Distance Measures for Surgical Process Models

S. Schumann1; U. Bühligen2; T. Neumuth1
1Universität Leipzig, Innovation Center Computer Assisted Surgery, Leipzig, Germany;
2University Medical Center, Department of Pediatric Surgery, Leipzig, Germany

Keywords
Workflow, computer-assisted surgery, surgical process model, distance measures

Summary
Background: The development of new resources, such as surgical techniques and approaches, results in continuous modification of surgery. To assess these modifications, it is necessary to use measures that quantify the impact of resources on surgical processes.

Objectives: The objective of this work is to introduce and evaluate distance measurements that are able to represent differences in the courses of surgical interventions as processes.

Methods: Hence, we present four different distance measures for surgical processes: the Jaccard distance, Levenshtein distance, Adjacency distance, and Graph matching distance. These measures are formally introduced and evaluated by applying them to clinical data sets from laparoscopic training in pediatric surgery.

Results: We analyzed the distances of 450 surgical processes using these four measures with a focus on the difference in surgical processes performed by novices and by experienced surgeons. The Levenshtein and Adjacency distances were best suited to measure distances between surgical processes. The measurement of distances between surgical processes is necessary to estimate the benefit of new surgical techniques and strategies.

Conclusion: The measurement of distances between surgical processes is necessary to estimate the benefit of new surgical techniques and strategies.

1. Introduction
The development of the digital operating room is an ever-progressive topic [1, 2]. The complexity of the systems and the concomitant challenges posed to surgeons and other medical and technical staff are rapidly increasing. To cope with these growing demands and to develop new and sensible systems, it is indispensable to assess, manage, and evolve the surgical processes that are intrinsic to patient care.

Additionally, new instruments, tools, and software systems related to surgical techniques are continuously being developed. These new techniques lead to modifications of surgical processes. To evaluate a new approach, it is necessary to quantify its impact on these processes. Thus, it is necessary to have a measurement that quantifies the distance between conventional and new approaches.

A measurement that quantifies the distance between processes can also be used in surgical training. The improvement of practiced tasks can be measured on the basis of the performed work steps by using distance measurements. In this manner, trained surgeons can assess their skill level and the possibility of improvement through advanced training sessions can be analyzed.

Process distance measures have the objective of quantifying variations between processes. Using these methods, it is possible to assess the similarity between sequences of surgical activities. The objective of this work is to introduce and evaluate distance measurements that are able to represent the differences between courses of surgical intervention, provided that the measurement is contextually sensible, correct, and objective. With these methods, it is possible to determine whether the variation of the course of surgical intervention when using a new tool or method is significant when compared with the conventional method.

In the past, analysis of surgical processes was based primarily on the recoding and appraisal of key figures [3, 4] in isolated aspects of surgical work steps, such as frequencies and durations, rather than on activity sequences. However, some approaches exist for analyzing surgical (trans-action) processes. For instance, Meng et al. [5] reported a method for extracting the chronological order of treated anatomical structures from the OR documentation to generate the surgical course of action. The activities are then represented as directed graphs, and similarities are computed using the Jaccard index of the catchwords. Multiple similar graphs are then transferred into a model using sequence alignment.

Neumuth et al. [6] developed a set of similarity metrics for surgical process models. These metrics address several dimensions of process compliance in surgery, including granularity, content, time, order, and frequency of surgical activities. The metrics were experimentally validated using multiple clinical data sets. The objective of these metrics was to facilitate a com-
The comparison of the accuracy of different data acquisition modalities that recorded the same process. Consequently, the comparison of surgical processes was based only on aligned pairs of activities. To align activities, registration maps were used to link the activities of surgical process models. In this paper, we develop distance measurements with the objective of comparing different processes with non-registered activity sequences. Thus, there is no need for registration mapping.

Nelson et al. [7] analyzed exchanges of surgical instruments during minimally invasive interventions to extract recurring patterns by using graph-theoretical approaches. Because significant similarities in the use of surgical instruments according to the type of the intervention were demonstrated, the authors issued a recommendation concerning the traying of the instruments based on their calculations. Furthermore, the positive influence of this traying was numerically simulated. Additionally, some approaches exist in the research field of business information systems and have recently attracted interest in the literature.

Dijkman et al. [8, 9], for instance, appraised methods for evaluating different process models from a pool of such models to identify the model that is most similar to the standard model. They presented three different metrics for the measurement of similarities: label matching, which compares labels and attributes; structural similarity, which, in addition to the labels, compares the topology of the process models; and behavioral similarity, which compares labels in addition to causal relations reflected in the model.

Cook and Wolf [10], in contrast, analyzed software processes with regard to their similarity to or deviation from the respective models, and they represented processes as sequences of actions. The actions themselves had no temporal extent; they were represented as single instants of time. Thus, a process was depicted as a stream of events. Therefore, the behavior of a process could be considered, but the structures and responsibilities that underlie the process were systematically disregarded. This study used string distance metrics as the similarity metric.

Another approach by van der Aalst et al. [11] emphasized the importance of a graduation of similarities by employing event logs that contained the typical behavior as the basis for comparison. Thus, the frequent sequences were automatically emphasized, whereas the less-frequent sequences were de-emphasized.

A method to capture the similarities and differences between process models, with the goal of performing process mining, process discovery, and process integration, was presented by Bae et al. [12, 13], who also introduced a delta comparability based on the Jaccard index to select the graphs that were not completely dissimilar to one another. Distance measurements were then applied to these graphs. The respective dependency graph of the underlying process was transferred to a normalized matrix. Then, the distance was measured with respect to the metric distance of the matrices. Such a process matrix is an Adjacency matrix and was subsequently used by Zha et al. [14] as the basis for the calculation of similarities. Furthermore, an algorithm was presented that efficiently performed the similarity evaluation. The results of this calculation were then validated using real-life processes. The last two approaches did not employ the Euclidian distance used in this study to calculate the metric distance of the matrices.

Based on graph-theoretical aspects, graph matching methods can also be employed. The maximum common subgraph algorithm was used and further developed by Bunke et al. [15–18]. Another method, the graph edit distance, is based on the closest distance of the nodes and edges of the operations ‘delete’, ‘insert’, and ‘exchange’. Furthermore, Bunke et al. demonstrated the equivalence of the graph edit distance and the maximum common subgraph algorithm using cost functions.

In contrast, Robles-Kelly and Hancock [19] introduced procedures from bioinformatics that converted graphs into sequences and juxtaposed strongly correlated nodes, thereby demonstrating that spectral seriation and the string edit distance are useful tools to expressively classify graphs.

Forestier [20] explored the automatic classification of a set of surgical processes based on the dynamic time warping (DTW) algorithm and demonstrated that this algorithm was able to automatically identify groups of senior and junior surgeons. The DTW is based on the Levenshtein distance, but in contrast to the other distance measures introduced in this paper, DTW not only is based on the activities performed during a surgical intervention but also minimizes time differences. Time issues are not relevant for the measurements presented here.

To recognize and evaluate surgical skills in minimally invasive surgery, some authors, e.g., Megali, Speidel, Reiley, and Darzi [21–24], have used motion data. Using the trajectory and kinematic data of the surgical instruments and a hidden Markov model, they evaluated performance in laparoscopic surgery training. To obtain data for analyzing the performance of surgical skills, it is sensible to use tracking technology. Thus, these methods can typically only use training scenarios. Claus et al. [25] additionally used video streams and analyzed them using a standardized analysis routine and a set of standard terms. Combined visual and haptic rendering of a training scene was used by Payande et al. [26] to assess surgical skills. The methods evaluated in this paper are not based on tracking techniques or motion data; rather, they are based on performed activities recorded by the ICCAS surgical workflow editor [27]. It is thus possible to assess not only training scenarios but also real intra-operative surgical processes.

Neumuth and Liebmann [28] designed and implemented a surgical workflow management system that can provide robust guidance for surgical activities, and they addressed the high inter-patient variability of surgical processes. With the methods in this paper, they would have been able to measure the variability and also the differences in complexity of different types of surgical interventions.

Another topic in medical care is the accuracy of diagnostic programs. Todd and Stamper [29, 30] evaluated methods for distance measurements between different cases of a disease to retrieve similar previously diagnosed cases from a database.

None of the mentioned approaches, however, was evaluated for structures as complex as surgical processes. In addition,
comparative analyses of the methods with regard to their applicability are still not available.

To measure the distances between surgical processes, we introduce several approaches. In this study, we adopt the Jaccard distance, Levenshtein distance, Adjacency distance, and Graph matching distance as measures of the distance between surgical process models. We start with a short general introduction regarding similarities and distances and formalize them for surgical process modeling. We introduce the four distance measures and evaluate their applicability in a study using training data from laparoscopic surgery.

2. Material and Methods

2.1 Surgical Process Modeling

Surgical processes consist of surgical activities and can be modeled using formal and structured languages [31]. Thus, surgical process models (SPMs) are obtained and are used to study, analyze, and optimize surgical processes and to evaluate technical support systems in the operating room [32]. The aggregate of the work steps of a single surgical case is referred to as an individual surgical process model (iSPM) [6]. Each activity in an iSPM is associated with a surgical work step in the underlying surgical process and can therefore be considered a symbol.

An iSPM can also be formally defined using basic formal languages. An iSPM X can hence be a string that is built from a concatenation of activities. Thus, X can be identified as a character string.

Surgical activities can be considered as a "symbol alphabet" such that the "alphabet" \( \Omega = \{\omega_1, \ldots, \omega_M\} \) is the set of all symbols that represent the mapping of activities as functions and distances or 2-tuples (pairs) of activities. Generally, a distance function \( d(\,\cdot\,\cdot) \) that represents the distance between pairs of activities is implicitly derived from the temporal information.

The difference in length \( |X| = n \) of two strings \( X \) and \( Y \) is calculated by \( |Y| - |X| \). The length of \( X \) is calculated by \( |X| = n + 1 \).

2.2 Distance Measurements

In the following section, four different approaches for quantifying distances between iSPMs will be introduced: the Jaccard distance, the Levenshtein distance, the Adjacency distance, and a distance based on Graph matching. These measurements were selected because they cover all similarity types. These types belong to sample sets, sequences, causal relationships, and structural relationships of the compared elements of processes. Dijkman [9] introduced analogous types of metrics for measuring the similarity of business process models. Many extended versions of distance measurements exist, including bipartite Graph matching [16]. In this paper, we choose a basic version. After a brief introduction of the theory, the distance measure will be illustrated using example iSPMs that are presented in Figure 1 as \( X_1, X_2 \) and \( X_3 \).

The terms similarity and distance are complementary notions in this context. The more similar two objects are, the smaller their distance is; thus, similarity and distance are inversely related. In a previous study [6], the objective was to propose metrics that express the degree of similarity between two SPMs obtained by different measurements. In contrast to the metrics that were presented there, we use the term distance in this work to emphasize the difference of the iSPMs represents different surgical cases. We illustrate the presentation of the metrics using the example sequences shown in Figure 1.

Figure 1. Example sequences \( X_1, X_2, \) and \( X_3 \) for distance computing.

2.2.1 Jaccard Distance

The Jaccard distance is based on a similarity known as the Jaccard index, which is sometimes called the Jaccard similarity coefficient. The Jaccard index is an average statistic that is primarily used for comparison of the similarity and diversity of different sample sets [34]. Even with a great amount of data, it is relatively easy to calculate.
The Jaccard index is computed as the ratio of the number of common elements and the size of the union. If two iSPMs $X_1$ and $X_2$ are being compared, the number of common elements is calculated using $|X_1 \cap X_2| = \sum_{m=1}^{M} \min[C_{m1}, C_{m2}]$, where the minima of the cardinalities of both iSPMs are summed. The union is the sum of the maxima of the cardinalities of the elements in the alphabet in both iSPMs, $|X_1 \cup X_2| = \sum_{m=1}^{M} \max[C_{m1}, C_{m2}]$.

Therefore, the definition of the Jaccard distance is

$$d_{\text{J}} = 1 - \frac{|X_1 \cap X_2|}{|X_1 \cup X_2|} = 1 - \frac{\sum_{m=1}^{M} \min[C_{m1}, C_{m2}]}{\sum_{m=1}^{M} \max[C_{m1}, C_{m2}]}$$  \hspace{1cm} (1)

In these calculations, not only the traditionally used sample sets but also the number of elements in the sample sets are considered. When a work step is performed for the first time, it is termed $A_1$; when the same work step is performed for the second time, it becomes $A_2$. The frequency of the elements in the sample sets is necessary for the calculation. The basis of calculation of the Jaccard distance in surgical processes is the cardinality of single work steps (e.g., cut) that are performed in both processes in comparison with the overall number of work steps in the iSPM.

A distinguished characteristic of the metric is its lack of sensitivity regarding symbol order: it does not consider the order in which the activities are performed. For instance, $X_1 = \{\text{pull, pull, hold, hold}\}$ and $X_2 = \{\text{hold, pull, hold, pull}\}$ both yield a Jaccard distance of $d_{\text{J}} = 0$. In traditional use, $X_1 = \{\text{pull, pull, hold}\}$ and $X_4 = \{\text{pull, hold}\}$ yield a null distance, but here, the frequency is also considered ($X_4 = \{\text{pull1, pull2, hold1}\}$ and $X_4 = \{\text{pull1, hold1}\}$; therefore, $d_{\text{J}} = 0.33$. An example is presented in Figure 2.

### 2.2.2 Levenshtein Distance

The Levenshtein distance [35] was explicitly developed to assess the similarity of strings. Therefore, it is often referred to as the string edit distance. At present, different versions of the Levenshtein distance are employed. The basis for the calculation is the minimum number of the operations insert, delete, and substitute needed to transform one string into another. An iSPM, as has been demonstrated earlier, can be considered as a string in which the single letters represent different surgical work steps. Thus, the Levenshtein distance can easily be applied to calculate the similarity of iSPMs.

Let $X_2$ be the i-th activity in an iSPM $X_1$. If $X_1 \neq X_2$, then $X_2$ can be replaced by $x_1$, which constitutes one substitution. In the deletion operation for an iSPM $X_1$ with $|X_1| = N$, the element $x_1$ must be eliminated. Still, $i \leq N$ must be true. The new iSPM $X_{\text{new}}$ then has length $|X_{\text{new}}| = N - 1$, and all the indices $j$ for which $i < j \leq N$ is true must be transposed to $j - 1$. The insertion operation works similarly. A new element $x_{\omega_m} \in \Omega$ is added to an iSPM $X_1$ of length $|X_1| = N$ at position $i$. Subsequently, all indices $j$ for which $i < j \leq N$ is true must be transposed to $j + 1$ such that $|X_{\text{new}}| = N + 1$. Cost functions can be added to weight different edit operations.

The editing operations $E$ and their order of succession are not definite. There are various possibilities for transforming one string into another. The Levenshtein distance measures the minimal number of such operations performed to transform iSPM $X_1$ into iSPM $X_2$:

$$d_L = \min_{E} \left| \sum_{x_1 \in X_1} E(x_1) \right|$$

where the minima of the cardinalities of the elements $X_1 \cap X_2$.

![Figure 2](image-url) Example calculation of the Jaccard distance for $X_1$ with $X_2$ and $X_1$ with $X_3$: $d_{\text{J}}^{12} = 0.05$ and $d_{\text{J}}^{13} = 0.40$

In the special case that $X_1 \cap X_2 = \emptyset$, the Levenshtein distance is $d_L^{12} = \max| |X_1|, |X_2| |$

The basis for the calculation of the Levenshtein distance thus comprises not only the number of activities but also their order in the surgical process model. Figure 2 presents an example of edit operations.

### 2.2.3 Adjacency Distance

Adjacency matrices are neighborhood matrices in the sense that they depict the consecutiveness of elements of a process or graph [14]. A matrix $\Omega \times \Omega$ in which the elements $\omega_m \in \Omega$ form the rows and columns is constructed. An Adjacency matrix is a square matrix of size $|\Omega| = M$. With regard to the iSPM, this matrix is constructed from all possible work steps of an intervention.

For a given string $X_1$, each element of the matrix represents the frequency $C_{mk}$ of the substring $Y = |\omega_k \cdot \omega_m|, 1 \leq m, k \leq M; m, k \in \mathbb{N}$ in $X_1$, where every substring has a length of 2, $|Y| = 2$. For a comparison of the two iSPMs $X_1$ and $X_2$, the Adjacency matrices $MX_1$ and $MX_2$ are constructed (Figure 3). For calculation of the Adjacency distance:

$$d_A = \sum_{x_1 \in X_1} \sum_{x_2 \in X_2} |C_{mx} - C_{mk}|$$

Example of Adjacency Distance:

- $X_1$ with $X_2$: $d_A = 2$
- $X_1$ with $X_3$: $d_A = 3$
S. Schumann et al.: Distance Measures for Surgical Process Models

adjacency distance, the Euclidian distance of the two matrices \(M_X^1\) and \(M_X^2\) is calculated using

\[
d_{A}^{12} = \sqrt{(s_{11} - t_{11})^2 + \ldots + (s_{MM} - t_{MM})^2}
\]

\(S_{mk} = C_{mk}\) in \(X_1\) and \(t_{mk} = C_{mk}\) in \(X_2\). Thus, transitions between activities and their frequency form the basis for the calculation of the Adjacency distance. The computed distances for the example given above are \(d_{A}^{12} = 3.32\) and \(d_{A}^{13} = 6.32\).

### 2.2.4 Graph Matching Distance

Graph matching uses methods from graph theory. The transitions between the different nodes of a graph can be calculated using an Adjacency matrix [15]. Furthermore, it is also possible to reverse the operation and construct a graph from an Adjacency matrix. If the matrix is constructed, as depicted in Section 2.2.3 with regard to the Adjacency distance, and transformed into a graph, then the resulting representation depicts the course of a procedure, including possible loops, iterations, predecessors, and successors. In a surgical process, retransformation of such a graph is not necessary for distance measurement and does not yield definite results.

If two iSPMs, \(X_1\) and \(X_2\), are compared, the Graph matching operation searches for the graph that constitutes the maximum common subgraph. For this aim, a new subgraph matrix \((SG)\) is constructed element-wise using the minimum of \(M_X^1\) and \(M_X^2\). \(|G|\) is the number of all edges within graph \(X_1\), which is constructed from iSPM \(X_1\). Figure 5 presents the respective graphs constructed from the matrices presented in Figure 4. In Graph matching, the distance is computed as follows:

\[
d_{G}^{12} = 1 - \frac{|SG|}{\max(|G_1|, |G_2|)} \tag{4}
\]

Here, the subprocess that both iSPMs have in common is used as the basis. Thus, behavior is the basis. A positive aspect of Graph matching is that it is independent of cost functions.

### 3. Evaluation Study

#### 3.1 Study Setup and Clinical Data Sets

We applied the four measures to clinical data sets of surgical processes to evaluate the application of the distance measures. The data sets consisted of iSPMs that were obtained during pediatric surgical training sessions. To evaluate our methods, we measured the iSPMs of novice and expert surgeons while they performed several complex exercises in laparoscopic training and assessed the distance between both groups. We assumed that the surgical processes are significantly more similar among expert surgeons than between novice surgeons.

The iSPMs were obtained by observation using the methods described by [36].
during training sessions in minimally invasive surgery in the Department of Pediatric Surgery of the University Medical Center Leipzig in 2011.

Members of two subject groups, one consisting of five novices and the other of five expert surgeons, each performed three different tasks (a cutting task, simple suturing, and complex suturing) by applying different surgical strategies. A sequence of work steps was needed to perform a task. A Pelvitrainer [37], which represented the abdomen, and silicone manikins, on which several tasks were performed, were used (Figure 6). The surgical strategies involved variations of applying different instrument types and incision points: single incision with angled laparoscopic instruments, triple incision with straight laparoscopic instruments, and triple incision with angled laparoscopic instruments. The subjects iterated each of the tasks and each of the incision/instrument combinations five times. Thus, the evaluation data set contained 450 iSPMs and a total of approximately 28,600 activities.

The work steps of the 450 training sessions were recorded using the ICCAS surgical workflow editor (27) and contained a high level of information. The decomposition of the surgical processes into single surgical motions followed the hierarchical decomposition approach of MacKenzie et al. [38, 39]. This very fine breakdown of surgical work steps can lead to double activities in the sense that one work step was performed several times consecutively, such as \( \star \ AA \ \star \) in iSPM \( X_2 \). Such double activities occur when the subject takes a break of less than three seconds between work steps.

The finite set of motions that was used in all iSPMs was \{push; grasp, successful; grasp, unsuccessful; hold; unclasp; cut; suture; change holding position; pull; arrange\}. An example representation of the ordered activities for the simple suturing task is presented in Figure 7.

In all available iSPMs, the activities that were performed by the right hand of the subject were ordered by time and then mapped onto symbols. Thus, the example from Figure 7 is transformed into \( X_1 \), as observed in Figure 1.

The main work steps for analyzing the data are described in Table 1. After data acquisition, activity ordering, and alphabet mapping, the iSPMs were grouped according to surgeons, surgical strategy, and tasks because only surgical processes regarding the same subject matter were compared. For example, it was not reasonable, from the clinical point of view, to compare the cutting task with the complex suturing task.

We subsequently calculated the average distance for all iterations of a task for each surgeon, and these average values were finally assessed to compare novice and expert surgeons. We performed a statistical analysis using the Mann-Whitney U test to estimate the significance of the difference between the two groups, and we used a significance level of \( \alpha = 0.05 \).

### 3.2 Results of the Evaluation Study

The objective of the evaluation study was to assess the distance measurements for the surgical process models of novice and expert surgeons. In total, 450 surgical process models were analyzed.
The results demonstrate that both the Levenshtein and the Adjacency distance were better suited for calculating the distances between surgical process models than the Jaccard distance and Graph matching distance. The Levenshtein distance performed slightly better than the Adjacency distance. However, with regard to practical application, the results should be regarded in a different manner. The contents emphasized are important of the choice of an appropriate measurement of distance. Each of the assessed methods was based on a different background. Therefore, it could also be sensible to use a distance measurement different from those presented here.

The Jaccard distance is based solely on the set of activities and does not consider their sequence. Thus, similar procedures also have similar frequencies. However, this aspect cannot be reversed; a similar frequency does not necessarily refer to the same procedure \( d'(AABB, BABA) = 0 \). In contrast, the Jaccard distance is quite simple and can be computed without much effort, which is useful, especially when a vast amount of data is present, e.g., in operations with many single work steps. The Jaccard distance does provide a clear distinction, but a rough allocation is possible. To determine the SPMs that best match the gold standard, this method can serve as a pre-selective approach. In such an approach, only the processes for which the distance is less than a previously set threshold are used for the actual distance measurement, thereby sparing resources. Therefore, the Jaccard distance can be used as a coarse means of classification.

The Levenshtein distance was calculated by assessing the minimum effort involved to transform one sequence of work steps into another sequence. Therefore, it is based not only on the frequency but also on the sequence of activities. Compared with the Jaccard method, this approach yields a better overview of the course of work required to perform a surgical task.

Transitions of activities are important for the Adjacency distance. Hence, the paired cohesiveness of activities provides the basis for measuring distance with Adjacency matrices. All possible sequences of two separate activities and their frequencies are well-behaved patterns for the workflow during the performance of a task. Thus, the Adjacency matrix is a good method for distance measurement.

The Graph matching distance uses methods from graph theory and considers sequences of activities. However, a disadvantage of this method is that some of the surgical processes only contain a single work step. Comparisons between these single work steps can lead to very high distances even when the real distance between these steps is low. For instance, in \( d'(AAA, A) = 1 \), the activity is repeated thrice or only once, respectively. Nevertheless, the distance calculated is 1. The other distance measures provide more reasonable results. Future studies should consider whether the Graph matching distance can be sensibly adapted to this problem.

Because no ground truth measure is available in the research field of SPMs for comparison with our metrics, we had to create a "common-sense approach". Therefore, the evaluation study was designed to analyze the modi operandi of the experienced surgeons and novices.

Statistical analysis of expert and novice surgeons was a valuable approach to identify the appropriate distance measures. The precise, and hence less varying, modus operandi of the experienced surgeons, compared with the more erratic and unstable methods of the novices, were used as the basis to identify the distance measurement that is best suited to quantify the distances between surgical processes. This assumption was supported by significant time differences; all tasks required longer overall durations when performed by the novice surgeons. These differences were significant \( p < 0.05 \) for both suturing tasks in all settings but in only one setting for cutting. Furthermore, the working strategies of the novice and expert surgeons were significantly different in terms of the number of activities performed.

These results demonstrate that more experienced surgeons performed short and standard processes that were very similar to one another, whereas novices tended to add

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data acquisition of the 450 different iSPMs with chronologically ordered activities</td>
</tr>
<tr>
<td>2</td>
<td>Mapping of all activities ( \Phi ) onto the alphabet ( \Omega )</td>
</tr>
<tr>
<td>3</td>
<td>iSPM grouping according to surgeon, strategy, and task</td>
</tr>
<tr>
<td>4</td>
<td>Calculation and averaging of distance measures between the iterations of each combination of strategy and task according to surgical experience (novice or expert surgeon)</td>
</tr>
<tr>
<td>5</td>
<td>Statistical test for each combination of strategy and task according to surgical experience (novice or expert surgeon)</td>
</tr>
</tbody>
</table>

4. Discussion

The results demonstrate that both the Levenshtein and the Adjacency distance were better suited for calculating the distances between surgical process models than the Jaccard distance and Graph matching distance. The Levenshtein distance performed slightly better than the Adjacency distance. However, with regard to practical application, the results should be regarded in a different manner. The contents emphasized are important of the choice of an appropriate measurement of distance. Each of the assessed methods was based on a different background. Therefore, it could also be sensible to use a distance measurement different from those presented here.

The Jaccard distance is based solely on the set of activities and does not consider their sequence. Thus, similar procedures also have similar frequencies. However, this aspect cannot be reversed; a similar frequency does not necessarily refer to the same procedure \( d'(AABB, BABA) = 0 \). In contrast, the Jaccard distance is quite simple and can be computed without much effort, which is useful, especially when a vast amount of data is present, e.g., in operations with many single work steps. The Jaccard distance does provide a clear distinction, but a rough allocation is possible. To determine the SPMs that best match the gold standard, this method can serve as a pre-selective approach. In such an approach, only the processes for which the distance is less than a previously set threshold are used for the actual distance measurement, thereby sparing resources. Therefore, the Jaccard distance can be used as a coarse means of classification.

The Levenshtein distance was calculated by assessing the minimum effort involved to transform one sequence of work steps into another sequence. Therefore, it is based not only on the frequency but also on the sequence of activities. Compared with the Jaccard method, this approach yields a better overview of the course of work required to perform a surgical task.

Transitions of activities are important for the Adjacency distance. Hence, the paired cohesiveness of activities provides the basis for measuring distance with Adjacency matrices. All possible sequences of two separate activities and their frequencies are well-behaved patterns for the workflow during the performance of a task. Thus, the Adjacency matrix is a good method for distance measurement.

The Graph matching distance uses methods from graph theory and considers sequences of activities. However, a disadvantage of this method is that some of the surgical processes only contain a single work step. Comparisons between these single work steps can lead to very high distances even when the real distance between these steps is low. For instance, in \( d'(AAA, A) = 1 \), the activity is repeated thrice or only once, respectively. Nevertheless, the distance calculated is 1. The other distance measures provide more reasonable results. Future studies should consider whether the Graph matching distance can be sensibly adapted to this problem.

Because no ground truth measure is available in the research field of SPMs for comparison with our metrics, we had to create a "common-sense approach". Therefore, the evaluation study was designed to analyze the modi operandi of the experienced surgeons and novices.

Statistical analysis of expert and novice surgeons was a valuable approach to identify the appropriate distance measures. The precise, and hence less varying, modus operandi of the experienced surgeons, compared with the more erratic and unstable methods of the novices, were used as the basis to identify the distance measurement that is best suited to quantify the distances between surgical processes. This assumption was supported by significant time differences; all tasks required longer overall durations when performed by the novice surgeons. These differences were significant \( p < 0.05 \) for both suturing tasks in all settings but in only one setting for cutting. Furthermore, the working strategies of the novice and expert surgeons were significantly different in terms of the number of activities performed.

These results demonstrate that more experienced surgeons performed short and standard processes that were very similar to one another, whereas novices tended to add...
more, sometimes unnecessary, work steps and were unsure of the correct course of actions, especially when compared with the targeted and practiced work sequences of their more experienced colleagues [40, 41].

There is good reason to believe that more experienced subjects have similar modi operandi but have nevertheless developed personal ‘trademarks’. Therefore, the standardized line of action of a well-experienced surgeon and the low amount of variation can be explained.

The novices also showed different approaches every time they had to perform a task. In addition, the approaches of the individual novices differed greatly from one another. In contrast, the experts were much more consistent in both regards.

To determine whether a sequence of work steps was performed by an expert or by a novice in this study, it was adequate to evaluate the whole sequence of work steps and its duration in the protocol. However, there was no information regarding the reason for longer or shorter durations. To learn about the differences between surgical processes, it is therefore not sufficient to only consider duration.

Using the methods evaluated in this paper, it is possible to determine the impact of resources, e.g., new techniques in surgery, new surgical instruments and the experience of the surgeon. The following use case can illustrate this possibility: a clinic has been offered new surgical instruments, which have the same application area but different modes of operation, by two companies. With the distance measurements, it is possible to quantify which of these instruments has a greater impact on the whole course of intervention in comparison with conventional courses. This comparison may help the clinic to choose the instrument that better suits the needs of the clinic.

A study concerning the interference of different incision methods and instruments in pediatric laparoscopic surgery was used as the basis for the data in this study. A next step would be to use data obtained from real surgical interventions. However, surgical activities are not the only aspects that are important for the course of the intervention. Other considerations, such as the surgical instrument used [41] or the anatomical structure treated, are also of interest.

This study only considered the functional perspective (activities). Extension of the focus to other perspectives, such as the

Table 2 Distances between experienced and novice surgeons

<table>
<thead>
<tr>
<th>setting</th>
<th>task</th>
<th>Jaccard distance</th>
<th>Levenshtein distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>expert</td>
<td>novice</td>
</tr>
<tr>
<td>angled laparoscopic instruments, triple incision</td>
<td>cutting task</td>
<td>0.50 ± 0.11</td>
<td>0.60 ± 0.13</td>
</tr>
<tr>
<td></td>
<td>simple suturing</td>
<td>0.39 ± 0.09</td>
<td>0.58 ± 0.06</td>
</tr>
<tr>
<td></td>
<td>complex suturing</td>
<td>0.38 ± 0.10</td>
<td>0.49 ± 0.07</td>
</tr>
<tr>
<td>angled laparoscopic instruments, single incision</td>
<td>cutting task</td>
<td>0.46 ± 0.20</td>
<td>0.53 ± 0.12</td>
</tr>
<tr>
<td></td>
<td>simple suturing</td>
<td>0.50 ± 0.08</td>
<td>0.51 ± 0.05</td>
</tr>
<tr>
<td></td>
<td>complex suturing</td>
<td>0.44 ± 0.06</td>
<td>0.37 ± 0.05</td>
</tr>
<tr>
<td>straight laparoscopic instruments, triple incision</td>
<td>cutting task</td>
<td>0.48 ± 0.22</td>
<td>0.50 ± 0.19</td>
</tr>
<tr>
<td></td>
<td>simple suturing</td>
<td>0.48 ± 0.13</td>
<td>0.48 ± 0.09</td>
</tr>
<tr>
<td></td>
<td>complex suturing</td>
<td>0.44 ± 0.08</td>
<td>0.51 ± 0.14</td>
</tr>
</tbody>
</table>

Agreement with assumption 33% 89%

Adjacency distance

<table>
<thead>
<tr>
<th>setting</th>
<th>task</th>
<th>expert</th>
<th>novice</th>
<th>p-value</th>
<th>expert</th>
<th>novice</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>angled laparoscopic instruments, triple incision</td>
<td>cutting task</td>
<td>1.3 ± 0.4</td>
<td>9.3 ± 5.7</td>
<td>p &lt; 0.001</td>
<td>0.88 ± 0.10</td>
<td>0.67 ± 0.13</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>simple suturing</td>
<td>9.1 ± 1.5</td>
<td>47.8 ± 11.5</td>
<td>p &lt; 0.001</td>
<td>0.51 ± 0.09</td>
<td>0.64 ± 0.08</td>
<td>p = 0.001</td>
</tr>
<tr>
<td></td>
<td>complex suturing</td>
<td>10.1 ± 2.7</td>
<td>38.1 ± 9.7</td>
<td>p &lt; 0.001</td>
<td>0.47 ± 0.11</td>
<td>0.56 ± 0.08</td>
<td>p = 0.033</td>
</tr>
<tr>
<td>angled laparoscopic instruments, single incision</td>
<td>cutting task</td>
<td>2.0 ± 1.2</td>
<td>5.4 ± 3.1</td>
<td>p &lt; 0.001</td>
<td>0.94 ± 0.06</td>
<td>0.63 ± 0.15</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>simple suturing</td>
<td>20.5 ± 5.2</td>
<td>44.1 ± 9.5</td>
<td>p &lt; 0.001</td>
<td>0.58 ± 0.10</td>
<td>0.57 ± 0.07</td>
<td>p &gt; 0.050</td>
</tr>
<tr>
<td></td>
<td>complex suturing</td>
<td>17.9 ± 4.7</td>
<td>41.6 ± 8.6</td>
<td>p &lt; 0.001</td>
<td>0.53 ± 0.09</td>
<td>0.45 ± 0.08</td>
<td>p &gt; 0.050</td>
</tr>
<tr>
<td>straight laparoscopic instruments, triple incision</td>
<td>cutting task</td>
<td>3.0 ± 2.3</td>
<td>6.5 ± 5.2</td>
<td>p &gt; 0.050</td>
<td>0.90 ± 0.13</td>
<td>0.56 ± 0.18</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>simple suturing</td>
<td>14.6 ± 5.9</td>
<td>19.8 ± 7.0</td>
<td>p &gt; 0.050</td>
<td>0.59 ± 0.16</td>
<td>0.61 ± 0.12</td>
<td>p &gt; 0.050</td>
</tr>
<tr>
<td></td>
<td>complex suturing</td>
<td>11.3 ± 2.5</td>
<td>24.7 ± 9.0</td>
<td>p &lt; 0.001</td>
<td>0.54 ± 0.12</td>
<td>0.62 ± 0.16</td>
<td>p &gt; 0.050</td>
</tr>
</tbody>
</table>

Agreement with assumption 78% 22%
operational perspective, which considers the instruments used, should be included in the distance analysis. Because the activities performed were closely related to the surgical instrument used, the conjunction of these two perspectives would answer another host of interesting issues, such as "How does a newly developed instrument X influence the activities NOT performed using the instrument?"

The cost matrix used to calculate the Levenshtein distance is of great importance. Occasionally, different costs for insertion and deletion operations relative to substitution operations are estimated. Prominent findings can be obtained from context-sensitive variation of the substitution operation. In surgical interventions, many activities are apparently different but serve the same goals. In these cases, a less cost-intensive substitution would be sensible. In addition, the surgical activities clean and rinse are semantically closer to one another than, for instance, clean and suture [31]. Assessment of the semantic similarity of surgical activities could be advantageously included in the distance analysis.

As mentioned in Chapter 3.1, it is possible that the same work step is performed several times consecutively. However, it is important that these occurrences are regarded as several distinct work steps and not as a single work step because only by making this distinction is it possible to measure the differences between the two methods. An example is cutting tissue with scissors. Some subjects may cut in one step, whereas other subjects may cut a shorter distance and then take a short break to relocate the scissors before resuming the cutting action. With this differentiation, it is possible to analyze which of these methods leads to shorter durations and/or to more accurate cutting results.

In this study, the comparison of surgical processes was based on the work steps performed, i.e., the activities themselves. The four presented distance measures were based on activities. Additional methods are available that consider other aspects. One of these aspects is the temporal extent of the activities, which has not been taken into consideration. However, the results of this paper demonstrate that straightforward and low-complexity methods that do not consider the duration of work steps are sufficient to measure distances between surgical processes. The employment of dynamic time warping [20] or other methods could also be investigated.

In a clinical environment, comparison of different surgical processes can serve many purposes. For instance, surgeons-in-training could be supported in learning well-established surgical methods. With the new method it is possible to analyze, and thus influence, the learning curves during training. The single test runs executed by the study participants can be recorded and compared with the gold standard for the intervention methodology. Using the similarity of the executed intervention with the standard as a basis, the progress of the surgeon can be extracted, which will enable recognition of the phases in which more training is needed and of the phases that have already been mastered and are therefore not in need of improvement.

5. Conclusions

The measurement of distances between surgical processes is necessary to estimate the benefit of new surgical techniques and strategies. Consequently, the development of new surgical assistance systems for the support of the surgeon should strongly emphasize the impact of the new system on the surgical process.

Measurement of distances between processes is also important in surgical training and education. The learning curves of students or the compliance of surgical experts with best practices can be quantified. In laparoscopic surgery, the skill level of the surgeon is very important for the success of the operation. Using distance measurements, the key aspects of surgical laparoscopic training can be expressed more precisely.

We have introduced four distance measures for surgical processes. The difference in the workflows between novices and experienced surgeons was the basic concept for evaluation. By applying the distance measures to 450 surgical processes from laparoscopic training in pediatric surgery, we demonstrated that the Levenshtein and Adjacency distances are best suited for measurement of distances between the activity sequences of surgical process models.

Acknowledgments

We thank the study participants from the Department of Pediatric Surgery at the University Medical Center Leipzig for their support. Many thanks are also owed to Dayana Neumuth for her valuable comments and discussion that led to improvement of the manuscript.

ICCAS is funded by the German Federal Ministry of Education and Research (BMBF) and the Saxon Ministry of Science and Fine Arts (SMWK) in the scope of the Unternehmen Region with grant number 03ZILN12.

References


