a robotic scrub nurse for the operating theatre," in *Proceedings of the IEEE RAS & EMBS Conference on Biomedical Robotics and Biomechatronics*, 2010, pp. 504 –509.
- [18] M. Treat, S. Amory, P. Downey, and D. Taliaferro, "Initial clinical experience with a partly autonomous robotic surgical instrument server," *Surgical Endoscopy*, vol. 20, pp. 1310–1314, 2006.