An Automatic Multi-metrics-based Evaluating Model of Surgeon’s Operating Performance for Interventional Surgical Robotic System

Zhijun Lin1, Shuxiang Guo1, Chuqiao Lyu1, Duohao Zhao1

1 Key Laboratory of Convergence Biomedical Engineering System and Healthcare Technology, The Ministry of Industry and Information Technology, School of Life Science, Beijing Institute of Technology, No.5, Zhongguancun South Street, Haidian District, Beijing 100081, China
E-mails: {guoshuxiang & linzhijun}@bit.edu.cn
*Corresponding author

Abstract—A quantitative operation performance assessment is essential for establishing surgeons’ learning curve and improving the techniques for operating surgical robots in percutaneous coronary intervention (PCI). Most of the existing evaluation methods only consider the influence of catheter and guidewire’s tip, and they are almost not automatic. However, the overall shape of guidewire and catheter have a great impact on surgical performance. In this study, an automatic evaluation model is introduced to deal with the above issue. Firstly, guidewire's offset, velocity and contact force are considered in this model. Additionally, the guidewire’s offset and velocity are the features extracted from the video sequence, and the contact force is the feature obtained by the force sensor. Secondly, analytic hierarchy process (AHP) is used to establish automatic evaluation model with ability of quantitative assessment based on the above three metrics. Lastly, the proposed model was evaluated by the experiments through a given task, which shows that the proposed model can train and improve the operation skills for interventional surgical robotic system.

Index Terms—Surgeon’s Operating Performance, Evaluation model, Interventional Surgical Robotic System, Multi-metrics.

I. INTRODUCTION

Cardiac cerebral diseases have been becoming the number one killer of human in the world. Because of its adventures like: small incisions and quicker recovery, Vascular interventional surgery (VIS) is the more effective treatment for these disease [1]. However, during the operation, doctors have to suffer from the long-term irradiation of X-ray. In order to decrease the radiological hazard, surgeons need to wear 20 kg of lead clothing [2]. Although the heavy lead clothing resist some radiations, it do bring some chronic diseases to doctors, like: Lumbar spondylosis. With the measurable progress of robotics, more and more robot becoming prevalent in the operation room [3][4]. Robotic and computer-assisted surgical systems have the potential to protect surgeons form radiation exposure [5][6].

Recently, many master-slave VIS robot systems have appeared, like: CorPath® Robot System (Corindus Robotics Inc., Waltham, MA, USA) [7], Sensei® Robotic System (Hansen Medical Inc., Mountain View, CA, USA) [8], Amigo® Robot System (Catheter Precision Inc., Ledgewood, NJ, USA) [9]. These robots are effective means for protecting surgeons form radiation and reducing operation mistakes [10][11]. Moreover, some researchers attempt to use force feedback module to help surgeons feel the danger quickly [12]. The study of operating vascular interventional robots in remote situations is of more significant value [13].

Nevertheless, surgical outcome, quality, and efficiency are impressionable to the surgeon’s acquisition and mastery of skills over the new surgical systems [14]. On the contrary, the surgical system’s usability impacts surgeons directly. Hence, there is a crying need better surgical systems to provide maximum efficiency. In order to attest that new surgical technologies can promote the surgeons’ performance in PCI required: (i) the assessment of the performance of the systems, (ii) the assessment of the ‘learning curve’ required for surgeons, and (iii) the establishment of benchmarks for comparative analysis [15]. In addition, the evaluation need to minish the effect of measurement errors and the affect of subjective factors to achieve objective and robust. Therefore a significant assessment tool need to evaluate surgeons’ operation performance impersonally and quantificationally.

Despite advances in computer system, surgical evaluation is still based on direct observation by experts [16]. The shortcoming of human subjective observation results in the non-reliance, inefficiency and unsteadiness of these methods. These is a vital need to promote a consistent and reliable tool to assess clinical performance objectively in order to protect patient safety. Martin et al. introduce the Objective Structured Assessment of Technical Skills (OSATS) objective framework which was utilized to validate design and test skill in subsequent surgical research [17]. More emphasis is being placed on competency-based training and earlier development of technical skills for new surgeons. OSATS is being replaced by more structured techniques, even these are still required the expert surgeon’s subjective evaluation [18]. The emergence of new technologies such as robotics makes it possible to evaluate clinical surgical performance objectively and automatically. Surgical robots like the Da Vinci surgical robot can record video of movements during surgery, and data-driven real-time surgical evaluation systems can be developed [19]. Contact force between guidewire and vessel was proposed for evaluating [20]. Metrics such as the path length, completion time, speed, depth perception, curvature, and smoothness [21] were used for the surgical assessment.
Based on the above background, surgical assessment model still has the following problems:

1). Most current evaluation methods only focus on the impact of guidewire tip on surgical performance, but ignore the overall impact of guidewire. Moreover, most of them only pay attention to one type of metrics, without comprehensive evaluation of multiple types of metrics.

2). The current evaluation model does not achieve fully automatic surgical evaluation and still requires surgeons to operate, which is easy to cause surgeons’ distraction and affect surgical performance.

The main contribution of this paper is an automatic evaluation model was proposed to automatically extract the required metrics from video sequences and evaluate surgical performance combined with contact force. The deviation error between guidewire and vascular centerline as a metric form video sequences was calculated to determine the overall influence of guide wire on surgical performance. It enriches the metric of the evaluation model and makes the evaluation result more objective and accurate.

The reminder of this paper is as follows. In Section II, three metrics for establishing assessment model are proposed. The detailed experimental design is described in Section III. Section IV discusses the combined performance rating model and preliminary experimental results validation the chosen approach. Section V concludes the paper with a reference to future work.

II. METHODS

In this section, the proposed three metrics for establishing assessment model are described in detail. Guidewire’s offset and velocity are the features extracted from the video sequence automatically, and the contact force is the feature obtained by the force sensor.

A. Guidewire Deviation Error

The guidewire deviation error metric is calculated as the root mean square error (RMSE) between the guidewire and the vascular centerline. This metric reflects the contact of the whole guidewire (not just the tip of the guidewire) with the vascular wall. Therefore the maximum RMSE and the mean RMSE of the surgeon's operation can offer pivotal information with regard to the usability of surgical robot and the performance of the surgeon.

The following two steps are required to calculate the REMS:

1) In order to obtain REMS, it is necessary to take the centerline of the experimental vascular platform as the standard. For a more realistic simulated angiogram image of blood vessels, we filled the simulated vascular platform with black clay, as shown in Fig. 1(a). The obtained image is segmented by threshold value, as shown in Fig. 1(b). The obtained image is close to the real two-dimensional DSA image.

The centerline of the simulated vascular platform was extracted by thinning algorithm. Refinement is to find the central axis or skeleton of a figure or stroke and replace the figure or stroke with its skeleton. In accordance with the most basic principle of image refinement to maintain image connectivity, the pixel width of the refined figure or stroke is 1. The thinning process is the process of peeling off the image layer by layer. With the thinning process, the image shrinks regularly. The extracted vascular centerline is shown in Fig. 1(c).

Fig. 1. Extraction process of model centerline. (a) Filled model (b) Processed image (c) Centerline diagram

2) Firstly, the image subtraction is carried out by using empty simulated vascular background, and the guide wire image of each frame is extracted, as show in Fig. 2.

Fig. 2. Guidewire mask (pixel value)

The RMSE is described in [24], which is shown as follows:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}
\]  

(1)
where \( \text{err} \) represents minimum distance from each pixel on the guidewire to the medial axis.

This paper will calculate the average REMS (aR) and maximum REMS (mR) in the experiment.

B. Force Feedback

Some uncertainties in robot-assisted surgery, such as surgeons nonproficiency or robot errors, can cause additional injuries during surgery. This type of damage in VIS is called iatrogenic vascular injury [25]. These errors can easily lead the guidewire to collide with the blood vessel wall, causing damage even to penetrate the blood vessel, or causing plaque to fall off so that lead to more severe blockage. Thus, the contact force between the guidewire tip and the vascular wall during the whole operation be recorded, which can directly reflect the performance of the operation process. For contact force, we record two indexes are average contact force and maximum contact force. The average contact force can embody the overall performance of the operation, and the maximum contact force reflects the risk situation during the operation.

ATI Force/Torque sensor (ATI, FT17123, ATI Industrial Automation, USA) used to record the surgeon’s manual and surgical robotic forces. The sensor measures output force and torques at the cartesian coordinates.

This paper will record the average force (aF) and maximum force (mF) in the experiment.

C. Movement Speed

The speed of the guidewire tip is the most intuitive embodiment of the operation. When the operation goes well, the guidewire tip can move quickly and smoothly. However, it is difficult to calculate the true speed from two-dimensional video images, and other sensors are needed to determine the true proportion. But in the case of this paper, it’s enough to compare pixel speeds for the same proportion of motion. In order to calculate the pixel speed, a process is adopted:

First, the guidewire tip needs to be captured and tracked. We can further judge whether the guidewire is moving by the position change of the corners. The corner represents a pixel in the image. A small change in any direction will bring a significant change in the corner. The Shi-Tomasi corner analyzes the eigenvalue of the autocorrelation matrix \( X \). If the smaller of the two eigenvalues are more significant than the minimum threshold, this pixel will be obtained as a strong corner. The Shi-Tomasi algorithm used in this paper is described as:

\[
\begin{align*}
E(u, v) & \approx [u, v]M \begin{pmatrix} u \\ v \end{pmatrix} \\
M & = \sum_{(x,y)} w(x, y) \begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix} \rightarrow R^{-1} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} R \\
ST & = \min(\lambda_1, \lambda_2)
\end{align*}
\]

where \( M \) represents the change of gray value when the window moves in all directions, \( I_x \) and \( I_y \) are the gradients in \( X \) and directions, and \( u \) and \( v \) represent the displacement of the moving window. \( R \) represents the rotation factor. It does not affect the variation component. Shi-Tomasi algorithm determines whether the current point is a corner by judging the minimum eigenvalue of matrix \( M \).

After the coordinate of the guide wire tip is obtained, the relative velocity of the guide wire tip is calculated by the change of the coordinate of the guide wire head per second.

This paper will calculate the average speed (aS) and maximum speed (mS) in the experiment.

III. EXPERIMENTAL DESIGN

In this section, the model we proposed will validated by experiments. First, experimental platform containing a master and slave vascular interventional surgery robot, camera and PCI trainer was designed. Then, data were collected by three experimenters with different experience to take two paths of different difficulty.

A. Experimental Platform

The guidewire operation video was collected from a PCI Trainer for Experts, (Medialpha Co., Ltd.). PCI trainer is a 2D training model for interventional surgeries, which contains many PCI operations. For example, stent, chronic total occlusion, etc. In PCI trainers, the thinnest vascular is the aortic arch, about 10 mm. A standard guidewire(J-shape) is used to simulate PCI operation in the phantom. As show in Fig. 3, the platform includes a camera, PCI trainer, operation interface and Master-Slave Robot. The size of the collected image is 640 \( \times \) 480.

![Fig. 3. Experimental platform](image)

The master-slave vascular interventional surgical robot system used in this paper includes a master manipulator, a slave manipulator and a communication system, as show in Fig.1 During the operation, the surgeon operates the master manipulator. The displacement sensor on the master manipulator collects the surgical displacement information of the surgeon and transmits it to the slave manipulator through the information communication system. Clamp the guidewire/catheter from the manipulator and push it along the
vessel to the target position. During the push, the impact force of the guidewire/catheter is transmitted to the master manipulator, where a force feedback device provides tactile perception for the surgeon. During the procedure, the doctor can use visual information from the camera to measure the status of the guidewire/catheter being pushed inside the blood vessel.

**B. Operating Process**

The experiments in this paper were designed to facilitate the acquisition of objective performance indicators of PCI. The repeatable task approach allows for randomized comparative trials. In order to evaluate ergonomic factors and equipment availability, the experimental task consisted of a series of motion tracks. Surgical procedures were performed manually and with a robotic system following a predetermined random route. Two difficult paths were chosen by human, representing the actual surgical task. The selected surgical route is shown in Fig. 4.

![Fig. 4. The two experimental approaches](image)

Cameras capture video at 20 frames per second, and these are the only videos used to calculate performance metrics for each trial.

**C. Participant**

The experimental data in this paper were acquired from the pilot experiment conducted by three operators with different experience. One of them had manual guide wire operation experience, another had robot guide wire operation experience, and the last one had no guide wire operation experience. The subjects were given oral and written instructions and then asked to follow the task as accurately as possible.

**IV. Result and Discussion**

Three different experiments were conducted to compare the differences in man–machine operation. The experimental paths are the three artificially defined difficulty paths mentioned above. The collected video data is intercepted and a frame of image is selected every second. The obtained data is analyzed based on the introduced metric, which facilitates intuitive comparison.

Different metrics in diverse experimental results are different. In order to objectively and comprehensively evaluate the performance of surgery, this paper introduced AHP to construct a rating scale based on the proposed indexes. There are many factors affecting the performance of surgical procedures, and it is difficult to obtain an ideal value for each indicator. This paper analyzes the obtained expert data and combines the results with experience to determine the ideal value and maximum acceptable threshold of each metric. TABLE I shows the data of two comparative experiments conducted by one of the operators.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Ideal Value</th>
<th>Robot</th>
<th>Hand</th>
</tr>
</thead>
<tbody>
<tr>
<td>aR</td>
<td>0</td>
<td>3.76</td>
<td>4.37</td>
</tr>
<tr>
<td>mR</td>
<td>0</td>
<td>9.64</td>
<td>8.97</td>
</tr>
<tr>
<td>aF</td>
<td>0</td>
<td>13.52</td>
<td>15.40</td>
</tr>
<tr>
<td>mF</td>
<td>0</td>
<td>51.55</td>
<td>47.16</td>
</tr>
<tr>
<td>aS</td>
<td>high</td>
<td>9.37</td>
<td>81.35</td>
</tr>
<tr>
<td>mS</td>
<td>high</td>
<td>21.26</td>
<td>136.52</td>
</tr>
</tbody>
</table>

In order to carry out AHP, this paper needs to normalize each metric. The contribution to performance is greatest from the ideal value of the specified indicator, and the further away from the ideal value, the contribution decreases exponentially. Equations (3) and (4) show the functions.

\[
\begin{align*}
\text{f}(mS,aS) &= \frac{10}{1 + \exp[-\alpha \cdot (\text{val} - \text{thresh})]} \quad (3) \\
\text{f}(aF,mF,aR,mR) &= 10 - \frac{10}{1 + \exp[-\alpha \cdot (\text{val} - \text{thresh})]} \quad (4)
\end{align*}
\]

where \(\alpha\) and \(\text{thresh}\) are constants, and their values are determined by the maximum acceptable value. \(\text{ideal}\) is the ideal value for each metric.

In order to realize the normalization of each metric, \(\alpha\) and \(\text{thresh}\) value of different indexes are set differently, as can be seen in the TABLE II.

<table>
<thead>
<tr>
<th>Metric</th>
<th>(\alpha)</th>
<th>(\text{thresh})</th>
<th>Metric</th>
<th>(\alpha)</th>
<th>(\text{thresh})</th>
</tr>
</thead>
<tbody>
<tr>
<td>aR</td>
<td>1.25</td>
<td>5</td>
<td>mF</td>
<td>0.125</td>
<td>50</td>
</tr>
<tr>
<td>mR</td>
<td>0.75</td>
<td>10</td>
<td>aS</td>
<td>0.05</td>
<td>50</td>
</tr>
<tr>
<td>aF</td>
<td>0.5</td>
<td>15</td>
<td>mS</td>
<td>0.01</td>
<td>75</td>
</tr>
</tbody>
</table>
After completing the normalization of each metric, it is necessary to determine the weight of each index. In this paper, AHP is used for index weight analysis.

First of all, we need to establish a multi-level hierarchical structure according to the object, and divide the system into several levels according to the different goals and the differences in the realization of functions. The second part is to determine the degree of correlation between adjacent elements in the hierarchical structure. By constructing two comparative judgment matrices and the mathematical method of matrix operation, we can determine the order of importance of an element in this hierarchy and its relative elements relative weight. The scale method for judging matrix elements is shown in the TABLE III.

**TABLE III**

<table>
<thead>
<tr>
<th>Scale</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Two factors are equally important</td>
</tr>
<tr>
<td>3</td>
<td>One is slightly more important than the other</td>
</tr>
<tr>
<td>5</td>
<td>One factor is significantly more important than the other</td>
</tr>
<tr>
<td>7</td>
<td>One is more strongly important than the other</td>
</tr>
<tr>
<td>9</td>
<td>One is extremely more important than the other</td>
</tr>
<tr>
<td>2,4,6,8</td>
<td>The median value of the above adjacent judgment</td>
</tr>
</tbody>
</table>

According to the above rules, this paper gives the judgment matrix based on the knowledge and experience of experts as shown in the TABLE IV.

**TABLE IV**

<table>
<thead>
<tr>
<th></th>
<th>aR</th>
<th>mR</th>
<th>aF</th>
<th>mF</th>
<th>aS</th>
<th>mS</th>
</tr>
</thead>
<tbody>
<tr>
<td>aR</td>
<td>1</td>
<td>1/2</td>
<td>1/4</td>
<td>1/6</td>
<td>1/3</td>
<td>1/5</td>
</tr>
<tr>
<td>mR</td>
<td>2</td>
<td>1</td>
<td>1/2</td>
<td>1/4</td>
<td>1/2</td>
<td>1/3</td>
</tr>
<tr>
<td>aF</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1/2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>mF</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>aS</td>
<td>3</td>
<td>2</td>
<td>1/3</td>
<td>1/3</td>
<td>1</td>
<td>1/2</td>
</tr>
<tr>
<td>mS</td>
<td>5</td>
<td>3</td>
<td>1/2</td>
<td>1/2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Finally, the consistency test is carried out on the judgment matrix. If the consistency test is passed, the weight of relative importance of all factors at a certain level to the overall goal can be calculated. The weights of various metrics are calculated as shown in the TABLE V.

**TABLE V**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Weights</th>
<th>Metric</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>aR</td>
<td>0.0444</td>
<td>mF</td>
<td>0.3412</td>
</tr>
<tr>
<td>mR</td>
<td>0.0793</td>
<td>aS</td>
<td>0.1117</td>
</tr>
<tr>
<td>aF</td>
<td>0.2353</td>
<td>mS</td>
<td>0.1878</td>
</tr>
</tbody>
</table>

Based on equations (3), (4), and the weights from Table III, the combined metric rating for the route 3 over the 3 trials is shown in TABLE VI and VII.

**TABLE VI**

<table>
<thead>
<tr>
<th></th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robot</td>
<td>8.25</td>
<td>7.96</td>
<td>8.47</td>
</tr>
<tr>
<td>Hand</td>
<td>6.93</td>
<td>6.95</td>
<td>6.74</td>
</tr>
<tr>
<td>Rating</td>
<td>4.649</td>
<td>6.734</td>
<td>4.352</td>
</tr>
</tbody>
</table>

**TABLE VII**

<table>
<thead>
<tr>
<th></th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robot</td>
<td>8.35</td>
<td>7.82</td>
<td>8.57</td>
</tr>
<tr>
<td>Hand</td>
<td>7.45</td>
<td>6.87</td>
<td>6.98</td>
</tr>
<tr>
<td>Rating</td>
<td>2.920</td>
<td>5.576</td>
<td>3.251</td>
</tr>
</tbody>
</table>

Through the data in the TABLE VI and VII, it can be found quantitatively that path 1 has a lower score than path 2, indicating that it is more difficult. This is consistent with experts' judgment of route difficulty before the experiment, proving that our model can effectively evaluate surgical performance. It was also observed that the performance of the
robotic surgery decreased more in the difficult vessels than in the manual surgery.

Among the three experiment-members selected in this paper, experiment-one has rich experience in manual surgery, experiment-three has experience in robot operation, and experiment-two has no experience in either. This paper found a phenomenon that having experience in one area can improve performance at another compared to being a complete novice. But robotic surgery is still less effective than manual surgery, especially in terms of speed and time.

V. CONCLUSION

In this paper, an automatic multi-metrics-based evaluation model were introduced for analysis and assessment in PCI. The final evaluation model can be used to automatically extract the desired features from the video sequence and quantitatively analyze surgical performance. The introduced assessment model assists in two key aspects:

1) The experiment proves that the evaluation model proposed in this paper can accurately evaluate the surgical performance quantitatively, and can be used to form a learning curve for doctors, help doctors perform postoperative analysis, and improve their surgical skills.

2) The evaluation model can realize automatic evaluation from guide wire motion video to final score. It allows the surgeons to concentrate on the surgery without having to perform other operations.

In the future work, the most important thing is to collect more widely the data of experts and machine operation, and adopt more advanced data analysis methods (such as machine learning) to build evaluation models. It can also be considered to use the model established in this paper to develop a real-time evaluation tool for surgical status and provide intuitive surgical status feedback to doctors.

REFERENCES


