Benchmarking Expert Surgeons' Path for Evaluating a Trainee Surgeon's Performance

Malik Anas Ahmad[*], Shamyl Bin Mansoor[†], Zohaib Amjad Khan[‡], Waqar Ahmad Aqeel[§], Shahroze Humayun Kabir[¶]
Smart Machines and Robotics Technology (SMART) Lab
School of Electrical Engineering & Computer Science (SEECS), National University of Sciences & Technology, Pakistan

Abstract

The affordance of independent learning is one of the most important advantages of computer simulators for surgical training. This advantage can get dull if the simulator does not provide the useful instructional feedback to the user and the instructor has to supervise and tutor the trainee while using the simulator. In fact the continued need of instructor feedback with most existing simulators is often cited as a primary reason for the reluctance of many medical schools to fully embrace simulator technology [Sewell 2007]. Thus the incorporation of relevant, intuitive metrics in a way that it provides a constructive feedback which facilitates independent learning is essential for the development of efficient simulators. Evaluating a trainee surgeon's performance as per trainer surgeon's desire is always a challenging problem in the development of minimal invasive surgery simulators. In this research we have proposed a novel metric for trainee surgeons' performance evaluation using machine learning algorithms.

Keywords: Virtual Reality, Performance Evaluation, Surgical Simulation, Machine Learning, Artificial Intelligence

1. Introduction

Virtual reality training simulators are used in different fields such as aviation, vehicle driving and medical disciplines such as anesthetics, surgery, and procedure-based medicine.

The advent of medical simulators started a new era for surgical education. Now technical skills are no longer learned in the Operation Room (OR) through a traditional apprenticeship model of training. Instead, the acquisition of new skills and development of basic surgical proficiency are moving to a simulated environment in the surgical skills laboratory. Basic surgical tasks and some advanced surgical techniques can be replicated in the skills virtual laboratory, allowing both trainees and practicing surgeons to gain proficiency in these skills [Montbrun and MacRae 2013].

Laparoscopic surgery (LS), a widely established procedure in Minimally Invasive Surgery (MIS) is being used nowadays for an increasing number of surgical interventions. Compared to open surgery, LS has multiple important benefits for the patients such as faster healing, shorter hospitals stays, and minimized risk of infection. However, there are a number of inherent particularities in the application of MIS techniques, such as visual projection of the 3D space onto a 2D display, restricted mobility, reduced force feedback, and the fulcrum effect. As a consequence, the technical skills required for laparoscopic surgery are different from those required for traditional open surgery. Adjustment of the hand-eye coordination by interpreting the three dimensional image from a two dimensional image along with overcoming the fulcrum effect are some of the skills specific to minimally invasive surgery [Nugent et al. 2013]. Moreover, a laparoscopic surgeon needs to have very good depth perception and advanced psychomotor skills (such as grasping, tracking and suturing), using long and thin surgical instruments inserted in the abdominal cavity via small incisions.

Complications can occur even when experienced surgeons who are well versed in open techniques and have a good knowledge of anatomy and pitfalls embrace new techniques. This fact heightened the concerns about the training of novices who lacked such a background in open surgery [Gallagher 2012]. Hence simulators are increasingly being recognized as a valuable tool for training. This is particularly true for the early part of the learning curve in laparoscopy and other minimally invasive surgical techniques.

In recent years virtual reality (VR) technology based surgical skill assessment has received major attention. It allows surgical trainees to acquire the essential skills required to perform a laparoscopic operation. VR simulators provide a secure educational paradigm where errors do not risk patients' safety and these systems are also able to provide constructive feedback about the trainees' performances. Moreover, VR simulators encompass a wide range of training tasks with different levels of difficulty, as well as scenarios of entire operations such as laparoscopic cholecystectomy. In addition to surgical skills training, VR simulators allow for skills assessment [Megali et al. 2006].

One of the most important functions of a simulation is to facilitate the effective and efficient training of skill outside the clinical situation thus minimizing the risk to the patient from at least part of the novice's learning curve. But what is skill? Failure by medicine to explicitly answer this question has been one of the major impediments to the development of good simulations and simulation-based training.

A simulator without performance evaluation is as good as an expensive video game. The main objective of evaluation metrics is to provide objective and immediate feedback to a trainee on his performance. This allows trainer to provide formative feedback to aid trainee in acquiring skill. All available Virtual Reality (VR) surgical simulators use execution time as a metric. Unfortunately time analyzed is at best crude and at worst a dangerous metric [Gallagher et al. 2012].

However, it is also vital to ensure that the tools used for both training and assessment actually measure what they purport to

[*]e-mail:anas.ahmad@seecs.edu.pk
[†]e-mail:shamyl.mansoor@seecs.edu.pk
[‡]e-mail:zohaib.amjad@seecs.edu.pk
[§]e-mail:10besewaqeel@seecs.edu.pk
[¶]e-mail:10beeskabir@seecs.edu.pk

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
VRCAI 2013, November 17 – 19, 2013, Hong Kong.
Copyright © ACM 978-1-4503-2590-5/13/11 $15.00

measure. It is necessary to demonstrate that the simulator modules have construct validity and that the training program results in a training benefit.

A high level of psychomotor skills is required to perform minimally invasive surgery (MIS) safely. To assure high quality of skills, it is important to be able to measure and assess these skills. For that, it is necessary to determine aspects that indicate the difference between performances at various levels of proficiency. Measurement and assessment of skills in MIS can best be done in an automatic and objective way.

The study in [Hofstad et al. 2012] investigates a set of nine motion-related metrics for their relevance to assess psychomotor skills in MIS during the performance of a labyrinth task. They used time, bimanual dexterity, path length, angular length, depth perception, response orientation, motion smoothness, number of sub-movements and average velocity as evaluation metrics. They found significant difference in the performance of three groups of users with different set of experiences in the field of laparoscopic surgery which showed the difference in their proficiency levels.

As compared to simple metrics like time and number of errors, motion tracking has been suggested to be a more sensitive performance metric for the assessment of surgical performance [Stefanidis et al. 2013].

Over the last few years, surgical skill evaluation has attracted the interest of various scientific groups due to the great demand for shifting from subjective scoring to a more objective and quantitative analysis that can provide important insight into the trainee's qualities. Experience discrimination by analyzing hand motion has received special attention in the last few years [Loukas et al. 2013]. Another way of assessment is analyzing the motion of the instrument with a path planning approach.

Usually the evaluation of surgical skills is based on the application of Hidden Markov Models (HMMs) and Support Vector Machines (SVM) for reasons such as task decomposition and recognition of surgical expertise. Surgical dexterity has been evaluated by applying HMMs. It is done by considering a hand motion signal as a stochastic process composed of state sequence that usually relates to a user-defined set of primitive movements [Megali et al. 2006]. In discussions with different groups of physicians and surgeons from around the world there appears to be a consensus that reaching an agreement on performance metrics is all but impossible [Gallagher and O'Sullivan 2012]. Rather than benchmarking on some abstract performance level reached by consensus in a committee, the training pass level is defined based on the performance levels of individuals who are actually very experienced at performing the procedure clinically.

Defining an optimal surgical performance is the real difficulty. So the key is to model the expert surgeons' skills which will help in assessing the skill of a novice surgeon [Megali et al. 2006].

In this study, we investigate the role of instruments' motion connectivity in the performance of a laparoscopic VR simulator. Two groups were considered: experienced surgeons and beginners. The connectivity pattern of each subject was evaluated by analyzing their instruments' motion signals.

2. Previous Work

The existing performance metrics for surgical simulations can be broadly classified into two types; the first type includes metrics that are common to all of the exercises where as the second category encompasses exercise specific metrics. The first type of metrics depicts the efficiency and ease of a trainee while handling MIS instruments and the level of trainee's hand eye coordination. The second type indicates the accuracy of the trainee while performing a certain task. These second type metrics measure how accurately the trainee has achieved a certain goal. Though they might not be the true indicators of the efficiency of a user's performance of a surgical exercise yet they certainly demonstrate the accuracy of the trainee for that particular exercise.

In [Hofstad et al. 2012] the authors have used orientation, smoothness, perception, velocity and time as metrics to measure the performance of a surgeon in order to differentiate between an expert and novice surgeon. They have further used them for mapping a learning curve of a novice surgeon. Other than measuring how good a novice surgeon is learning, these parameters have also been used for online performance evaluation, like it was used for pattern cutting exercise of a glove by [Pellen et al. 2009]. Other than basic skills training these generic metrics have also been applied to assess advanced skills like those involved in cholecystectomy [Aggarwal et al. 2009].

In MIS simulators, application of machine learning for the purpose of performance evaluation of a user is not very common. Rather evaluation is generally based on the surface area, time and path length [Bajka et al. 2010]. In [Loukas et al. 2013] the authors have used multivariate auto aggressive models to find a connectivity pattern to make an expert benchmark so that novice surgeons can be evaluated in comparison with the expert score. Four parameters; values, covariance, number of nonzero multivariate autoregressive (MAR) coefficients, and area under the coherence spectra were extracted to compare a hand motion analysis of surgical groups with different levels of experience.

In [Megali et al. 2006] Hidden Markov Model (HMM) was applied to define a model of expertise and objective model to evaluate the performance of novice surgeons during laparoscopic surgical simulator training. The kinematic data describing the movements of surgical instruments was processed and HMM was used to define an expert model that describes expert surgical gesture. Subsequently, this expert model was used as a reference in the definition of an objective metric for performance evaluation of someone with different ability.

Though uncommon, usage of machine learning techniques for performance evaluation in a surgical simulator is not unique. B. Allen et al. [Allen et al. 2010] have used support vector machines in three exercises of peg transfer, pass rope and cap needle for classifying the expert surgeons and novices' performance.

Considering the unavailability of an optimal path for a surgical procedure, basis of our metric is to train the trainee surgeons as per expert surgeons' desire. So keeping that in mind we took expert trainer surgeons' instrument paths (in a specific type of simulation exercise) as the optimal paths. A machine learning algorithm named as Artificial Neural Network (ANN) was trained on them. Afterwards the new trainee surgeon's performance was judged on how good he followed the expert trainer surgeons' paths. The amount of deviation from the expert trainer surgeons' paths was penalized by calculating the Euclidian distance in between desired/expected location and trainee surgeon's instrument location.

Artificial neural network (ANN) is a widely used machine learning technique. One of its many applications is path planning. It has a great ability to learn non linear paths.

The power of our technique is that instead of finding an optimal path in between multiple expert surgeons' path, it tries to learn all of them except on those rare points where they intersect each other.

3. Methods & Technique

We used a machine learning technique which is commonly applied in path planning of mobile robots. First we discretized the virtual space into Cartesian coordinates. Then each location of the instrument at a certain time was taken as a new instance and its next location as the resultant instance. Then we saved these two locations as pairs in a log file. Then we applied some preprocessing on these pairs of the current and subsequent locations. An ANN was then trained for the paths followed by the expert surgeons' instrument while performing the virtual exercises. A typical node graph of ANN is shown in Figure 1 where each node represents an artificial neuron and arrows represent connections between these neurons.

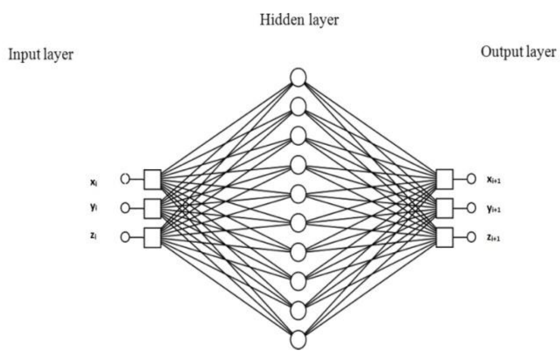

Figure 1: *Depiction of an Artificial Neural Network (ANN).*

While evaluating a trainee surgeon, we fed the instrument's location coordinates of trainee surgeon into our trained ANN. The ANN as trained resulted into a forecasted location which our instrument should have reached. Euclidian distance between the ANN's forecasted location coordinates and the actual subsequent location coordinates of the trainee's instrument was considered as a penalty for deviating from the expert surgeon path.

In following subsections, we have first explained the maintenance of the instruments' path log followed by the incorporation of the time invariance, training & forecasting method, penalization method and score standardization.

3.1 Surgeons' Instruments Path Logs

Initially we asked expert trainer surgeons to perform some surgical simulation exercises. Their instruments' locations were saved in a log to maintain a record of the paths followed by the expert trainer surgeons. In these logs instruments' locations were recorded with a fixed time period of 0.05 seconds. We logged the trainee surgeons' instruments' paths in a similar way. As we discretized the virtual space in Cartesian coordinates, location of the instruments was in the form of three values, i.e. horizontal (x), vertical (y), and depth (z).

3.2 Time Invariance

As mentioned in Section 2, time is widely used in the simulators as a metric for speed. So in an effort to emphasize more on the better following of the path we made speed irrelevant in our new metrics.

The path logs recorded the location of the instrument after every 0.05 seconds time. To make the path logs time invariant we deleted all of the location coordinates which were consecutively repeating themselves. This time invariance was incorporated in both trainee and trainer surgeons' path logs.

3.3 Training

We used one of the supervised learning methods named as Elman network method of Artificial Neural Network (ANN) for learning multiple expert surgeons' path. We provided labeled training data to this supervised way of machine learning. The training examples were in the form of pairs of an input object and desired output object.

The input and output objects were both vectors with three members each. Each location of the expert trainer surgeons' instruments' time invariant path was taken as a separate training example's input vector and its subsequent location coordinates as output vector. So ANN was trained on each location except the last one.

3.4 Results and Penalties

To evaluate the trainee surgeon's performance, instruments' location coordinates from his time invariant path log were given as an input vector object to the trained network. The trained network gave a three member output vector object. The output vector contains the forecasted location coordinates which it expects the trainee surgeon to take his instrument to.

Once the output vector object was retrieved, the Euclidian-distance between the instruments' forecasted and the original location coordinates was calculated. After calculating the deviation for each move, absolute mean of the deviation was taken as the penalty for deviation from the benchmarked path. The bigger the penalty a user had the more he/she had deviated from the path. This whole process is illustrated with an example in Figure 2.

3.5 Score Standardization

The core of this concept is the mimicking of the expert surgeon's evaluation. A comparison on how good our algorithm's evaluation has mimicked the expert surgeons' evaluation will exhibit the success of our efforts since the goal is to benchmark the expert surgeons' performances. In order to compare the performance of our algorithm's marking scheme in comparison to the expert surgeons' marking we standardized these two different types of evaluation schemes. For this standardization we calculated the z-score for each individual. In following we have explained how the standardization was calculated.

Let P^{Xj} represent a metric with trainee surgeon's number j= {1, 2, 3, 4, 5} and group title X= {AS,SS} where "AS" represents algorithm's score where as "SS" represents the expert surgeon's score. The score z corresponding to metric P for a user j is then given by:

$$z = \frac{P^{Xj} - \bar{P}}{\sigma} \quad (1)$$

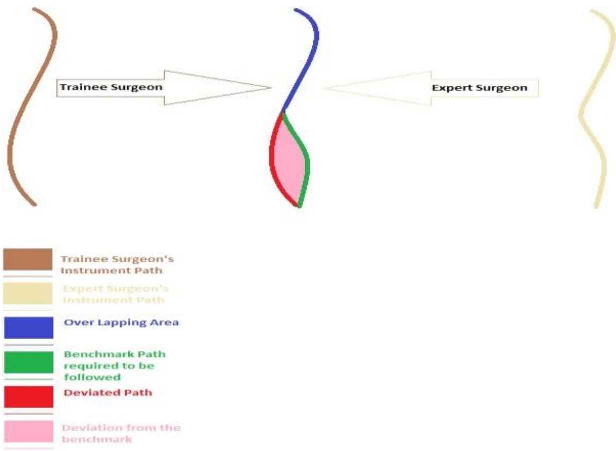

Figure 2: *Use of an expert surgeon's path as a benchmark for evaluating a trainee's performance.*

Where $\overline{P}$ is the mean of parameter P of all the users and σ is the corresponding standard deviation. Using independent z-scores we can then calculate a standardized z-score for each user using

$$z'_j = 1 - \frac{z_j}{z_{max}} \qquad (2)$$

Raw values of both algorithms' evaluated results for two different surgical training exercises are shown in Table 1 and 3. Their standardized values are shown in Table 2 and 4 respectively.

4. Results

We used Simulation Open Framework Architecture (SOFA) [Allard et al. 2007] to build the laparoscopic surgical simulation training exercises of Peg transfer and Grasping.

During the Grasping exercise shown in Figure 3, a user has to iteratively grasp a peg and place it in a basket without touching the floor and the other instrument, where as the location of the basket and the peg keeps on changing in a random order during the course of an exercise.

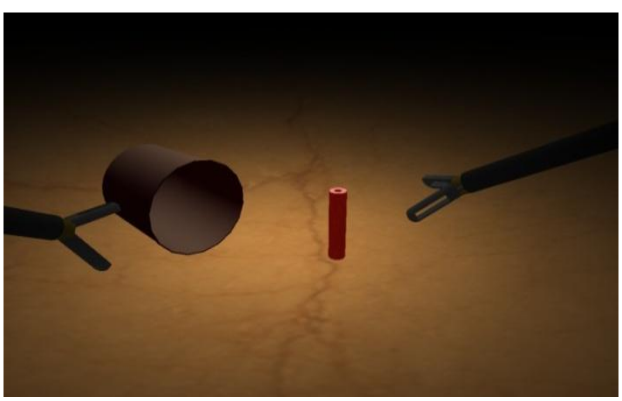

Figure 3: *Grasping exercise.*

During the Peg transfer exercise shown in Figure 4 a user has to pick four disks placed at one side of the box and place them on the respective peg stands at the other side of the box.

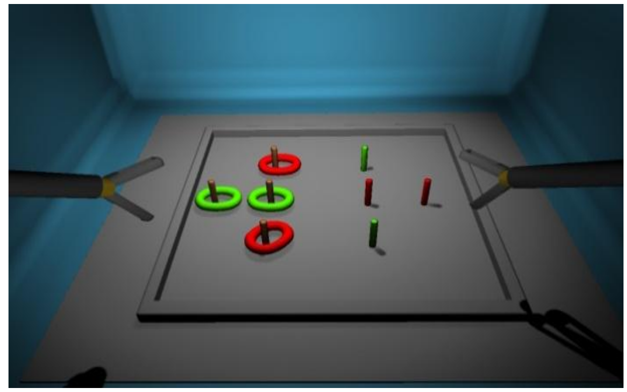

Figure 4: *Peg transfer exercise.*

Along with SOFA we used Matlab® to implement machine learning algorithms. We used the custom designed and indigenously manufactured hand manipulator named as Al-Zahrawi™ for these two exercises. The hand manipulators are shown in Figure 5.

Our algorithm was applied in one of the periodically held workshops of Laparoscopic Surgical training in Holy Family Hospital, Rawalpindi, Pakistan in the month of April, 2013. During that workshop we requested four expert surgeons to perform the peg transfer and grasping exercise multiple times. Meanwhile we maintained the log of their instruments' paths. An ANN was trained on these paths using the Elman method as described in C part of the Section III. Learning rate for the training was 0.05 with tolerance rate of 0.001. After training, the trained network validated the learning with an error of 0.08 average Euclidean distances (approximately 0.046 mm) for grasping exercise and 0.12 average Euclidian distances (approximately 0.07 mm) for peg transfer exercise. The error is quite low and tolerable.

After wards two of the expert surgeons manually examined, instructed and evaluated the trainees. All of the novice surgeons were asked to practice on the simulator for few times before recording their logs so that they may get used to the simulator. Afterwards their performance was evaluated by our metric in comparison to benchmarked expert surgeons' path. Meanwhile the expert surgeons evaluated the novice surgeon's performance and

Figure 5: *Al Zahrawi™ Hand Manipulators.*

marked them in between 1-10 where 1 being the best and 10 being the worst. Since our algorithm's rating is penalty-based whereas surgeons gave their marks out of 10, both the scores were later standardized to give a fair comparison. Once the novice surgeons felt comfortable with the use of our simulator, each of them was asked to complete the exercise once for which his/her scores were calculated.

The raw and standardized scores for 4 novice surgeons who performed peg transfer exercise are shown in Table 1 and Table 2 respectively. The standardized scores were calculated based on Score standardization method described earlier. The left hand column of each table shows the rating of trainees as given by the expert surgeons. The right hand column shows the scores given to the trainees calculated using our method. The standardized values in Table 2 have been plotted in Figure 6 for comparison between our algorithm's score and expert surgeon's marking. As can be seen from the figure, our algorithm's performance for Peg Transfer Exercise was quite encouraging. The average of the difference between standardized score marked by our algorithm and the expert surgeon was found to be only 1 percent.

Table 3 and Table 4 show the raw and standardized scores for 5 trainees respectively, who performed the grasping exercise. Like in case of peg transfer, the standardized scores based on both expert surgeons' markings and our algorithm have been plotted which are shown in Figure 7. From the figure, it is quite obvious that our algorithm's performance was not very encouraging for the grasping exercise. There was a one clear shoot out (highlighted in green in Tables 3 and 4 and in Figure 7) while evaluating one trainee's performance, but other than that the average of the difference between standardized score marked by our algorithm and the expert surgeon was 21 percent (without the one abnormal shoot out). One of the reasons behind this difference is the lack of training for this exercise because as mentioned before during grasping exercise the target object do not always appear at a fixed location rather it iteratively changes its reappearance location in a random order. This random appearance changes the course of the instrument motion, thus creating multiple possibilities of an instrument's trajectory. To resolve this issue of versatility we will need more examples from the expert trainer surgeon.

Of the two basic skill exercises peg transfer has more critical importance of the path followed by the instrument as the trainee has multiple target locations to reach while avoiding more obstacles as compared to the grasping exercise. In peg transfer exercise our

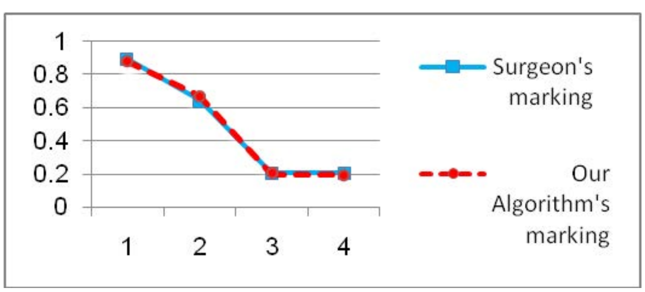

Figure 6: *Standardized Peg Transfer Exercise Evaluation score.*

Table 1 : *Raw Peg Transfer Exercise Evaluation score*

Average of Expert Surgeon's Rating (out of 10)	Our algorithm's ratings
7	20.07
6.25	16.42
5.25	10.21
5.25	10.04

Table 2 : *Standardized Peg Transfer Exercise score*

Expert Surgeon's Rating	Our algorithm's ratings
0.89	0.88
0.64	0.68
0.21	0.21
0.21	0.20

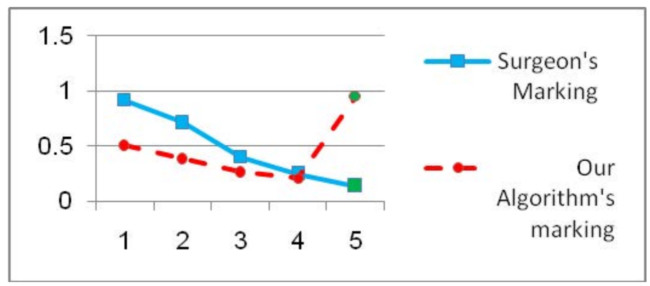

Figure 7: *Standardized Grasping Exercise Evaluation score.*

Table 3: *Raw Grasping Exercise Evaluation score*

Average of Expert Surgeon's Rating (out of 10)	Our algorithm's ratings
7	0.53
6.5	0.42
6	0.29
5.75	0.23
5.5	1.15

Table 4 : *Standardized score for Grasping Exercise*

Expert Surgeon's Rating	Our algorithm's ratings
0.92	0.51
0.72	0.39
0.40	0.27
0.25	0.21
0.14	0.95

algorithm's evaluation pattern and the expert surgeons' evaluation patterns matched more closely as compared to those in the grasping exercise.

5. Conclusion & Future Work

Based on our proposed method multiple trainer surgeons can teach the art of surgery to multiple students at the same time in a more convenient way. The intention of this paper is the introduction of a novel intelligent metric. There is definite room for further validation and improvement in this metric by training it on a bigger and versatile data set. The proposed approach provides a suitable basis for instrument motion analysis of surgical trainees and could be utilized in future VR simulators for skill assessment.

The future of surgery is non-invasive and robotic surgery. For an autonomous robotic surgery, we will need to train the autonomous or semi-autonomous equipment/device. A human surgeon gets trained by practising and benchmarking his trainer surgeon's work along with some improvisation. Our work is just imposing this effort in a more automated way by minimizing the fatigue on the trainer surgeon's part where as same efforts can be evolved into training a robot for autonomous or semi-autonomous surgery or other medical procedures.

The results can be improved with the inclusion of more training examples. One possible future approach to explore is to analyze multiple locations at a time for trajectory analysis of the critical surgical manoeuvres. Too much of a dictation in this metric could prove to be a discouragement for the improvised learning of a trainee surgeon, To cater this we can include the allowance or less penalization of the trainee surgeon on performing in a way predicted by genetic algorithms. Other than the path, inclusion of angular movement can also help incorporating the fulcrum effect. This work is a sound proof that the incorporation of path based intelligent metric can help in automating the classical method of apprenticeship in a convenient way.

Acknowledgement

The authors wish to thank Dr Asif Zafar Malik and his team at Holy Family Hospital, for their support in this research. This work was supported by a grant from National ICT R&D Fund, Ministry of Information Technology, Government of Pakistan.

References

AGGARWAL, R., CROCHET, P., DIAS, A., MISRA, A., ZIPRIN, P., AND DARZI, A. 2009. Development of a virtual reality training curriculum for laparoscopic cholecystectomy. In *British Journal of Surgery*, *96*(9), 1086-1093.

ALLARD, J., COTIN, S., FAURE, F., BENSOUSSAN, P. J., POYER, F., DURIEZ, C., DELINGETTE, H., AND GRISONI, L. 2007. Sofa - an open source framework for medical simulation. In *Medicine Meets Virtual Reality, MMVR 15*.

ALLEN, B., NISTOR, V., DUTSON, E., CARMAN, G., LEWIS, C., AND FALOUTSOS, P. 2010. Support vector machines improve the accuracy of evaluation for the performance of laparoscopic training tasks. In *Surgical endoscopy*, *24*(1), 170-178.

BAJKA, M., TUCHSCHMID, S., FINK, D., SZÉKELY, G., AND HARDERS, M. 2010. Establishing construct validity of a virtual-reality training simulator for hysteroscopy via a multimetric scoring system. In *Surgical endoscopy, 24*(1), 79-88.

GALLAGHER, A. G. 2012. Metric-based simulation training to proficiency in medical education: What it is and how to do it, In *The Ulster Medical Journal*, vol. *3* (July), 107-113.

GALLAGHER, A. G., AND O'SULLIVAN, G. C. 2011. *Fundamentals of Surgical Simulation: Principles and Practice*. Springer London.

HOFSTAD, E. F., VÅPENSTAD, C., CHMARRA, M. K., LANGØ, T., KUHRY, E., AND MÅRVIK, R. 2012. A study of psychomotor skills in minimally invasive surgery: what differentiates expert and nonexpert performance. In *Surgical endoscopy*, 1-10.

LOUKAS, C., ROUSEAS, C., AND GEORGIOU, E. 2013. The role of hand motion connectivity in the performance of laparoscopic procedures on a virtual reality simulator, In *Medical & Biological Engineering & Computing*, 1-12.

MEGALI, G., SINIGAGLIA, S., TONET, O. AND DARIO, P. 2006. Modelling and Evaluation of Surgical Performance Using Hidden Markov Models, In *IEEE Transactions on Biomedical Engineering*, vol. 53, 10 (October), 1911-1919.

MONTBRUN, S. D. AND MACRAE, H. 2013. Simulation and Minimally Invasive Colorectal Surgery, In *Seminars in Colon and Rectal Surgery*, vol. 24 (March), 53-60

NUGENT, E., SHIRILLA, N., HAFEEZ, A., O'RIORDAIN, D. S., TRAYNOR, O., HARRISON, A. M., AND NEARY, P. 2013. Development and evaluation of a simulator-based laparoscopic training program for surgical novices. In *Surgical Endoscopy*, vol. *27*(1), 214-221.

PELLEN, M. G., HORGAN, L. F., BARTON, J. R., AND ATTWOOD, S. E. 2009. Construct validity of the ProMIS laparoscopic simulator. In *Surgical endoscopy*, *23*(1), 130-139.

SEWELL,C. 2007. *Automatic Performance Evaluation in Surgical Simulation*, Diss. Stanford University.

STEFANIDIS, D., YONCE, T. C., KORNDORFFER JR, J. R., PHILLIPS, R., AND COKER, A. 2013. Does the Incorporation of Motion Metrics Into the Existing FLS Metrics Lead to Improved Skill Acquisition on Simulators? A Single Blinded, Randomized Controlled Trial. In *Annals of surgery*, *258*(1), 46-52.