A SIMULATION-BASED ASSESSMENT SYSTEM FOR COMPUTER ASSISTED SURGICAL TRAINER

Minsik Hong
Department of Electrical and Computer Engineering
University of Arizona
1230 E Speedway boulevard, Tucson, AZ, USA
mshong@email.arizona.edu

Jerzy W. Rozenblit
Department of Electrical and Computer Engineering
University of Arizona
1230 E Speedway boulevard, Tucson, AZ, USA
jr@ece.arizona.edu

Allan J. Hamilton
Department of Surgery
University of Arizona
1501 N. Campbell avenue, Tucson, AZ, USA
allan@surgery.arizona.edu

ABSTRACT

An assessment system is presented to provide objective assessment results for laparoscopic surgery skills training. Six innovative evaluation metrics are introduced in the design of the proposed system. Like a video game, the system suggests *achievable* goals that are used to define evaluation metrics for a trainee who is performing a particular training task. To implement the proposed system, a design method of the hierarchical fuzzy system is used. The simulation results and the prototype illustrate the feasibility of the proposed evaluation approach.

Keywords: Surgical Training, Objective Assessment, Hierarchical Fuzzy Inference System.

1 INTRODUCTION

Simulation-based training is widely used in various fields such as first-aid training, aviation, and military. For laparoscopic surgery skills training, many training simulators that range from cost effective trainers (eoSurgical 2016; Jaber 2010) to sophisticated virtual reality (VR) simulators (Simbionix 2016; Stylopoulos et al. 2004) have been proposed. Low cost devices do not provide any guidance feedback and objective assessment results. In contrast, several VR trainers provide not only assistance information using graphical interface and haptic feedback (e.g., force feedback) to a trainee but they also give feedback on performance outcomes.

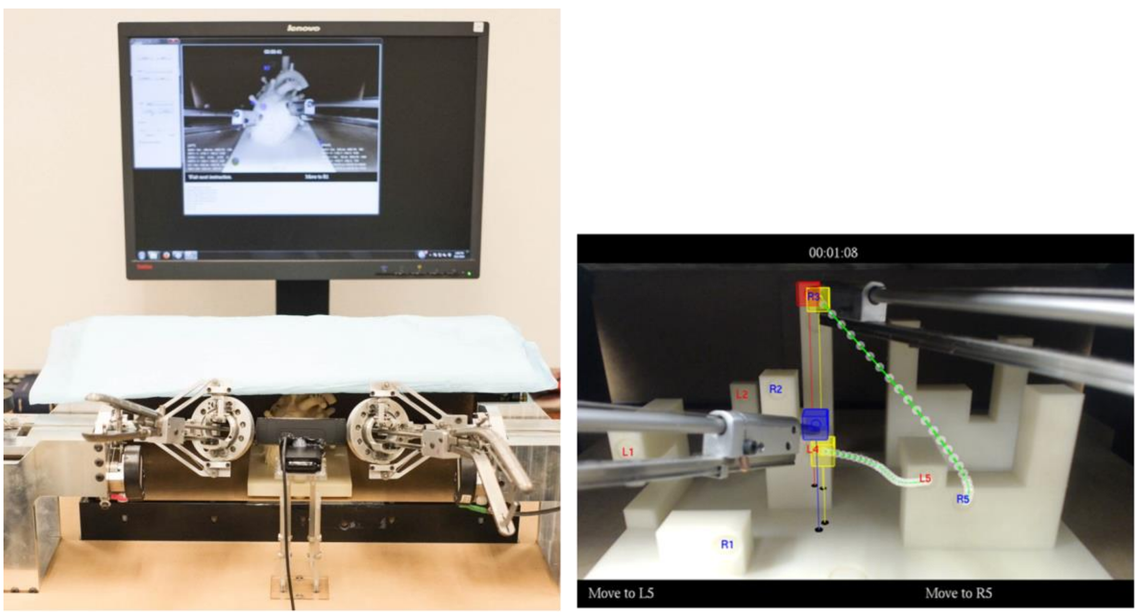
To provide a better training environment, Computer Assisted Surgical Trainer (CAST) (Rozenblit et al. 2014) has been developed by incorporating several key features from both cost effective kits and the VR simulators. CAST provides realistic training environment with visual and haptic guidance to trainees. It consists of hardware, software, and electro-mechanical components. There are two mechanical fixtures to hold real surgical instruments, an exchangeable cassette to provide various practice scenarios, a web camera to display a training scenario, and motors with electronics to support haptic feedback and well as visual guidance.

SpringSim-MSM 2017, April 23-26, Virginia Beach, VA, USA

There are four software modules, *opt*MIS, *opt*Viz, *opt*Guide, and *opt*Assessment. The *opt*MIS is a collision free and shortest path generator to provide recommended trajectories for a training scenario (Napalkova et al. 2014). The *opt*Viz (Rozenblit et al. 2014) and *opt*Guide (Hong and Rozenblit 2016) are responsible for visual guidance and haptic guidance, respectively. Figure 1 illustrates the CAST system and an example of visual guidance. The *opt*Assessment quantifies objectively trainee's competency (Riojas et al. 2011). To improve the *opt*Assessment, in this paper, we propose a simulation-based assessment system.

Objective assessment metrics for laparoscopic surgery skills training have been proposed by several researchers (Chang et al. 2016; Cotin et al. 2002; Kowalewski et al. 2014; Maithel et al. 2006; Retrosi et al. 2015; Reiley et al. 2011; Oropesa et al. 2014; Ritter and Scott 2007). Some proposed to design a scoring system to classify trainees' proficiency of surgical movements (Reiley et al. 2011; Oropesa et al. 2014). Others proposed assessment metrics to design a better training curriculum and a training system for medical students (Cotin et al. 2002; Ritter and Scott 2007).

In (Oropesa et al. 2014), three classifiers (linear discriminant analysis, support vector machine, and adaptive neuro-fuzzy inference systems) were developed to investigate the performance of the proposed classifiers that categorize participating groups. Several participants (i.e., 42 people divided into an expert group and a novice/intermediate group) were enrolled to verify the proposed classifiers. Also, the classifier's results were compared with pre-categorized groups to verify the methods. In (Ritter and Scott 2007), the authors designed a new training curriculum based on proficiency levels. To design the new curriculum, the authors hired five experts and asked them to perform 5 fundamentals of laparoscopic surgery (FLS) (FLS 2016) tasks several times. Based on the measured completion time for the training tasks, the authors proposed the curriculum for novice trainees.



(a) (b)

Figure 1: (a) Computer Assisted Surgical Trainer (CAST) and (b) an example of visual guidance.

In this paper, we propose a novel assessment system using a hierarchical fuzzy inference system (HFIS) with several assessment metrics. The remainder of this paper is organized as follows: the assessment metrics are presented in Section 2. In Section 3, the proposed hierarchical fuzzy assessment system is described. Simulation and implementation result are given in Section 4. In Section 5, discussion and conclusion are presented.

2 OBJECTIVE ASSESSMENT METRICS

Movement of economy (path length) and completion time are widely used as common metrics for laparoscopic surgery skills evaluation. Additional considerations are peak speed, average speed, and idle time (Riojas et al. 2011; Oropesa et al. 2014). We select several metrics from those measures and introduce a feature from video games to design the hierarchical fuzzy assessment system for the CAST system.

Generally, a video game has four basic elements – play, pretending, goal(s), and rules (Adams 2014). Among them, a goal of the game is highly related to the assessment system. In (Moreno-Ger et al. 2008), adaptation and assessment were presented as requirements for an educational game design. By providing a goal to a user while performing a game, the game engine can evaluate progress (e.g., achieving a goal) and provide a next goal based on the assessment results.

Like a game, CAST provides a goal to a trainee and evaluates the trainee's progress. The most important action of laparoscopic surgery training is moving an instrument accurately with reasonable speed. Therefore, CAST suggests a target speed with a recommended path as goals for a trainee to move an instrument while performing a particular training scenario. Based on these two basic goals, we consider six evaluation metrics as follows.

Average speed ratio: The proposed simulation-based assessment system provides a recommended movement to a user. By using average speed, the CAST system provides a "mission" to a trainee (e.g., "Please move an instrument from position A to position B with 5cm/s speed"). As mentioned above, in laparoscopy training, one of the training goal is to learn how to maneuver a surgical instrument with reasonable speed (i.e., not too fast and not too slow). If the average speed of the instrument movement is too slow while performing real surgery, it may mean that it takes more time to complete the surgery even though a surgeon minimizes the chance of erroneous moves. In contrast, if the average speed is too fast, it may cause critical mistake like hitting organs and tissues even though the operation is finished quickly. To evaluate a trainee's performance in terms of the speed, an average speed ratio is proposed as follows:

$$Speed\ Ratio(SR) = \frac{Reference\ average\ speed\ - Actual\ Average\ speed}{Reference\ average\ speed}$$

where the reference average speed is provided by the CAST system as a target goal and the actual average speed is calculated by capturing the movement of the instrument tip. The range of SR is $[-L, L]$ where L is a constant value. Using this metric, we can evaluate a trainee's movement. An example of the evaluation is as follows:

$$evaluation\ result = \begin{cases} Good\ movement\ but\ fast\ movement & if\ \ SR < 0 \\ Good\ movement & if\ \ SR \to 0 \\ Bad\ movement & if\ SR \to L. \end{cases}$$

Completion time ratio: One of the basic assessment metrics for laparoscopic surgery training is the completion time. For example, hands-on exams of the fundamentals of laparoscopic surgery (FLS) have their own maximum time limit. Whenever a trainee performs a particular practice scenario, CAST measures completion time for this metric. This completion time is compared to a recommended completion time using the formula:

$$Time\ Ratio(TR) = \frac{Actual\ completion\ time\ - Reference\ completion\ time}{Reference\ completion\ time}$$

where $-L \leq TR \leq L$ and L is a constant value. Like using speed ratio, we can evaluate a trainee's movement as follows:

$$evaluation\ result = \begin{cases} Good\ movement\ but\ fast\ movement & if\ TR < 0 \\ Good\ movement & if\ TR \rightarrow 0 \\ Bad\ movement & if\ TR \rightarrow L. \end{cases}$$

Economy of movement (Path length) ratio: In CAST system, *opt*MIS provides a recommended path to assist a trainee. The trainee maneuvers a surgical instrument to complete a training practice. By comparing the actual user's traversed path length with a recommended path length, we can assess a trainee's performance.

$$Path\ Length\ Ratio(PLR) = \left|\frac{ActualPathLength\ - ReferencePathLength}{ReferencePathLength}\right| \geq 0$$

If PLR is close to zero (i.e., $PLR \rightarrow 0$), then we can conclude that a trainee's performance is good. However, if PLR is a large value (i.e., $PLR \rightarrow L$ where L is a constant value), then a trainee's performance is less strong.

Idle time ratio: In order to complete a practice within a time limit, it is necessary for a trainee to move a surgical instrument with reasonable and constant speed. If movement speed is too slow, we consider this a "stop" (i.e., holding an instrument) motion. The stop motion may indicate that a user has difficulty to maneuver an instrument. In this paper, we assume that a user's performance is not good if there are many stop. To evaluate this, the formula below is used.

$$Idle\ Time\ Ratio(ITR) = \frac{Idle\ Time}{Reference\ completion\ time} \geq 0$$

where $Idle\ Time$ is measured when the speed is less than a certain value (e.g., $speed < 0.3cm/s$). Like PLR, if ITR is close to zero, then we can say that a trainee performs well. Otherwise, we can say that a trainee has some difficulties while performing a training task.

Deviation ratio: An ideal movement of the surgical instrument is precisely traversing a recommended path generated by *opt*MIS. If a trainee is a novice, he or she may have a lot of difficulties and this causes large deviations from the recommended path. In actual surgery, huge deviation may cause life critical issues. Therefore, we consider a deviation ratio as an evaluation metric. To calculate the deviation ratio, we introduce a good movement counter and a bad movement counter. Generally, it is challenging to reach the ideal movement (i.e., deviation is zero). In order to take into account the general case, we set a reasonable distance bound (ε) to increase the bad movement counter. The instrument tip position (p_{tip}) is captured every sampling period. Also, using this updated tip position, the nearest point (p_{near}) on the recommended path is determined and it is used to calculate the deviation ($|p_{tip} - p_{near}|$). Each counter is updated as follows.

$$update\ counters: \begin{cases} increase\ good\ movement\ counter & if\ deviation < \varepsilon \\ increase\ bad\ movement\ counter & if\ deviation \geq \varepsilon \end{cases}$$

where $deviation$ is the Euclidian distance between the recommended path and the instrument tip. Based on these two counters, the deviation ratio is calculated as follows.

$$Deviation\ Ratio(DR) = \frac{\#\ of\ bad\ movement\ count}{\#\ of\ good\ movement\ count + \#\ of\ bad\ movement\ count} \geq 0$$

If the number of bad movements count is zero, then the deviation ratio is zero and it shows that a user's performance is good. However, if a user makes huge deviations, then the deviation ratio is a large value and it represents bad performance.

Direction profile ratio: This metric is responsible for assessing movement of the instrument by using direction vectors. Two vector sets, recommended direction vectors and actual instrument's direction vectors, are used to calculate the direction profile ratio. Figure 2 illustrates an example of the direction vectors. This direction profile indicates deviation and moving direction. For this metric, cosine of angle is calculated using two vectors (i.e., $\cos(\theta) = |\vec{r}||\vec{a}|/\vec{r} \cdot \vec{a}$ where $\vec{r}$ and $\vec{a}$ are a recommended direction vector and an actual instrument tip's direction vector, respectively). If the instrument tip traverses right on the path, the cosine of the angle is close to 1. In contrast, if the instrument tip deviates from the path, then the cosine of the angle is close to 0. Also, if the instrument tip moves backward direction, the value is close to -1. While a trainee performs a particular task, CAST system captures an instrument tip movement and calculates the cosine of the angle. The direction profile ratio is formed as follows.

$$Direction\ profile\ Ratio(DPR) = \frac{\sum_{i=1}^{n}(\cos\theta_{ref} - \cos\theta_i)}{\sum_{i=1}^{n}\cos\theta_{ref}}$$

where θ_{ref} is a target angle, θ_i is an angle between a recommended direction vector and an actual tip direction vector for i^{th} sample, and n is the number of samples. For instance, if a target angle is 0 (i.e., the goal of the task is to traverse on a recommend path) and DPR is close to zero, then it represents good movement. However, if DPR is close to 1 with same target angle, then it represents bad movement.

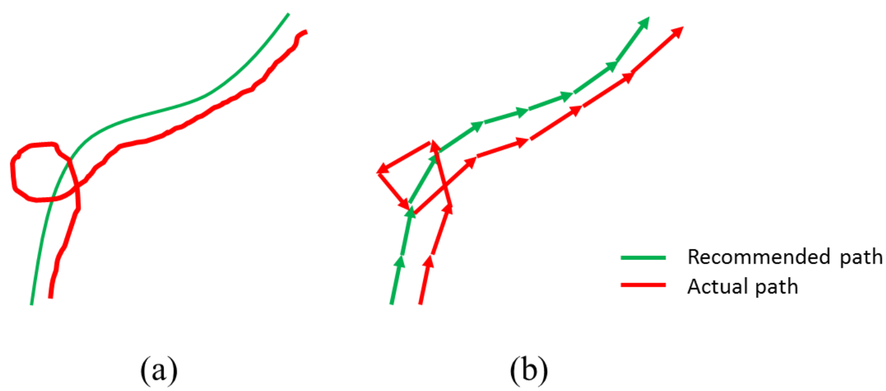
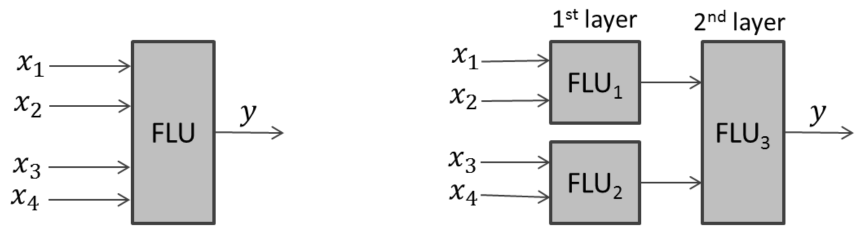


Figure 2: An example of the direction profile: (a) recommended path and actual path and (b) direction vectors for both recommended path and actual path.

3 HIERARCHICAL FUZZY ASSESSMENT SYSTEM

For the existing *opt*Assessment, five metrics (movement economy ratio, movement direction profile, peak speed width, continuity of movement, and completion time) were used to implement a scoring system (Riojas et al. 2011). To imitate human's judgement, a fuzzy logic inference system was used for the scoring system. The rule base of this fuzzy system consists of single-input and single-output rules. The weakness of this rule base is that it is difficult to take into account interrelationships among metrics. To overcome this drawback, multi-input and single-output fuzzy inference system may be used. However, this system also has a weakness, such as increasing the number of rules dramatically if there are a lot of inputs with several membership functions.

To overcome those shortcomings, we propose to use a hierarchical fuzzy logic inference system (HFLIS) (Lee et al. 2003) with the metrics specified in Section 2. The advantage of the HFLIS is that it is possible to reduce the number of rules dramatically without performance degradation. For instance, if four evaluation metrics are used and each metric is characterized by five membership functions, 625 (5^4) rules are needed for the rule base. However, if HFLIS is used for the same metrics with the same membership functions, only 75 (3×5^2) rules are needed for the rule base. The first layer's outputs are used as inputs of the second layer. In this case, the output of the first layer is characterized by five membership functions. Figure 3 illustrates examples of using the single layer fuzzy inference system and using the HFLIS.



- 4 input variables
- 5 membership functions
- 5^4 Rules

(a)

- 4 input variables
- 5 membership functions for all FLUs
- 3×5^2 Rules

(b)

Figure 3: (a) Single layer fuzzy logic system vs. (b) Hierarchical fuzzy logic system (Lee et al. 2003).

Our proposed hierarchical scoring system shown in Figure 4 consists of five fuzzy logic units (FLUs) that have their own fuzzifier, defuzzifier, rule base, and a fuzzy inference engine. There are 6 inputs and a single output to provide an evaluation score. For the inputs of the scoring system, proper membership functions are modeled to characterize for all layer's inputs. In section 2, ranges for the six evaluation metrics are presented (e.g., $-L \leq TR \leq L$). We use limited ranges of inputs to design the proposed scoring system based on human's knowledge. For instance, if the actual path is twice as long as the reference path, we can clearly conclude that the performance is not good. Therefore, the region of interest for PLR is [0, 1]. Similarly proper membership functions are designed based on human knowledge. Three linguistic terms (Good (G), Normal (N), and Bad (B)) are used to define inputs for PLR, ITR, DRD, and PR with trapezoidal membership functions (Figure 5-(a)). These three terms are also used for inputs of FLU_4 and FLU_5 with triangular membership functions (Figure 5-(b)). The output of FLU_1 and the output of FLU_2 are inputs of FLU_4. Also, the output of FLU_4 and the output of FLU_3 are inputs of FLU_5. Similarly, four linguistic terms (Fast (F), Good (G), Normal (N), and Bad (B)) are used to define SR and TR with trapezoidal membership functions (Figure 5-(c)).

For an output variable (Y_k) for each FLU_k where k is FLU index ($k = 1,2,3,4, and\ 5$), singleton membership functions are used to characterize the output fuzzy sets; Y_k is restricted to $[0,\ 1]$. Fuzzy IF-THEN rules for all FLUs are as follows:

$$R_{k,l}\text{: If } X_k^1 \text{ is } F_{k,l}^1 \text{ and } X_k^2\ is\ F_{k,l}^2, then\ Y_{k,l}\ is\ G_{k,l}$$

where k is FLU index, l is the rule number, X_k^1 and X_k^2 are inputs for FLU_k, $F_{k,l}^1$ and $F_{k,l}^2$ are linguistic terms, $Y_{k,l}$ is the l^{th} rule output for FLU_k, and $G_{k,l}$ is a constant output value ($0 \leq G_{k,l} \leq 1$). Three linguistic terms are used to define antecedent variables for FLU_1, FLU_2, FLU_4, and FLU_5. Four linguistic terms are used for FLU_3. Therefore, the number of rules is 52 ($4\times 3^2 + 4^2$).

To generate a score, rule-bases are designed as follows: if an instrument's tip movement is good, then a large value (i.e., close to 1) is assigned to provide a high score. If the movement is bad, a small value (i.e., close to 0) is assigned to report that the performance is not good. Fast movement may cause mistakes like hitting an object. Therefore, we assign a penalty in this case, which is applied in the rule table for FLU_3. Based on these considerations, all rule tables are designed as presented in Table 1, Table 2, and Table 3.

A singleton fuzzifier, product inference engine, and center average defuzzifier are used to calculate Y_k as given by the following equation.

$$Y_k = \frac{\sum_{l=1}^{m} Y_{k,l}\mu_{k,l}(X_k^1)\mu_{k,l}(X_k^2)}{\sum_{l=1}^{m} \mu_{k,l}(X_k^1)\mu_{k,l}(X_k^2)}$$

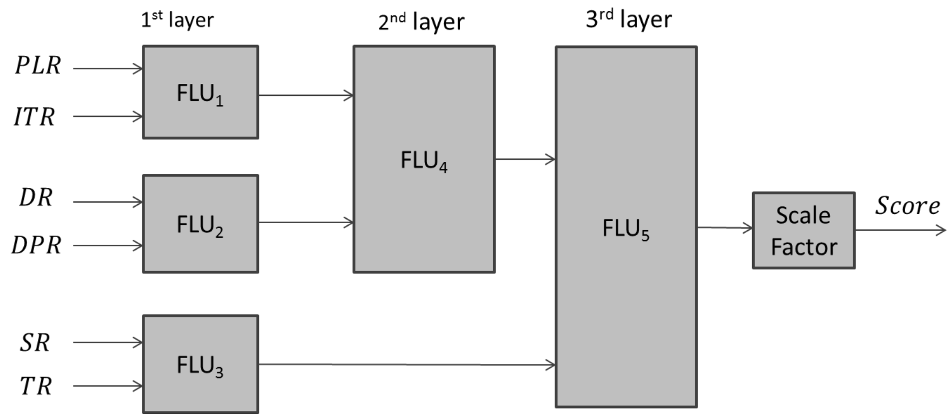


Figure 4: Hierarchical fuzzy systems for a new *opt*Assessment.

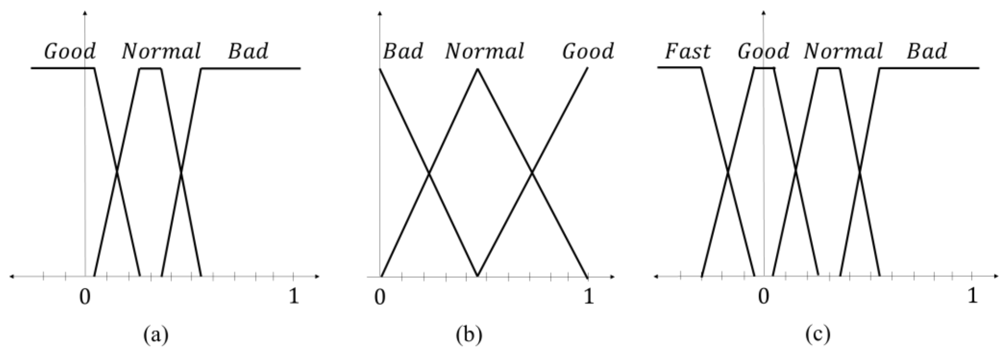


Figure 5: (a) Membership functions for PLR, ITR, DRD, and PR, (b) membership functions for inputs of FLU_4 and FLU_5, and (c) membership functions for SR and TR.

where $\mu_{k,l}(X_k^1)$ and $\mu_{k,l}(X_k^2)$ are the degrees of membership functions for two inputs, X_k^1 and X_k^1, respectively, $Y_{k,l}$ is the singleton output value, and m is the number of rules for each FLU ($m = 9$ when $k = 1,2,4, and$ 5 and $m = 16$ when $k = 3$).

The final score is generated based on Y_5, where Y_5 is a bounded value from 0 to 1 ($0 \le Y_5 \le 1$). To provide a score that ranges from 0 to 100, a constant scaling factor is used.

Table 1: Rule table for FLU_1 and FLU_2

x_1 \ x_2	Good	Normal	Bad
Good	1.00	0.70	0.30
Normal	0.70	0.50	0.15
Bad	0.30	0.15	0.00

Table 2: Rule table for FLU_3

x_1 \ x_2	Fast	Good	Normal	Bad
Fast	0.9	0.8	0.6	0.2
Good	0.8	1.0	0.7	0.3
Normal	0.6	0.7	0.5	0.15
Bad	0.2	0.3	0.5	0.0

Table 3: Rule table for FLU_4 and FLU_5

x_1 \ x_2	Bad	Normal	Good
Bad	0.00	0.15	0.30
Normal	0.15	0.50	0.70
Good	0.30	0.70	1.00

4 SIMULATION AND EXPERIMENTAL RESULTS

To verify the proposed assessment system, a simple hand-eye coordination task was designed. In the training scenario, a trainee must traverse the space along the recommend trajectories to touch multiple targets (e.g., four targets labeled by R1, R2, R3, and R4) using the right instrument as shown in Figure 1-(b).

The goal for this task is given by the CAST system to assist a trainee. For instance, CAST suggests average moving speed (e.g., 1.0cm/s), completion time (e.g., 7.1 seconds), path length (e.g., 7.07 cm), idle time duration (e.g., 0 seconds), a distance bound (e.g., $\varepsilon = 0.5cm$) for a deviation, and a target angle (e.g., $\theta_{ref} = 0°$) for direction profile to move from R1 to R2, and to touch R2. While performing this task, actual movements of the right instrument tip are collected every 50 milliseconds by using encoders. Figure 6 illustrates the simulation results for two cases. The blue and red lines represent a recommended path and the actual instrument tip movement, respectively. The example depicts two different scenarios – good movement and bad movement. Intuitively, the system has to assign a high score for the good movement and report a poor grade for the bad movement. The simulation results show this concept well. Even though the average speed is relatively fast vis a vis the goal speed, other metrics clearly show that the overall performance is bad (see Figure 6-(a)). Therefore, the proposed assessment system reports lower score (9.7715) as the evaluation result. In contrast, Figure 6-(b) illustrates that the overall performance is good and the score (89.6829) is close to 100.

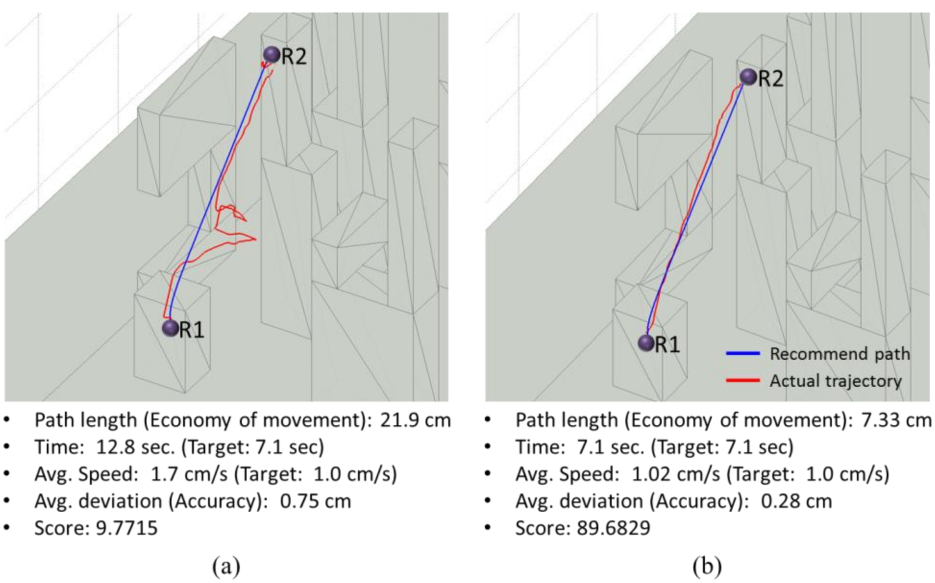


Figure 6: Simulation results for (a) bad performance and (b) good performance.

Based on these simulation results, the proposed assessment system is implemented in the CAST system. Figure 7 depicts the prototype of the implementation. There are four targets with three recommended paths. The system provides goals for a trainee. For each sub task (e.g., move to R2 from R1), the proposed system reports individual assessment results. To provide a user-friendly interface, the assessment report also contains actual training results with the recommended goals (e.g., actual completion time is 6.25 seconds while moving from R2 to R3. The recommend completion time is 2.75 seconds.). For the final score, the system presently takes the average score from all scores for sub-tasks (e.g., final score = 32.93 = (32.86+25.57+40.36)/3). We will investigate the better method to provide the final score (e.g., The median value may be used instead of the average score.).

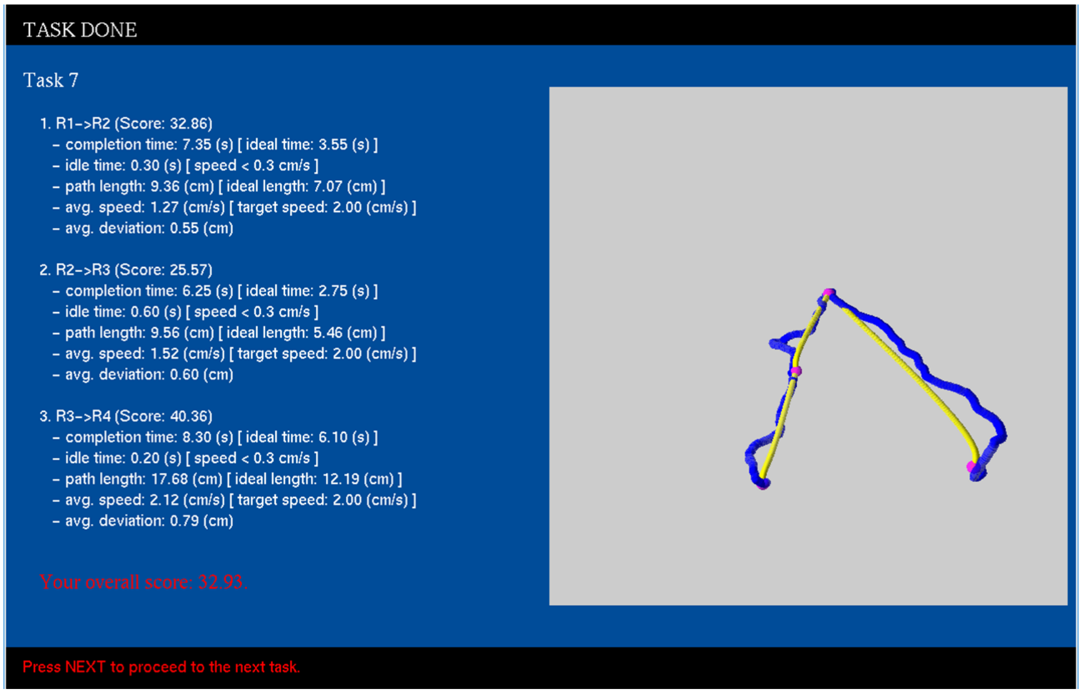


Figure 7: Prototype assessment report by the proposed assessment system**.**

5 DISCUSSION AND CONCLUSION

In this paper, we have explained a design method of the assessment system based on a hierarchical fuzzy system. To design the proposed system, we have introduced six assessment metrics. Due to using HFLIS, we can easily add more evaluation metrics if these are necessary. For example, if we consider motion smoothness and average acceleration as new metrics, we only design an individual FLU for two metrics and place the new FLU into the first layer. In this case, we can reuse the upper layer's FLUs to build the entire HFLIS. Also, if we use an adaptive technique to determine membership functions and linguistic terms with lots of training data (e.g., results from a human subjects study), it may be possible to design a more accurate evaluation system.

The proposed assessment system provides an achievable goal to a trainee while he or she performs a particular training task. This goal enables a trainee to understand the training progress clearly by comparing outcomes with the provided goal. As for the future work, we will consider how to design an achievable goal based on experimental data that is collected by conducting a human subjects study. Also, the proposed system will be used to design a task generator that creates a new task once an exercise at hand has been completed. For instance, if a trainee has any difficulty while performing a particular task, the evaluation score may be low. In this case, the task generator might generate a somewhat easier training set. Finally, we will improve the graphic user interface to visualize assessment results more intuitively.

The simulation results and the prototype implementation indicate that the proposed system is feasible to assess laparoscopic surgery skills. The key contributions of this work are introducing innovative evaluation metrics with achievable goals and designing HFLIS that has a potential to introduce adaptive features. This work can be applied to any applications that need assessment score by redefining several metrics with new goals.

ACKNOWLEDGMENTS

This work has been supported by the National Science Foundation grant no. 1622589 “Computer Guided Laparoscopy Training” and the Raymond J Oglethorpe Foundation.

REFERENCES

Adams, E. 2014. “Fundamentals of Game Design”. *Pearson Education*.

Chang, J., D. C. Banaszek, J. Gambrel, and D. Bardana. 2016. “Global Rating Scales and Motion Analysis Are Valid Proficiency Metrics in Virtual and Benchtop Knee Arthroscopy Simulators”. *Clinical Orthopaedics and Related Research*, vol. 474, no. 4, pp.956-964.

Cotin, S., N. Stylopoulos, M. Ottensmeyer, P. Neumann, et al. 2002. “Metrics for Laparoscopic Skills Trainers: The Weakest Link!”. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 35-43.

eoSurgical. 2016, http://www.eosurgical.com. Accessed Dec. 2, 2016.

FLS. 2016, http://www.flsprogram.org. Accessed Dec.2, 2016.

Hong, M. and J. W. Rozenblit. 2016. “A Haptic Guidance System for Computer Assisted Surgical Training Using Virtual Fixtures”. *In 2016 IEEE International Conference on Systems, Men, and Cybernetics (SMC)*, pp. 2230-2235.

Jaber, N. 2010. "The Basket Trainer: A Homemade Laparoscopic Trainer Attainable to Every Resident". *Journal of Minimal Access Surgery* vol. 6, no. 1, pp. 3-5.

Kowalewski, T. M., L. W. White, T. S. Lendvay, I.S. Jiang, R. Sweet, A. Wright, B. Hannaford, and M. N. Sinanan. 2014. “Beyond task time: automated measurement augments fundamentals of laparoscopic skills methodology”. *Journal of Surgical Research*, vol. 192, no. 2, pp.329-338.

Lee, M. L., H. Y. Chung, and F. M. Yu. 2003. “Modeling of Hierarchical Fuzzy Systems”. *Fuzzy sets and systems* vol. 138, no. 2, pp. 343-361.

Maithel, S., R. Sierra, J. Korndorffer, P. Neumann, S. Dawson, M. Callery, D. Jones, and D. Scott. 2006. “Construct and face validity of MIST-VR, Endotower, and CELTS”. *Surgical Endoscopy and Other Interventional Techniques*, vol. 20, no. 1, pp.104-112.

Moreno-Ger, P., D. Burgos, I. Martínez-Ortiz, J. L. Sierra, et al. 2008. “Educational Game Design for Online Education”. *Computers in Human Behavior* vol. 24, no. 6, pp. 2530-2540.

Napalkova, L., J. W. Rozenblit, G. Hwang, A. J. Hamilton, and L. Suantak. 2014. "An Optimal Motion Planning Method for Computer-Assisted Surgical Training". *Applied Soft Computing* vol. 24, pp. 889-899.

Oropesa, I., P. Sánchez-González, M. K. Chmarra, P. Lamata, R. Pérez-Rodríguez, et al. 2014. “Supervised Classification of Psychomotor Competence in Minimally Invasive Surgery Based On Instruments Motion Analysis”. *Surgical endoscopy* vol. 28, no. 2, pp. 657-670.

Reiley, C. E., H. C. Lin, D. D. Yuh, and G. D. Hager. 2011. “Review of Methods for Objective Surgical Skill Evaluation”. *Surgical endoscopy* vol. 25, no. 2, pp. 356-366.

Retrosi, G., T. Cundy, M. Haddad, and S. Clarke. 2015. “Motion Analysis–Based Skills Training and Assessment in Pediatric Laparoscopy: Construct, Concurrent, and Content Validity for the eoSim Simulator”. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, vol. 25, no. 11, pp.944-950.

Riojas, M., C. Feng, A. J. Hamilton, and J. W. Rozenblit. 2011. "Knowledge Elicitation for Performance Assessment in a Computerized Surgical Training System". *Applied Soft Computing* vol. 11, no. 4, pp. 3697-3708.

Ritter, E. M. and D. J. Scott. 2007. "Design of a Proficiency-based Skills Training Curriculum for the Fundamentals of Laparoscopic Surgery". *Surgical innovation* vol. 14, no. 2, pp.107-112.

Rozenblit, J. W., C. Feng, M. Riojas, L. Napalkova, A. J. Hamilton, M. Hong, P. Berthet-Rayne, P. Czapiewski, G. Hwang, J. Nikodem, and A. Shankaran. 2014. "The Computer Assisted Surgical Trainer: Design, Models, and Implementation". *Proc. of the 2014 Summer Simulation Conference, Modeling and Simulation in Medicine*, pp. 211-220.

Simbionix. 2016, http://www.simbionix.com. Accessed Dec.2, 2016.

Stylopoulos, N., S. Cotin, S. K. Maithel, M. Ottensmeyer, P. G. Jackson, R. S. Bardsley, P. F. Neumann, D. W. Rattner, and S. L. Dawson. 2004. "Computer-enhanced Laparoscopic Training System (CELTS): Bridging the Gap". *Surgical Endoscopy and Other Interventional Techniques* vol. 18, no. 5, pp.782-789.

AUTHOR BIOGRAPHIES

MINSIK HONG is a Ph. D. candidate at the University of Arizona. He received a Master of Science degree in Electrical and Computer Engineering from POSTECH, Republic of Korea. His research interests are robotics, control system, fuzzy theory, and modeling and simulation for medical devices. His email address is mshong@email.arizona.edu.

JERZY W. ROZENBLIT is University Distinguished Professor, Raymond J. Oglethorpe Endowed Chair in the Electrical and Computer Engineering (ECE) Department, with a joint appointment as Professor of Surgery in the College of Medicine at the University of Arizona. During his tenure at the University of Arizona, he established the Model-Based Design Laboratory with major projects in design and analysis of complex, computer-based systems, hardware/software codesign, and simulation modeling. He presently serves as Director of the Life-Critical Computing Systems Initiative, a research enterprise intended to improve the reliability and safety of technology in healthcare and life-critical applications. His email address is jr@ece.arizona.edu.

ALLAN J. HAMILTON, MD, FACS is Harvard trained physician, a professor of Neurosurgery at the University of Arizona. Dr. Hamilton was elected a Fellow of the American College of Surgeons in 1994. In 1995, Dr. Hamilton was promoted to Chief of Neurosurgery and became the Chairman of the entire Department of Surgery in 1998. He currently holds a tenured professorship in Neurosurgery, as well as additional professorships in the Departments of Psychology, Radiation Oncology, and the School of Electrical and Computer Engineering. He is Executive Director of the Arizona Simulation Technology and Education Center in the College of Medicine, University of Arizona. His email address is allan@surgery.arizona.edu.