Sequential surgical signatures in micro-suturing task

Arnaud Huaulmé, Kanako Harada, Germain Forestier, Mamoru Mitsuishi,
Pierre Jannin

HAL Id: hal-01791310

https://hal.science/hal-01791310

Sequential surgical signatures in micro-suturing task

**Arnaud Huaulmé · Kanako Harada ·
Germain Forestier · Mamoru Mitsuishi ·
Pierre Jannin**

Abstract

Purpose Surgical processes are generally only studied by identifying differences in populations such as participants or level of expertise. But the similarity between this population is also important in understanding the process. We therefore proposed to study these two aspects.

Methods In this article, we show how similarities in process workflow within a population can be identified as sequential surgical signatures. To this purpose, we have proposed a pattern mining approach to identify these signatures.

Validation We validated our method with a data set composed of seventeen micro-surgical suturing tasks performed by four participants with two levels of expertise.

Results We identified sequential surgical signatures specific to each participant, shared between participants with and without the same level of expertise. These signatures are also able to perfectly define the level of expertise of the participant who performed a new micro-surgical suturing task. However, it is more complicated to determine who the participant is, and the method correctly determines this information in only 64% of cases.

Conclusion We show for the first time the concept of sequential surgical signature. This new concept has the potential to further help to understand surgical procedures and provide useful knowledge to define future CAS systems.

A. Huaulmé · P. Jannin
Univ Rennes, Inserm, LTSI - UMR_S 1099, F35000, Rennes, France
E-mail: arnaud.huaulme@univ-rennes1.fr

K. Harada · M. Mitsuishi
Department of Mechanical Engineering, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

G. Forestier
MIPS, University of Haute-Alsace, Mulhouse, France

1 Introduction

We all have our own habits that depend on our past. For example, some people take a shower when they wake up, while others prefer to take a shower before going to sleep. Although all surgical procedures are unique because of the patient's anatomical characteristics, they do not escape to this rule because of the habits and experience of the surgical team. The surgical process modeling methodology, which was introduced around 15 years ago [1, 2], could be used to study these "habits". A surgical process model describes a surgical procedure at different levels of granularity [2]. For example, a surgical intervention can be divided into successive phases corresponding to the main periods of the intervention. A phase is composed of one or more steps. A step is a sequence of activities used to achieve a surgical objective. An activity is a physical action performed by the surgeon. Each activity is broken down into different components, including the verb of action, the target involved in the action (usually an anatomical structure) and the surgical instrument used to perform the action. Lower granularity levels are closer to kinematic data, such as surgemes and dexemes [3, 4]. A surgeme was defined as a surgical motion with explicit semantic meaning, composed by dexemes. A dexeme is a numerical representation of the performed physical motion. Surgical Process Models (SPMs) have been developed for three main purposes: (1) Formalize surgical knowledge, (2) Evaluate surgical skills and systems, (3) Assist the surgeon in surgical intervention.

A SPM can be acquired manually from observations [5] or automatically thanks to recent advances in automatic recognition of phases [6, 7], steps [8, 9] and activities [10, 11]. These SPMs have recently been used to identify different surgical behaviors, such as those depending on surgical sites [12, 13], surgical skills [14], types of procedures used [15] and surgical expertise levels [12, 16, 13].

In these studies, the analysis is generally done by underlining differences between two or more populations, using one or several information, such as the surgical duration [14, 16], the number of activities [14, 16] or sequence based metrics [12]. Recently, [13, 17] showed that sequences were highly discriminatory.

In this paper, we introduce the concept of sequential surgical signatures: sequences of phases, stages or activities being common within a more or less homogeneous population. To demonstrate this concept, we propose an approach that is an extension of a method presented in [13].

2 Material and methods

The aim of this paper is to identify sequential surgical signatures in the context of micro-surgical suturing training task (see subsection 2.1). For this, we used a pattern mining method presented in subsection 2.2.

2.1 Data

The data set was collected at the Tokyo University Hospital. It consists of seventeen micro-surgical suture tasks of a 0.7mm artificial blood vessel performed using a master-slave robotic platform [18]. Figure 1 shows snapshots of this task. The data set included 4 participants with different levels of surgical expertise and robotics skills. Two of them, called experts, are surgeons but novice roboticians, the other two, called engineering students, have no surgical skills, but are expert roboticians. Each participant conducted between 3 and 6 trials, according to their availability. This explains why there were in total 7 trials made by surgeons, and 10 by engineering students. The average suture duration is about 3 minutes. For each test, the video was recorded at 30Hz. Thanks to these videos, both hands were annotated manually, at the level of granularity of the activities, using the software "Surgery Workflow Toolbox [annotate]" [19]. The suture task is relatively simple to describe if the chosen granularity is superficial. Indeed, the stain consists in taking the needle, passing through the two artificial blood vessels and making 3 knots. But, such a description cannot capture variations between participants. Thus, we have broken down each gesture as much as possible in order to better describe the progress of the task. Thus, we are able to take into account the gestures that are repeated several times before being completed, as well as the intra-participant variabilities. Table 1 summarizes the number of trials, average duration and average number of activities per hand for each participant.
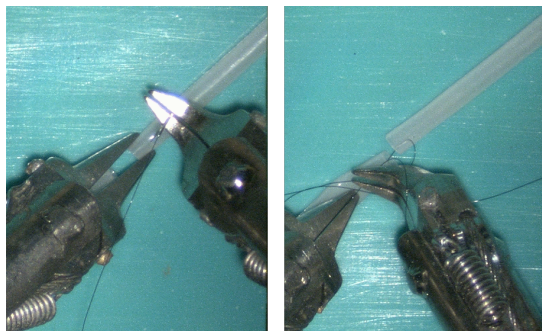

Fig. 1: Snapshots of micro-surgical suture tasks of a 0.7mm artificial blood vessel performed using a master-slave robotic platform.

Table 1: Description of the 17 trials according to participants

Participant Id		1	2	3	4
Expertise level		Student	Student	Expert	Expert
Number trial		4	6	3	4
Mean duration		4'08	5'46	2'45	1'48
Mean number	Left Hand	37	52	31	32
of activities	Right Hand	45	52	39	27

The output of the surgical process annotation is a sequential list of phases, steps and/or activities performed by the participant's left and right hands. In order to analyze both hand sequences, we preprocessed the data by a step called synchronization. It consists of dividing, step a in figure 2, the activity from one hand into two parts when on the other hand an activity changes of status (begin or end). Then, the activities of both hands are grouped together in the same sequence, and when no activity is present on one of the hands, the emptiness is supplemented by an activity, called "Idle", representing this absence of activity, step b in figure 2.
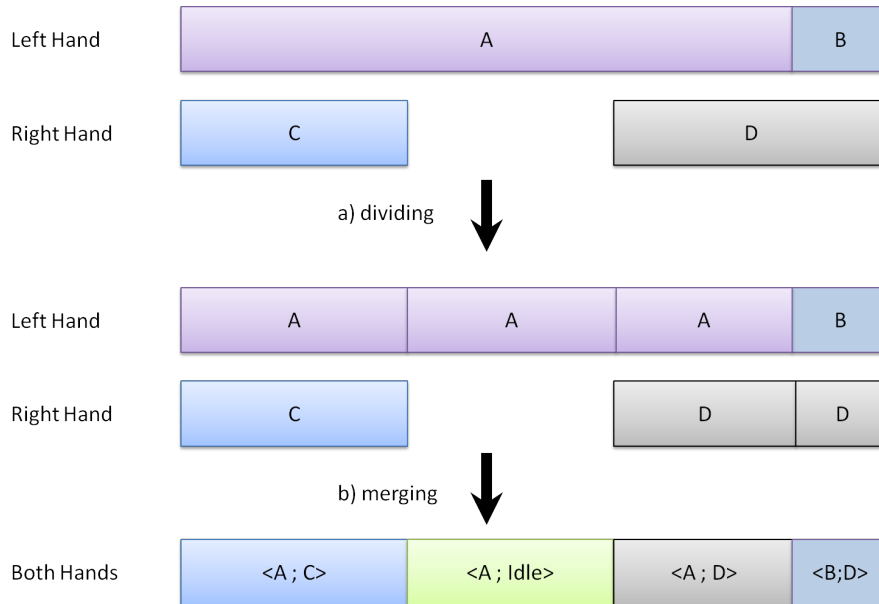

Fig. 2: The process of synchronizing the sequences of the left and right hand to create a sequence with both hands. Each letter corresponds to an activity, Idle means that no activity is executed by a hand.

As a reminder, an activity is composed of three components: the verb of action, the target and the surgical instrument. To improve readability, we do

not use information from the surgical instrument as only one surgical instrument was used in all trials. Thus, we will note the activities of both hands as follows: "$< verb_{left}, target_{left} >; < verb_{right}, target_{right} >$".

2.2 Methods

In this paper, we propose to extend a method published in [13] and use it as a means to identify sequential surgical signatures. In summary, the method consists of finding the longest frequent patterns in sequences, i.e. identifying the longest sequence of activities (2 or more) which are present at least min_fr times in all sequences, where min_fr is a predetermined threshold. The original method is composed of three steps:

- Step 1: establish a vocabulary of frequent activities;
- Step 2: generate possible frequent patterns of length k, thanks to the frequent patterns of length $k-1$ and frequent activities;
- Step 3: determine if possible frequent patterns are really frequent and compute the longest frequent patterns of size $k-1$.

Steps 2 and 3 are repeated to extend the patterns until no new frequent patterns of size k are found. At each loop, the longest frequent patterns of size $k-1$ are added to the longest frequent patterns of smaller sizes.

The extension consists in removing all patterns that are composed of less than min_length activities. This step assumes that the shorter patterns do not have enough discriminating power to be interesting. Finally, for the other patterns, we determine if they are sequential surgical signatures or not by checking if they are shared within a more or less homogeneous population. The figure 3 summarizes the complete process for a simple example with the following parameters: frequency threshold $min\_fr = 2$ and the length threshold $min\_length = 3$.

To classify sequences, we use the Shared Longest Frequent Sequential Pattern metric (SLFSP metric) developed in [13] to make a hierarchical clustering with the average-link approach using UPGMA algorithm (Unweighted Pair Group Method with Arithmetic Mean) [20]. SLFSP metric as defined as:

$$SLFSP(A, B) = \frac{|shared_{A,B}|}{|patterns_A| + |patterns_B| - |shared_{A,B}|} \quad , \qquad (1)$$

where A and B are 2 sequences, $|shared_{A,B}|$ is the number of shared longest frequent sequential patterns between A and B, and $|patterns_A|$ and $|patterns_B|$ are respectively the number of longest frequent patterns of A and B.

3 Validation studies

We propose three validation studies, first of all, to verify the usefulness of the additional step (subsection 3.1). The aim of the second study (subsection 3.2)
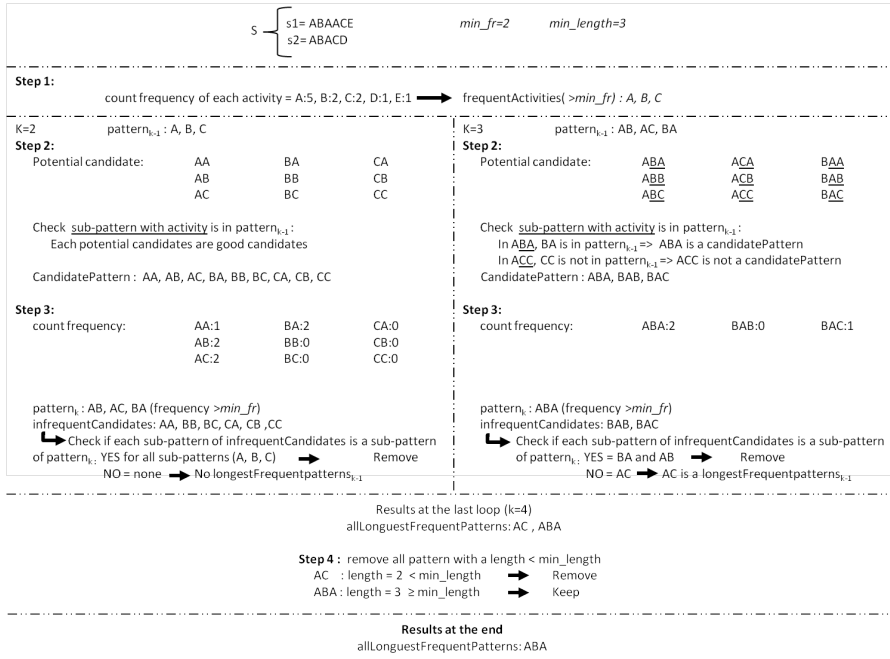
Fig. 3: Sequential pattern discovery method procedure for a simple example. S is a set of activity sequences, min_fr is a frequency threshold and min_length is the pattern length threshold.

is to identify sequential surgical signatures according to the participants and their level of expertise, but also shared between different populations. Finally, we use sequential surgical signatures to predict from which populations a new sequence belongs to (subsection 3.3).

3.1 Classification according to sequential signatures

The objective of this first study is to ensure that the evolution of the method produces better results than the original, or in the worst case, that it does not deteriorate them. To do this, we try to classify the sequences by level of expertise and participants using both methods.

For the first study we tested different parameter values, varying them in the following way:

- Frequency threshold: $min\_fr \in [2, 7]$ for both methods. We have not tested for frequency thresholds superior to 7, because in these cases it would not have been possible to have patterns present only among experts (only 7 trials are made by experts);
- Length threshold: $min\_length \in [3, 10]$ for proposed method.

3.2 Analysis of sequential surgical signatures

In this study, we use the method on all the data to identify sequential surgical signatures. Based on optimal results of the first study, we selected, for this study and the next one, the following parameters:

- Frequency threshold: $min\_fr = 3$;
- Length threshold: $min\_length = 3$.

3.3 Prediction of belonging to a population

In this study, we determine who is the participant who performs a new sequence and his or her level of expertise, thanks to the signatures present in his or her sequence. To do this, we conduct a leave one out cross-validation study. We trained our model on all sequences except one. For all the longest frequent patterns, we determined whether this pattern was an indicator of sequential surgical signature and the percentage of sequences where this pattern is present. For the remaining sequence, we checked the presence of all signatures. With this signature list, we were able to determine the metadata of the remaining sequence. If some signatures are specific to contradictory populations, we have determined the belonging for the remaining sequence based on the highest probability of belonging defined as follows:

$$\frac{1}{n}\sum_{i=1}^{n}\%present_i * \frac{nb\_Pattern\_specific}{nb\_Pattern\_With\_Contradictory\_Meta\_Data} \qquad (2)$$

4 Results

In this section, we present the results of the studies respectively in subsection 4.1, 4.2 and 4.3.

4.1 Classification according to sequential signatures

Tables 2 and 3 summarize the accuracy of both methods to distinguish between the levels of expertise, and respectively between the participants, for different parameter configurations. To determine this accuracy we use a distance of 0.6 to define clusters.

   The proposed method gives better results than the original for the same frequency threshold when the length threshold is 3. The only exception is for a frequency threshold of 7, where the accuracy of expertise classification is the same (94.12%) and lower for the classification of participants (76.47% for the original method versus 64.71% for the proposed one). When the length threshold increases the classification accuracy decreases or stays stable. Figure 4

min_fr		2	3	4	5	6	7
original		58.82%	88.24%	94.12%	94.12%	94.12%	94.12%
min_length	3	76.47%	100%	100%	100%	100%	94.12%
	4	58.82%	100%	100%	94.12%	88.24%	88.24%
	5	58.82%	94.12%	88.24%	82.35%	47.06%	64.71%
	6	82.35%	94.12%	76.47%	70.59%	52.94%	64.71%
	7	76.47%	88.24%	52.94%	52.94%	52.94%	52.94%
	8	76.47%	64.71%	47.06%	52.94%	52.94%	52.94%
	9	82.35%	58.82%	47.06%	47.06%	47.06%	35.29%
	10	70.59%	47.06%	47.06%	47.06%	47.06%	Na

Table 2: Accuracy of the classification of expertise using both methods and the SLFSP metric for a cutting threshold of 0.6. Na: not applicable

min_fr		2	3	4	5	6	7
original		52.94%	88.24%	82.35%	82.35%	82.35%	76.47%
min_length	3	76.47%	94.12%	88.24%	88.24%	88.24%	64.71%
	4	58.82%	94.12%	88.24%	88.24%	58.82%	58.82%
	5	58.82%	88.24%	52.94%	70.59%	41.18%	41.18%
	6	76.47%	88.24%	70.59%	58.82%	17.65%	41.18%
	7	64.71%	82.35%	41.18%	35.29%	35.29%	35.29%
	8	64.71%	52.94%	41.18%	35.29%	35.29%	35.29%
	9	82.35%	47.06%	17.65%	29.41%	29.41%	23.53%
	10	70.59%	35.29%	17.65%	29.41%	29.41%	Na

Table 3: Accuracy of the classification of participant using both methods and the SLFSP metric for a cutting threshold of 0.6. Na: not applicable

(a) Accuracy of expertise classification
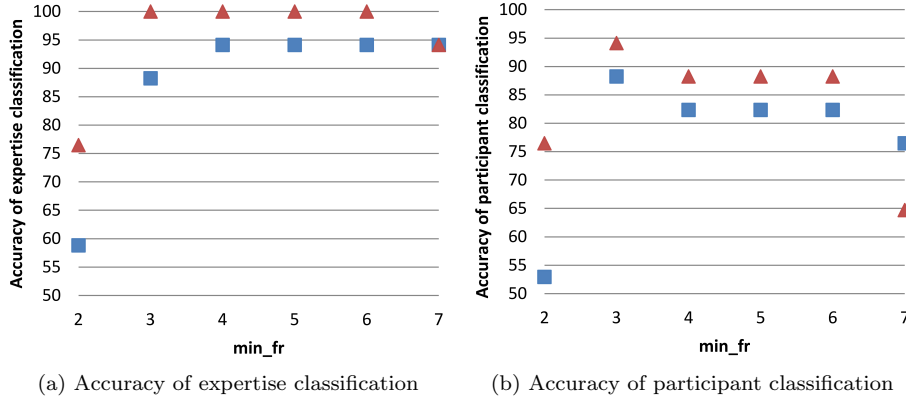
(b) Accuracy of participant classification

Fig. 4: Accuracy of expertise (a) and participant (b) classification for original method (blue square) and extended one with a $min\_lenght = 3$ (red triangle) according to different value of frequency threshold.

summarizes the results of the original method and the best of the extended method ($min\_length = 3$).

For both, the parameters which give the optimal results for classifications are for a frequency threshold of 3. The original method gives 88% of accuracy for expertise and participant classification, whereas the extended one gives 100% of accuracy for expertise classification and 94.12% for participant classification. The results are better for the extended method, even if we have less information than the original one. Indeed, with the optimal parameters, we found 97 patterns for the original method, and only 76 of them for the extended method, i.e. a 22% decrease in information.

The classification results are shown in figure 5 for the original method, and in figure 6 for the extended one. In these figures, the ordinate corresponds to the distance between sequences, and each leaf corresponds to the sequence ID. This ID is composed of the participant ID for the hundreds and the trial number. Thus, leaf 402 corresponds to the second attempt of the participant 4.

When we cut dendrogram of figure 5 at a distance of 0.6, we can define 4 different clusters:

- C^1: a cluster which gathers all trial of participant 1 together;
- C^2: a cluster which gathers all trial of participant 2 together except the first trial (201);
- C^3: a cluster which gathers all trial of participant 3 together;
- C^4: a cluster which gathers all trial of participant 4 together except the fourth trial (404).

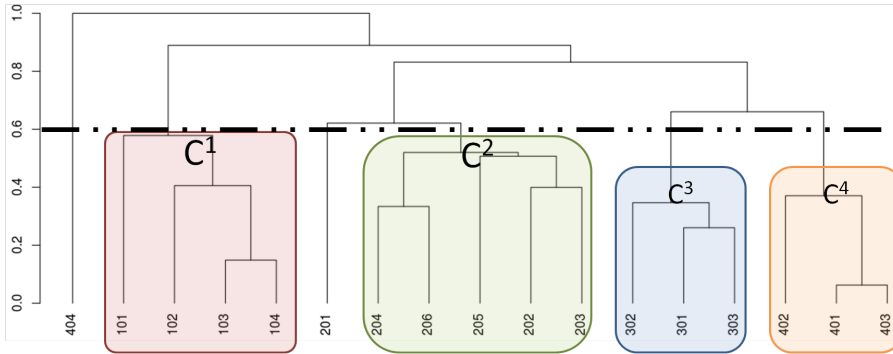Fig. 5: Dendrogram of the classification using the original method and the SLFSP metric.

With the same distance to define cluster (0.6), in dendrogram of figure 6 we can define 3 different clusters:

- C^1: a cluster which gathers all trial of participant 1 together;
- C^2: a cluster which gathers all trial of participant 2 together;
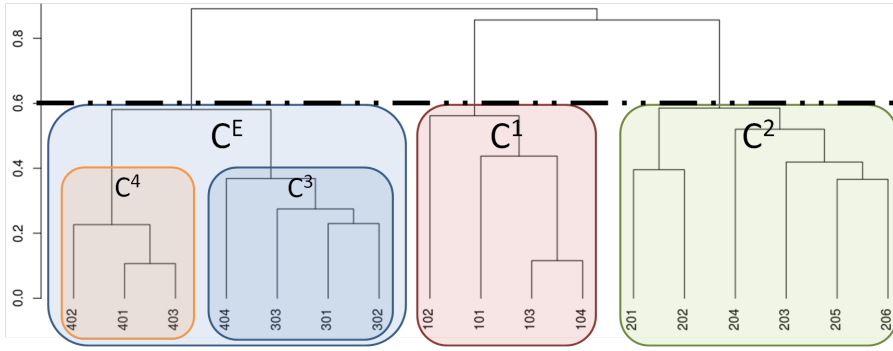- C^E: a cluster which gathers all trial of expert participant together.

Fig. 6: Dendrogram of the classification using the extended method ($min\_lenght = 3$) and the SLFSP metric.

This last cluster could be divided into two sub-clusters C^3 and C^4. C^3 bringing together the majority of the trials of participant 3 and participant4 respectively for C^4. Only participant 4's trial 404 is grouped with participant 3's trials.

4.2 Analysis of sequential signatures

We looked more closely at the longest frequent patterns. With our parameters ($min\_fr = 3$ and $min\_length = 3$)), 76 longest patterns composed of 3 or more activities were found. On these 76 patterns, 56 are specific to one of the following metadata: participant 1, participant 2, participant 3, participant 4, student or expert. Table 4 summarizes the number of patterns specific to each type of metadata, the number of patterns that are more frequent than the threshold (Present 4+) or whose length is greater than or equal to 5 activities (length $\geq$5). A pattern specific to a participant is only found in the sequences executed by this participant. Whereas a pattern specific to a level of expertise is found in the sequences performed by the two participants with this level of expertise. In parentheses, we have the result in proportion to the number of trials for the column "Nb Patterns" and in proportion to the number of pattern for the other columns. The length of the longest patterns in each category ranges from 10 to 14 activities.

4.3 Prediction of belonging to a population

Results of the leave one out cross validation for predicting population affiliation are summarized in the table 4. Our model is able to perfectly predict the expertise in all cases (accuracy of prediction and correct prediction of 100%). But it is more difficult to predict the participant, the model gives the par-

Table 4: The number of patterns specific to each metadata and the number of those patterns that are present at least 4 times or composed of 5 or more activities. For each metadata we also provided the length of the longest patterns

. In brackets, it is the number of patterns by the number of trials for column "Nb Patterns" and the proportion by numbers of patterns for other columns.

	Nb patterns	Present 4+	length ≥ 5	Longest patterns
Participant 1	12 (3)	1 (0.09)	5 (0.42)	10
Participant 2	22 (3.66)	5 (0.23)	8 (0.36)	14
Participant 3	0 (0)	0 (0)	0 (0)	0
Participant 4	3 (0.75)	0 (0)	1 (0.25)	12
Student	13 (1.3)	1 (0.08)	8 (0.61)	14
Expert	6 (0.86)	4 (0.66)	5 (0.83)	11
Total	56 (3.3)	11 (0.19)	27 (0.48)	14

ticipant's information for only 83% of the sequences and makes many errors (correct predictions in 64% of the cases).

Table 5: Percentage of metadata predicted thanks to sequential signatures and the accuracy of this prediction.

	Expertise	Participant
Prediction	100 %	82.35 %
Accuracy of prediction	100 %	64.29 %

5 Discussion

5.1 Method

In this article, we introduced for the first time the concept of sequential surgical signatures. And we have demonstrated this concept using a pattern exploration method. We have decided to ignore patterns that are shorter than a predetermined threshold (min_length) by deleting them at the end of the method. Another approach would have been to directly find the most frequent patterns with the length of min_length, but in this case, we would have had a large number of results after step 2, which would have caused many unnecessary tests in step 3. As a reminder, to find the longest frequent patterns of length k, it is necessary to find frequent patterns of $k + 1$. In this way, with the example shown in figure 3, if we try to find the longest frequent patterns with the length of 3, we have to find frequent patterns of size 4. Thus, with 3 frequent activities, the second step would give 84 candidate models of size 4 (3^4). Whereas with our method, we do 3 times step 2 but for a total of 12 candidate patterns (9 for k=2, 3 for k=2 and 0 for k=4).

5.2 Classification according to sequential signatures

In this first study, we validated the utility of the extended method for different parameters values. In most cases, not taking the shortest signatures into account increases the classification rate. Although, for the optimal parameters, the accuracy of classification improved (94 versus 88% for participants and 100 versus 88% for expertise), this improvement is not significant. However, this classification was carried out with only 78% of the data available using the other method. Thus, it has been shown that our method gives similar results with fewer data. Thus, the hypothesis that the shorter patterns do not have enough discriminating power to be interesting is verified.

5.3 Analysis of sequential signatures

Our method is able to distinguish sequences according to the level of expertise and the participant. This differentiation cannot be made by the length of the longest patterns since they all have lengths between 10 and 14 activities regardless the category. We can also find many signatures specific to the level of expertise and each participant, except for participant 3 (Table 4). Even if we did not find any signature that only appears in this participant does not mean that there is no signature in trials of this participant. The 6 signatures noted as being specific to the experts are present in the trials of the both expert participants. In order to detect a signature specific to participant 3, it should be present in each of the participant's trials because only three have been performed, which corresponds exactly to our min_fr threshold. It is highly improbable that all trials have been proceed in the same way, especially since we did not take into account the signatures composed of 2 activities. To identify the sequential surgical signatures of participant 3, we need to collect more data.

The number of sequential surgical signatures found by each category depends on the number of trials, for example, even though we found fewer sequential surgical signatures for participant 1 than participant 2, 12 versus 22, when we count the average number of signatures per trial, the difference is less significant: 3 versus 3.66. As shown in Table 4, participants 1 and 2 have more sequential surgical signatures specific to participants 1 and 2 than sequential surgical signatures specific to their level of expertise (3 and 3.66 compared to 1.3). On the other hand, for participants 3 and 4, it is the opposite, there are more sequential surgical signatures specific to their level of expertise than for themselves (0.86 versus 0 and 0.75). This could be interpreted by the fact that the experts' participants are more consistent when they perform a task and their signatures are composed of more activities than the students' participants. This hypothesis is confirmed by the proportion of signatures present more often than the threshold for expert participants (66%) than for student participants (8%), but also by the proportion of signatures composed of many activities (83% for experts versus 61% for students).

In the 6 expert sequential surgical signatures (table 4), one of them attracted our attention because of its number of activities (11) and the number of sequences where this signature is present (5 out of 7 expert sequences). This signature, notated $signature_1$, is presented at table 6.

Table 6: Signature present in 5 expert sequences.

Activity id	Left hand	Right hand
1	Hold, Long wire strand	Idle
2	Hold, Long wire strand	Catch, Long wire strand
3	Hold, Long wire strand	Hold, Long wire strand
4	Idle	Hold, Long wire strand
5	Catch, Long wire strand	Hold, Long wire strand
6	Hold, Long wire strand	Hold, Long wire strand
7	Hold, Long wire strand	Idle
8	Make a loop, Long wire strand	Make a loop, Long wire strand
9	Hold, Long wire strand	Idle
10	Hold, Long wire strand	Catch, Short wire strand
11	Pull, Long wire strand	Pull, Short wire strand

However, a signature may be the marker for something other than a population of individuals. This is the case, for example, of the following signature, noted $signature_2$ (Table 7), which are composed of 3 activities and shared between 5 sequences, 2 performed by students and 3 by experts:

Table 7: Signature present in 3 expert sequences and 2 student sequences.

Activity id	Left hand	Right hand
a	Pull, Long wire strand	Pull, Short wire strand
b	Hold, Long wire strand	Idle
c	Hold, Long wire strand	Catch, Short wire strand

These two examples are interesting for multiple reasons:

- $Signature_1$ illustrates the full knot tying process without unnecessary activities;
- $Signature_2$ reflects a mistake independent of the level of expertise. In each case, the participant tried to tie the knot by pulling the two strands of the wire (activity a), but dropped the short wire strand (activity b) so had to catch again the short wire strand (activity c);
- Both signatures provide hints to improve or facilitate the execution of the task by informing us how an expert performs the task and which mistakes should not be made;
- These both types of signature, coupled with real time activity detection methods, can be used for automatically analysis of surgical workflow and thus providing relevant information for situation aware systems.

A video representation of each of these two signatures is available as supplementary material. In these videos, the animated process was realized thanks to Disco software [21], and each video was synchronized to start each activity at the same time.

5.4 Prediction of belonging to a population

Our method also showed that sequential surgical signatures could be used to determine which population a new sequence belongs to. These initial results need to be complemented by more data that would not only increase the number of participants in each population, but also the number of different populations.

6 Conclusion

In this article, we introduced the concept of sequential surgical signatures and demonstrated their usefulness in classifying surgical sequences and their ability to determine by which individual a sequence was performed. This could be interesting in order to provide an automatic and objective skill assessment system.

The identification of sequential surgical signature could provide leads for understanding surgical skills and consequently useful pedagogical guidance for trainees.

Conflict of Interest

The authors declare that they have no conflict of interest.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

This articles does not contain patient data.

References

1. P. Jannin, M. Raimbault, X. Morandi, and B. Gibaud. Modeling Surgical Procedures for Multimodal Image-Guided Neurosurgery. In Wiro J. Niessen and Max A. Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2001*, number 2208 in Lecture Notes in Computer Science, pages 565–572. Springer Berlin Heidelberg, October 2001.
2. F. Lalys and P. Jannin. Surgical process modelling: a review. *International Journal of Computer Assisted Radiology and Surgery*, 9(3):495–511, September 2013.
3. C. E. Reiley and G. D. Hager. Task versus Subtask Surgical Skill Evaluation of Robotic Minimally Invasive Surgery. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2009*, Lecture Notes in Computer Science, pages 435–442. Springer, Berlin, Heidelberg, September 2009.
4. F. Despinoy, D. Bouget, G. Forestier, C. Penet, N. Zemiti, P. Poignet, and P. Jannin. Unsupervised trajectory segmentation for surgical gesture recognition in robotic training. *IEEE Transactions on Biomedical Engineering*, 63(6):1280–1291, 2016.
5. T. Neumuth, G. Strau, J. Meixensberger, H. U. Lemke, and O. Burgert. Acquisition of Process Descriptions from Surgical Interventions. In S. Bressan, J. Kng, and R. Wagner, editors, *Database and Expert Systems Applications*, number 4080 in Lecture Notes in Computer Science, pages 602–611. Springer Berlin Heidelberg, September 2006.
6. N. Padoy, B. Tobias, H. Feussner, M.-O. Berger, and N. Navab. On-line Recognition of Surgical Activity for Monitoring in the Operating Room. pages 1718–1724, 2008.
7. N. Padoy, T. Blum, S.-A. Ahmadi, H. Feussner, M.-O. Berger, and N. Navab. Statistical modeling and recognition of surgical workflow. *Medical Image Analysis*, 16(3):632–641, 2010.
8. L. Bouarfa, P. P. Jonker, and J. Dankelman. Discovery of high-level tasks in the operating room. *Journal of Biomedical Informatics*, 44(3):455–462, June 2011.
9. A. James, D. Vieira, B. Lo, A. Darzi, and G.-Z. Yang. Eye-Gaze Driven Surgical Workflow Segmentation. *Medical Image Computing and Computer-Assisted Intervention MICCAI 2007*, pages 110–117, October 2007.
10. S.-Y. Ko, J. Kim, W.-J. Lee, and D.-S. Kwon. Surgery task model for intelligent interaction between surgeon and laparoscopic assistant robot. *International Journal of Assitive Robotics and Mechatronics*, 8(1):38–46, 2007.
11. F. Lalys, D. Bouget, L. Riffaud, and P. Jannin. Automatic knowledge-based recognition of low-level tasks in ophthalmological procedures. *International Journal of Computer Assisted Radiology and Surgery*, 8(1): 39–49, April 2012.

12. G. Forestier, F. Lalys, L. Riffaud, D. Louis Collins, J. Meixensberger, S. N. Wassef, T. Neumuth, B. Goulet, and P. Jannin. Multi-site study of surgical practice in neurosurgery based on surgical process models. *Journal of Biomedical Informatics*, 46(5):822–829, October 2013.

13. A. Huaulmé, S. Voros, L. Riffaud, G. Forestier, A. Moreau-Gaudry, and P. Jannin. Distinguishing surgical behavior by sequential pattern discovery. *Journal of Biomedical Informatics*, 67:34–41, 2017.

14. L. Riffaud, T. Neumuth, X. Morandi, C. Trantakis, J. Meixensberger, O. Burgert, B. Trelhu, and P. Jannin. Recording of Surgical Processes: A Study Comparing Senior and Junior Neurosurgeons During Lumbar Disc Herniation Surgery:. *Operative Neurosurgery*, 67:ons325–ons332, December 2010.

15. T. Neumuth, R. Wiedemann, C. Foja, P. Meier, J. Schlomberg, D. Neumuth, and P. Wiedemann. Identification of surgeon\individual treatment profiles to support the provision of an optimum treatment service for cataract patients. *Journal of Ocular Biology, Diseases, and Informatics*, 3(2):73–83, June 2010.

16. C. Cao, C. L. MacKenzie, and S. Payandeh. Task and motion analyses in endoscopic surgery. pages 583–590, 1996.

17. G. Forestier, F. Petitjean, P. Senin, F. Despinoy, and P. Jannin. Discovering Discriminative and Interpretable Patterns for Surgical Motion Analysis. In *Conference on Artificial Intelligence in Medicine in Europe*, pages 136–145. Springer, 2017.

18. M. Mitsuishi, A. Morita, N. Sugita, S. Sora, R. Mochizuki, K. Tanimoto, Y.M. Baek, H. Takahashi, and K. Harada. Master-slave robotic platform and its feasibility study for micro-neurosurgery: Master-slave robotic platform for microneurosurgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 9(2):180–189, June 2013.

19. C. Garraud, B. Gibaud, C. Penet, G. Gazuguel, G. Dardenne, and P. Jannin. An Ontology-based Software Suite for the Analysis of Surgical Process Model. In *Proceedings of Surgetica'2014*, pages 243–245. Chambery, France, December 2014.

20. R. R. Sokal and C. D. Michener. A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin*, 28:1409–1438, 1958.

21. Process Mining and Automated Process Discovery Software for Professionals - Fluxicon Disco. URL `https://fluxicon.com/disco/`.