Assessing system operation skills in robotic surgery trainees

Rajesh Kumar[1], **Amod Jog**[1], **Anand Malpani**[1], **Balazs Vagvolgyi**[1], **David Yuh**[4], **Hiep Nguyen**[2], **Gregory Hager**[1], and **Chi Chiung Grace Chen**[1]

[1]Department of Computer Science, Johns Hopkins University, 3400 N Charles Street, Baltimore, MD 21218

[2]Department of Urology, Children's Hospital, Boston, MA

[3]Department of Obstetrics and Gynecology, Johns Hopkins Hospital, Baltimore, MD

[4]Division of Cardiac Surgery, Johns Hopkins Hospital, Baltimore

Abstract

Background—With increased use of robotic surgery in specialties including urology, development of training methods has also intensified. However, current approaches lack the ability of discriminating *operational* and *surgical* skills.

Methods—An automated recording system was used to longitudinally (monthly) acquire instrument motion/telemetry and video and for 4 basic surgical skills -- suturing, manipulation, transection, and dissection. Statistical models were then developed to discriminate the human-machine skill differences between practicing expert surgeons and trainees.

Results—Data from 6 trainee and 2 experts was analyzed to validate the first ever statistical models of operational skills, and demonstrate classification with very high accuracy (91.7% for masters, and 88.2% for camera motion) and sensitivity.

Conclusions—We report on our longitudinal study aimed at tracking robotic surgery trainees to proficiency, and methods capable of objectively assessing operational and technical skills that would be used in assessing trainee progress at the participating institutions.

Keywords

Robotic surgery training; objective skill assessment; automated skill assessment

Introduction

Minimally invasive surgery has seen a rapid transformation over the last two decades with the introduction and approval of robotic surgery systems [1,2]. Robotic surgery is the dominant treatment for localized prostate cancer [5] with approximately 75,000 robotic procedures being performed in the USA every year, up from only 18,000 procedures in 2005[3]. Advances in tools and techniques have also established minimally invasive surgery as a standard of care in other specialties including, cardiothoracic surgery[4], gynecology[5], and urology [6]. Over 1700 da Vinci surgical systems (Intuitive Surgical, Inc., CA) were in surgical use by late 2010, up from 1395 systems in 2009 [2].

However, robotic surgical training in academic medical centers still uses the Halstedean apprenticeship model where a mentor's subjective assessment forms the main evaluation.

Corresponding Author: Rajesh Kumar, PhD, Department of Computer Science, Johns Hopkins University, Hackerman Hall 117, 3400 North Charles St., Baltimore MD 21218, USA, rajesh@jhu.edu Phone: 410-516-6708 Fax: 206-260-3189.

This subjective assessment during the course of training relies on analysis of recorded training videos, direct procedure observations, and discussions between the mentor and the trainee. It is challenging for both the mentor and the trainee to identify specific limitations and objective recommendations, or skills where the trainee may have achieved significant expertise. The need for objective assessment in robotic surgery training has long been evident [7], along with the need for development and standardization of the environment in which the skills are imparted [8–10]. Consequently, the development of training methods for robotic surgery has also intensified [11–14] and training and assessment have become an active area of research [15–17].

Robotic surgery skills consist of separate human-machine interface operational skills (operating the masters and console interface), and surgical technique and procedural skills. However, current robotic training approaches so far lack both the ability and separate criteria for assessing and tracking man-machine interactions.

We are addressing these limitations systematically in a National Institutes of Health funded multi-center study of robotic surgery training where we aim to track robotic surgery trainees to human surgery proficiency, subject only to continued availability and access to the subjects. Here we focus primarily on discriminating proficiency in human-machine interface skills that are new to all trainees, and with a complex machine such as the da Vinci surgical system, are also a major reason for steep learning curves in robotic surgeons. Below, we establish that there are significant skill differences between experts who are comfortable using the system, and trainees of varying experience who are still learning to use the system. Such a test has many important uses, for example, to identify when a trainee has reached a competent skill level in system operation skill to graduate to more complex training task that emphasizes the skills where a trainee is still deficient. This customization, which would require extensive oversight, and subjective and structured assessment by a mentor, is not currently possible.

Materials and Methods

We utilize the Intuitive Surgical da Vinci surgical system [1] as our experimental platform. The da Vinci consists of a surgeon's stereo viewing console with a pair of master manipulators and their control systems, a patient cart with a set of patient side manipulators, and an additional cart for the stereo endoscopic vision equipment. Wristed instruments are attached and removed from the patient side manipulators as needed for surgery. The slave surgical instrument manipulators are teleoperated with the master manipulators, by associating an instrument using the foot pedals and switches on surgeons console.

In addition to providing an immersive teleoperation environment, the da Vinci system also streams all motions and events via an Ethernet application programming interface (API) [18]. The API provides transparent and immediate access to instrument motion and system configuration information. We record all 334 variables containing Cartesian position and velocity, joint angles, joint velocities, torque data at 50Hz, as well as events for robotic arms, console buttons, and foot pedals generated during a training procedure in an unhindered manner. Such recording previously needed devices could not be easily incorporated into the surgical and training infrastructure [20] without impacting surgical or training workflow.

By contrast, our data collection system (Figure 1) is designed to collect data from a surgical training procedure without any modification of the procedure workflow, and using our system, without on-site supervision. We also record high quality synchronized video from the stereoscopic camera at real-time frame rates (30Hz). The archiving workstation does not

need any special training to operate and can be left connected without affecting the system operation.

This experimental data, de-identified at the source by a unique subject identifier (approximately 20GB for each training session), is stored in a secure archive according to our Johns Hopkins Institutional Review Board protocol. The unique identifiers permit longitudinal analysis, such as computation of learning curves.

Our training protocol consists of a set of minimally invasive surgery benchmarking tasks (Figure 2). This includes assessment of manipulation skills, and benchmarking of suturing, transection, and dissection skills performed approximately monthly on training pods (The Chamberlain Group, Inc.) commonly used for robotic surgery training. These tasks also form part of the Intuitive Surgical robotic surgery training practicum [19].

The manipulation task involves moving rubber rings to pegs placed around the entire robotic workspace. Subjects also perform suturing (3 interrupted sutures along an I-defect using 8--10cm length of Vicryl 3-0 suture), transect a pattern on a transection pod, and mobilize an artificial vessel buried in a gel phantom using blunt dissection. In addition to the motion and video data above, we also record the trainee's practice and training hours between these benchmarking sessions via a survey.

Recruitment and Status

Our subjects are robotic surgery residents and fellows from four institutions - Johns Hopkins, Children's Hospital, Boston, Stanford/VA Hospitals, and University of Pennsylvania. Subjects (residents, fellows, practicing clinicians) had varying amounts of surgical training, but typically no prior robotic surgery experience. We aim to track the continued progress of our trainees limited only by their availability at the participating institutions, and if possible, to human surgery proficiency upon completion of their training. Practicing clinicians (experts) from these institutions were recruited to provide ground truth data for computing proficiency levels of performance measures.

Objective Assessment

We are investigating automated assessment for differentiating experts and non-experts in addition to the manual structured expert assessment. Previous similar studies (e.g. [21–22]) have used preliminary measures to identify skill with an emphasis only on comparing users of different skill levels to the experts. The motion data from the da Vinci API has also been previously used to classify skill using statistical machine learning methods. These studies [11,16–17] have primarily focused on recognizing the surgical task being performed.

The motion data from the API is contains 334 total dimensions acquired at 50 Hz. We used appropriately sampled API data for the relevant instruments or master handles for further statistical analysis. This sampled API data is then used for training and validation of separate models of human-machine operation skills using Support Vector Machines (SVM). A SVM uses a kernel function and an optimization algorithm to finds a hyperplane with optimum separation between the two classes. Given ground truth labels for the trials, sensitivity and accuracy of the classifiers can be directly computed as performance measures.

Our expert collaborators have assigned OSATS scores [23] to these task executions to verify facility with robotic operation. The OSATS global rating scale consists of six skill-related variables in operative procedures that are typically graded on a 5 point Likert-like scale. Experts (OSATS > 13) were well separated from other trainees (OSATS <10) at the start of the training. This provides the ground truth for training and validating the statistical methods above. We investigated the following human-machine interactions – a) managing the master

workspace - due to motion scaling the master handle workspace in robotic surgery needs frequent adjustment, b) the camera field of view adjustment, and c) maintaining the instruments safely in the field of view of the camera.

The master workspace is adjusted by pressing a foot-pedal (clutch) which disconnects the instruments from the master manipulators, allowing the master handles to be repositioned. The current generation of the da Vinci surgical system (the Si model) also contains switches on the master handles for independent hand adjustment. For adjusting the camera field of view, the camera clutch, a different foot pedal allows constrained motion of the masters as if they were attached to edges of the screen. As before, upon completion of a camera move, the teleoperation relationships are recomputed and teleoperation mode is restored.

Proficiency with these short duration dexterous system configuration tasks (often less than a second long) requires significant familiarity with system interfaces. Note also that the motion during these events relates only to operational skills on the platform. Lastly, these tasks relate mostly to simple and short motions, back to a comfortable hand position or to a more favorable view of the camera. During these events, the instruments are disconnected from the masters, and the user is focusing only on operation skills. Therefore, they are ideal for assessing operative skills. Our 4 skill benchmarking laboratories above take approximately 30 minutes to complete and many adjustments are needed during each task providing multiple observations of these events for each training session.

Figure 3 shows camera motion during the manipulation training task, and Figure 2 shows the workspace adjustment (clutch on/off) highlighted on the motion trajectory of the master handles. We treat camera on/off, and clutch on/off as four separate events, and focus on master motion during and near these events. We generated fixed size feature vectors from the API motion data corresponding to these events and use supervised classification for assessment of each operation skill. Fixed size high-dimension feature vectors permit a range of supervised classification methods for estimating skill levels, including Support Vector Machines (SVM) [24]. SVMs have also been used in the art with motion data. For example, with feature vectors containing force and torque data, a hybrid HMM-SVM model was used to segment a teleoperated peg in-hole task into 4 states in [25] with 84% on line and 100% for off line segmentation accuracy. For our expert vs. trainee classification, a binary SVM classifier is directly applicable. During validation of such a trained classifier, held- out observations may be directly classified to obtain performance measures as:

$$accuracy = true\ positives/total\ observations\ classified \tag{2}$$

$$precision = true\ positives/total\ positive\ classifications \tag{3}$$

$$recall = true\ positives/(true\ positives + false\ negatives) \tag{4}$$

For each of the operational skills above, these performance measures allows us to assess when a trainee has achieved proficiency by training a classifier with the experts as the positive class and evaluating the trainee's performance with the trained classifier. We validated these methods using longitudinal data from subjects in our study including 2 expert users, and 6 trainees of varying skill levels for master workspace adjustment and 8 trainees for camera motion.

For master clutch operation, we sampled 2 seconds (100 observations) or more of motion (134 observations) around each master clutch event, and then post-processed these segments

to extract a feature vector containing Cartesian position (6 DOF), Cartesian velocity (6 DOF), and the gripper angle (1 DOF) for both left and right hand master motion. When concatenated, this gives a fixed length (2600, and 3484 dimensions) feature vector for each observed event. Each observation is assigned the skill class (expert or novice) of the user performing this task. Over the 8 users in the experimental dataset, we obtain a total of 3998 events. Further dimension reduction can be performed given this high dimensions of the feature vector, however, here we just used the high-dimension data. We then trained binary SVM classifiers with polynomial kernel functions with the feature vectors extracted above using the SVM implementation in the Bioinformatics toolbox of MATLAB (Mathworks Inc, Natick, MA) on a windows workstation (Core i5, 4GB RAM). A subset of the experimental data above was used as a training dataset for training the classifier, and the remaining for validation.

Camera motions are even shorter than workspace adjustment motions, therefore, we segment only 0.5 seconds (25 observations) of motion around each event and then post-process these segments as before to extract a feature vector containing Cartesian position (6 DOF), Cartesian velocity (3 DOF), for left hand, right hand, and the camera. Hand orientation is constrained by the system during camera motion, so orientation velocities were not included in the feature vector. When concatenated, this gives a fixed length (675 dimensions) feature vector for each observed event. This feature vector is also assigned the skill class (expert or trainee) of the user performing this task. As before, we use the concatenated feature vectors without dimensionality reduction.

Results

Master Workspace Adjustment

The experimental dataset of 3998 events contains 352 expert events, 3646 novice clutch events. Initial experiments used smaller portions of the data for training the SVM classifier to select experimental parameters. Starting with 50% held-out data, the classifier was trained on the other half of the data (1999 samples), followed by training on 90% of the data. We also performed a 10-fold cross-validation on the experimental data set to assess statistical significance. The SVM classifier was tested by performing 10 different tests, each using approximately 10% of randomly selected data for testing, and the remaining for training. These tests were performed both for 100, and 134 observations. Using a quadratic kernel, the classifier correctly classified 91.75% of the observations using 50% of the data, rising to 92% with training on 90% of the data. Higher order polynomials did not significantly improve these results.

With the larger number of observations (134), the classifier performance improved to 93.20% with training on 50% of the data. Increasing this training dataset size to 70% and 90% of the data resulted in precision rates of 94.99% and 95.70%. As durations of the workspace adjustment events is limited, a longer time period would result in portions of the task manipulation with instruments being included, and therefore motion vectors longer than 134 were not tested in these experiments.

Camera Manipulation Skills

The experimental dataset of 483 separate sampled camera motion trajectories contains 63 expert samples. Leave-one-out validation with 10% of the data for showed promising results (87% accuracy and precision, 100% recall). Varying the held-out data size did not significantly change the results. The classifier correctly classified 88.16% of the data, with a 100% recall during the 10-fold cross-validation. As before, higher order polynomial kernels did not significantly alter the classifier performance. Camera motions are very short events -

typically less shorter than a second in duration. A longer time period results in portions of the task manipulation with instruments being included, and therefore motion vectors longer than 0.5 seconds were not tested.

Unsafe Motion and Collisions

Analysis of unsafe instrument motions (e.g. those not visible to the trainee) is ongoing. Figure 4 shows the instrument motions during the training task (Blue–left, Green–right instrument). The red portions represent where an instrument was not visible in the camera field of view. An adaptive threshold already distinguishes trainees from experts, and statistical analysis of the motion data is currently in progress.

Discussion

We describe our novel unsupervised data collection infrastructure for robotic surgery training the da Vinci surgical system, and the an initial evaluation of construct validity of system skill classification based on the collected data. This infrastructure is in use for capturing training data at four different training centers (Johns Hopkins, University of Pennsylvania, Children's Hospital, Boston and Stanford).

Training issues have received prior attention of many researchers, industry, and various surgical education societies and several efforts are now underway both using simulators [16–17], and on data obtained from real surgical systems [11,21–22,26–27]. Literature also contains several attempts at designing training tasks and measuring skills in robotic surgery. Such research typically only uses inanimate and bench top experiments simulating basic minimally invasive skills (e.g. suturing) and was designed keeping the corresponding analysis in view, often with special equipment not commercially available. Prior assessment has also used elementary measures [16–17,22].

In [22], Judkins et al report on the comparison of 5 experts and 5 novices using ad-hoc inanimate tasks using trends in pre and post-training evaluation with aggregate measures such as task completion time, total instrument distance, and velocity. They note that after only 10 trials trainees and experts performed the experimental tasks were indistinguishable using their elementary metrics. 10 trials is clearly not adequate training for basic surgical skills, even for trained users.

Robotic surgery also uses a complex man-machine interface and the complexity of this interface contributes significantly towards the long learning curves even for laparoscopically trained surgeons. These man-machine interface skill effects can be completely captured using the telemetry available from the robotic system, and with appropriate tasks and measures, separate learning curves can be identified. This work presents the first results for such operational skills - master workspace adjustment, and camera manipulation - to assess when a user may be classified as being comfortable with field of view reconfiguration. Over data collected from 8 real surgical trainees, we achieve >90% accuracy in classifying whether a system interaction was performed by an expert or a user of another skill level. With 10 real surgical experts and trainees, we also achieve >87% accuracy, and 100% sensitivity in separating users classes based on camera motions. We also identify other key operational skills e.g. collisions, or unsafe motion. Analysis of our data for these additional skills is ongoing.

These tools are being integrated into system skill training assessment that will allow us to emphasize training experiments to a trainee's needs. In related publications [26], we have reported learning curves based on the aggregate measures considered above, the structured assessment by experts, and their correlation. In a parallel study [27], we are also extending

and applying this assessment for simulation based training. Work in preparation for publication also includes a consensus document describing the role of technology in robotic surgery assessment based on a workshop held at World Robotic Symposium, 2011, and reports of assessment from residency based robotic surgery curricula for urology and otolaryngology. We are also working to assign skill levels beyond the two class separation performed here, and to develop multi-class classification methods. Additional validation, and utilization of validated methods in proficiency based training is also currently being investigated

Acknowledgments

References

1. Guthart, G.; Salisbury, J. The intuitive telesurgery system: overview and application; IEEE International Conference on Robotics and Automation; San Francisco, CA. USA: 2000. p. 618-621.

2. Frequently asked questions the intuitive surgical inc. website. [Online]. Available: http://www.intuitivesurgical.com/products/faq/index.aspx.

3. Shuford, M. Proc. Baylor University Medical Center. Vol. vol. 20. 2007. Robotically assisted laparoscopic radical prostatectomy: a brief review of outcomes; p. 354-356.

4. Rodriguez E, Randolph Chitwood W. Outcomes in robotic cardiac surgery. Journal of Robotic Surgery. 2007; vol. 1(no. 1):19–23.

5. Boggess J. Robotic surgery in gynecologic oncology: evolution of a new surgical paradigm. Journal of Robotic Surgery. 2007; vol. 1(no. 1):31–37.

6. Aron M, Koenig P, Kaouk J, Nguyen M, Desai M, Gill I. Robotic and laparoscopic partial nephrectomy: a matched-pair comparison from a high-volume centre. BJU International. 2008; vol. 102:86–92. [PubMed: 18336600]

7. Moorthy K, Munz Y, Sarker S, Darzi A. Objective assessment of technical skills in surgery. British Medical Journal. 2003; vol. 327(no. 7422):1032–1037. [PubMed: 14593041]

8. Bann S, Datta V, Khan M, Ridgway P, Darzi A. Attitudes towards skills examinations for basic surgical trainees. International Journal of Clinical Practice. 2005; vol. 59(no. 1):107–113. [PubMed: 15707474]

9. Szalay D, MacRae H, Regehr G, Reznick R. Using operative outcome to assess technical skill. The American Journal of Surgery. 2000; vol. 180(no. 3):234–237.

10. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative bench station examination. The American Journal of Surgery. 1997; vol. 173(no. 3):226–230.

11. Lin H, Shafran I, Yuh D, Hager G. Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions. Computer Aided Surgery. 2006; vol. 11(no. 5):220–230. [PubMed: 17127647]

12. Ro, CY.; Toumpoulis, IK.; Ashton, RC., Jr; Imielinska, C.; Jebara, T.; Shin, SH.; Zipkin, JD.; McGinty, JJ.; Todd, GJ.; Derose, JJ, Jr.. Studies in Health Technology and Informatics, Proc. MMVR. Vol. vol. 111. 2005. A novel drill set for the enhancement and assessment of robotic surgical performance; p. 418-421.

13. Sarle R, Tewari A, Shrivastava A, Peabody J, Menon M. Surgical robotics and laparoscopic training drills. Journal of Endourology. 2004; vol. 18(no. 1):63–67. [PubMed: 15006057]

14. Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. Journal of the American College of Surgeons. 2001; vol. 193(no. 5):479–485. [PubMed: 11708503]

15. Moorthy K, Munz Y, Sarker S, Darzi A. Objective assessment of technical skills in surgery. British Medical Journal. 2003; vol. 327(no. 7422):1032–1037. [PubMed: 14593041]

16. Kenney PA, Wszolek MF, Gould JJ, Libertino JA, Moinzadeh A. Face, Content, and Construct Validity of dV-Trainer, a Novel Virtual Reality Simulator for Robotic Surgery. Urology. 2009

17. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. British Journal of Surgery. 2004; vol. 91(no. 2):146–150. [PubMed: 14760660]

18. DiMaio SL, Hasser C. The da vinci research interface, miccai workshop on systems and architectures for computer assisted interventions. MIDAS journal. 2008 [Online]. Available: http://hdl.handle.net/10380/1464.

19. Rosen J, Solazzo M, Hannaford B, Sinanan M. Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden Markov model. Journal of Image Guided Surgery. 2002; vol. 7(no. 1):49–61.

20. Curet M, Mohr C. The intuitive surgical robotic surgery training practicum. Intuitive surgical, inc., Training document. 2008

21. Narazaki K, Oleynikov D, Stergiou N. Objective assessment of proficiency with bimanual inanimate tasks in robotic laparoscopy. Journal of Laparoendoscopic & Advanced Surgical Techniques. 2007; vol. 17(no. 1):47–52. [PubMed: 17362179]

22. Judkins T, Oleynikov D, Stergiou N. Objective evaluation of expert and novice performance during robotic surgical training tasks. Surgical Endoscopy. 2009; vol. 23(no. 3):590–597. [PubMed: 18443870]

23. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M. Objective structured assessment of technical skill (OSATS) for surgical residents. British Journal of Surgery. 1997; vol. 84(no. 2):273–278. [PubMed: 9052454]

24. Burges C. A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery. 1998; vol. 2(no. 2):121–167.

25. Castellani, A.; Botturi, D.; Bicego, M.; Fiorini, P. Hybrid HMM/SVM model for the analysis and segmentation of teleoperation tasks; Proc. IEEE International Conference on Robotics and Automation; 2004. p. 2918-2923.

26. Yuh, D.; Jog, A.; Kumar, R. Automated skill assessment for robotic surgical training; 47th Annual Meeting of the Society of Thoracic Surgeons; 2011. (abstract/poster).

27. Jog, A.; Itkowitz, B.; Liu, M.; DiMaio, S.; Curet, M.; Kumar, R. Towards integrating task information in skills assessment for dexterous tasks in surgery and simulation; International Conference on Robotics and Automation (ICRA); 2011.
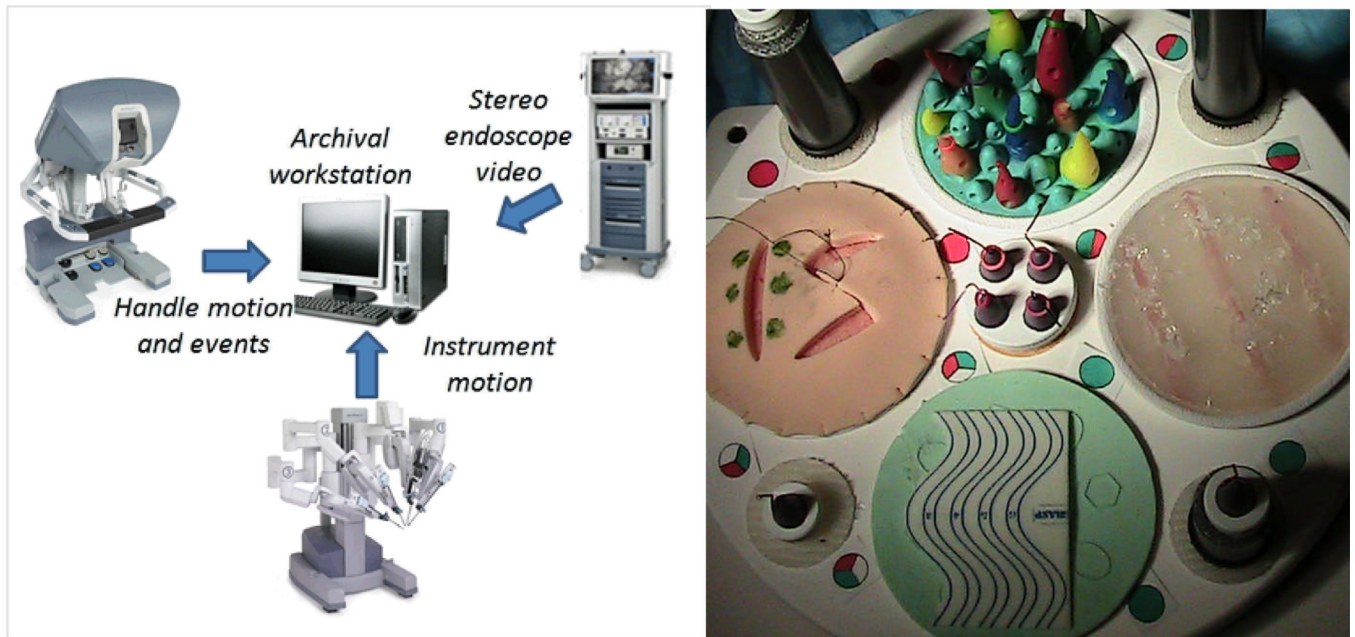
Figure 1.
Information flow for the JHU/VISR archival system for the da Vinci system(left), and the benchmarking task pod (right). A demonstration pod (not used for benchmarking), the dissection pod, transection pod and the suturing pod (clockwise, respectively). The posts for the manipulation task are in the center and on the periphery.
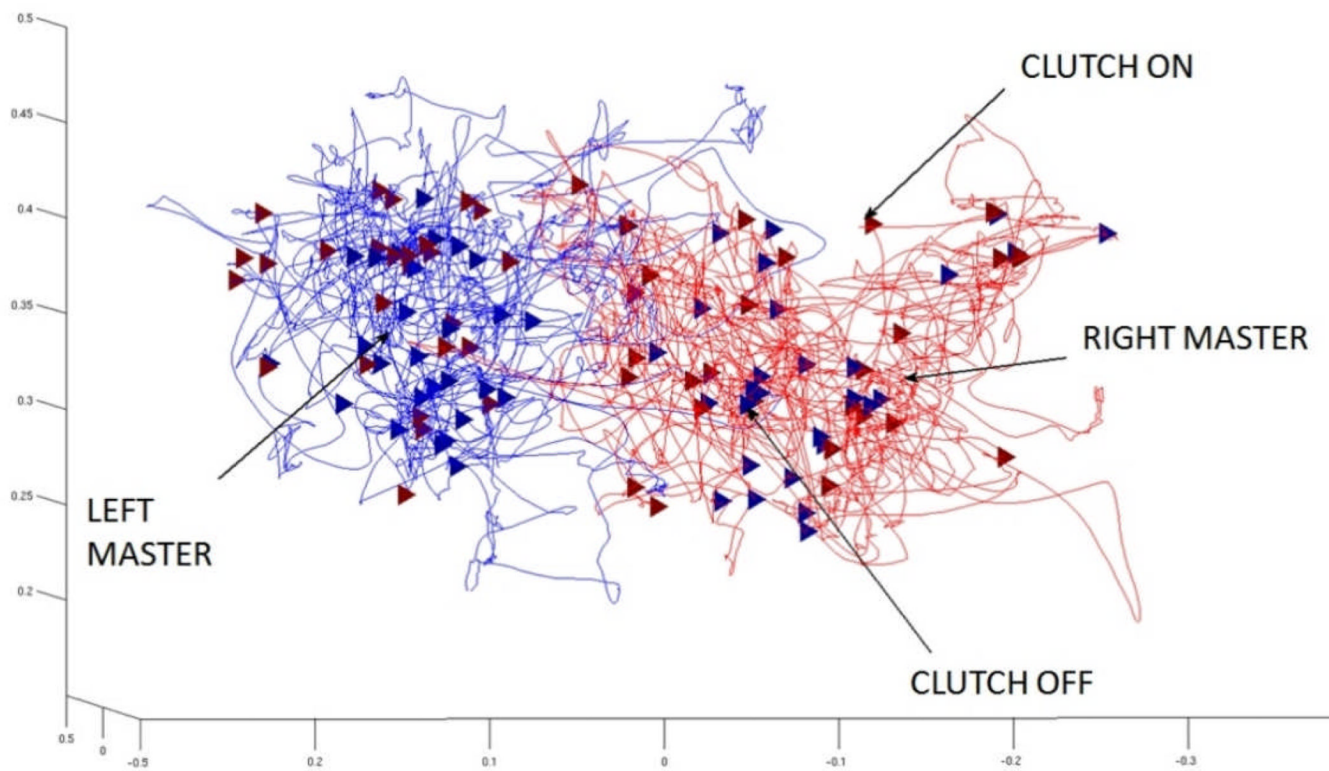
Figure 2.
Master motion during a manipulation task with an overlay of master clutch events. The blue triangles represent start of a workspace adjustment (clutch pedal pressed), and the red triangles the end of the reconfiguration (clutch pedal depressed).
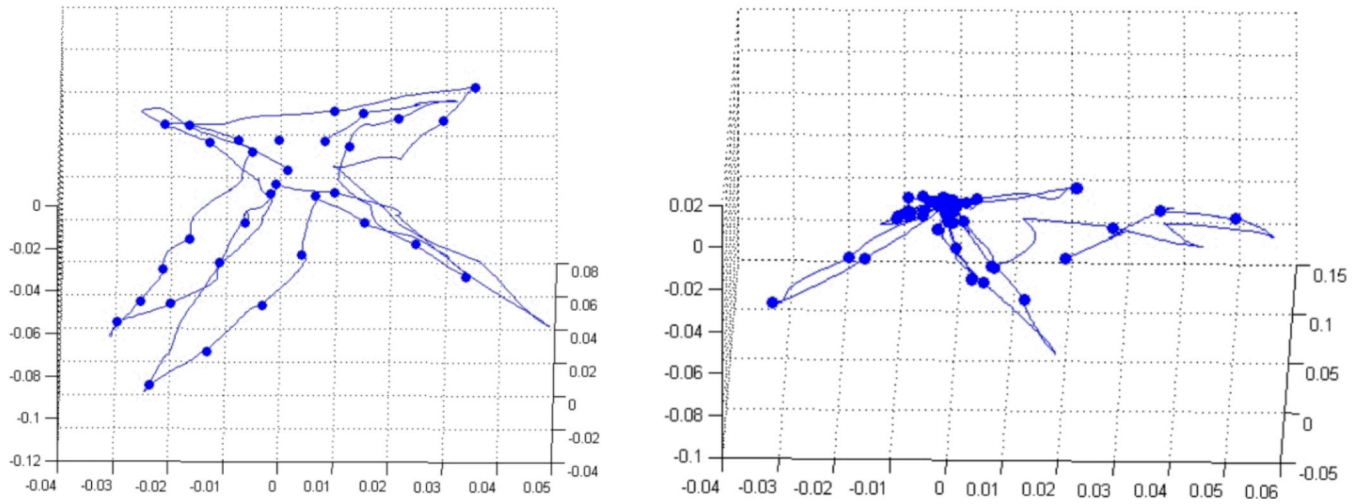
Figure 3.
Expert (left) and trainee (right) endoscopic camera Cartesian trajectories during the manipulation task. Expert manipulate the field of view to keep the instruments visible at all times, while novices camera use is less structured. The points represent start of a camera motion.
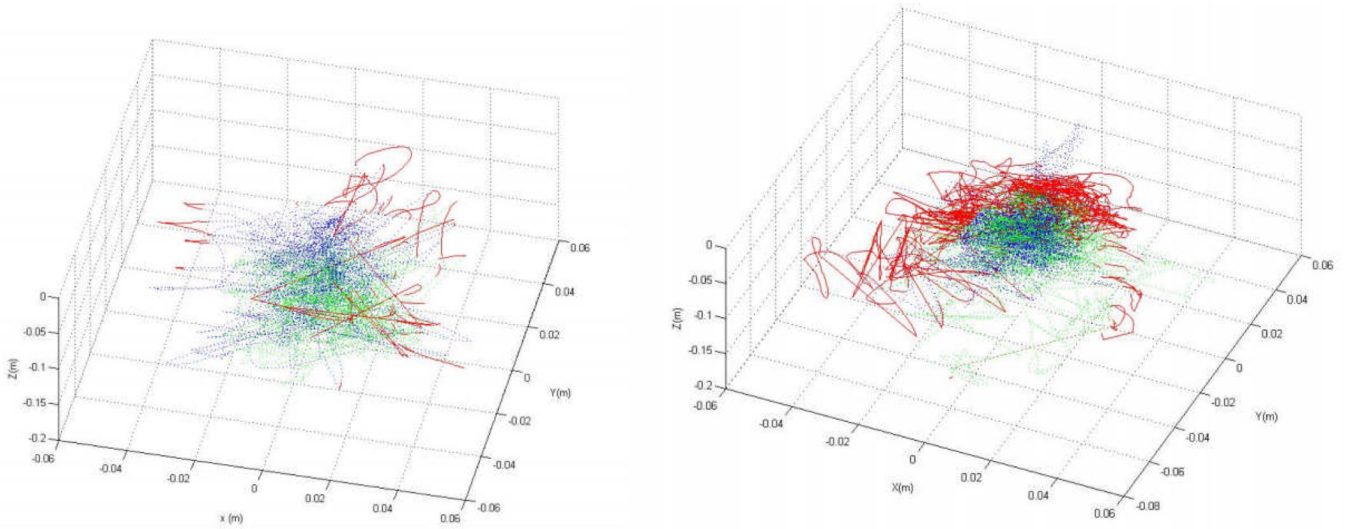
Figure 4.
Expert (left) and trainee (right) instrument Cartesian trajectories for left (blue) and right (green) instruments during a training task. The red portions represent where an instrument was manipulated unsafely outside the field of view of the endoscopic camera.