JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS): A Surgical Activity Dataset for Human Motion Modeling

Yixin Gao[1], S. Swaroop Vedula[1], Carol E. Reiley[1], Narges Ahmidi[1],
Balakrishnan Varadarajan[2], Henry C. Lin[3], Lingling Tao[1], Luca Zappella[4],
Benjamín Béjar[5], David D. Yuh[6], Chi Chiung Grace Chen[7], René Vidal[1],
Sanjeev Khudanpur[1] and Gregory D. Hager[1]

[1] Johns Hopkins University, Baltimore, MD 21210
[2] Google Inc., Mountain View, CA 94043
[3] Intuitive Surgical, Inc., Sunnyvale, CA 94086
[4] Metaio GmbH, 80335 Munich, Germany
[5] Swiss Federal Institute of Technology in Lausanne, 1015 Lausanne, Switzerland
[6] Yale University School of Medicine, New Haven, CT 06520
[7] Johns Hopkins Medical Institutions, Baltimore, MD 21224

Abstract. Dexterous surgical activity is of interest to many researchers in human motion modeling. In this paper, we describe a dataset of surgical activities and release it for public use. The dataset was captured using the *da Vinci* Surgical System and consists of kinematic and video from eight surgeons with different levels of skill performing five repetitions of three elementary surgical tasks on a bench-top model. The tasks, which include suturing, knot-tying and needle-passing, are standard components of most surgical skills training curricula. In addition to kinematic and video data captured from the *da Vinci* Surgical System, we are also releasing manual annotations of surgical gestures (atomic activity segments), surgical skill using global rating scores, a standardized cross-validation experimental setup, and a C++/Matlab toolkits for analyzing surgical gestures using hidden Markov models and using linear dynamical systems. We refer to the dataset as the JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS) to indicate the collaboration between Johns Hopkins University (JHU) and Intuitive Surgical Inc. (ISI), Sunnyvale, CA, on collecting these data.

1 Introduction

Studying dexterous human motion is important for at least three reasons. First, insights on how humans acquire dexterity in motion can be applied to facilitate skill acquisition. Second, skill may be objectively assessed using automated technology. Third, dexterous human motion may be partially or completely automated. Partial automation involves human machine collaboration where humans and robots perform separate parts of the task, for example [8].

Surgery involves dexterous human motion. The eventual goal for studying surgical motion is to improve safety and effectiveness of surgical patient care.

Poor surgical technical skill has been shown to be associated with higher post-surgical patient complications including death [3]. Surgical technical errors were the most common reason for post-surgical complications including re-operation and re-admission [9]. Thus, improving how surgeons acquire technical skill can positively impact safety and effectiveness of surgical patient care.

1.1 The Language of Surgery project

We believe that surgical motion is analogous to human language because it is a composition of elementary activities that are sequentially performed with certain constraints. Consequently, surgical motion can be modeled using techniques that have successfully been applied for analyzing human language and speech. We define the *Language of Surgery* as a systematic description of surgical activities or proceedings in terms of constituents and rules of composition. Based upon [6], the language of surgical motion involves describing specific actions that surgeons perform with their instruments or hands to achieve an intended surgical goal.

We study surgical activity as an example of dexterous human motion within the Language of Surgery project at the Johns Hopkins University. The overall goals for the project are to establish archives of surgical motion datasets, procedures and protocols to curate and securely store and share the datasets, develop and evaluate models to analyze surgical motion data, develop applications that use our models for teaching and assessing skillful motion to surgical trainees, and conduct research towards human machine collaboration in surgery.

Surgical activity data may be considered to encompass several types of variables related to human activity during surgery such as surgical tool motion (kinematics), video, log of events happening within and beyond the surgical field, and other variables such as surgeon's posture, speech, or manual annotations.

The objective for this paper is to release and describe a dataset we compiled within one of our studies on skilled human motion where surgeons performed elementary tasks on a bench-top model in the laboratory using the *da Vinci* Surgical System [5] (dVSS; Intuitive Surgical, Inc., Sunnyvale, CA). The surgical tasks included in this dataset are typically part of a surgical skills training curriculum. We refer to the dataset being released as the "JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS)" to indicate that this dataset was collected through a collaborative project between the Johns Hopkins University (JHU) and Intuitive Surgical Inc. (ISI).

1.2 The dVSS and its research interface

Within the Language of Surgery project, several studies have been conducted where we captured surgical activity data using the Application Programming Interface (API) of the dVSS. The dVSS is a tele-robotic surgical system that provides surgeons with enhanced dexterity, precision and control. The dVSS has been widely used to perform minimally invasive procedures in urology, gynecology, general surgery, and cardiothoracic surgery [5]. The dVSS is comprised of a master-side console with two master tool manipulators (MTMs) that are operated by the surgeon, a patient-side robot with three patient-side manipulators

(PSMs) and a camera arm with a stereo endoscope. During surgery, the surgeon tele-operates the PSMs by manipulating the MTMs.

The research interface (da Vinci API) [4] of the dVSS offers a convenient tool for data capture in real-time via an ethernet API. Through a collaborative agreement with Intuitive Surgical, Inc., we used the research interface of the dVSS to collect motion data from the MTMs and PSMs, as well as the camera pose, video data, and system events and converted them into an accessible format, as detailed in the next section.

2   Dataset

2.1   Surgical tasks

The JIGSAWS includes data on three elementary surgical tasks performed by study subjects (surgeons) on bench-top models. All three tasks (or their variants) are typically part of surgical skills training curricula. The three tasks include:

– **Suturing (SU)**: The subject picks up needle, proceeds to the incision (designated as a vertical line on the bench-top model), and passes the needle through the "tissue", entering at the dot marked on one side of the incision and exiting at the corresponding dot marked on the other side of the incision. After the first needle pass, the subject extracts the needle out of the tissue, passes it to the right hand and repeats the needle pass three more times.
– **Knot-Tying (KT)**: The subject picks up one end of a suture tied to a flexible tube attached at its ends to the surface of the bench-top model, and ties a single loop knot.
– **Needle-Passing (NP)**: The subject picks up the needle (in some cases not captured in the video) and passes it through four small metal hoops from right to left. The hoops are attached at a small height above the surface of the bench-top model.

The subjects were not allowed to move the camera or apply the clutch while performing the tasks. Figure 1 shows snapshots of the three surgical tasks.
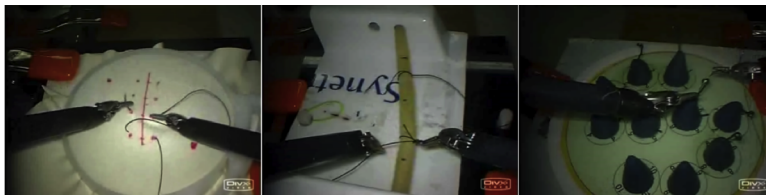
Fig. 1. Snapshots of the three surgical tasks in the JIGSAWS (from left to right): suturing, knot-tying, needle-passing.

2.2   Subjects

The JIGSAWS includes data from eight subjects, indexed as {"B", "C", "D", "E", "F", "H", "I"} with varying robotic surgical experience. Subjects D and E

reported having more than 100 hours, subjects B, G, H, and I reported having fewer than 10 hours, and subjects C and F reported between 10 and 100 hours of robotic surgical experience. All subjects were reportedly right-handed.

2.3   Task repetitions

All subjects repeated each surgical task five times. We refer to each of the five instances of a study task performed by a subject as a "trial". Each trial is on average about two minutes in duration for all three surgical tasks. We assigned each trial a unique identifier in the form of "Task_UidRep". For example, "Knot_Tying_B001" in the dataset files indicates the first repetition of subject "B" for the knot tying task. The JIGSAWS consists of 39 trials of SU, 36 trials of KT, and 28 trials of NP. The data for the remaining trials (1 for SU, 12 for NP, and 4 for KT) are unusable because of corrupted data recordings.

2.4   Data description

Kinematic data The JIGSAWS consists of three components: kinematic data, video data, and manual annotations. We captured the kinematic data from the dVSS using its API at 30 Hz. The left and right MTMs, and the first and second PSMs (PSM1 and PSM2, also referred as the right and left PSMs in this dataset), are included in the dataset. The motion of each manipulator was described by a local frame attached at the far end of the manipulator using 19 kinematic variables, therefore there are 76 dimensional data to describe the kinematics for all four manipulators listed above. The 19 kinematic variables for each manipulator include Cartesian positions (3 variables, denoted by xyz), a rotation matrix (9 variables, denoted by R), linear velocities (3 variables, denoted by $x'y'z'$), angular velocities (3 variables, denoted by $\alpha'\beta'\gamma'$, where $\alpha\beta\gamma$ are Euler angles), and a gripper angle (denoted by θ). All kinematic variables are represented within a common coordinate system. Table 1 describes the details of the variables included in the kinematic dataset. The kinematic data for the MTMs, PSMs, and the video data were synchronized with the same sampling rate.

Table 1. Kinematic data variables.

Column indices	Number of variables	Description of variables
1-3	3	Left MTM tool tip position (xyz)
4-12	9	Left MTM tool tip rotation matrix (R)
13-15	3	Left MTM tool tip linear velocity ($x'y'z'$)
16-18	3	Left MTM tool tip rotational velocity ($\alpha'\beta'\gamma'$)
19	1	Left MTM gripper angle velocity (θ)
20-38	19	Right MTM kinematics
39-41	3	PSM1 tool tip position (xyz)
42-50	9	PSM1 tool tip rotation matrix (R)
51-53	3	PSM1 tool tip linear velocity ($x'y'z'$)
54-56	3	PSM1 tool tip rotational velocity ($\alpha'\beta'\gamma'$)
57	1	PSM1 gripper angle velocity (θ)
58-76	19	PSM2 kinematics

Video data We captured stereo video from both endoscopic cameras of the dVSS at 30 Hz and at a resolution of 640 x 480. The video and kinematic data were synchronized such that each video frame corresponds to a kinematic data frame captured at the same instant of time. The videos in the dataset being released are saved as AVI files encoded in four character code (FOURCC) format with the DX50 codec. The video files named "capture1" and "capture2" were recorded from the left and right endoscopic cameras, respectively. Figure 2 shows a snapshot of corresponding left and right images for a single frame. The dataset does not include calibration parameters for the two endoscopic cameras.

Fig. 2. Screenshots of the corresponding left and right images for a single frame in the suturing task.

2.5 Manual annotations

Surgical activity annotation A key feature of the JIGSAWS is the manually annotated ground-truth for atomic surgical activity segments called "gestures" or "surgemes" [6,11]. A surgical gesture is defined as an atomic unit of intentional surgical activity resulting in a perceivable and meaningful outcome. We specified a common vocabulary comprised of 15 elements as detailed in table 2, to describe gestures for all three tasks in the dataset through consultation with an experienced cardiac surgeon with an established robotic surgical practice.

Table 2. Gesture vocabulary

Gesture index	Gesture description
G1	Reaching for needle with right hand
G2	Positioning needle
G3	Pushing needle through tissue
G4	Transferring needle from left to right
G5	Moving to center with needle in grip
G6	Pulling suture with left hand
G7	Pulling suture with right hand
G8	Orienting needle
G9	Using right hand to help tighten suture
G10	Loosening more suture
G11	Dropping suture at end and moving to end points
G12	Reaching for needle with left hand
G13	Making C loop around right hand
G14	Reaching for suture with right hand
G15	Pulling suture with both hands.

One individual annotated the data for surgical gestures by watching the videos in consultation with the surgeon. Each annotation includes the name of the gesture, and the start and end frames in the video. The video frames correspond to the frames in the kinematic data (one-to-one mapping) because we synchronized them using a common global time. All frames for each trial have been assigned a gesture label except a couple of frames at the end of a trial after the task was finished. Because we used a common vocabulary of gestures for all three surgical tasks, not all gestures in the vocabulary are used to annotate each surgical task. Although some gestures may be observed in more than one task, the background environment differs across tasks. For example, the "Pushing needle through tissue (G3)" gesture may be observed both in SU and NP tasks. But the needle is passed through an artificial tissue in SU and through a metal hoop in NP.

Surgical skill annotation The JIGSAWS also includes manual annotation for surgical technical skill. A gynecologic surgeon with extensive robotic and laparoscopic surgical experience watched each video and assigned a global rating score (GRS) using a modified objective structured assessments of technical skills (OSATS) approach [7]. The OSATS approach is a standard and structured method to assess surgical technical skills. We modified the original OSATS approach to exclude items that are not applicable to the surgical tasks we studied, for example, use of assistants. The modified GRS provided as part of the JIGSAWS is a sum of scores on six elements, each scored on a Likert scale of 1 through 5. The six elements are described in Table 3. The GRS represents a measure of technical skill over the entire trial. The annotating surgeon was masked to the identity of the subject who performed the surgical task.

3   Additional utilities

We are releasing two utilities along with the JIGSAWS for others interested in studying surgical activity or skill with these data:

1. An experimental setup that can be used to evaluate automatic surgical gesture recognition and skill assessment methods. The experimental setup includes two cross-validation schemes to account for the structure of the dataset:
   – **Leave-one-supertrial-out (LOSO)**: Supertrial i is defined as the set consisting of the i-th trial from all subjects for a given surgical task. In the LOSO setup for cross-validation, we created five folds with each fold comprising of data from one of the five supertrials. The LOSO setup can be used to evaluate the robustness of a method by leaving out the i-th repetition for all subjects.
   – **Leave-one-user-out (LOUO)**: In the LOUO setup for cross-validation, we created eight folds, each one consisting of data from one the eight subjects. The LOUO setup can be used to evaluate the robustness of a method when a subject is not previously seen in the training data.

Table 3. Elements of modified global rating score

Element	Rating scale
Respect for tissue	1-Frequently used unnecessary force on tissue; 3-Careful tissue handling but occasionally caused inadvertent damage; 5-Consistent appropriate tissue handling;
Suture/needle handling	1-Awkward and unsure with repeated entanglement and poor knot tying; 3-Majority of knots placed correctly with appropriate tension; 5-Excellent suture control
Time and motion	1-Made unnecessary moves; 3-Efficient time/motion but some unnecessary moves; 5-Clear economy of movement and maximum efficiency
Flow of operation	1-Frequently interrupted flow to discuss the next move; 3-Demonstrated some forward planning and reasonable procedure progression; 5-Obviously planned course of operation with efficient transitions between moves;
Overall performance	1-Very poor; 3-Competent; 5-Clearly Superior;
Quality of final product	1-Very poor; 3-Competent; 5-Clearly Superior;

2. A C++/Matlab toolkit that can be used to analyze the kinematic data or video data. The toolkit contains three main tools:

   – A grammar-constrained hidden Markov model (HMM) to analyze the kinematic data: This tool implements an HMM framework with the states modeled as Gaussian mixture models, as described in [17], where each gesture can be modeled using one elementary HMM and a trial is modeled as a composite HMM by concatenating the constituent gesture-specific elementary HMMs. Grammatical constraints can be applied on decoding to simplify the search.

   – A sparse-representation based HMM to analyze the kinematic data: This tool implements an HMM with the observations modeled as a sparse linear combination of atoms from a dictionary, as described in [14]. In contrast to the grammar-constrained HMM tool, here the gestures are modeled as the states in an HMM, and each trial can be viewed as an instance from an HMM.

   – A multiple kernel learning framework with linear dynamical system (LDS) and bag-of-features (BoF) to analyze the video data: This tool implements the three methods described in [2,18]: 1) using an LDS to model a video clip corresponding to a surgical gesture, where the observations are the video frames; 2) a BoF approach where a dictionary of visual words is learned and each image is represented by a histogram over the

dictionary; 3) a multiple kernel learning framework to optimally combine the LDS and BoF approaches.

4 Logistics

4.1 Accessing the dataset

The JIGSAWS can be downloaded without a fee from the Language of Surgery website: *http://cirl.lcsr.jhu.edu/jigsaws*. Access to the dataset requires a free registration, including provision of a valid email address so that we can inform investigators about any issues or corrections to the dataset.

4.2 Data organization

The JIGSAWS is available for download as zip files. Users are free to choose among tasks and data types, including kinematics or videos, together with transcriptions for surgical gestures annotations and a meta_file of self-proclaimed skill and overall OSATS score for each tasks. In addition, users can optionally download the experimental setup with standardized cross-validation folds we used in our analyses, and a C++/Matlab toolkit using this experimental setup for surgical gesture modeling. Partitioned downloads are enabled.

5 Prior work with the JIGSAWS

The JIGSAWS has been used for several research studies with two major areas of focus - surgical activity recognition and skill assessment. Techniques related to speech recognition, control theory, and computer vision have been developed and applied in the research studies using this dataset. We worked with various features extracted from the dataset, for example, features composed of a subset raw kinematics or converted kinematics, features extracted from videos, or features combining information from both kinematics and video, to classify (assign class label with known segment boundaries), or recognize (assign class label with unknown segment boundaries) surgical activities, and to evaluate dexterous skill. Below, we briefly summarize the prior work with the JIGSAWS.

In an early study, [12] performed gesture classification using a 3-state elementary HMM to model each gesture in a trial after applying linear discriminant analysis (LDA) to the continuous kinematic data for dimensionality reduction. Later on, [11] applied an HMM on discrete spectrum features extracted from a short time Fourier transform of the velocity features in the kinematic data with the goal of evaluating surgical skill in tasks and sub-tasks. In [10] studied the relation between subtasks and skill using the same features. [17] accomplished a surgical recognition task (jointly identifying gesture boundaries and assigning gesture labels) by applying an HMM to features derived from modeling the kinematic data (after an optional LDA) using Gaussian mixture models (GMMs). The topology for HMMs in [17] was entirely data-derived. These methods were enhanced in [16] through the development of various statistical models

and machine learning algorithms including factor-analyzed HMMs, switching linear dynamical systems (S-LDS) in supervised, semi-supervised, and unsupervised contexts. Other research conducted using kinematic data includes applying Gaussian mixture regression to reproduce expert motion trajectories as described in [13]. [14] combined sparse dictionary learning with HMMs to recognize surgical gestures and assess skill. Using only three-dimensional Cartesian positions of the PSM tool-tips, [1] discretized a Frenet frame representation of the data to descriptively encode tool motion trajectories and applied a string motif-based method to classify surgical gestures.

Computer vision techniques have also been applied to model surgical activity using the dataset. [6] describes methods to detect surgical gestures. [2] describe applications of three different video-based approaches to classify surgical gestures: LDS for modeling video clips corresponding to gestures, BoF approach on spatio-temporal features extracted from videos, and multiple kernel learning to combine the LDS and BoF approaches. [18] extended previous research to exploit both kinematic and video data to classify surgical gestures. Finally, [15] applied Markov-semi-Markov conditional random fields to features extracted from both kinematic and video data for surgical gesture recognition.

6 Conclusion

We have described details for a dataset named JIGSAWS on dexterous surgical motion performed using the *da Vinci* Surgical System in a training laboratory. In addition to the data directly captured from the surgical robot, we also included manual annotations for surgical gestures and skill. Our dataset will be of interest to investigators studying dexterous human motion.

7 Acknowledgment

References

1. N. Ahmidi, Y. Gao, B. Béjar, S. S. Vedula, S. Khudanpur, R. Vidal, and G. D. Hager. String motif-based description of tool motion for detecting skill and gestures in robotic surgery. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, Nagoya, Japan, 2013. Springer.
2. B. Béjar, L. Zappella, and R. Vidal. Surgical gesture classification from video data. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, page 34–41, Nice, France, 2012. Springer.

3. J. D. Birkmeyer, J. F. Finks, A. O'Reilly, M. Oerline, A. M. Carlin, A. R. Nunn, J. Dimick, M. Banerjee, and N. J. Birkmeyer. Surgical skill and complication rates after bariatric surgery. *New England Journal of Medicine*, 369(15):1434–1442, 2013.

4. S. DiMaio and C. Hasser. The *da Vinci* research interface. *The MIDAS Journal - Systems and Architectures for Computer Assisted Interventions (MICCAI 2008 Workshop)*, 2008.

5. G. Guthart and J. Salisbury, J. The Intuitive™telesurgery system: Overview and application. In *IEEE International Conference on Robotics and Automation*, volume 1, pages 618–621 vol.1, 2000.

6. H. C. Lin. *Structure in Surgical Motion*. dissertation, Johns Hopkins University, 2010.

7. J. Martin, G. Regehr, R. Reznick, H. MacRae, J. Murnaghan, C. Hutchison, and M. Brown. Objective structured assessment of technical skill (OSATS) for surgical residents. *British Journal of Surgery*, 84:273–278, 1997.

8. N. Padoy and G. Hager. Human-machine collaborative surgery using learned models. In *IEEE International Conference on Robotics and Automation*, pages 5285–5292, May 2011.

9. S. E. Regenbogen, C. C. Greenberg, D. M. Studdert, S. R. Lipsitz, M. J. Zinner, and A. A. Gawande. Patterns of technical error among surgical malpractice claims: an analysis of strategies to prevent injury to surgical patients. *Ann. Surg.*, 246(5):705–711, Nov 2007.

10. C. E. Reiley and G. D. Hager. Decomposition of robotic surgical tasks: An analysis of subtasks and their correlation to skill. In *Modeling and Monitoring of Computer Assisted Interventions (M2CAI) - MICCAI Workshop*, 2009.

11. C. E. Reiley and G. D. Hager. Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pages 435–442. Elsevier, 2009.

12. C. E. Reiley, H. C. Lin, B. Varadarajan, S. Khudanpur, D. D. Yuh, and G. D. Hager. Automatic recognition of surgical motions using statistical modeling for capturing variability. In *Medicine Meets Virtual Reality*, volume 132, pages 396–401, 2008.

13. C. E. Reiley, E. Plaku, and G. D. Hager. Motion generation of robotic surgical tasks: Learning from expert demonstrations. In *Engineering in Medicine and Biology*. Elsevier, 2010.

14. L. Tao, E. Elhamifar, S. Khudanpur, G. D. Hager, and R. Vidal. Sparse hidden markov models for surgical gesture classification and skill evaluation. In *Information Processing in Computer-Assisted Interventions*, volume 7330, pages 167–177. Springer Berlin Heidelberg, 2012.

15. L. Tao, L. Zappella, G. D. Hager, and R. Vidal. Surgical gesture segmentation and recognition. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, Nagoya, Japan, 2013. Springer.

16. B. Varadarajan. *Learning and Inference Algorithms for Dynamic System Models of Dextrous Motion*. dissertation, Johns Hopkins University, 2011.

17. B. Varadarajan, C. E. Reiley, H. C. Lin, S. Khudanpur, and G. D. Hager. Data-derived models for segmentation with application to surgical assessment and training. In *Medical Image Computing and Computer-Assisted Intervention –MICCAI 2009*, pages 426–434. Elsevier, 2009.

18. L. Zappella, B. Béjar, G. D. Hager, and R. Vidal. Surgical gesture classification from video and kinematic data. *Medical Image Analysis*, 17:732 – 745, 2013.