学　位　論　文　内　容　の　要　約

博士の専攻分野の名称　　博士（情報科学）　　氏名　海老名 光希

学　位　論　文　題　名

Development of machine learning-based skill assessment system for laparoscopic surgery using motion-capture

（モーションキャプチャを用いた腹腔鏡手術のための機械学習による技量評価システムの開発）

Laparoscopic surgery has become widespread due to its minimal invasiveness to the patient. However, the difficulty of surgery, such as limited visual information and inconsistent hand-eye coordination, makes it difficult to acquire the technique quickly. Moreover, the recent restriction of working hours and shortage of surgeons have resulted in insufficient training time, and the establishment of efficient training methods and objective surgical skill analysis is becoming urgent. In general, surgeons have developed their surgical skills through on-the-job training (OJT) in clinical settings and off-the-job training (Off-JT) using organ models or animal organs. Although OJT is the most effective, there are medical safety concerns, and surgeons must have a certain level of skill before OJT. Against this background, to promote effective skill proficiency, there are several studies for surgical skill evaluation in off-JT. However, these studies have mainly focused on dry-lab training using an organ model with simple training tasks, which limits skill analysis. Although the measurement in wet-lab training using animal organ model has been reported, other core surgical procedures such as tissue dissection and vascular clip attachment to the pedicle have not been adequately analyzed. Automated skill assessment methods using machine learning have also been reported, but these studies were limited to simple classification and cannot provide a basis for evaluation and lack specificity. In order to avoid unintentional organ damage, not only the movement of the surgical instrument but also the operating force is important. However, no studies were found that measured the grasping point at the grasper of forceps or analyzed the force measurement data in combination with the forceps movement data. Therefore, this study developed the surgical instrument measurement system and performed skill analysis through measurement experiment in wet lab training, and developed an automatic skill assessment system that can present the evaluation basis.

In this system, the surgical instrument motion was measured by optical motion capture (MoCap) system OptiTrack Prime41 manufactured by NaturalPoint, Inc. Each instrument has two marker sets, and the positional relationship between the instrument was used to calculate tip position, gripper opening ratio, sheath rotation angle, and orientation. By attaching individual pattern marker sets, this system can measure the movement of multiple surgical instruments simultaneously. The grasping force/point of the grasper was calculated from the measurements of two strain gauges attached to the back of the grasper and recorded with the instrument movement. The validation experiment confirmed that this system can measure surgical training with sufficient accuracy for skill analysis (average positional error: within 1 mm, average grasping force error: 0.019 kgf, and average grasping point error: 0.59 mm). Note that the average grasping force/point error was calculated from data excluding the root of the grasper ($L < 7.5$ mm) and low loads ($<$ 100 g). This result is considered reasonable because grasping at the root side is rarely used in normal surgery.

The measurement experiments were conducted in wet lab simulation training, and 143 cases of the lymphadenectomy and the renal parenchyma suturing were recorded. The measurement frequency was set to 30 Hz for the grasping force/point measurement and that was set to 120 Hz for the measurement with the practical surgical training measurement system described below. Missing values were linearly interpolated and smoothed by the Savitzky-Golay filter. The questionnaire survey on the operational feel of the instrument confirmed that the markers attached on the instrument handle have little effect on the operational feeling.

This study also developed the measurement system for practical surgical training, where there are many obstacles. The OptiTrack V120: Trio manufactured by NaturalPoint, Inc. was adopted to measure the movement of surgical instruments because of its simple configuration. This system can perform measurement only one unit, and there is no need for dynamic calibration, and easy to install an operation room. Because one marker set was removed to improve measurement stability, this system only measures tip position and orientation. The validation experiment was performed by simultaneously measurement with the previous measurement system OptiTrack Prime41. The experimental result showed that the developed system has sufficient measurement success rate and accuracy (approximately 85 to 95 % measurement success rate and 2 to 4 mm positional error depending on the instrument). The questionnaire survey on the operational feel of the instrument confirmed that the markers attached on the instrument handle have little effect on the operational feeling. To further improve measurement accuracy, surgical instrument equipped with inertial measurement unit (IMU) and distance measurement sensors were also newly developed. In laparoscopic surgery, the surgical instrument is inserted into the patient's abdomen via a trocar, and its motion is limited to 4 degrees of freedom (DoF): rotation around the insertion point (3DoF) and translation along the insertion axis (1 DoF). Thus, this system calculates the instrument tip position from the orientation measured by the IMU and the distance from the instrument to the insertion point measured by the time-of-flight (ToF)-based distance sensor while MoCap measurements are missing. The validation experiments confirmed that this system improved the overall average position error from 4.2 mm to 3.6 mm, and that of MoCap missing period improved from 11.4 mm to 7.9 mm.

The measurement experiments were conducted on 36 cases of laparoscopic radical nephrectomy with human cadaver. The measurement frequency was set to 120 Hz. Since this system measures instrument motion from only one direction, it is susceptible to occlusion, and the measured data contains noise and outliers. Therefore, this study developed three outlier removal methods based on partial measurement success rate, region of interest (ROI) and Kalman filter. Outliers and missing values were linearly interpolated and smoothed by the Savitzky-Golay filter. Experimental result confirmed that the system could measure with measurement success rate of more than 90%. The questionnaire survey on the operational feel of the instrument also confirmed that the markers attached on the instrument handle have little effect on the operational feeling.

Based on the measurement data of wet-lab training, surgical skill analysis was conducted. The subjects divided into three groups: novice, intermediate, and expert, according to the total score of Global Operative Assessment of Laparoscopic Skills (GOALS) which was rated by two skilled urologists. GOALS is a type of surgical technique assessment tool that rates a surgeon's technique on five items, with 1 to 5 points for each item, for a total score of 5 to 25 points. Some kinematic indices representing the motion characteristics were calculated, and the differences between three groups were examined by Kruskal-Wallis test and Wilcoxon's rank sum test. Principal component analysis was also performed using significant indices of the statistical test. In the lymphadenectomy task, the analysis showed that efficiency of the movement of grasping forceps and the scissors forceps mainly contributed to the skill difference, and the operating speed of the instruments secondarily contributed. The same tendency was also observed in the suturing task. The skill analysis based on the indices calculated from automatic feature calculator, tsfresh, was also performed. In the previous studies, the indices characterizing instrument movements were defined according to the hypotheses made by the analysts, and it was possible that some characteristics have not been extracted. To address these issues, this study used tsfresh, an automatic feature calculator for time series data, to compute a variety of indices from measured instrument motion without being bound by preconceived notions. The analysis result shows that in addition to the previously confirmed efficiency and speed-related factors, vibration, complexity of operation, and irregular motion were also involved in the skill differences. The skill analysis on the grasping force/point measurement data was also performed. The composite indices were calculated from the force/point measurement data, and statistical test and PCA were performed. The analysis results indicated that the stability of the grasping force and the overexert at the beginning of the task mainly contributed to the skill difference.

The machine learning-based surgical skill assessment system was developed. The classification model and the regression model based on the total GOALS score were constructed. In the classification model, three machine learning algorithms: support vector machine (SVM), principal component-based support vector machine (PCA-SVM), and gradient boosting decision tree (GBDT) was used. In the regression model, five algorithms: support vector regression (SVR), PCA based SVR (PCASVR), ridge regression, partial least squired regression (PLSR) and GBDT were used. The model accuracies were validated using nested and repeated 10-fold cross validation, and accuracies were compared by Friedman's test and Wilcoxon's signed rank sum test. Model validation showed that the classification model can assess surgeon skill with a median accuracy of 75.0% (classification) for the lymphadenectomy task, and 85.5% for the suturing task. The regression model, the median of the mean absolute error was 2.36 for the lymphadenectomy task (score range: 5 to 25) and 1.13 for the suturing task (score range: 4 to 20). By combining the above constructed evaluation models with SHAP, one of the explainable artificial intelligence (AI) methods, an automatic surgical skill feedback system was developed. In addition to the three-level evaluation and GOALS score estimated by above evaluation model, this system can provide the contributions of each indices to the estimation result by automatically generated comments. Explainable AI was also applied to skill analysis, which revealed the factors that most contributed to skill differences among the indices generated by the automatic feature calculator. The result of the analysis showed that operation stability and efficiency mainly contributed to the lymphadenectomy task, and the path length of both hands and the vibration of the left hand were especially important in the suturing task.

This study verbalized the surgical skills of expert surgeons and developed the skill evaluation system that provides quantitative and concrete feedback immediately after surgical training. Future work is to verify the effectiveness of the system through verification experiments and to improve this system through questionnaires. The use of image recognition AI is also considered as future work. Image recognition will make it possible to measure the movement of the instrument more robustly, and by recognizing biological tissue, it can be used to evaluate organ damage. In addition, automatic surgical procedure segmentation for practical laparoscopic training involving multiple surgical procedures for more detailed skill analysis is also considered as future work.