Feasibility of object detection for skill assessment in central venous catheterization

Olivia O'Driscoll[a], Rebecca Hisey[a], Matthew Holden[b], Daenis Camire[c], Jason Erb[d], Daniel Howes[d], Tamas Ungi[a], Gabor Fichtinger[a]

[a]Laboratory for Percutaneous Surgery, Queen's University

[b]School of Computer Science, Carleton University

[c]Department of Anesthesiology, Queen's University

[d]Department of Critical Care Medicine, Queen's University

ABSTRACT

Purpose: Computer-assisted surgical skill assessment methods have traditionally relied on tracking tool motion with physical sensors. These tracking systems can be expensive, bulky, and impede tool function. Recent advances in object detection networks have made it possible to quantify tool motion using only a camera. These advances open the door for a low-cost alternative to current physical tracking systems for surgical skill assessment. This study determines the feasibility of using metrics computed with object detection by comparing them to widely accepted metrics computed using traditional tracking methods in central venous catheterization. **Methods:** Both video and tracking data were recorded from participants performing central venous catheterization on a venous access phantom. A Faster Region-Based Convolutional Neural Network was trained to recognize the ultrasound probe and syringe on the video data. Tracking-based metrics were computed using the Perk Tutor extension of 3D Slicer. The path length and usage time for each tool were then computed using both the video and tracking data. The metrics from object detection and tracking were compared using Spearman rank correlation. **Results:** The path lengths had a rank correlation coefficient of 0.22 for the syringe ($p<0.03$) and 0.35 ($p<0.001$) for the ultrasound probe. For the usage times, the correlation coefficient was 0.37 ($p<0.001$) for the syringe and 0.34 ($p<0.001$) for the ultrasound probe. **Conclusions**: The video-based metrics correlated significantly with the tracked metrics, suggesting that object detection could be a feasible skill assessment method for central venous catheterization.

Keywords: Object detection, computer-assisted skill assessment, central venous catheterization

INTRODUCTION

Feedback is a critical component of medical trainee education and is shown to decrease complication rates [1]. Currently, skill assessment is largely performed by expert onlookers through means like checklists, global rating scales, and entrustment scores. Checklists measure the trainee’s compliance to a set of intervention-specific steps [2]. Global rating scales are based on intervention-specific markers of trainee performance [3]. Entrustment scores, reflect the expert reviewer’s confidence in the trainee to complete the procedure independently [4]. Each of these methods requires trainees to be continuously supervised by experts, which represents a substantial time burden on their behalf. Indeed, physicians widely cite limited time as primary obstacle to delivering feedback to trainees [5]. This limits trainee development because it is shown that trainees cannot gauge their own skill levels without this expert feedback [6]. Further, assessor-based skill assessment methods are vulnerable to subjective interpretation. This, combined with the time commitment required of the experts, contributes to the numerous shortcomings of assessor-based skill assessment.

Recent research has focused on developing methods of computer-assisted skill assessment that do not require expert observers. Such systems are based on quantitative metrics regarding tool handling and can more readily give the trainees feedback to improve their performance. The current gold standard for computer-assisted skill assessment requires tracking surgical tool motion with physical sensors to compute metrics that correlate with trainee skill. Since it is shown that objective feedback of technical skills is crucial to the structured learning of surgical skills, computer-assisted skill assessment methods further trainees' training through objective feedback without the time burden on expert observers [7].

Computer-assisted skill assessment methods that use physical tracking systems can take various forms, including optical and electromagnetic (EM) [8]. Using these systems, skill assessment metrics are measured with six degrees of freedom (DOF), and the metrics can include tool velocity and position. Specific work by Clinkard et al. used EM tracking to compute metrics for CVC, including the path lengths and usage times of the needle and ultrasound probe [9].

Medical simulation validity has numerous aspects, including: face validity, construct validity, predictive validity, and content validity. Face validity refers to the simulation's realism [10]. Construct validity relates to the simulator's ability to distinguish trainee participants from expert ones [10]. Predictive validity refers to the simulator's ability to predict performance in a real clinical setting [11]. Content validity relates to the simulation's representativeness of what it aims to measure [12]. Current approaches that depend on physical tracking sensors are of decreasing practicality. First, physical tracking sensors are expensive and can be burdensome to set up and calibrate. Second, the physical sensors change the feeling of the tools in the trainees' hands, and their overall experience, thereby reducing face validity. Finally, they reduce predictive validity by impeding the direct measurement of how competence translates from the simulation environment to patient care in a real clinical setting, where there are no physical sensors. Therefore, future work in the field of computer-assisted skill assessment must emphasize methods that are independent of physical tracking sensors.

To this end, more recent computer-assisted skill assessment methods use deep learning to identify each tool as it is being used. Hisey et al. used an image classification network detect tools used in central venous catheterization (CVC) using RGB webcam data [13]. The image classification network was integrated into a CVC training platform called Central Line Tutor, built on 3D Slicer (www.slicer.org): a free, open-source software platform designed for medical informatics, image analysis, and visualization. Central Line Tutor uses an RGB webcam and image classification to measure workflow compliance in CVC. However, its inability to quantify tool handling and motion is listed as a limitation.

In contrast to image classification, object detection allows us to identify the region of the image where surgical tools may be located. While the methods proposed in this paper focus on CVC, they could potentially apply to all ultrasound-guided interventions. Previous work on this specific project trained a Faster Region-based Convolutional Neural Network (Faster R-CNN) on a limited video dataset without expert participants [14]. The network detected the tools used in the procedure with reasonable accuracy. However, the skill assessment metrics computed in the previous study were not compared to any ground truth skill assessment metrics, so no skill assessment was performed.

This study focuses particularly on CVC, a procedure in which a catheter is inserted into a major vein, such as the internal jugular. The procedure, which is shared across many medical disciplines, is complicated by the fact that it is mandated to be performed under ultrasound guidance as well as its complex workflow [15]. Consequently, novice complications can be as high as 35%, whereas experienced clinicians have less than half of this complication rate [16,17]. The association between surgical skill and clinical outcomes stresses the need for continuous feedback and practice for trainees.

While previous computer-assisted skill assessment methods computed performance metrics with six DOF, we hypothesized that we could achieve a comparable skill assessment using a 3 DOF measurement that is most readily computed from a single RGB camera. Therefore, we measure if a low-cost, video-based tracking method could provide a comparable skill assessment with only three DOF. We analyze the feasibility of using video-based metrics by comparing them against the current gold standard metrics computed using EM tracking systems that have previously been validated against trainee skill. Using an expansive dataset containing both expert and novice participants, we assess whether it is possible to capture a similar assessment of skill using only a webcam, as opposed to an expensive tracking system.

METHODS

We obtained tracking and video recording of novices and experts performing CVC on the Central Line Tutor setup shown in Figure 1. The dataset recording setup consisted of an ultrasound machine, an EM tracker, a phantom, a webcam, and a display monitor. The ultrasound machine was connected to a computer, displaying the Central Line Tutor interface in 3D Slicer.

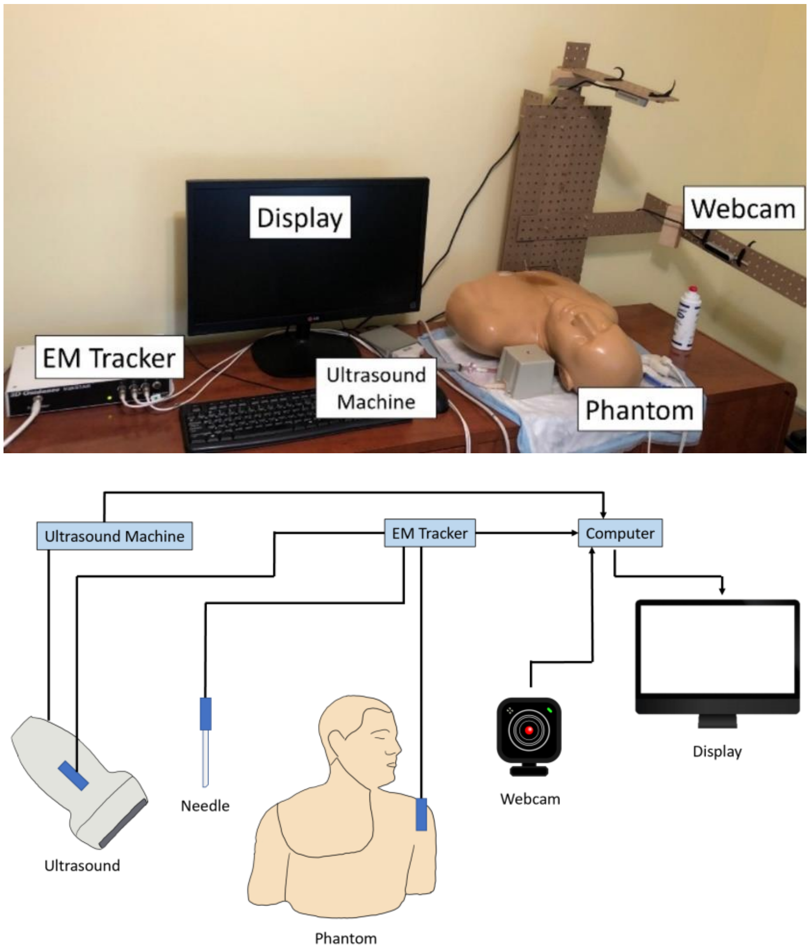

Figure 1. Dataset recording setup.

Since path length and usage time have previously been shown to correlate with skill in CVC, we compute both using our video-based object detection method and the EM tracking-based method presented by Clinkard et al. [9]. The EM tracking-based method then serves as a ground truth by which to assess the feasibility of our video-based method.

2.1 Video-based methods

An object detection network was used to compute bounding box predictions for the video-based metrics on the syringe and ultrasound probe. Specifically, Faster Region-Based Convolutional Neural Network (Faster R-CNN) was used [18]. In a Faster R-CNN, the input images are first passed to a Resnet-50. The Resnet-50 was pretrained on the ImageNet dataset to improve model performance via transfer learning. Then, a region proposal network generates region proposals, from which the ROI pooling layer extracts a fixed-length feature vector. These feature vectors are then classified according to the training classes. Additionally, CVC tool kits only contain one copy of each tool, so model accuracy was further improved by restricting the model to one prediction per class per frame.

The video-based path length for each class was defined the sum of Euclidean distances between the center of bounding boxes in sequential frames, measured in pixels. It was computed by Equation 1, where N is the total number of frames, i is the frame number, and d is the distance between sequential bounding boxes.

$$path\ length = \sum_{i=0}^{N} d(i, i+1) \qquad (1)$$

Video-based usage time was defined as Equation 2, which is the number of frames in which each tool was present, represented by n, divided by the number of frames per second. Both the video-based usage times and the path lengths were computed using the network's predicted bounding boxes.

$$usage\ time = \frac{n}{frame\ rate} \tag{2}$$

While Clinkard et al. tracked the needle tip, this study tracked the syringe instead. The syringe is larger and a more distinct color, so it is better suited to object detection. Because the needle tip is rigidly fixed to the syringe, we view tracking the syringe as analogous to tracking the needle tip.

2.2 Physical tracking methods

The metrics using EM tracking were computed using Perk Tutor (www.PerkTutor.org), an extension for training image-guided interventions in 3D Slicer [13]. We used the usage time and path length metrics. To obtain the usage time, we first defined a cube model in 3D Slicer that encapsulated the virtual model of the phantom, as shown in Figure 2. The usage time metric was then defined as the time the tool was in contact with the cube model. The tools' path lengths were defined as the total distance in millimeters travelled by the tool during the procedure.

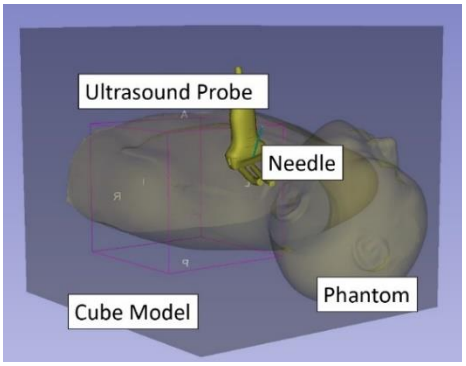

Figure 2. The 3D Slicer scene.

2.3 Experiments

We recorded video and tracking data of four novices and four experts performing CVC using the Central Line Tutor setup. The novices were medical students with no previous experience performing the procedure. The experts were attending anesthesiologists who were all certified to perform the procedure independently. Each participant performed 13 CVC trials on a venous access phantom, totaling 104 recordings. The video data was recorded with two RGB-depth cameras, however for this study only the RGB video recorded from the side camera was used. The recordings were framed with only the tools, phantom, and participants' hands in view to preserve anonymity, as seen in Figure 3.

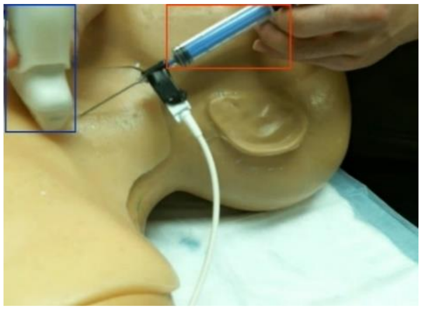

Figure 3. Ground-truth examples of the ultrasound probe (blue) and the syringe (red)

We separated the video recordings into individual frames in 3D Slicer using a module designed for deep learning training image collection (https://github.com/SlicerIGT/aigt/tree/master/DeepLearnLive). The entire visible portion of both the ultrasound probe and the syringe were annotated with a bounding box as they appeared in each frame of the recordings, meaning both tools could be present in the same frame. Figure 3 exemplifies how the classes were annotated. The dataset totaled 156,751 annotations.

The annotated images from the video data were used to train a Faster R-CNN in a leave-two-user-out cross-validation scheme, in which all 13 videos from both a single medical student and anesthesiologist comprised the test set. Since the probe is used more during the procedure than the syringe, it had greater representation in the dataset. We randomly down sampled the probe so there would be equal representation of both tools in the dataset.

The network's performance was measured using mean average precision (mAP), which is a standard network performance metric in object detection studies and challenges. MAP combines both the accuracy of the bounding box and its classification to quantify network performance. Intersection over union (IOU) is used to measure the accuracy of the bounding boxes. It is calculated as the area of overlap between the ground truth and predicted bounding boxes, divided by their combined area. In this study, an IOU greater than 0.5, as well as a correct classification, was considered a true positive sample. Correctly classified samples with bounding box IOUs less than or equal to 0.5 were considered false positives. False negatives were ground truth bounding boxes that the network did not detect. The mAP was calculated as the area under the precision/recall curve, where precision is the number of true positives divided by the sum of true and false positives. Recall is the number of true positives divided by the sum of true positives and false negatives.

A Spearman rank correlation was performed between all the metrics computed by the EM tracking method and the video-based method after skill assessment metrics were obtained with both methods. Both the usage time and path length metrics are generally correlated with the procedure time. Therefore, we also compared the correlations between different metrics for both tracking methods to test how well the video-based metrics captured the corresponding EM-tracked metrics. A Spearman rank correlation was chosen over a Pearson correlation because moving from a six DOF to three DOF system would result in a non-linear correlation.

For comparison on the same axes, both the EM-tracked and video-based metrics' values were normalized from zero to one by Equation 3, for every value, x_i , in the entire set of the specific performance metrics, x. The unnormalized values were used for the correlation.

$$x_{norm} = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{3}$$

RESULTS AND DISCUSSION

The syringe path lengths had a rank correlation coefficient of 0.22 (p<0.03). The syringe's usage times had a rank correlation coefficient of 0.37 (p<0.001). Figures 4 and 5 show the change in the syringe's normalized skill assessment metric values over 13 trials for novice and expert participants. The probe usage times had a rank correlation coefficient of 0.34 (p<0.001). The probe path lengths had a rank correlation coefficient of 0.35 (p<0.001). Figures 6 and 7 shows the change in skill assessment metrics for the ultrasound probe across 13 trials for novice and expert participants. Table 1 shows the correlations between the EM-tracked and video-based skill assessment metrics.

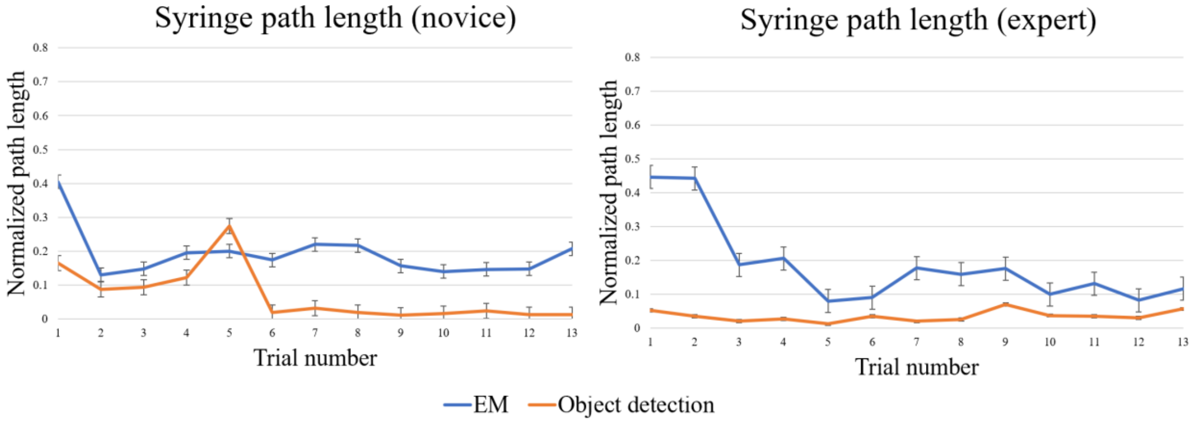

Figure 4. The normalized syringe path lengths as measured by both EM and object detection.

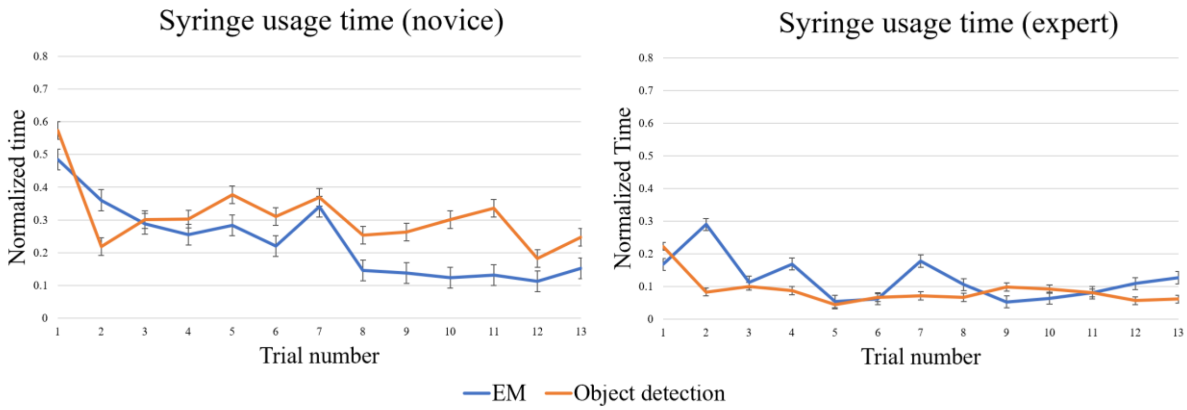

Figure 5. The normalized syringe usage times as measured by both EM and object detection.

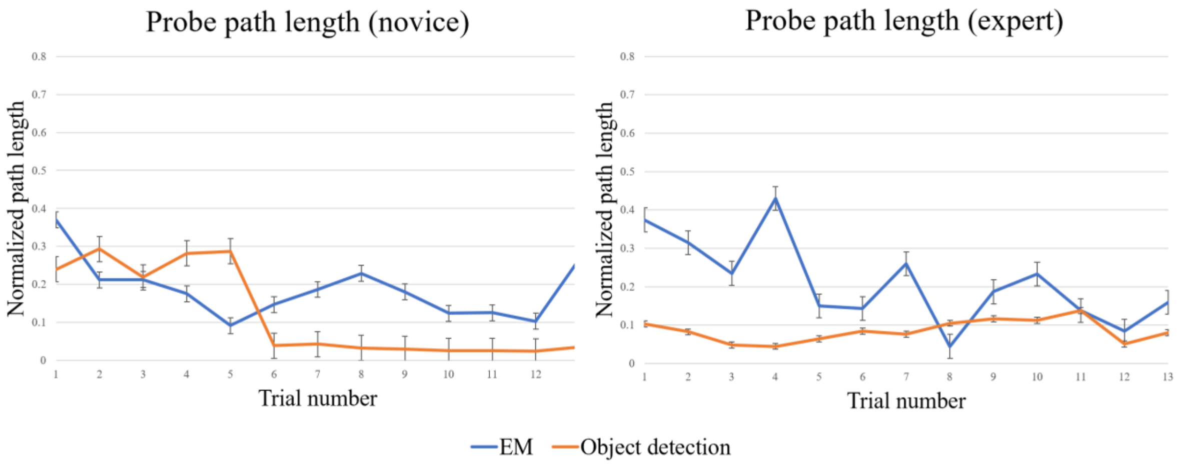

Figure 6. The normalized probe path lengths as measured by both EM and object detection.

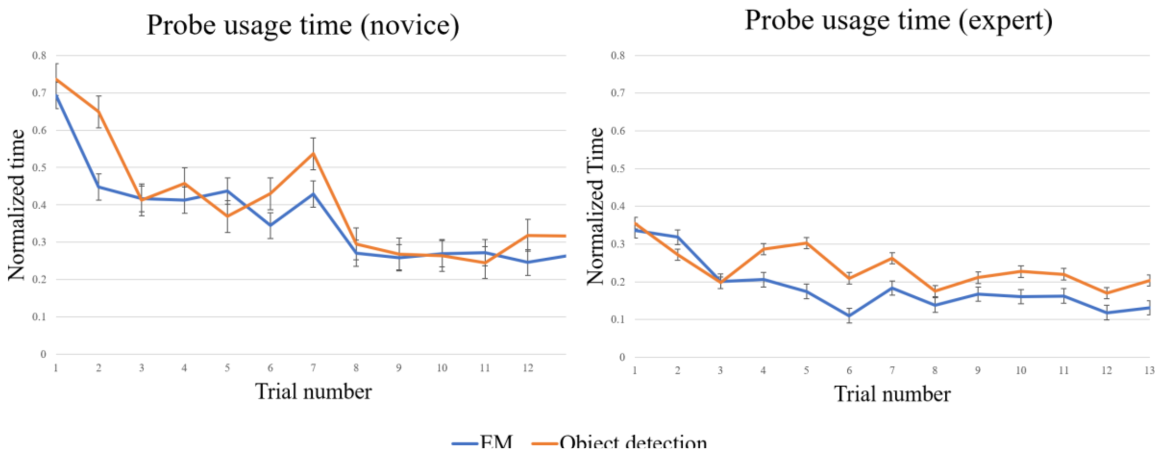

Figure 7. The normalized probe usage times as measured by both EM and object detection.

Video-based usage time and EM-tracked path length did not have significant correlation for the syringe (-0.09) or for the ultrasound probe (-0.01). Video-based path length and EM-tracked usage time did not have significant correlation for the syringe (0.17) or for the ultrasound probe (-0.05).

Table 1. The correlation between the EM-tracked and video-based metrics

Tool	Metric	ρ	P-value
Syringe	Path length	0.22	<0.03
	Usage time	0.37	<0.001
Probe	Path length	0.35	<0.001
	Usage time	0.34	<0.001

Table 2. The correlation between the alternately paired metrics

Tool	EM-tracked metric	Video-based metric	ρ	P-value
Syringe	Path length	Usage time	-0.09	>0.35
	Usage time	Path length	0.17	>0.08
Probe	Path length	Usage time	-0.01	>0.92
	Usage time	Path length	-0.05	>0.62

The network's mAP was 0.84 ± 0.14. The syringe path length computed with object detection showed the weakest correlation with the corresponding EM metric. This result can likely be attributed to the three DOF setup's inability to track rotational motion and motion away from the camera's frame of view. Future work will seek to address this shortcoming while still maintaining the low-cost appeal of our methods

CONCLUSIONS

This project sought to determine if an inexpensive, video-based method was feasible for computer-assisted skill assessment. There was a significant correlation between the three DOF, in-plane, metrics from video and the six DOF metrics generated by EM tracking. EM tracking has previously been validated against trainee skill, so this significant correlation reveals that object detection is likely feasible as a skill assessment method. Since the correlation was stronger between the matching metrics versus the alternately paired metrics, the results further suggest that the video-based metrics are indeed measuring the same attributes of skill as are measured with the EM-tracked metrics.

This study reveals that using a Faster R-CNN on video data has the potential to provide a similar assessment of skill to the current gold standard of computer-assisted skill assessment while improving the face, content, and predictive validity of these systems. These results are encouraging and show promise that an inexpensive camera can provide a similar assessment of skill comparable to current expensive, bulky, six DOF tracking systems. Now that object detection has been shown to be a feasible skill assessment method, the next steps in this project will involve integrating object detection into Central Line Tutor to provide both a training and a skill assessment platform.

NEW OR BREAKTHROUGH WORK TO BE PRESENTED

This study determines the feasibility of using video to surgical skill assessment metrics in CVC using tracking and video data from both novice and expert participants. Comparing video-based metrics with established metrics based on tracking information reveals that object detection could be a feasible skill assessment method. Previous work trained a Faster Region-based Convolutional Neural Network (Faster R-CNN) on a limited video dataset without expert participants [14]. The network detected the tools used in the procedure with reasonable accuracy. However, the skill assessment metrics computed in the previous study were not compared to any ground truth skill assessment metrics, so no skill assessment was performed.

ACKNOWLEDGEMENTS

This work was funded in part by NIH/NIBIB and NIH/NIGMS (via grant 1R01EB021396-01A1), by CANARIE's Research Software Program, the Natural Sciences and Engineering Research Council of Canada grant RGPIN-2020-05582 and is supported as a Collaborative Health Research Project (CHRP #127797) by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes of Health Research (CIHR). R. Hisey is supported by the Queen Elizabeth II Graduate Scholarships in Science and Technology (QEII-GSST). G. Fichtinger is supported as a Canada Research Chair.

REFERENCES

[1] Shelesky, G., D'Amico, F., Marfatia, R., Munshi, A. and Wilson, S.A. Does weekly direct observation and formal feedback improve intern patient care skills development? A randomized controlled trial. Family Medicine-Kansas City, 44(7) 486 (2012).

[2] Weersink, K., Hall, A. K., Rich, J., Szulewski, A., Dagnone, J. D., "Simulation versus real-world performance: A direct comparison of emergency medicine resident resuscitation entrustment scoring," Advances in Simulation 4(1) (2019).

[3] Gray, J. D., "Global Rating Scales in residency education," Academic Medicine 71(1) (1996).

[4] Howley, L. D., "Performance Assessment in Medical Education," Evaluation &amp; the Health Professions 27(3), 285–303 (2004).

[5] Yarris, L. M., Linden, J. A., Gene Hern, H., Lefebvre, C., Nestler, D. M., Fu, R., Choo, E., LaMantia, J., Brunett, P., "Attending and resident satisfaction with feedback in the emergency department," Academic Emergency Medicine 16 (2009).

[6] Tuck, K. K., Murchison, C., Flores, C., Kraakevik, J., "Survey of residents' attitudes and awareness toward teaching and student feedback," Journal of Graduate Medical Education 6(4), 698–703 (2014).

[7] Reiley, C. E., Lin, H. C., Yuh, D. D., Hager, G. D., "Review of methods for objective surgical skill evaluation," Surgical Endoscopy 25(2), 356–366 (2010).

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

[9] Clinkard, D., Holden, M., Ungi, T., Messenger, D., Davison, C., Fichtinger, G., McGraw, R., "The development and validation of hand motion analysis to Evaluate competency in Central LINE CATHETERIZATION," Academic Emergency Medicine 22(2), 212–218 (2015).

[10] Mick, P. T., Arnoldner, C., Mainprize, J. G., Symons, S. P., Chen, J. M., "Face validity study of an artificial temporal bone for simulation surgery," Otology &amp; Neurotology 34(7), 1305–1310 (2013).

[11] Gardner, A. K., Kosemund, M., Martinez, J., "Examining the feasibility and predictive validity of the Sagat tool to assess situation awareness among medical trainees," Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare 12(1), 17–21 (2017).

[12] Rutherford-Hemming, T., "Determining content validity and reporting a content validity index for simulation scenarios," Nursing Education Perspectives 36(6), 389–393 (2015).

[13] Hisey, R. J., "Computer-assisted workflow recognition for central venous catheterization," thesis (2019).

[14] O'Driscoll, O., Hisey, R., Camire, D., Erb, J., Howes, D., Fichtinger, G., Ungi, T., "Object detection to compute performance metrics for skill assessment in central venous catheterization," Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling (2021).

[15] Soni, N. J., Reyes, L. F., Keyt, H., Arango, A., Gelfond, J. A., Peters, J. I., Levine, S. M., Adams, S. G., Restrepo, M. I., "Use of ultrasound guidance for central venous catheterization: A national survey of Intensivists and Hospitalists," Journal of Critical Care 36, 277–283 (2016).

[16] Taylor, R. W., Palagiri, A. V., "Central Venous Catheterization," Critical Care Medicine 35(5), 1390–1396 (2007)

[17] Kumar, A., and Alwin, C. "Ultrasound guided vascular access: efficacy and safety," Best Practice & Research Clinical Anaesthesiology, 299-311 (2009)

[18] Ungi, T., Sargent, D., Moult, E., Lasso, A., Pinter, C., Mcgraw, R. C., Fichtinger, G., "Perk Tutor: An Open-Source Training Platform for Ultrasound-Guided Needle Insertions," IEEE Transactions on Biomedical Engineering 59(12), 3475–3481 (2012).