

Towards Data-Driven Hierarchical Surgical Skill Analysis

Bin Li¹, Bèrènice Mettler¹, and Timothy M. Kowalewski²

(1) Department of Aerospace Engineering and Mechanics and (2) Department of Mechanical Engineering, University of Minnesota, Minneapolis, MN, 55455, USA
{lix1778, mettler, timk}.umn.edu

Abstract. This paper evaluates methods of hierarchical skill analysis developed in aerospace to the problem of surgical skill assessment and modeling. The analysis employs tool motion data of Fundamental of Laparoscopic Skills (FLS) tasks collected from clinicians of various skill levels at three different clinical teaching hospitals in the United States. Outcomes are evaluated based on their ability to provide relevant information about the underlying processes across the entire system hierarchy including control, guidance and planning.

1 Introduction

Over 32,000 deaths and \$9B in losses are annually attributed to avoidable surgical errors [1], highlighting the need to ensure quality by mandating proficiency benchmarks in standardized training and credentialing [2] of surgical trainees. Modern procedural trainees and graduates prove inadequately prepared and in need of additional training [3,4]. Yet this is not viable with the faculty-intensive and time-demanding, subjective methods in use today, particularly given the steady influx of ever-changing technologies into the operating room. Automated, objective assessments of skill are needed. To ultimately address such a broad issue, simple performance scores on tasks such as task time or path length will not suffice, the entire, complex spectrum of human skill must be treated.

1.1 Surgical skill analysis

Prior work [5,6] has made considerable advances in task segmentation and skill classification for surgical contexts, particularly [7,8,9,10,11]. These approaches tend to focus on a specific subtask or modality (e.g., robotics vs. manual laparoscopy). While they succeed in providing valuable metrics to discriminate skill or procedural context, they do not extend directly to hierarchical human skill constructs like perception, planning, and cognition that are ultimately vital to this area [12]. No comprehensive framework exists that can successfully tie these many disparate attributes. Universal metrics of skill, proposed in [13], provided an early approach to such task-agnostic metrics and produced datasets to ultimately evaluate such metrics, however, it yielded little progress towards

such a goal. We herein introduce a different approach based on invariants. The approach makes it possible to delineate between key processes of the hierarchical control and sensory system. We provide preliminary evaluation for laparoscopy.

1.2 Alternative Skill Model and Analysis Framework

More recently researchers have grown interested in a more formal dynamics and control based theory of perception and action. The notable examples include Warren’s control theory of dynamics of action and perception [14,15]. More comprehensive models that capture the closed-loop interaction have been proposed in the aerospace field in the form of multi-loop models. The loops are organized hierarchically starting with the low-level attitude stabilization, to tracking, and ultimately goal directed maneuvering [16,17,18]. These models and efforts suggest that comprehensive skill evaluation requires accessing and using information across the different levels of the system hierarchy; not just the performance “outputs” but also the various internal processes and if possible should encompass the “inputs” to the system such as perception and attention.

The multi-loop framework provides a rigorous, deterministic basis for measurement, evaluation and modeling of skills. Figure 1 shows the primary loops in a multi-loop model [19]. This hierarchical multi-loop model suggests that operators learn feedback structures across multiple levels.

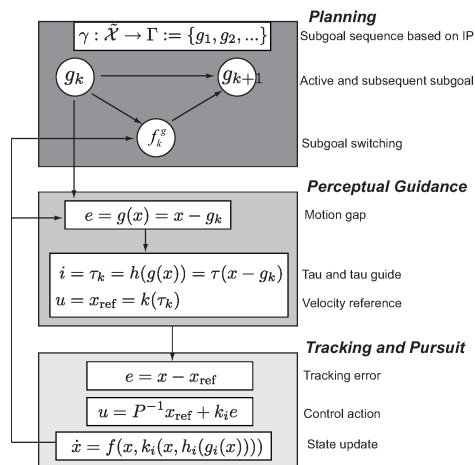


Fig. 1. Hierarchic multi-loop model of human guidance behavior. The top level describes the planning level based on the decomposition of the task and environment in terms of interaction patterns. The plan is codified based on a subgoal sequence g_k . The currently active subgoal defines the reference for the perceptual guidance. The latter extracts the current motion gap which is used to determine a state reference trajectory x_{ref} . At the lowest level, a tracking feedback system implements the desired motion.

The concept of “interaction pattern” is introduced in [20] following investigation of human guidance behavior using experiments with miniature remote control helicopters. These patterns are based on invariants of the closed-loop interactions. The significance of these invariants is that they describe what principles human pilots or operators use to break down complex guidance problems into a sequence of smaller, tractable ones. Preliminary studies based on piecewise affine (PWA) model identification methods suggest that the equivalence classes can be further decomposed into distinct dynamic modes, which provides deeper insight into lower-level control strategies. The higher-level interaction patterns combined with the lower-level dynamic modes provide the building blocks needed to codify the behavior across the entire hierarchy, from the lower-level control, guidance and perception, all the way to higher-level planning, adaptation and learning [19]. This paper investigates how this framework can be used for surgical skill evaluation.

2 Experimental Setup

This study employed the dataset collected in [21] which used the Electronic Data Generation and Evaluation (EDGE) platform (Simulab Corp. Seattle, WA), Fig. 2(a). This consists of 22.7 hours of synchronized video and tool motion data of Fundamental of Laparoscopic Skills (FLS) tasks collected from clinicians of various skill levels at three different clinical teaching hospitals in the United States. FLS has been shown to correlate to operating room performance [22]. We herein incorporate only a small part of this data that also provides ratings by faculty clinicians via blinded video review to establish valid categories of skill.

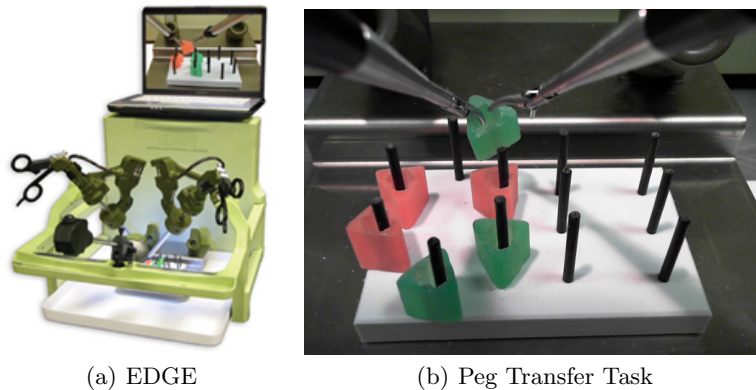


Fig. 2. The EDGE platform (a) and screen shot of the FLS Peg Transfer task (b) used for this work.

2.1 Task Description

In this paper, we use only the Peg Transfer task where clinicians use Maryland Graspers to transfer blocks in minimal time and with minimal drops. The blocks must be picked up by one hand with a laparoscopic tool and then transferred mid-air to the other hand tool.

2.2 Data Overview and Group Selection

The motion data (tool tip position, orientation, grasp angle and grasp force for both hands) sampled at 30Hz. Three skill groups were selected based on a combination of criteria in Table 1. Complete details are available in [13].

A set of six complete Peg Transfer task instances were arbitrarily selected from unique subjects among the three geographically distinct sites to represent each of the three skill groups.

Table 1. Summary of criteria used to select each set of iterations and its intended purpose. N refers to the total count of iterations of each set.

Group	N	Criteria
Expert (Exp)	6	Practicing laparoscopists (over 100 lapr. procedures): surgeons' and fellows' best FLS-scoring logs with 3/5 or greater average OSATS video review scores.
Intermediate (Int)	6	15th percentile of FLS scores about midpoint FLS score determined between lowest Expert FLS score and highest Novice FLS score [.59, .73].
Novice (Nov)	6	All logs below 15 th -percentile FLS score.

3 Surgical performance analysis

3.1 Task-Level Statistical Analysis

The objective of statistical analysis at the task level is to provide general characteristics of operator skill in an intuitive context. A mapping of the surgical movements into a probability distribution in the speed-curvature space was adopted as initial technique to assess general characteristics of operator's performance. This analysis provides both the maneuver envelope and the dominant states in the behavior [23]. The dominant states are defined to be the most frequently

visited states which may serve as the transition quasi-equilibrium between maneuvers. The probability distribution of trajectory points in the speed-curvature plane is shown in Figure 3. The expert group exhibits a larger envelope and more condensed dominant states.

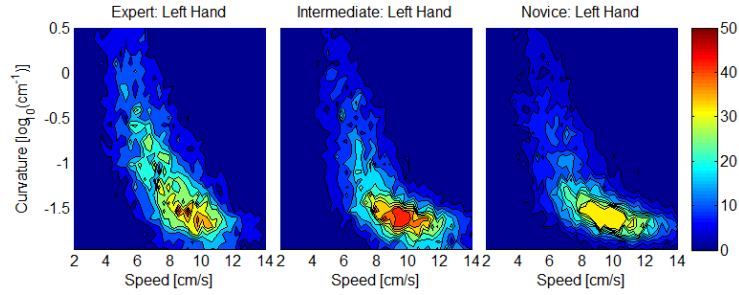


Fig. 3. 2D distribution of speed and curvature

Probability distributions generated over the same event space can be compared via a symmetric Kullback-Leibler divergence $D(p||q) = [D_{KL}(p||q) + D_{KL}(q||p)]/2$. A leave-one-(surgeon)-out permutation analysis was performed for all three skill groups. The resulting divergence among the populations of possible group probability distributions is shown in Table 2(a). Each entry in the table, $e_{m,n}$, is the mean and standard deviation of $D(P_{grp_m \setminus subj_i} || P_{grp_n \setminus subj_j}), \forall i \neq j$. Low values along the diagonal indicate that group distributions remain consistent and remain distinct from other groups even if individual surgeons are removed, whereas the low standard deviations (in parenthesis) indicate little change in overall distribution due to the removal of a single surgeon. However, the classification power of such a distributional approach is poor, see Table 2(b). This shows how such broad statistical approaches succeed in providing general characteristics of operator behavior in an intuitive context (the task variables), however, such generalization is too broad to be used for classification of very specific runs or individuals.

3.2 Kinematic classification

The kinematic classification method introduced in [23] is based on the concept of control and attention workload. The method uses a library of motion primitives of different attention load levels. The assumption is that experts favor motions of low control and attention load. Simpler motions, such as rectilinear and uniform (non-accelerated) motion, are easier to implement and more predictable, and therefore demanding less attention. These motions would allow the maneuvers to be more efficient and more consistent in multi-trial operations. Moreover, given the limited information processing capacity of human, simpler motions would allow for more cognitive processes, in particular, planning and decision making. Parsing trajectories under human control into sub-level sequences of motion primitives provides insights into the organization principles which is an

Table 2. Leave-one-(surgeon)-out (a) permutation analysis for symmetric KL-Divergences of group distributions and (b) cross-validation confusion matrix for classification success for individual surgeons.

(a) Permutation Analysis, mean (std dev)				(b) Classifier Performance			
	Alternate			Predicted			
	Exp	Int	Nov	Exp	Int	Nov	
Exp	0.18 (0.10)	0.82 (0.03)	0.78 (0.05)	3	2	1	
Int	0.81 (0.02)	0.14 (0.04)	0.81 (0.01)	0	4	2	
Nov	0.86 (0.05)	0.69 (0.05)	0.11 (0.02)	1	2	3	

essential aspect of human spatial control skills. The metrics derived from the segmentation, include the frequency of motion primitives and the mean duration of each segments. The kinematic classification results are shown in Figures 4 and 5.

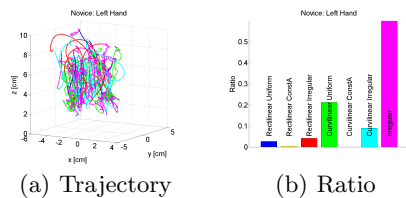


Fig. 4. Novice kinematic classification

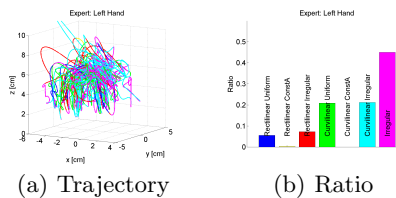


Fig. 5. Expert Kinematic classification

3.3 Dynamic clustering

The dynamic clustering method is based on the assumption of hierarchic organization of behavior [23]. It assumes that humans tend to adopt limited number of strategies in dealing with complex interactions taking place between organism, task and environment elements. The interaction involves the whole system of processes including perception, cognition and motor control. With extensive practice, these interactions will exhibit patterns that are manifestation of processes used to reduce the attention load and facilitate the organization of behaviors. Therefore capturing and describing these patterns are significant to investigate skills across the comprehensive hierarchy of processes.

In the dynamic clustering method, the dynamics of human-agent-environment interaction is described with Piece-Wise Auto-Regressive eXogenous (PWARX) model given in a parametric state-space form. Although the closed-loop interaction dynamics is always non-linear, it assumes that each interaction pattern describes an invariant in human's behavior that manifests as a quasi-equilibrium

in the dynamics. Therefore, the interaction patterns can be captured using a PWARX model in the form described in [23] and identified with different set of parameters:

$$\begin{bmatrix} x \\ y \\ v_x \\ v_y \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ a_{31} & 0 & a_{33} & 0 \\ 0 & a_{42} & 0 & a_{44} \end{bmatrix} \begin{bmatrix} x \\ y \\ v_x \\ v_y \end{bmatrix}_k + \begin{bmatrix} 0 \\ 0 \\ b_3 \\ b_4 \end{bmatrix} \quad (1)$$

where Δt is the sampling time [23]. Each interaction pattern is identified as a model of different set of parameters. The PWARX results are shown in Figures 6 and 7, where three clusters are identified for both novice and expert groups. To make the analysis more intuitive, the PWARX parameters are transformed into speed, normal and tangential acceleration in the ellipsoid plot.

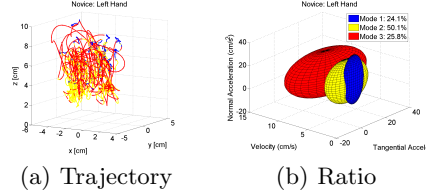


Fig. 6. Dynamic clustering of novices

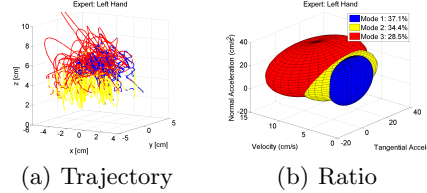


Fig. 7. Dynamic clustering of experts

4 Hierarchical Skill Assessment Results

4.1 Subgoal closure

As suggested in the introduction, dynamics-based clustering reveals spatial organization abilities of expert surgeons in performing Peg Transfer task, and makes it possible to delineate the different phases of the task and therein analyze specific performance characteristics. In Figure 7(a), the spatial organization of the behavior of expert surgeons is closely correlated with the three phases in Peg Transfer task:

1. Starting phase (cluster mode 2) coincides with the surgeons picking up the blocks. The movement during this phase follows a medium velocity range.
2. Maneuvering phase (cluster mode 1) coincides with the surgeons moving the gripped blocks to the central area of the board. There is no restriction on the movement during the maneuvering phase, and the objective of the phase is to be as fast as possible. Therefore the surgeons adopt high velocity and the accelerations span a large range.

- Interception phase (cluster mode 3) coincides with the blocks being transferred in the air between two hands of laparoscopic tool. This phase is critical in that it requires a large amount of coordination effort for both hands.

For each phase of the task, expert surgeons adopt very consistent strategy. In contrast, the maneuvers of novices are less consistent. During the starting phase, novices sometimes drive the control to high velocity, which penalizes the accuracy. The lack of consistency in the strategy also demands more attention load to plan for new trials and to handle the range of conditions. More attempts are required for novices to successfully pick up the blocks. In the maneuvering phase, the laparoscopic tool frequently slows to a lower velocity, penalizing the completion speed of the task.

4.2 High-level planning

To facilitate the accomplishment of complex task, humans divide the task into subtasks. In [20], Kong and Mettler have shown that subtasks exploit invariants in the dynamic environment interactions. The invariants in the human behavior emerge through extensive practice ostensibly as a result of the assimilation of coordinated movement and perceptual processes in procedural memory. These interaction pattern can then be used as a unit of behavior for the larger organization. High-level planning can therefore be assessed from the organization of interaction patterns. Effective planning allows using interaction patterns that take advantage of the dynamic interaction between human’s motor skills and task elements that also reduce the attention load. For this reason, spatial organization of the interaction pattern is an important measure of the surgeons’ planning skill. To quantify the spatial organization, the Cartesian coordinates of trajectory points are classified using a Fisher classifier based on the tags obtained in PWA clustering. The misclassification ratio is then used as the measure of spatial organization, as shown in Table 3.

Table 3. Skill metric in high-level planning

Spatial organization [%]	Expert	Intermediate	Novice
Complete Groups	17.9	29.6	38.6
Leave-One-Out Mean(std dev)	13.3 (4.4)	27.1 (5.6)	35.0 (6.5)

5 Conclusion

The results underscore the limitations of simple outcome measures such as those obtained from kinematic characteristics (see Section 3.1 and 3.2) and on the other hand demonstrates the discriminative power of dynamical characteristics obtained here using a PWARX model. The latter provides a more detailed segmentation and insights into the dynamic make-up of the behavior and their

spatial organization. This more detailed information provides correlation with important procedural movement stages, yet it requires no prior, high-level knowledge about the task to be implemented. Finally, the spatial characteristics of the segmented performance data provides a measure of the ability to organize the different stages of behavior in a manner which is consistent with the spatial and dynamic constraints of the task and operator skills. These results demonstrate that dynamical segmentation techniques can access attributes across the entire process hierarchy and provide the foundation to a more comprehensive skill analysis and modeling framework. Future work will immediately extend this approach to more FLS tasks and ultimately to different tasks in laparoscopic, robotic, and open surgery and incorporate gaze characteristics.

References

1. C. Zhan and M. Miller, "Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization," *JAMA: the journal of the American Medical Association*, vol. 290, no. 14, p. 1868, 2003.
2. L. Sturm, J. Windsor, P. Cosman, P. Cregan, P. Hewett, and G. Maddern, "A systematic review of skills transfer after surgical simulation training," *Annals of surgery*, vol. 248, no. 2, p. 166, 2008.
3. S. G. Mattar, A. A. Alseidi, D. B. Jones, D. R. Jeyarajah, L. L. Swannstrom, R. W. Aye, S. D. Wexner, J. M. Martinez, S. B. Ross, M. M. Awad *et al.*, "General surgery residency inadequately prepares trainees for fellowship: Results of a survey of fellowship program directors." *Annals of surgery*, vol. 258, no. 3, pp. 440–449, 2013.
4. J. J. Coleman, T. J. Esposito, G. S. Rozycki, and D. V. Feliciano, "Early subspecialization and perceived competence in surgical training: Are residents ready?" *Journal of the American College of Surgeons*, vol. 216, no. 4, pp. 764 – 771, 2013.
5. C. E. Reiley, H. C. Lin, D. D. Yuh, and G. D. Hager, "A review of methods for objective surgical skill evaluation," in *Surgical Endoscopy*, July 2010.
6. B. Varadarajan, C. Reiley, H. Lin, S. Khudanpur, and G. Hager, "Data-derived models for segmentation with application to surgical assessment and training," vol. 5761, pp. 426–434, 2009.
7. R. Kumar, A. S. Jog, A. Malpani, B. P. Vagvolgyi, D. D. Yuh, H. Nguyen, G. D. Hager, and C. Chen, "Assessing system operation skills in robotic surgery trainees," *The International Journal of Medical Robotics and Computer Assisted Surgery*, 2012.
8. L. Zappella, B. Béjar, G. D. Hager, and R. Vidal, "Surgical gesture classification from video and kinematic data," *Medical Image Analysis*, vol. 17, pp. 732 – 745, 2013.
9. 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*, Springer. Nogyo, Japan: Springer, 2013.
10. L. Tao, E. Elhamifar, S. Khudanpur, G. D. Hager, and R. Vidal, *Sparse Hidden Markov Models for Surgical Gesture Classification and Skill Evaluation*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2012, vol. 7330, pp. 167–177.

11. L. Tao, G. D. Hager, and R. Vidal, "Segmentation and recognition of surgical gestures from kinematic and video data," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*, Springer. Nogyo, Japan: Springer, 2013.
12. A. Gallagher, E. Ritter, H. Champion, G. Higgins, M. Fried, G. Moses, C. Smith, and R. Satava, "Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training," *Annals of surgery*, vol. 241, no. 2, p. 364, 2005.
13. T. Kowalewski, "Real-time Quantitative Assessment of Surgical Skill," Ph.D. dissertation, May 2012.
14. W. Warren and B. Fajen, "From optic flow to laws of control," in *Optic flow and beyond*. Kluwer Academic Publishers, 2004, pp. 307–337.
15. W. Warren, "The dynamics of perception and action." *Psychological Review*, vol. 113, no. 2, p. 358, 2006.
16. R. K. Heffley, "A model for manual decelerating approaches to hover," in *Proceedings of the Fifteenth Annual Conference on Manual Control, AFFDL-TR-79-3134*, 1979, pp. 545–554.
17. R. Hess and B. McNally, "Automation effects in a multiloop manual control system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 16, no. 1, pp. 111–121, Jan. 1986.
18. M. White, G. Padfield, and R. Armstrong, "Progress in measuring simulation fidelity using an adaptive pilot model," in *Proceedings of the 60th Annual Forum of the American Helicopter Society*. American Helicopter Society, 2004.
19. B. Mettler and Z. Kong, "Hierarchical model of human guidance performance based on interaction patterns in behavior," in *2nd International Conference on Application and Theory of Automation in Command and Control Systems*. London, UK: ATACCS, 2012.
20. Z. Kong and B. Mettler, "Modeling human guidance behavior based on patterns in agent-environment interactions," *IEEE Transaction on Human-Machine Systems*, vol. 43, no. 4, pp. 371–384, July 2013.
21. T. Kowalewski, L. White, J. Carlis, E. Santos, R. Sweet, T. Reihisen, and D. F. Keefe, "Exploratory Visualization of Surgical Training Databases for Improving Skill Acquisition," *IEEE Computer Graphics and Applications*, vol. 32, no. 6, pp. 71–81, 2012.
22. G. Sroka, L. Feldman, M. Vassiliou, P. Kaneva, R. Fayez, and G. Fried, "Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room—a randomized controlled trial," *The American Journal of Surgery*, vol. 199, no. 1, pp. 115–120, 2010.
23. B. Li and B. Mettler, "Application of Trajectory Segmentation Techniques for Operator Skill Evaluation," in *American Control Conference*, 2013.