Vision based Decision-Support and Safety Systems for Robotic Surgery

Suren Kumar[*]
PhD Candidate
surenkum@buffalo.edu

Madusudanan Sathia Narayanan*
PhD Candidate
ms329@buffalo.edu

Sukumar Misra[†]
Surgical Intern
smisra@buffalo.edu

Sudha Garimella[‡]
Assistant Professor
sgarimella@buffalo.edu

Pankaj Singhal[§]
Director of Robotic Surgery
psinghal@buffalo.edu

Jason J. Corso[¶]
Assistant Professor
jcorso@buffalo.edu

Venkat Krovi*
Asssociate Professor
vkrovi@buffalo.edu

ABSTRACT

Computer-vision based safety and decision-support methods in robotic surgical systems (such as the da Vinci) offer many advantages from implementation-, robustness- and cost-perspectives. This manuscript seeks to survey the state-of-art techniques in the use of computer-vision and video-understanding in robotic-surgeries; which leads naturally to examination of the potential of these methods to enforce and enhance the safety in multiple layers of a typical (autonomous or non-autonomous) surgical work-flow. Our primary focus is on addressing issues related to: (i) multiple tool tracking and detection; (ii) description of semantic attributes; and (iii) investigation of surgical tool motions for skill and dexterity using only the endoscopic video streams. Implementing such a video-understanding framework is not only immediately useful in augmenting visual feedback for surgeons but also is a big step in closing the loop for future automated surgeries.

In this work, we first present a novel multiple tool detection framework using sparse and dense optical flow in a causal manner by learning detectors for individual tools in real surgical scenarios [6]. Secondly, Bayesian filtering techniques are applied to the video-streams to identify semantic attributes [5] (such as tool open/close, stained with blood etc.) that can be used to provide feedback to surgeons or robots of the future. Finally, we discuss our current progress in automated surgical skill assessment [10] using video-based methods.

*Dept. of MAE, SUNY Buffalo NY 14260 USA
†Millard Fillmore Suburban Hospital NY 14221 USA
‡Dept. of Pediatrics, SUNY Buffalo NY 14214 USA
§Kaleida Health System, Buffalo NY 14221 USA
¶Dept. of CSE, SUNY Buffalo NY 14260 USA

Categories and Subject Descriptors

I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Motion, Tracking, Object Recognition, Color*; I.5.4 [**Pattern Recognition Applications**]: Computer Vision

General Terms

Measurement, Performance, Standardization

Keywords

Robotic surgery, video-understanding, automated tool tracking, video-based skill assessment, attribute identification

1. INTRODUCTION

The emerging classes of IT-enabled medical devices mark a significant paradigm shift in the modern era. The erstwhile passive human operated tools are being replaced by active systems. Nowhere is this more evident than in the new wave of computer integrated surgical system technology [26], commonly referred to as "surgical robots". Modular and embedded computing, sensing and communications, and deep integration with physical elements and processes allow these new *cyber- physical systems* (CPS) to currently achieve unmatched levels of functionality, adaptability, and effectiveness in the present [14].

However, as with any other software or hardware computing systems, the new generation of surgical robots remain susceptible to system/ component failures, information security breaches as well as faults (like 'bugs' and 'viruses'). Even where outright/ hard failures may be forestalled by frequent maintenance and replacements, the systems still depend critically on the awareness of the human-in-the-loop to forestall/ correct soft failures effectively. For example, the cables in the surgical instruments have finite life due to wear-and-tear. These progressively degenerate leading to compromised end-effector positioning capability (which is assumed to be correctable by the operating surgeon). There has been a growing awareness and interest, in both research and commercial arenas, to incorporate multiple layers of safety, autonomous and automated decision support systems to enhance overall capabilities of such systems.

Of all the current methods, video- (image-) based methods offer a technically and economically viable option for deployment within a real clinical setting [13]. Nonetheless, many approaches in the vision literature remain unsuitable for realizing real-time applications (either due to computational constraints or failure to cope with physical world complexities). Hence, there is a need for easy-to-implement, robust and potentially automatable alternatives. **From this perspective, the aims of this work is two-fold: (i) to review advances in developing safety critical and decision-support systems based on video- and image- understanding for surgical applications; and (ii) to discuss our recent progress in tool tracking, skill evaluation and semantic attribute labeling using video-based analysis and detection methods.**

Specifically, we implement object detection [6] and attributes labeling [5] methods for automated multiple tool tracking and semantic annotation of video-frames from challenging surgical videos. Challenges arise in tool tracking due to partial/ full occlusion of tools because of organs, blood stains, residual tissue, rinsing fluids or tissue cauterization smoke [29]. Other challenges arise from tools exiting the field of view or at times due to error amplification by compliant articulated chains. At the other end of the spectrum are the whole host of errors introduced by the human operators, especially due to lack of experience in the use of such complex tools. Hence, we will also discuss our work on video-based motion analysis to assess human surgical tool motions especially from the perspective of improving the safety. In particular, 3D kinematic motion capture were used on the stereo-feeds to extract trajectories for various points of interest [10]. These trajectory estimates lend themselves well to quantitative evaluation of user-specific characteristics (such as repeatability, stability and robustness) as a measure of skilled sensorimotor expertise.

The rest of the manuscript is organized as follows: Section II presents an overview of traditional object detection methods leading to a investigation of past and current work in tool detection and semantic labeling of scenes from videos. This section will also summarize some of the current surgical skill evaluation methods available from the literature. Section III discusses our state-of-art tool detection and semantic labeling framework for surgical applications. Section IV presents our experimental setup and evaluates the predictive performance of our tool tracking and semantic labeling algorithms when used with real surgical videos. This section will also cover some of the surgical- skill assessment results based on our micro-motion analysis technique. Finally, Section V summarizes our efforts and outlines possible extensions for the future work in this area.

2. LITERATURE REVIEW

2.1 Surgical Tool Tracking

Tracking surgical tools in general has been used for a wide range of applications including safety, decision-support as well as skill assessment. Predominant approaches have been using fiducial markers on the tool, using color marker and thresholding in HSV space to find tool [7], attaching light emitting diodes to the tip of instruments and then detecting these markers in endoscopic images [12], color coding tips of surgical instruments and using a simple color segmentation [29]. Using such marker-based methods for tracking in surgical videos has issues with manufacturing, bio-compatibility and additional instrumentation.

Other approaches that do not necessarily modify the tool itself include using color space for classification of pixel into instruments and organs, doing shape analysis of these classification labels and then predicting the location of tool in next frame using an Auto Regressive Model [27]. McKenna et al [18] use similar method for classification of surgical tool but uses particle filter to track instruments in a video. These approaches are limited to detecting a tool when a significant area of tool is present in an image frame and there is a good distinction between its background and instruments in color space. Recently there has been some work on locating specific landmarks on the surgical tools, learning a Random Forest based classifier to classify these landmarks from images and using an Extended Kalman Filter (EKF) to smooth the tool poses [22]. However this method requires knowledge of 3D CAD model of tool and extensive image labeling for a single tool.

2.2 Surgical Skill Assessment

Medical education has long relied on subjective or in some cases semi- quantitative (like Likert-scale based) [28] evaluations from experts due to lack of reliable, accurate and stable objective and quantitative performance metrics [11]. Key challenges to assessment and accreditation of surgeons in such scenarios include (i) creating clinically relevant scenarios and settings; and (ii) developing uniform, repeatable, stable, verifiable performance metrics; at manageable financial levels for ever increasing cohorts of trainees [1].

The growth of computer integration in minimally invasive surgery (MIS) especially in the form of robotic MIS (rMIS) [17, 2] now offers a unique set of opportunities to comprehensively address this situation, especially by providing alternatives for expert evaluators to 'monitor' the trainees. With such instrumented trainers, a variety of physical variables can now be transparently monitored in both simulated and real-life operational settings [15, 25].

In recent times, the standardized objective methods were studies using such trainers to assess technical skills for use in surgical programs. The Objective Structured Assessment of Technical Skills (OSATS) as well as Objective Structured Clinical Examination (OSCE) emphasizes the quantitative assessment processes without entirely relying on expert evaluators, but on the hardware systems (such as Imperial College Surgical Assessment Device (ICSAD) and Advanced Dundee Endoscopic Psychomotor Trainer (ADEPT)). However, as reported in [22], the robotic encoders and sensors are not very accurate and precise; and induce additional errors that may offset our analysis drastically. This again emphasizes the use of vision based techniques to reduce the errors in surgical tracking and assessment and to enable better decision-making and safety critical systems.

Besides, several studies in the recent past showed that segmenting the surgical videos into sub-tasks (defined as surgemes in [16, 21]) can aid in automated performance and skill assessment [24]. One of the most relevant work concerning this aspect is the automated motion recognition using Hidden Markov Modeling (HMM) [23, 24] for simulated surgical tasks using da Vinci Trainer (dV-Trainer). It is essential to define these building blocks in a generalized way to allow vision and motion algorithms to learn to perform automated primitive segmentation, analysis of complex procedures and finally, establish meaningful metrics for skill and expertise. In our experiments, we adopted a new micro-motion analysis method, commonly used in industrial engineering practice, for surgical task discretization and developed a comprehensive library of sub-tasks to validate our performance and dexterity metrics [10]. Even though the type of metrics to be used for skill assessment is not of interest here, it will be briefly discussed to outline the importance of motion analysis and video based methods that help automatedly derive the objective metrics from motion estimates compared to cumulative or averaged values that tend to mask out the class-specific discriminative characteristics.

The role of techniques used for video-based understanding have been clearly outlined in this section. Finally, we wish to envision some of the ideas presented in this manuscript by combining several aspects (tool tracking, detection, semantic labeling and skill assessment) under common vision-based safety critical and decision support system framework and eventually, close the loop as much as possible for the future automated robotic surgeries.

3. METHODS

3.1 Tool Tracking

The method that is implemented in this work is summarized in Figure 1. We learn different detectors for each type of surgical tool using state-of-the-art object detector [6]. This detector essentially captures the object shape by using Deformable Part Models (DPM) consisting of star-structured pictorial structure model which links root of an object to its parts using deformable springs. Hence this model captures articulation which is invariably present in surgical tool and allows for learning an detector for different tool configurations. We annotated surgical tools in real surgical videos and learn a Latent Support Vector Machine (LSVM) classifier by extracting Histogram of Oriented Gradients (HOG) from annotated bounding box. This type of learning classifiers is highly generalizable and extensible enabling one to find tools in videos without making any restrictive assumptions about the type, shape, color of tool, view etc.

A simple tracker was used based on extracting dense optical flow [3] and predict the bounding box in next frame. Optical flow measures apparent motion of a pixel between two images assuming that its brightness remains constant in both images. The tracking is initiated by using the detections with confidence measure above a given threshold. In each frame, the optical flow with the previous frame is obtained for all the pixels belonging to the desired bounding box and its location in next frame is approximated by mean flow of all pixels. Since velocity of each pixel is available, it is possible to estimate the scenarios such as tool leaving the frame of view, by verifying if the mean bounding box velocity has a significant directional component perpendicular to the frame of view.

3.2 Attribute Labeling

Knowledge of tool attributes such as open/closed, grasping tissue/without it, cauterizing tool condition etc., is an essential step in the path towards autonomous robotic surgery. To the best of our knowledge, no prior work has attempted to find such attributes directly from surgical videos. Furthermore, this type of attribute labeling can add a second layer of safety to account for lack of experience from surgeons and any instances of miscommunication between master and slave end of such systems.

To obtain such attributes, we are inspired from work done in computer vision literature [5] which identifies shape, part and material attributes for different types of objects by learning Support Vector Machine (SVM) classifier [4] using visual features. In this paper, we demonstrate our algorithm to identify only tool open/closed attribute, but should be extensible to any type of attributes of interest. We use Pyramid Histogram of Oriented Gradients (PHOG) which captures the shape of an object at various scales. After extracting the features from the image of a tool, we use a Probabilistic SVM (PSVM) classifier [20] that estimates the probability of tool being in open or closed states. Unlike standard two class SVM classifier, which only gives a rigid classification decision by locating a test point in feature space w.r.t to a separating hyperplane, PSVM gives a probability measure based on the distance from the separating hyperplane.

In contrast to the attributes research in vision community which is mostly based on images, in robotic surgery one has access to a real-time stream of video data as well. This data can be leveraged to ensure the smoothness of transition in attributes as can be seen in tool open/ closed states as observed in images. We use Bayesian tracking to ensure smoothness of attribute labeling. Let us denote the state of the tool at time t as $x_t \in \{0, 1\}$ where 0 represents closed state and vice versa. Probabilistic SVM classifier gives us $P(x = 0|y_t)$ and $P(x = 1|y_t)$ for an image observation y_t at time step t. However since the decision of classifier is only based on current image observation, it does not include any prior information about the current state of tool attribute. We use Bayesian tracking to maintain this smoothness by conditioning prior probability using data from τ previous frames. Using Bayes rule at time step t and assuming all observations are independent and current observation only depends on current state, one can write,

$$P(x_t = 1|y_{t-\tau}, .., y_t) = \frac{P(y_t|x_t = 1)P(x_t = 1|y_{t-\tau}, .., y_{t-1})}{P(y_{t-\tau}, .., y_t)} \tag{1}$$

where $P(x_t = 1|y_{t-\tau}, .., y_{t-1})$ models the prior probability of tool being in open state given observations only upto current time step, $P(y_t|x_t = 1)$ is the likelihood of current observation given that tool is open. We can write similar equation for tool being in closed state $x_t = 0$. Since we are only interested in the classification decision we ignore the denominator which will be same for $P(x_t = 1|y_{t-\tau}, .., y_t)$ and, $P(x_t = 0|y_{t-\tau}, .., y_t)$. We model prior probability by taking frames τ times prior to current observation to model smoothness inherently present in motion of an surgical tool.

$$\begin{aligned} P(x_t = 1|y_{t-\tau}, .., y_t) &\propto P(y_t|x_t = 1)P(x_t = 1|y_{t-\tau}, .., y_{t-1}) \\ &\propto P(x_t = 1|y_t)P(x_t = 1|y_{t-\tau}, .., y_{t-1}) \\ &\propto P(x_t = 1|y_t) \prod_{i=t-\tau}^{t-1} P(x_i = 1|y_i) \end{aligned}$$

$$\propto \prod_{i=t-\tau}^{t} P(x_i = 1|y_i) \tag{2}$$

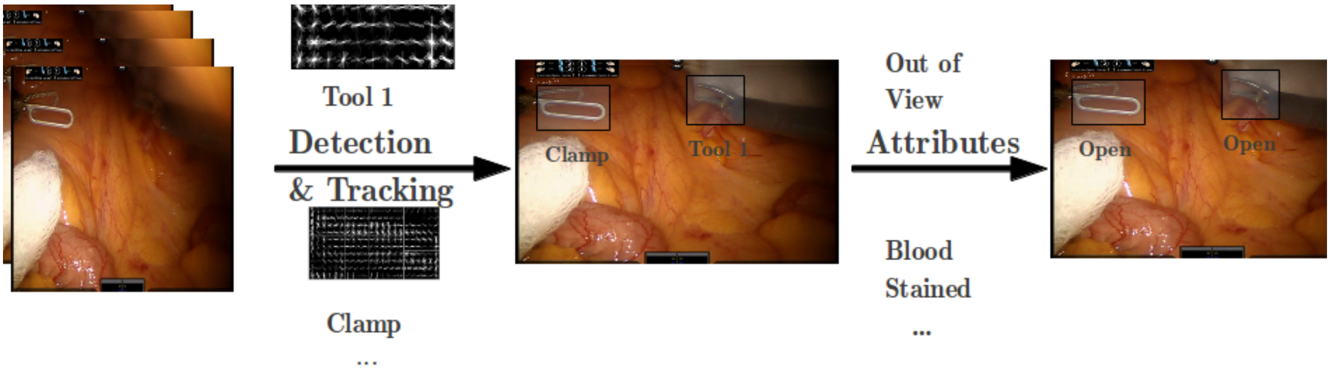


Figure 1: System pipeline of proposed algorithm

Figure 2: Intuitive Surgical's da Vinci SKILLS Simulator Si (dVSS-Si)

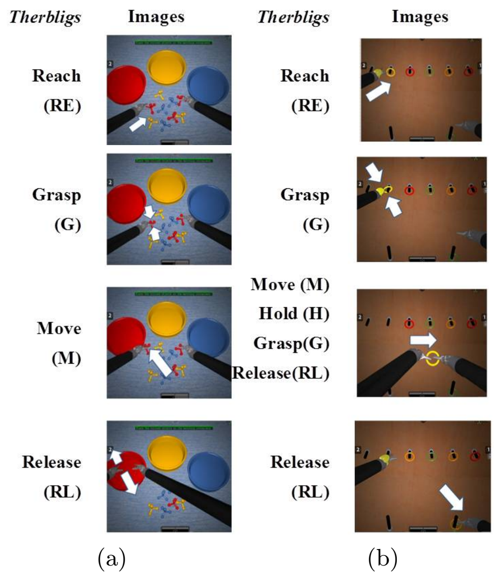


Figure 3: Robotic Surgical Tasks: (a) Pick and Place, (b) Peg Board

In Equation 2, the probabilities at each time step $P(x_i = 1|y_i)$ are directly given by probabilistic SVM classifier. This in effect weights the classification decision by not only the current observation but also by the decision in previous frames.

4. RESULTS

4.1 Experimental Setup

As our objective is to demonstrate the application of video understanding methods to provide multiple layers of safety and decision- support (tracking, detection, labeling and skill training) for generic surgical procedures, only the video feeds were used as input in evaluating the algorithms for each of these aspects. Using a standardized test-bed, the da Vinci Surgical System-Si (dVSS-Si) with its SKILLS simulator system [8] as in Fig. 2 proved to be useful in: (i) recording of stereoscopic video images to evaluate our tool tracking and labeling algorithms; (ii) bench-marking the performance of different surgeons for temporal-, intra- and inter-subject comparative surgical skill analyses and evaluations. For testing our tool tracking and detection algorithms, real surgical videos were used. These set of videos were first processed manually to generate necessary ground truth data against which output of our algorithms could be validated. Since, only experts are allowed to conduct real surgeries, all of our preliminary automated tool tracking and semantic labeling use surgical video recordings by experts only. However, the efforts are currently being undertaken in order to conduct physical training tasks using box trainers or dVSS including surgeons and trainees from a varied set of expertise population. Our preliminary results on skill assessment include SKILLS simulator tasks – pick and place and peg board as in Fig. 3, for which a set of two subjects were selected from each expertise category (2 each in expert, intermediate and novice) to conduct comparative skill evaluation.

4.2 Detection and Tracking

We collected and annotated a large set of surgical videos for specifically two kinds of tools as in Fig.4-6. Parts based detector is able to detect different types of tools in varying lighting, articulation, partial occlusion and appearance variations. Although detector gives true positives as well as false positives, we initialize tracker only with detections with high confidence score. Also detector does not give results over all the input frames and generally has higher detection probability only over shapes of tool seen in training data. Hence tracking is necessary to fill these gaps and maintain identity of tools over time. Tracking results on two different videos are shown in Figures 4-5. In Figure 4, two different types of surgical tools are tracked using a monocular input and tracker is efficiently able to track both these tools. In Figure 5, one type of surgical tool is tracked using a monocular input and tracker is found robust to appearance changes due to tissue, blood and motion blur. We have collected and annotated large set of surgical videos for two kinds of tools. Our detector is able to detect different types of tools in varying lighting, articulation, partial occlusion and appearance variations. Figure 6 shows tracking of a surgical tool in presence of strong specular reflection and smoke.

4.3 Attribute Labeling

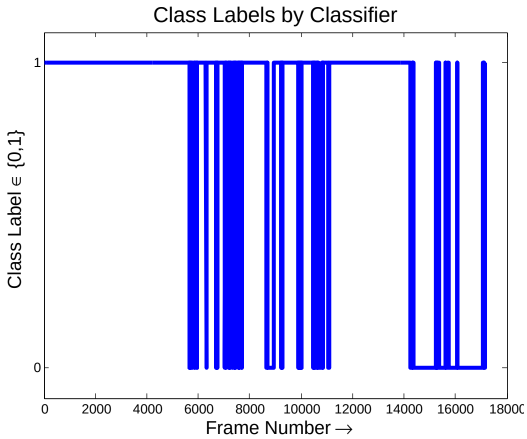


Figure 7: Classification results for Attributes Labeling

We learned the SVM classifier by using the ground truth data obtained from manual annotations of the bounding boxes of each tool and the corresponding attribute (open/closed). We then tested the baseline SVM classification by determining the class using the probabilistic measure $x_t = i$ if $P(x_t = i|y_t) > P(x_t = j|y_t), i, j \in \{0, 1\}$. Using the baseline classifiers results in an accuracy of 77.74%. The overall classification results are shown in Figure 7. After bayesian smoothing accuracy improves to 79.52% by choosing the parameter τ to be 100 and their is more smoothness in label assignment as can be seen from Figure 8.

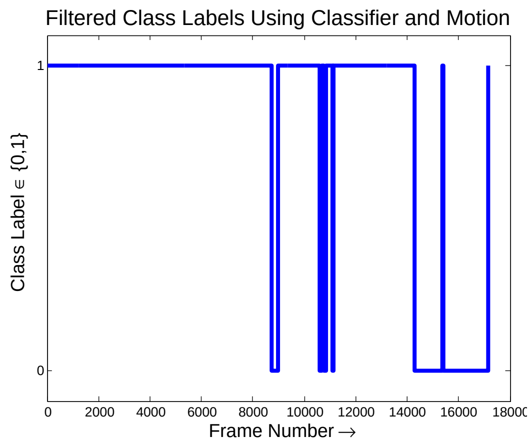


Figure 8: Classification and smoothing results for Attributes Labeling

4.4 Skill Assessment

Traditional motion studies are captured in a process chart [19] where *'Therbligs'* are hierarchically grouped into work elements and then ultimately into meaningful tasks, which has proved adequate to offer a primary discretization of industrial manipulation tasks. At its core, the process chart captures the type of *'Therblig'* (discrete-state) at each instant in time for each tool in form of a table as shown in Fig. 9. Enhancements to this basic process chart now involve taking advantage of bilateral symmetry (left/ right) or increased discretization or agglomeration of tasks as well as performing varying levels of statistical analyzes based on the collected and processed data.

Left hand Description	Sym	Time (ABS)	Time (Rel)	Time (Rel)	Time (ABS)	Sym	Right hand Description
Reach	RE	**1.5**	1.5	1.6	**1.6**	RE	Reach
Grasp	G	**1.8**	0.3	0.1	**1.7**	G	Grasp
Move	M	**2.5**	0.7	0.9	**2.6**	M	Move
Release	RL	**2.8**	0.3	0.2	**2.8**	RL	Release
Reach	RE	**3.5**	0.7	1.6	**4.4**	RE	Reach
Avoidable delay	AD	**4.4**	0.9	0.1	**4.5**	G	Grasp
....	....	....	....	....	....	....	....
....	....	....	....	....	....	....	....
End of Task							

Figure 9: Sample Process Chart for Left and Right Hand Side Surgical Tools

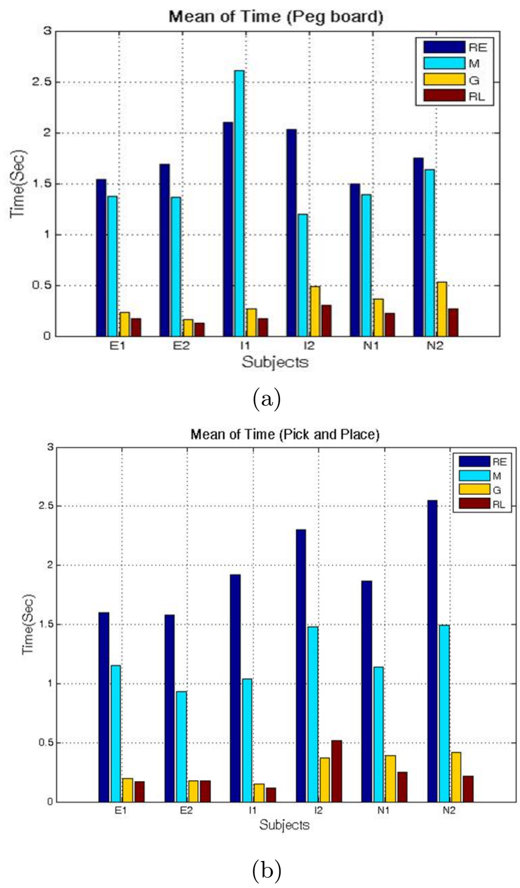


Figure 10: Histograms of Mean TTC for each *'Therbligs'* of (a) Pick-and-Place; and (b) Peg Board Tasks

For the chosen set of tasks, pick-and-place (PnP) and peg-

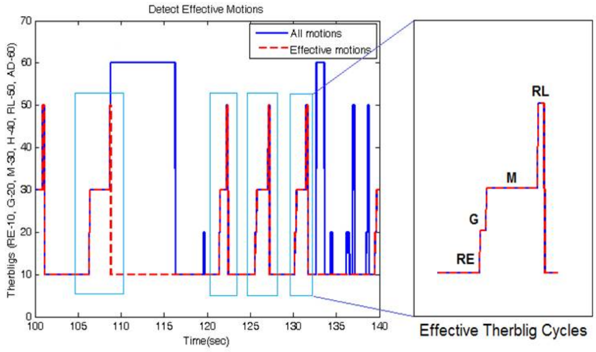


Figure 11: Micromotion *Therblig* Analysis for Effective Motion Detection

board (PegI), only a subset of complete set of *'Therbligs'* are required by users to complete the exercise. In particular, PnP tasks need 4 *'Therbligs'* - Reach (RE), Grasp (G), Move (M) and Release (RL) while the PegI tasks require a total of 5 elements- four of those as earlier with a Hold (H) *'Therblig'*. The recorded videos for all the cases were manually processed and ground truth data (two hand chart) were obtained in form of text files using *'Therblig'* labeling software developed in our lab. These data files were then analyzed for each subject, each task and each *'Therblig'* based on the distributions of time to task completion. In order to anonymize the subject information, the following symbols were assigned during our analysis - experts (E_1 and E_2), intermediates (I_1 and I_2) and novices (N_1 and N_2). An immediate observation of the final results based on this analysis reveals that higher task complexity magnifies the discriminative skill characteristics between experts and novices. The bar charts of the cumulative time instances spent by each subject in each of the *'Therblig'* confirming the same for both the tasks is shown in Fig.10.

The segmentation process has enabled us now to break down the continuous video streams into a time-series discrete state vector (*'Therbligs'*). Our next objective is to use these as our input to develop meaningful performance-metrics in terms of ineffective motions and dexterity. Since, the training tasks are going to be standard, it is easy to determine the ideal discrete state trajectories and form a template for each task. By comparing the template with the actual trajectory for each cycle, it is then possible to estimate the total instances when a sequence was missed or repeated unsuccessfully as shown in Fig.11. Similarly, dexterity can be represented as cumulative time instances that each user has used both the tools to perform useful sub-tasks simultaneously. Pre-processing of *'Therblig'* cycle data is required to eliminate all the ineffective motions prior to computing the dexterity metric. The corresponding performance metrics computations for both the experimental tasks are listed in Table. 1.

In our prior work [11], we also demonstrated the applicability of this method to real surgical scenarios wherein *'Therblig'* based analysis is conducted in order to decipher the expert's tool motion signature. In future, our objective is to validate the method over a variety of real surgical scenarios and improve the overall accuracy of skill predictions by combining the automated motion tracking and attribute labeling framework discussed earlier.

Table 1: Ineffective and Synchronized Motions for Pick-and-Place and Peg Board Tasks

	Pick-and-place			Peg Board I		
	AD	I*	S*	AD	I*	S*
E_1	25.1	44.8	57.3	17.8	27.9	65.4
E_2	10.9	23.8	80.8	18.4	25.1	64.9
I_1	28.9	40.3	52.3	25.9	41.8	50.6
I_2	6.1	21.0	88.8	36.2	39.9	56.9
N_1	6.9	22.2	90.2	26.4	34.1	55.0
N_2	15.1	29.2	73.5	33.6	45.8	37.7

I: Inefficient motions, *S: Synchronized motions*

5. DISCUSSION

Vision-based tracking and video-understanding algorithms are flexible and hence, enable ease of implementation within the existing infrastructure. Overall robustness and accuracy of this method could be easily observed in the tool tracking results and could be improved further. Importantly, our algorithm was able to detect the tools precisely even in occluded and partially occluded environments which was otherwise not possible with many other existing methods that assume clear tool tracking conditions. For the semantic attribute labeling, the performance is expected to improve further with generation of additional extensive ground truth data for various attributes (such as blood stains, presence of foreign substances, excessive motions etc) as we demonstrated it to work for tool open/ closed states. In this preliminary study, a reasonable performance accuracy was obtained (about 80 %) for such attributes labeling but with extensive benchmark data and cross-validation testing a more robust and accurate labeling could be achieved. At the same time, the results for surgical skill assessment shows the reliability of the kinematics-based video analysis method. The method was also employed to obtain surgical signatures of experts from real surgical cases [9]. However, the method still needs further refinement to handle variations in subjects and tasks, and provide a reasonable and accurate assessment under diverse scenarios. This is currently being addressed to an extent by using a customized (sensorized) physical trainer setup as well as from a da Vinci system using its own API for recording the kinematic data.

6. REFERENCES

[1] A. Amodeo, A. Linares Quevedo, J. V. Joseph, E. Belgrano, and H. R. H. Patel. Robotic laparoscopic surgery: cost and training. *The Italian Journal Of Urology And Nephrology*, 61(2):121–128, 2009.

[2] D. B. Camarillo, T. M. Krummel, and J. J. K. Salisbury. Robotic technology in surgery: Past, present, and future. *The American Journal of Surgery*, 188(Supplement 1):2–15, 2004.

[3] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[4] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[5] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *IEEE*

Conference on Computer Vision and Pattern Recognition (CVPR), pages 1778–1785, 2009.

[6] P. Felzenszwalb, R. Girshick, and D. McAllester. Cascade object detection with deformable part models. In *IEEE Conference on Computer vision & Pattern Recognition (CVPR)*, pages 2241–2248, 2010.

[7] M. Groeger, K. Arbter, and G. Hirzinger. Motion tracking for minimally invasive robotic surgery. *Medical Robotics, I-Tech Education and Publishing*, pages 117–148, 2008.

[8] $Intuitive - SurgicalandInc.$. da Vinci Surgical Robot: Si and SKILLS Simulator, 2012.

[9] S.-K. Jun, M. S.-Narayanan, P. Agarwal, A. Eddib, P. Singhal, S. Garimella, and V. Krovi. Minimally invasive surgical skill assessment by video-motion analysis. In *The Hamlyn Symposium on Medical Robotics*, 2012.

[10] S.-K. Jun, M. S.-Narayanan, P. Agarwal, A. Eddib, P. Singhal, S. Garimella, and V. Krovi. Robotic minimally invasive surgical skill assessment based on automated video-analysis motion studies. In *4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics*, pages 25–31, 2012.

[11] S.-K. Jun, M. S.-Narayanan, P. Singhal, S. Garimella, A. Eddib, and V. Krovi. Evaluation of robotic minimally invasive surgical skills using motion studies. In *Proc. Workshop on Performance Metrics for Intelligent Systems*, PerMIS, pages 198–205, New York, NY, USA, 2012.

[12] A. Krupa, J. Gangloff, C. Doignon, M. de Mathelin, G. Morel, J. Leroy, L. Soler, and J. Marescaux. Autonomous 3-d positioning of surgical instruments in robotized laparoscopic surgery using visual servoing. *IEEE Transactions on Robotics and Automation*, 19(5):842–853, 2003.

[13] A. Krupa, M. d. Mathelin, C. Doignon, J. Gangloff, G. Morel, L. Soler, J. Leroy, and J. Marescaux. Automatic 3-d positioning of surgical instruments during robotized laparoscopic surgery using automatic visual feedback. In *Proc. 5th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) -Part I*, pages 9–16, London, UK, UK, 2002. Springer-Verlag.

[14] G. S. Lee and B. Thuraisingham. Cyberphysical systems security applied to telesurgical robotics. *Comput. Stand. Interfaces*, 34(1):225–229, Jan. 2012.

[15] M. A. Lerner, M. Ayalew, W. J. Peine, and C. P. Sundaram. Does training on a virtual reality robotic simulator improve performance on the da vinci surgical system? *Journal of Endourology*, 24(3):467–72, 2010.

[16] H. C. Lin, I. Shafran, D. Yuh, and G. D. Hager. Towards automatic skill evaluation: detection and segmentation of robot-assisted surgical motions. *Computer Aided Surgery: Official Journal Of The International Society For Computer Aided Surgery*, 11(5):220–230, 2006.

[17] M. J. Mack. Minimally invasive and robotic surgery. *JAMA: The Journal of the American Medical Association*, 285(5):568–572, 2001.

[18] S. McKenna, H. Charif, and T. Frank. Towards video understanding of laparoscopic surgery: Instrument tracking. In *Proc. of Image and Vision Computing, New Zealand*, 2005.

[19] B. Niebel and A. Freivalds. *Methods, Standards, and Work Design*. McGraw-Hill, 10th edition, 2003.

[20] J. Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74, 1999.

[21] C. E. Reiley, E. Plaku, and G. D. Hager. Motion generation of robotic surgical tasks: Learning from expert demonstrations. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 967–970, 2010.

[22] A. Reiter, P. Allen, and T. Zhao. Feature classification for tracking articulated surgical tools. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pages 592–600, 2012.

[23] J. Rosen, J. D. Brown, L. Chang, M. N. Sinanan, and B. Hannaford. Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete markov model. *IEEE Transactions on Biomedical Engineering*, 53(3):399–413, 2006.

[24] J. Rosen, B. Hannaford, C. G. Richards, and M. N. Sinanan. Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. *Biomedical Engineering, IEEE Transactions on*, 48(5):579–591, 2001.

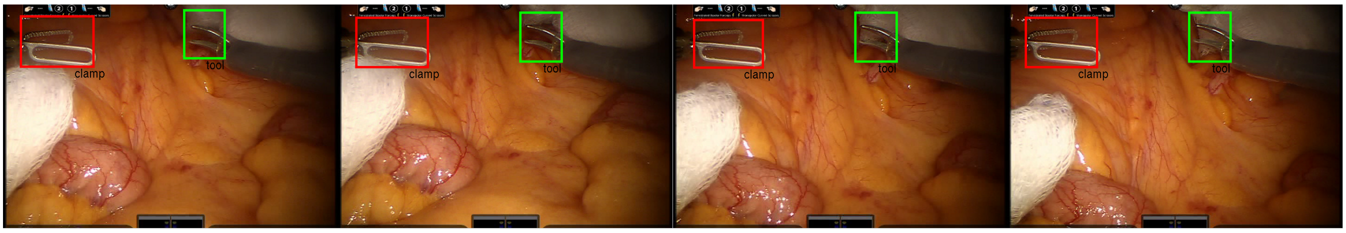
[25] M. S.-Narayanan, X. Zhou, S. Garimella, W. Waz, F. Mendel, and V. Krovi. Simbiopsies: An augmented reality training simulator for needle biopsies. In *The Hamlyn Symposium on Medical Robotics*, 2011.

[26] R. Taylor. Computer-integrated surgical systems and technology engineering research center (CISST ERC), 1997.

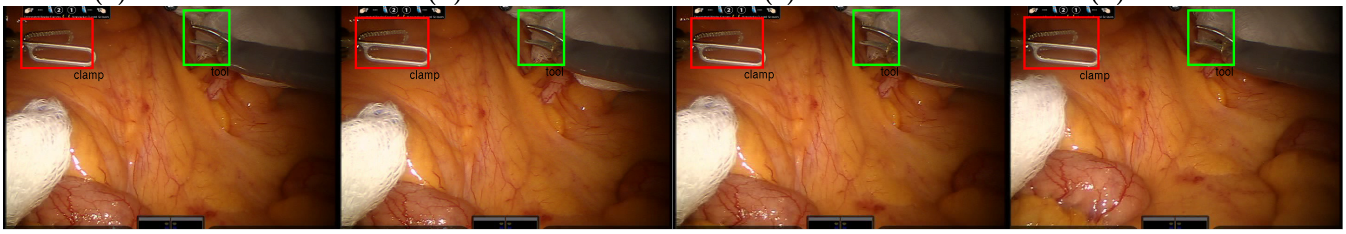
[27] D. Uecker, Y. Wang, C. Lee, and Y. Wang. Laboratory investigation: Automated instrument tracking in robotically assisted laparoscopic surgery. *Computer Aided Surgery*, 1(6):308–325, 1995.

[28] P. D. van Hove, G. J. M. Tuijthof, E. G. G. Verdaasdonk, L. P. S. Stassen, and J. Dankelman. Objective assessment of technical surgical skills. *British Journal of Surgery*, 97(7):972–987, 2010.

[29] G. Wei, K. Arbter, and G. Hirzinger. Automatic tracking of laparoscopic instruments by color coding. In *CVRMed-MRCAS'97*, pages 357–366. Springer, 1997.

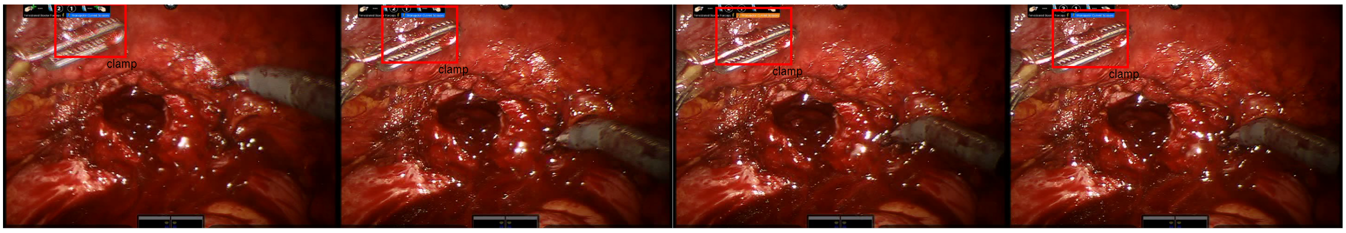


(a) ***Frame 25*** (b) ***Frame 50*** (c) ***Frame 75*** (d) ***Frame 100***



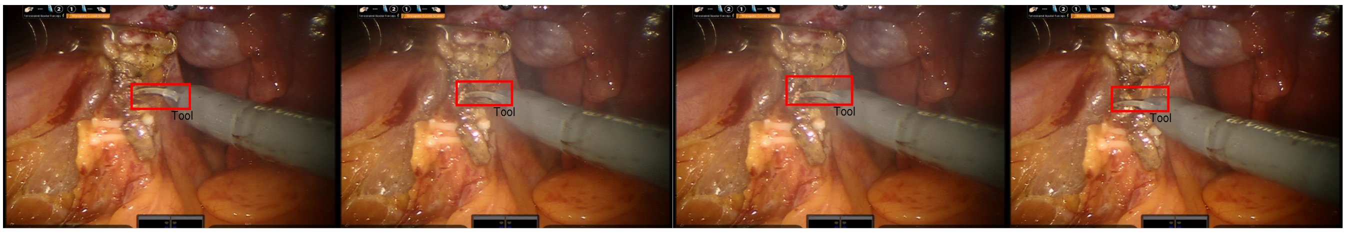
(e) ***Frame 125*** (f) ***Frame 150*** (g) ***Frame 175*** (h) ***Frame 200***

Figure 4: Tracking results for “Tool” and “Clamp” on a surgical operation video. (Please view in color)



(a) ***Frame 25*** (b) ***Frame 50*** (c) ***Frame 75*** (d) ***Frame 100***

Figure 5: Tracking results for “Clamp” on a surgical operation video. (Please view in color)



(a) ***Frame 39*** (b) ***Frame 48*** (c) ***Frame 57*** (d) ***Frame 93***

Figure 6: Tracking results for “Tool” on a surgical operation video. (Please view in color)