



HAL
open science

Non-Linear Temporal Scaling of Surgical Processes

Germain Forestier, Francois Petitjean, Laurent Riffaud, Pierre Jannin

► **To cite this version:**

Germain Forestier, Francois Petitjean, Laurent Riffaud, Pierre Jannin. Non-Linear Temporal Scaling of Surgical Processes. *Artificial Intelligence in Medicine*, 2014, 62 (3), pp.143-152. 10.1016/j.artmed.2014.10.007 . hal-00862978v3

HAL Id: hal-00862978

<https://hal.science/hal-00862978v3>

Submitted on 20 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Non-linear temporal scaling of surgical processes

Germain Forestier^{1,*}, François Petitjean², Laurent Riffaud^{3,4}, Pierre Jannin³

¹ MIPS EA 2332, University of Haute-Alsace, Mulhouse, France

² Faculty of Information Technology, Monash University, Melbourne, Australia

³ INSERM MediCIS, Unit U1099 LTSI, University of Rennes 1, Rennes, France

⁴ Department of Neurosurgery, Pontchaillou University Hospital, Rennes, France

Abstract

Objective. Surgery is one of the riskiest and most important medical acts that is performed today. Understanding the ways in which surgeries are similar or different from each other is of major interest. Desires to improve patient outcomes and surgeon training, and to reduce the costs of surgery, all motivate a better understanding of surgical practices. To facilitate this, surgeons have started recording the activities that are performed during surgery. New methods have to be developed to be able to make the most of this extremely rich and complex data. The objective of this work is to enable the simultaneous comparison of a set of surgeries, in order to be able to extract high-level information about surgical practices.

Material and Method. We introduce non-linear temporal scaling (NLTS): a method that finds a multiple alignment of a set of surgeries. Experiments are carried out on a set of lumbar disc neurosurgeries. We assess our method both on a highly standardised phase of the surgery (closure) and on the whole surgery.

Results. Experiments show that NLTS makes it possible to consistently derive standards of surgical practice and to understand differences between groups of surgeries. We take the training of surgeons as the common theme for the evaluation of the results and highlight, for example, the main differences between the practices of junior and senior surgeons in the removal of a lumbar disc herniation.

Conclusions. NLTS is an effective and efficient method to find a multiple alignment of a set of surgeries. NLTS realigns a set of sequences along their intrinsic timeline, which makes it possible to extract standards of surgical practices.

Supplementary material. The computer code implementing the proposed methods.

Keywords: Temporal Analysis, Dynamic Time Warping, Surgical Process Modelling, Surgical Technical Skills, Surgery

1. Introduction

More than half a million surgeries are performed every day worldwide [1], which makes surgery one of the most important component of global health care.

Competing demands are motivating a better understanding of surgical processes: surgical procedures are getting more complex [2], residents now have to be trained while performing less procedures [3], the surgical interventions have to be more and more justified [4] and the procedures have to cost less money [5], *et cætera*. A better understanding of surgical practices is the key component to addressing these issues. surgical process modelling

(SPM) is the general process that aims at understanding surgeries, in order to improve the quality of care and the training/assessment of surgeons.

This article addresses the issue of analysing sets of surgeries. Let us consider an example related to the training of neuro-surgeons. Surgical training is critical to ensure a smooth expertise transition between senior and junior surgeons. This training is generally provided in a one-on-one scheme between a junior surgeon and his or her senior, which makes this process extremely expensive and time-consuming. Let us give a brief example showing how a better understanding of senior surgical practice (*i.e.*, understanding sets of senior surgeries) can make it possible to improve and optimise the training of surgeons. Let us assume that the analysis of a set of senior surgeries tells us that all senior surgeons perform one step of the surgery in the exact same way. This information can be exploited by directing the training of junior surgeons towards a standard and stereotyped practice for this step. By contrast, if one part is very patient-specific, the analysis of a the same set will show that the behaviour of senior surgeons is less standard. The training can then be directed towards

*Corresponding author – ENSISA - Université de Haute Alsace, 12 rue des freres Lumiere, 68093 Mulhouse, France, Tel.:+33 3 8933 6963, Fax: +33 3 8942 3282

Email addresses: germain.forestier@uha.fr (Germain Forestier), francois.petitjean@monash.edu (François Petitjean), laurent.riffaud@chu-rennes.fr (Laurent Riffaud), pierre.jannin@univ-rennes1.fr (Pierre Jannin)

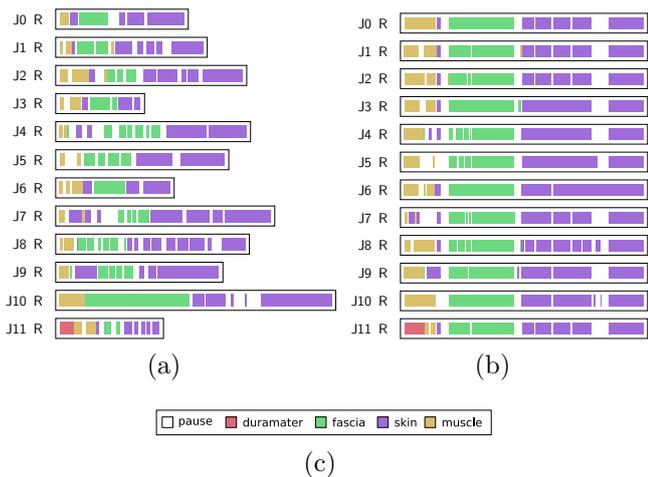


Figure 1: Example of the use of NLTS on a set of 12 sequences representing the successive anatomical structures targeted by the surgeon during the closure phase. (a) Sequences scaled using our NLTS approach. (c) Legend of the different anatomical structures.

a patient-specific care.

The idea that a set of observations may leverage more information than any individual observation is not new. The most famous example set dates back at least a century ago, when Francis Galton noted that the crowd at a county fair accurately guessed the weight of an ox when their individual guesses were averaged [6]. Galton realized that the average was closer to the ox’s true weight than the estimates of most crowd members, and also much closer than any of the separate estimates made by cattle experts.

Analysing a set of observations in \mathbb{R} has long been studied and is now very well understood. However, analysing and simultaneously comparing a *set of sequences* is much more challenging, because of the particular properties that are induced by time (*e.g.*, periodicity, symmetry, autocorrelation). Computational biologists have long known that the key to understanding a set of sequences is to find their *multiple alignment*. To illustrate this, Figure 1 presents the closure phase of 12 recordings of surgeries, as well as their multiple alignment (computed using the method proposed in this article). This example demonstrates how simultaneously aligning a set of surgeries (*i.e.*, computing their multiple alignment) directly highlights critical elements of the closure phase. We can for example observe the closure of the surgical route layer by layer (*muscle* \rightarrow *fascia* \rightarrow *skin*), and also that almost all surgeons pause before the final stitches.

Two important issues however prevent from using the methods developed in computational biology for the multiple alignment of surgical sequences:

1. Surgeries are very particular sequences: surgeries are sequences of surgical activities over time, whereas DNA/RNA sequences encode biological structures (with no reference to time). This makes surgeries

exhibit very different properties, such as autocorrelation: it is likely that a surgeon using a scalpel at time t will still be using it at time $t+1$, while observing the letter A in a DNA sequence does not increase the likelihood of observing it at the next position in the sequence. This is why time series dedicated measure, such as dynamic time warping (DTW)[7], have proved to be extremely relevant for the analysis of surgical processes [8, 9]. Multiple alignment methods that have been developed in computational biology, are however dedicated to the Levenshtein distance [10]. New methods have thus to be developed to enable multiple alignment dedicated to DTW, in order to unlock the value of surgical datasets.

2. Multiple sequence alignment is NP-complete [11], which prevents its computation for more than a few short sequences. As surgical datasets are rapidly growing, efficient and effective heuristics have to be developed to simultaneously analyse sets of recordings of surgeries.

This work, capitalising on our recent discoveries for time series averaging [12, 13], introduces *non-linear temporal scaling* (NLTS): a multiple alignment method for surgical processes. We show that NLTS effectively supports the extraction of high-level knowledge, by realigning the surgeries along their *intrinsic timeline*, *i.e.*, along the standard sequencing of actions that occur over the course of the different surgeries.

We consider surgeries as sequences of activities that are performed by the surgeon during the surgery. Mehta et al. [14] proposed to represent surgical activities as a triplet composed of an *action*, an *anatomical structure* and an *instrument*. For example, the surgeon can *cut* the *skin* using a *scalpel*. In this paper, we embrace this representation and use its formalisation introduced in [8]. In general, activities that are performed by both hands are recorded, as well as the use of the microscope. Figure 2 illustrates a set of 6 surgeries.

Without loss of generality with regard to the issues that are addressed by surgical processes modelling, we use the training of surgeons as the main theme to support our explanations for the development of the article.

This article addresses the issue of being able to compare a set of surgical processes; Section 2 describes this issue. Section 3 introduces our method NLTS. Section 4 details experiments carried out on neurosurgeries. Section 5 discusses these results and shows that our method makes it possible to support the extraction of high-level knowledge about the standards of neurosurgeries. We use the example of the training of neurosurgeons to illustrate that our approach captures interesting knowledge about the surgeries. Section 6 concludes this paper and presents some future work.

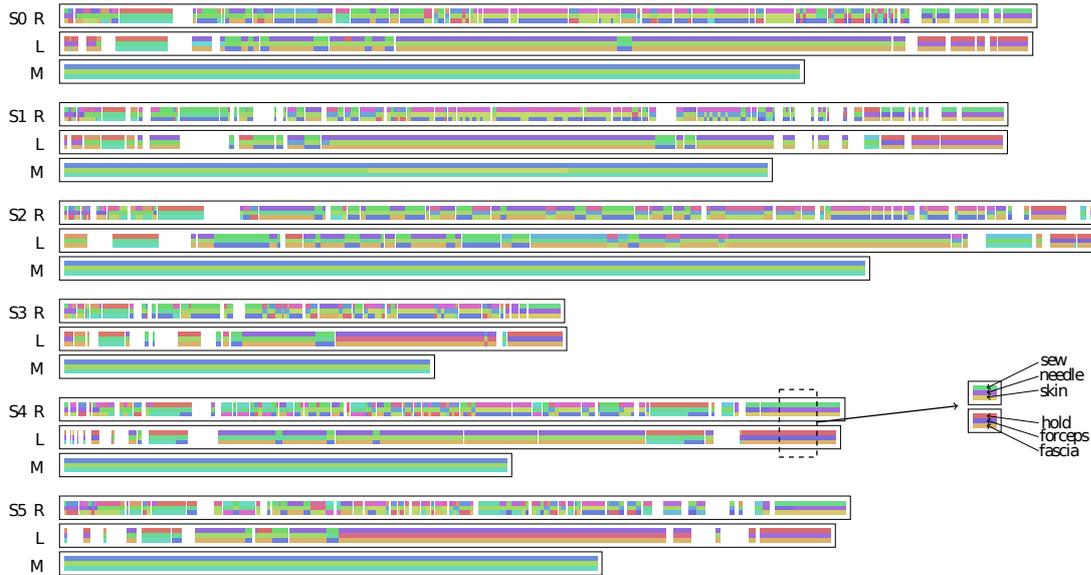


Figure 2: A set of surgeries recorded from the OR by a neurosurgeon. For every surgery, three sequences of activities are displayed. They represent activities related to the right hand (R), left hand (L), and microscope (M).

2. Problem statement

History of surgical process modelling. Several initiatives have already been proposed to better understand surgical practice, mainly focusing on the assessment of surgeries. The first approach was to assess surgeries with regard to patient outcomes [15]. In addition to requiring a long-term follow-up with the patient, this method is very dependent upon the patient and the conditions of the surgery. It consequently cannot be used for an objective assessment of surgeries.

Human grading techniques have been proposed to improve the objectivity of the assessment [16]. Junior surgeons are evaluated by their seniors with regard to a list of surgical skills. These tests are however subjective depending on the evaluator [17].

Time-motion approaches have been introduced to improve objectivity and to automate the information acquisition on a surgery [14, 18]. The idea was to use statistical information like the average duration of the surgery or the number of actions performed by the surgeon. These methods are very objective and easy to record. However, they do not provide enough information about the standard practices during the surgery. Assessing junior surgeons on such criteria can also be misleading. For example, senior surgeons are on average faster than junior surgeons [19]. It is obviously very undesirable for the junior to try to speed up their surgery without having reached the dexterity and experience of senior surgeons.

Recording surgeries has recently gained interest, either using sensor devices, or directly by an observer in the operating room (OR). An universal and adaptable recording scheme has been introduced [20]. It shows how to decompose a surgical intervention into manual work steps. This

data contains a lot of information, since it is much closer to the reality in the OR than, for instance, a record of only the number of actions performed during the surgery. An example of recorded surgeries is given in Figure 2. It is usually assumed that this data is rich enough for the knowledge that scientists want to extract. Automating the analysis of such a dataset is however very challenging. Surgical process modelling (SPM) is the field that aims at unlocking this issue [20–26].

Analysing a set of surgeries. In this article, we focus on the understanding of the similarities and differences that take place in a set of surgeries. We highlighted above that this is challenging because every surgery is different from another one. Yet senior surgeons successfully train junior surgeons every year. Surgery is indeed a very standardised practice that can be taught. The problem is that surgery is standardised at a high-level (phases, steps of each phase, ways of performing each step, *et cetera*). The challenge is to be able to support the recognition of these standard practices from the low-level description of the data, *i.e.*, from the actions that are performed by the surgeon.

From the data analysis perspective, the first step to unlock this issue is to be able to compare two surgeries in a consistent way. For a measure to be *consistent*, it has to provide a graduated evaluation of how similar two surgeries are. Similarity measures for surgeries were first studied in [27]. Dynamic time warping (DTW) is based on the Levenshtein distance and was introduced for speech recognition in the 1970s [7]. We recently demonstrated that DTW is sound for surgical processes comparison [8, 9]. DTW operates non-linear distortions on the time-axis, in order to find the best alignment of the two sequences. DTW optimally realigns (or stretches) sequences with each

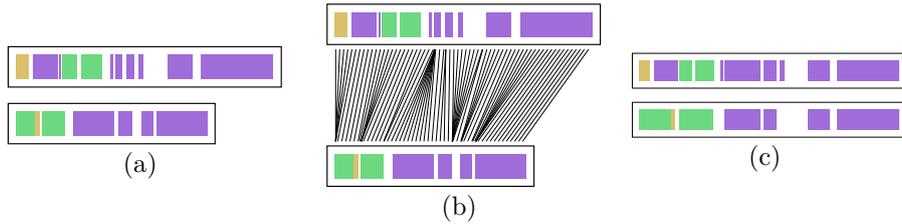


Figure 3: Stretching two sequences with DTW. Each color corresponds to one anatomical structure. (a) Original sequences. (b) DTW-alignment. (c) DTW-scaled sequences.

other, which makes it possible to compare them along their intrinsic timeline. We use the term intrinsic timeline to describe the fact that each surgery has its own timeline that is independent of the ones of other surgeries. Figure 3 illustrates this process on the anatomical structures that are targeted during two surgeries.

A lot of information can be retrieved very easily from the DTW-realigned sequences (variability, transitions, phases, *et cætera*). Improving the understanding of surgical processes would require a set of surgeries to be simultaneously realigned. DTW is unfortunately able to align a pair of sequences only. This is exactly the scientific issue that this article unlocks. Next section will introduce a method that makes it possible to re-align a set of sequences, in order to be able to provide high-level information about surgical processes.

3. Method: non-linear temporal scaling

In this section, we present our method for the non-linear temporal scaling (NLTS) of surgeries. We start by giving some notations and presenting our method NLTS. Then, we detail how this set can be analysed to extract high-level information about the surgeries.

Notations

Let $\mathbb{S} = \{S_1, \dots, S_N\}$ be the original set of N sequences (surgeries). Let E be the space of states of sequences in \mathbb{S} as:

$$E = \bigcup_{n=1}^N \bigcup_{\ell=1}^{\text{length}(S_n)} S_n(\ell) \quad (1)$$

3.1. NLTS

Optimally aligning a set of sequences under time warping has long been studied in computational biology. It is known as the multiple sequence alignment of the set, and is often considered the “Holy Grail” in computational biology [28]. Multiple sequence alignment is NP-complete [11], which prevents its computation for more than a few short sequences.

In this section, we propose *Non-linear temporal scaling* (NLTS): a method to align (or scale) a set of sequences under time warping. Our method builds on *compact multiple*

alignment [13], which was recently introduced to enable the averaging of large set of homogeneous time series. In addition, scalable methods exist for the definition of an average sequence under time warping [12].

Non-linear temporal scaling (NLTS) starts with a set of sequences and performs as follows:

1. Compute the average sequence \bar{S} of the set of sequences \mathbb{S} .
2. Compute the compact multiple alignment of \mathbb{S} from \bar{S} .
3. Unpack every column of the compact multiple alignment to its maximum width.

Algorithm 1 details the computation of NLTS; Figure 4 illustrates the approach; we describe its steps in the following paragraph.

Computation of the average sequence. The first step consists of computing the average sequence \bar{S} of \mathbb{S} (line 1). To this end, we use DTW barycenter averaging (DBA) [12, 29], but other methods like COMASA [13] could be used depending on time requirements. The computation of the average sequence is detailed in Algorithm 2. Note that DBA is initialised with the medoid sequence (Algorithm 3).

Computation of the compact multiple alignment. The compact multiple alignment is computed by aligning the average sequence \bar{S} to every one of the sequences of \mathbb{S} independently. This alignment is performed by the function `assocDTW`. This function simply returns, for every element of the first sequence, the elements of the second sequence that have been linked to it by DTW. In our case, it consists of finding which elements of the s^{th} sequence have been associated with the ℓ^{th} element of \bar{S} (line 5). These elements are then stored in `elements[s][ℓ]`. Moreover, we store in `widths[ℓ]` the maximum number of elements that have been associated with every ℓ^{th} element of \bar{S} , *i.e.*, in every column of the corresponding compact multiple alignment (lines 6–8).

Unpacking the compact multiple alignment. The last part of the algorithm unpacks the compact multiple alignment.

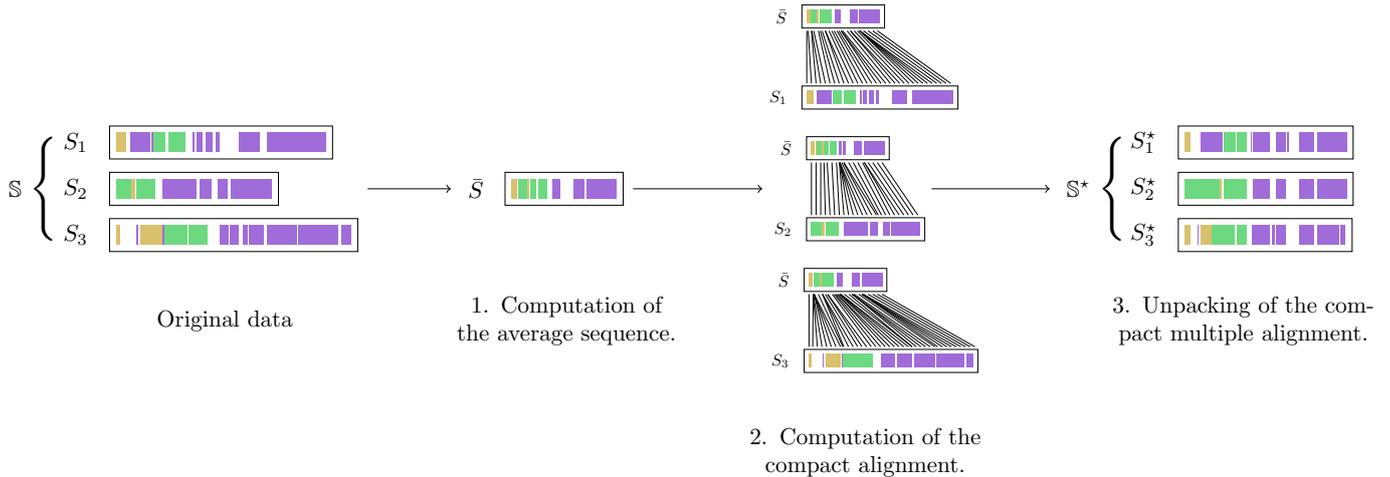


Figure 4: Illustration of the different steps of the NLTS method.

The compact multiple alignment provides a set of sequences that is consistently aligned. However, the first column can, for example, hold four elements of the first sequence (e.g. $\langle \mathbf{aaab} \rangle$), one element of the second sequence (e.g. $\langle \mathbf{a} \rangle$), and two elements of the third sequence (e.g. $\langle \mathbf{ab} \rangle$). As a result, for every column ℓ of the alignment, we scale every subsequence $elements[s][\ell]$ to the maximum number of elements contained in this column, *i.e.*, $widths[\ell]$ (line 15). In the latter example, it would correspond to stretch the single element of the second sequence to four elements ($\langle \mathbf{a} \rangle \rightarrow \langle \mathbf{aaaa} \rangle$) and the two elements of the third sequence to four elements as well ($\langle \mathbf{ab} \rangle \rightarrow \langle \mathbf{aabb} \rangle$). A web application is available here¹ which allows the reader to try NTLTS easily. Let us note that the unpacking step uses the maximum width as reference as it guarantees that the scaling does not lead to any loss of information, because other functions would lead to a “contraction” of one of the aligned sub-sequences. Other functions such as the mean or the median would indeed lead to two different elements to have to be merged, which would lose information about the original sequences.

Result. The algorithm outputs a set of sequences \mathbb{S}^* , which correspond to a multiple sequence alignment of \mathbb{S} . \mathbb{S}^* aims at supporting a detailed temporal analysis of the set of sequences with regard to their sequencing. Note that the information about the specific time at which the elements of the sequences is not available from the visual representation any more. All sequences of \mathbb{S}^* have the same length.

3.2. Analysing a set of non-linearly scaled sequences

NLTS makes it possible to realign a set of sequences, by finding a multiple sequence alignment of the set. In this section, we propose different methods that can be used to analyse this aligned set of sequences.

Probability distribution of the states over time. Let \mathbb{S}^* be the set of sequences scaled with NLTS. We want to study the distribution of the different states in every position (or column) of the multiple alignment. To this end, we define the (discrete) probability distribution of the states over the sequencing induced by the multiple alignment.

Definition 1. $S_{\hat{p}}^* = \langle \hat{p}_1, \dots, \hat{p}_L \rangle$ is the sequence of probability with maximum likelihood estimates \hat{p}_ℓ over E defined by:

$$\hat{p}_\ell : E \rightarrow [0, 1] \subset \mathbb{R}$$

$$e \mapsto \frac{|\{\mathbb{S}^*(\ell)=e \mid \forall S^* \in \mathbb{S}^*\}|}{N}$$

For example, to create the set of states for the position $\ell = 1$, we create a set the first states of all sequences in \mathbb{S}^* . If we are considering the targeted anatomical structures of three surgeries, the first states distribution could be $\{skin, skin, skin\}$ meaning that the first state “skin” is identical for all three surgeries. This sequence of probability $S_{\hat{p}}^*$ thus informs about the distribution of the states over time. The more similar these states are at position ℓ , the more standardised the action ℓ is. For example, the set of states $\{skin, skin, skin\}$ is very standardised (all the surgeon are targeting the skin at ℓ) but the set of states $\{skin, fascia, muscle\}$ is heterogeneous (all the surgeons are targeting a different structure at ℓ).

Entropy of the states. Entropy is a measure of uncertainty of a random variable. In our case, the entropy of every state of $S_{\hat{p}}^*$ gives information about how diverse the behaviour of surgeons is at state ℓ of the surgery. The entropy is null when all the surgeons perform the same action in the set. The entropy is maximal when every surgeon performs a different action. This can be used as a measure of the predictability of the action for models of surgeries. This entropy – $H(\hat{p}_\ell)$ – is defined as:

$$H(\hat{p}_\ell) = - \sum_e^E \hat{p}_\ell(e) \cdot \log_b(\hat{p}_\ell(e)) \quad (2)$$

¹<http://germain-forestier.info/src/aim2014/> (Accessed: 10 October 2014)

Algorithm 1 Non-linear temporal scaling

Require: $\mathbb{S} = \{S_1, \dots, S_N\}$ **Let** $\mathbb{S}^* = \{S_1^*, \dots, S_N^*\}$ be the resulting set of scaled sequences**Let** $\text{mean}(\cdot)$ return the average sequence of a set for DTW**Let** $\text{assocDTW}(S, T)$ return the elements associated mapping built by DTW from S to T **Let** $\text{scale}(S, n)$ return the uniform scaling [30] of the sub-sequence S to n elements

```
1:  $\bar{S} \leftarrow \text{mean}(\mathbb{S})$ 
2:  $L \leftarrow \text{length}(\bar{S})$ 
3:  $\text{widths}[L] \leftarrow [0, \dots, 0]$ 
4: for  $s \leftarrow 1$  to  $N$  do
5:    $\text{elements}[s] = \text{assocDTW}(\bar{S}, S_s)$ 
6:   for  $\ell \leftarrow 1$  to  $L$  do
7:      $\text{widths}[\ell] \leftarrow \max(\text{widths}[\ell], \text{size}(\text{elements}[s][\ell]))$ 
8:   end for
9: end for
10: for  $s \leftarrow 1$  to  $N$  do
11:    $S_s^* \leftarrow \langle \rangle$ 
12:   for  $\ell \leftarrow 1$  to  $L$  do
13:      $\text{targetLength} \leftarrow \text{widths}[\ell]$ 
14:     for  $n \leftarrow 1$  to  $\text{targetLength}$  do
15:        $S_s^* \leftarrow S_s^* \cdot \text{scale}(\text{elements}[s][\ell], \text{targetLength})$ 
16:     end for
17:   end for
18: end for
19: return  $\mathbb{S}^*$ 
```

Algorithm 2 mean

Require: $\mathbb{S} = \{S_1, \dots, S_N\}$ **Let** DBA be the averaging sequence method introduced in [12]**Let** I be the number of iterations**Let** mean be the returned average sequence of \mathbb{S} $\text{mean} \leftarrow \text{medoid}(\mathbb{S})$ **for** $i \leftarrow 1$ to I **do** $\text{mean} \leftarrow \text{DBA}(\text{mean}, \mathbb{S})$ **end for****return** mean

$H(\hat{p}_\ell)$ makes it possible to locate the states or phases of the surgeries, for which the behaviour of surgeons is standard or heterogeneous.

Entropy based encoding of \mathbb{S}^ .* Entropy has a direct connection with compression. This is because encoding an element e with probability $p(e)$ requires $-\log_2(p(e))$ bits. The length of the compressed string encoding \mathbb{S}^* on a state-by-state basis is given by:

$$\mathcal{L}_{\mathbb{S}^*} = \sum_{\ell=1}^L \sum_{n=1}^N -\log_2(\hat{p}_\ell(S_n^*(\ell))) \quad (3)$$

$$= \sum_{\ell=1}^L N \cdot \left(-\sum_e \hat{p}_\ell(e) \cdot \log_b(\hat{p}_\ell(e)) \right) \quad (4)$$

$$= N \cdot \sum_{\ell=1}^L H(\hat{p}_\ell) \quad (5)$$

This length $\mathcal{L}_{\mathbb{S}^*}$ gives information about the general uncertainty on \mathbb{S}^* . It evaluates the predictability of the behaviours of surgeons over the whole surgeries.

4. Experiments

We illustrate our approach on a neurosurgical dataset. We focus in particular on the training of neurosurgeons, by comparing the surgical practices of junior and senior surgeons. This theme is actually central in surgical process modelling. The complexities involved in operating on the human body means that the initial training of a surgeon usually takes more than 10 years, and requires extensive one-on-one instruction from a senior surgeon. After that initial training, surgeons still require several further years of experience to themselves reach a senior level. We focus on the comparison of junior *vs* senior surgical practices, in order to improve neurosurgeon training.

Algorithm 3 medoid

Require: $\mathbb{S} = \{S_1, \dots, S_N\}$ **Let** *medoid* be the returned medoid of \mathbb{S} **Let** *inertia* $\leftarrow \infty$ **for** S in \mathbb{S} **do** *sqrDist* $\leftarrow 0$ **for** T in \mathbb{S} **do** *sqrDist* $\leftarrow \text{sqrDist} + \text{DTW}(S, T)^2$ **end for** **if** *sqrDist* $<$ *inertia* **then** *inertia* $\leftarrow \text{sqrDist}$ *medoid* $\leftarrow S$ **end if****end for****return** *medoid*



Figure 5: Recording of the data in the OR.

4.1. Material

The dataset used in this paper is composed of 24 lumbar disc herniation surgeries and was recorded by Laurent Riffaud (MD) and neuro-surgeon, while he was visiting the Neurosurgery Department of the Leipzig University Hospital, Germany. Figure 5 gives an example of how the recording process takes place in the OR. In our case, a second neurosurgeon (foreground) records the activities of the operating one (next to the microscope).

We recorded the activities of both hands of the surgeon, as well as the use of the microscope. We treat the microscope as a particular instrument, as it is often used in combination to other activities. It is a useful information that gives some nuance about the performed surgical task, such as the targeted anatomical structure (*e.g.*, ligament), if the microscope is re-positioned at a particular time of the surgery, or even if the activity is performed without the use of the microscope as it is for example the case at the very end of the surgery.

The surgeries involved 10 male and 14 female patients, with a median age of 52 years. These lumbar disc surgeries are divided into three main steps: (1) approach of the disc, (2) discectomy and (3) closure. The herniated disc is approached *via* a posterior intermyolaminar route. The dis-

cectomy includes the dissection and removal of the disc. A hemostasis step might also be performed before the closure. The patients were operated on by five junior and five senior surgeons. Senior surgeons have performed at least a hundred removals of lumbar disc herniation. All the junior surgeons have passed more than two years of their residency program but have not performed more than a hundred removals of lumbar disc herniation. We recorded the surgeries so that the dataset contains equal representation of junior and senior surgeons. Twelve patients were operated upon by a senior surgeon with the help of his or her junior, and for the other 12 it was the other way around, with the junior surgeons taking the lead. In junior recordings, the assistance of the senior surgeon was mainly provided verbally. When the senior did take the lead, it was always to guide the junior through a simple step, and never for a long series of activities. We observed that these interventions were however very limited and did not strongly influence the recordings. During all junior recordings, the closure step was performed by the junior alone.

4.2. Results

Our results are divided into two main parts. The first aims at understanding our approach on a simple case: the closure phase of the surgery. The second part illustrates the results of NLTS on the whole surgery. These results are discussed in the next section.

4.2.1. Closure phase

Figure 6 presents the experiments that have been carried out on twelve surgeries during the closing phase. These results aim at giving some intuition about how our method helps access high-level information about the surgery. These twelve surgeries correspond to the ones that were performed by junior surgeons (with the aid of their seniors). Figure 6 illustrates the sequence of anatomical structure that were involved in the activities of the surgeon during the surgery. Note that the scaling process took into account the complete information, *i.e.*, action, anatomical structure and instrument for both hands and the microscope. These 9-dimensional sequences are then scaled and Figure 6 presents the sequence of anatomical structure that were involved in the activities performed with the main operating hand (right hand for right-handed surgeons and conversely).

Figure 6(a) presents the original (non-scaled) data. Figure 6(b) presents the set of sequences that has been scaled using our NLTS approach. We also compare the results of NLTS to the one of the Euclidean state-of-the-art method, namely *uniform scaling*. This method stretches the sequences uniformly over time. The work by [30] gives more details about this procedure. The results of this approach are depicted in Figure 6(c).

Figure 7 depicts the comparative study of the evolution of \hat{p} (Definition 1) over time, for the anatomical structures

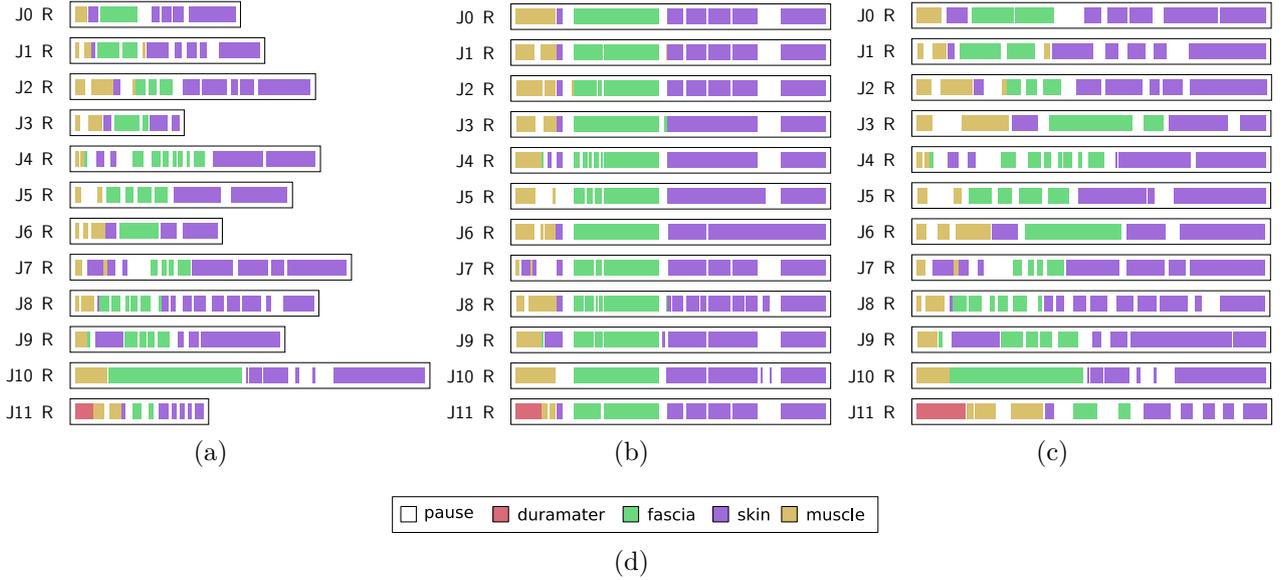


Figure 6: Comparative scaling process of a set of surgeries on the closure phase of the surgery. (a) Original sequences. (b) Sequences scaled using our NLTS approach. (c) Sequences scaled using uniform scaling. (d) Legend.

that are involved in the surgery. Figure 7(a) depicts this evolution for NLTS that can be compared to the results obtained with uniform scaling in Figure 7(b). For both cases, the first chart gives the evolution of the entropy ($H(\hat{p}_\ell)$), while the second chart gives the distribution of the anatomical structures that are involved through the surgery. The aim is to inform about how standard the surgery is at different stages.

4.2.2. Whole surgery

This results extend the previous ones, but at the level of the complete surgery. The methodology remains the same: the 9-dimensional sequences are aligned with NLTS, and we present results for the anatomical structures on which actions of the main hand are performed. We propose to illustrate the interest of our approach on the comparative analysis senior *vs* junior. Even if the information of the different surgical phases is present in the dataset for each surgery, we did not use it for processing the whole surgeries. This information was not used for the analysis, because our aim was to stress the ability of the method to work without supervision. In this experiment, the method is only relying on the sequence of performed activities to find correspondence between the different surgeries. Figure 8 presents the evolution of the targeted anatomical structures. We compare junior and senior behaviours scaled with NLTS, respectively in Figures 8(a) and 8(b). A similar analysis is conducted on the other components of the surgical activity, *i.e.*, surgical instrument (Figure 9) and action performed (Figure 10). Table 1 compares the length $\mathcal{L}_{\mathbb{S}^*}$ required to encode junior and senior sets of surgeries.

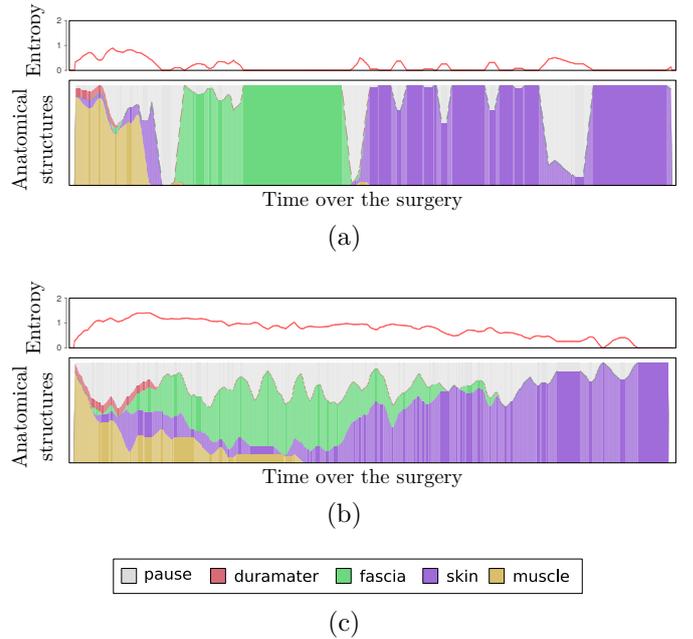


Figure 7: Distribution of the states of the re-aligned set of sequences computed on the data presented in Figure 6. (a) NLTS of the closure phase. (b) Uniform-scaling of the closure phase. (c) Legend. For both figures, we depict the evolution of the state entropy $H(\hat{p}_\ell)$ over time.

Table 1: Entropy based encoding of \mathbb{S}^*

	Senior	Junior
$\mathcal{L}_{\mathbb{S}^*}$	5,206 bits	8,051 bits

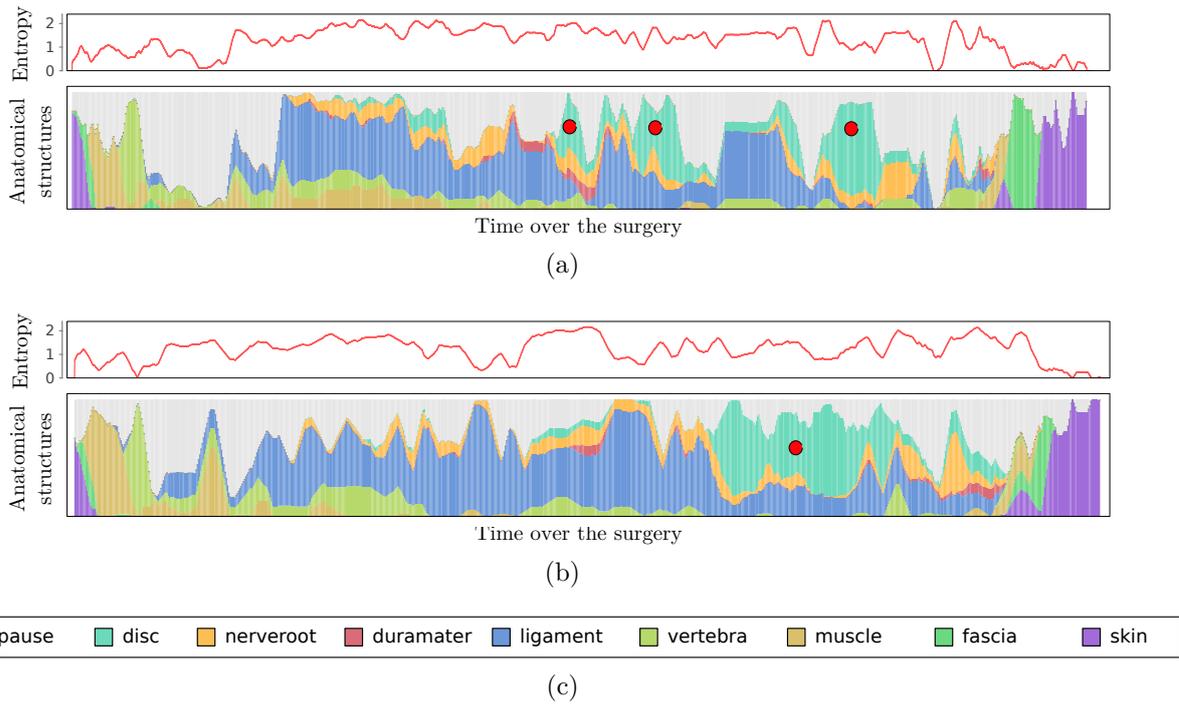


Figure 8: Comparison between junior and senior surgeons on the targeted anatomical structures over time of the surgery. (a) NLTS on junior surgeries. (b) NLTS on senior surgeries. (c) Legend.

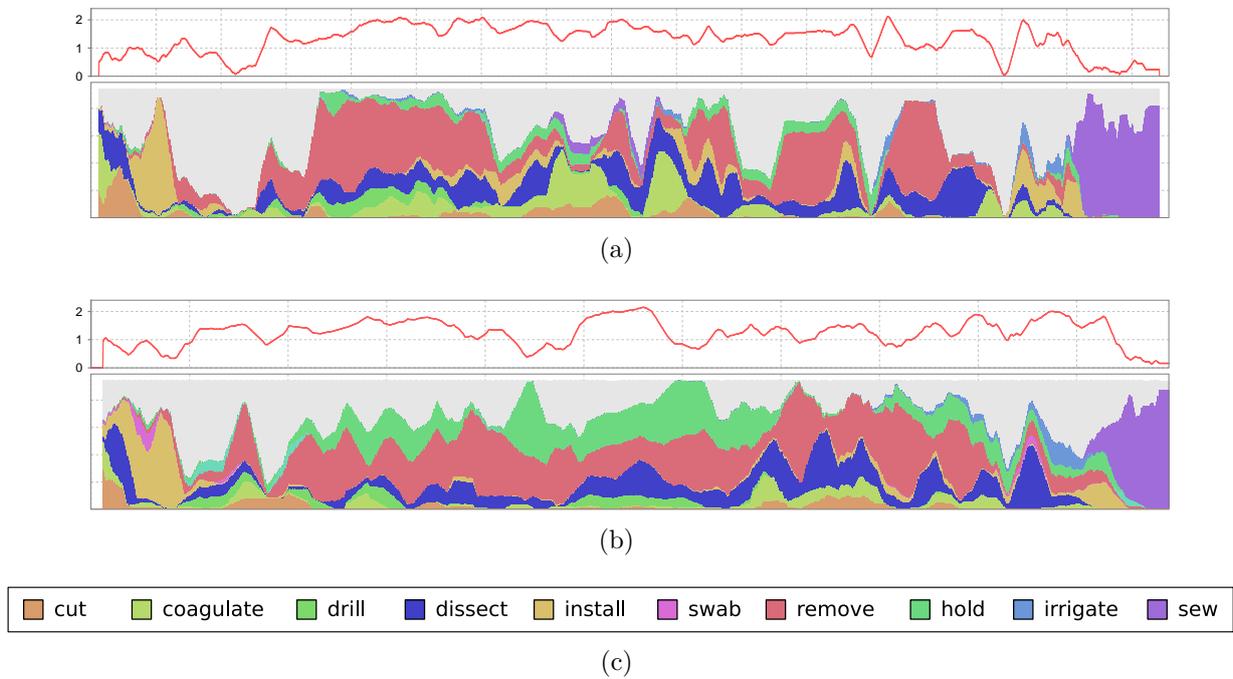


Figure 9: Comparison between junior and senior surgeons on the surgical instruments used over time of the surgery. (a) NLTS on junior surgeries. (b) NLTS on senior surgeries. (c) Legend.

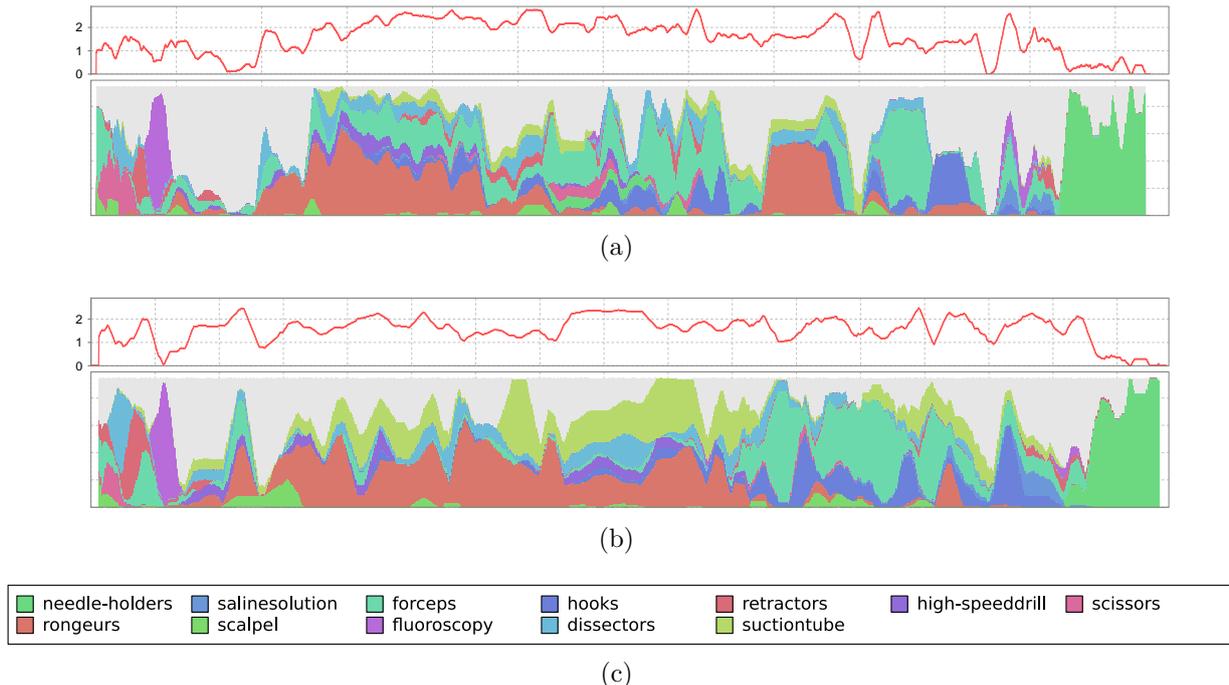


Figure 10: Comparison between junior and senior surgeons on actions performed over time of the surgery. (a) NLTS on junior surgeries. (b) NLTS on senior surgeries. (c) Legend.

5. Discussion

5.1. Closure phase

The visual analysis of original data depicted in Figure 6(a) is difficult, since they have neither the same length, nor a common timeline. Figure 6(b) depicts the results of our NLTS approach, which aims at unlocking this issue. NLTS makes it possible to realign the sequences on their intrinsic timeline. This allows us to consistently assess the way in which the anatomical structures are involved during the surgery.

Comparatively, the state-of-the-art uniform scaling approach (Figure 6(c)) does not take into account any correspondence between the sequences. The timeline corresponds to the percentage of time spent in the surgery, and the results provides only poor information about the surgeries.

Figure 7 presents the comparative results of NLTS on the closure phase of the surgery. We can see in Figure 7(a) that the information about the general sequencing of actions is directly accessible. From the chart, we can for example observe a sequence of anatomical structures that are involved during the closure phase: *muscle* \rightarrow *fascia* \rightarrow *skin*. This is typical of the closure phase: the surgical route is closed layer by layer. The corresponding entropy trend gives also interesting information about the variability of this phase. We can directly identify that the variability of surgical practices is located at the start of the closure phase. The start of the closure phase is indeed often influenced by patient-specific information like his or her anatomy. We can also observe that once the fascia

is targeted, the surgery is then much more stereotyped, because the targeted anatomy is less variable.

Comparatively, Figure 7(b) shows that the uniformly scaled set of sequences provides little information. This information is moreover often misleading the interpretation, because the different anatomical structures are mixed. As a result, it is impossible to use uniform scaling to deduce any standard sequencing of actions of surgeries. The corresponding entropy suggests that the closure phase is very variable/uncertain, while this phase is the most standardised phase of the surgery. The different anatomical structures have indeed to be closed in a specific order.

5.2. Whole surgery

Figure 8 illustrates the sequence of anatomical structures for the entire surgery. We compare the surgeries performed by senior *vs* junior surgeons using NLTS.

We can directly observe in Figures 8(a) and 8(b) the sequence *skin* \rightarrow *fascia* \rightarrow *muscle* \rightarrow *vertebra* at the start of the surgery. This sequence is present both for junior surgeons and for senior surgeons, which demonstrates that this practice is highly standardised.

NLTS can also be used as a tool to evaluate the differences between junior and senior about their surgical practice. First, we can observe that senior behaviour is more homogeneous than the behavior of junior surgeons. Over time, the anatomical structures that are targeted by senior surgeons are almost always the same, *i.e.*, the majority of surgeons perform the same action at the same time of the surgery (Figure 8(b)). To the contrary, Figure 8(a) shows

that junior-operated surgeries exhibit a strong heterogeneity. This heterogeneity can be explained by their lack of experience and dexterity. This visual observation is confirmed by the lengths of the compressed scaled sets of sequences. Table 1 shows that encoding the scaled sets of sequences for senior surgeons requires 35 % less space than for junior surgeons: surgeries performed by junior surgeons are much less predictable than the set of senior surgeries.

This representation makes it possible to finely compare surgical practices. Let us for example explain the phase of the removal of the hernial disc, which corresponds to the most characteristic and riskiest part of the surgery (red dots in Figure 8). We can see that senior surgeons operate on the herniated disc in one phase only. The actions performed on the disc are more spread out when junior surgeons are operating. We can actually observe a strong heterogeneity in the way the junior surgeons perform this phase. Most of junior surgeons operate on the disc in three steps with actions on the ligament in-between. Junior surgeons actually often start working on the disc and have to go back to the ligament, in order to better operate the herniated disc. This can be explained by their lack of experience.

A similar analysis can be conducted on the other components of the surgical activity, *i.e.*, surgical instrument (Figure 9) and action performed (Figure 10).

Regarding the instrument used during the surgery, we can observe that the behaviour of junior surgeons (Figure 9(a)) is very heterogeneous compared to the one of senior surgeons (Figure 9(b)); this is particularly the case during the approach to the disk. This observation can be easily explained by the tendency of junior surgeons to often switch from one instrument to another, because of their lower dexterity. Detecting patterns about surgical instruments would certainly make it possible to improve the organisation of the OR and the coordination of the surgical team.

Regarding the sequencing of actions performed by the surgeon, we can observe that the opening and the closure phases are similar between junior (Figure 10(a)) and senior (Figure 10(b)), which echoes the previous observations during the closure phase. It is indeed standard to *cut* and then *dissect* during the opening phase, and latter to *sew* during the closure phase. However, the heterogeneity of the junior behavior can also be witnessed from this result, especially in the middle of the intervention (see red box in Figure 10(a)).

6. Conclusion and general discussion

Modelling and understanding surgical procedures is an important challenge. DTW is consistent for the comparison of two surgeries but cannot be used to capture the similarities of a whole set of surgeries. In this paper, we introduced non-linear temporal scaling (NLTS). NLTS takes advantage of recent discoveries about scalable time

series averaging and makes it possible to realign a set of sequences, in the same way as DTW does for two sequences.

NLTS is unsupervised: given a set of surgeries, it provides a set of realigned surgeries that maximized the intra-group similarity. This is a major difference with classical SPM approaches that are based on graphical models (*e.g.*, [23]). Graphical model approaches require to define the structure of graphical model: the number of nodes, their roles, their connections and possibly their directions. In addition, once the structure has been defined, learning the parameters of the model (as well as drawing predictions from it) often requires to automatically recognise in which state the surgery currently is, as well as when the surgery is transitioning from one state to another. These decisions heavily rely on background knowledge and parameters' tuning. By contrast, NLTS makes it possible to realign the sequences in an unsupervised manner.

We also introduce different methods based of information theory that can be used to analyse this aligned set of sequences. Our experiments on neurosurgeries (lumbar disc herniation removal) showed that NLTS makes it possible to support the identification of standards of surgical practice. We used the training of neurosurgeons as the common theme for our experiments and showed how NLTS allows us to identify discriminant differences between the practices of junior and senior surgeons.

We believe that NLTS constitutes a useful addition to the SPM toolbox, and that it will support a broad range of applications, from the direct analysis of sets of surgeries, to the construction of more elaborated models (such as dynamic Bayesian networks).

Supplementary materials

NLTS package: Java package containing the source code for the proposed method. (Java ARchive file) – <http://germain-forestier.info/src/aim2014/nlts.jar> (Accessed: 10 October 2014)

Acknowledgement

This research was supported by the Australian Research Council under grant DP120100553 and the French-Australia Science Innovation Collaboration Grants (FASIC) PHC Grant #32571NA. The authors would like to thank all the surgeons of the Neurosurgery Department of the Leipzig University Hospital, Germany involved in this work. The authors would like to thank Dr Rebecca Giblin for her comments on the initial manuscript.

References

- [1] Haynes, A.B., Weiser, T.G., Berry, W.R., Lipsitz, S.R., Breizat, A.H.S., Dellinger, E.P., et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *New England Journal of Medicine* 2009;360(5):491-499.

- [2] Stitzenberg, K., Sheldon, G.. Progressive specialization within general surgery: adding to the complexity of workforce planning. *Journal of the American College of Surgeons* 2005;201(6):925–932.
- [3] Hutter, M., Kellogg, K., Ferguson, C., Abbott, W., Warshaw, A.. The impact of the 80-hour resident workweek on surgical residents and attending surgeons. *Annals of surgery* 2006;243(6):864.
- [4] McPherson, K., Bunker, J.. Costs, risks and benefits of surgery: a milestone in the development of health services research. *Journal of the Royal Society of Medicine* 2007;100(8):387–390.
- [5] Schuhmann, T.. Hospital financial performance trends to watch. *Healthcare financial management* 2008;62(7):59–66.
- [6] Galton, F.. *Vox populi*. *Nature* 1907;75(1949):450–451.
- [7] Sakoe, H., Chiba, S.. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing* 1978;26(1):43–49.
- [8] Forestier, G., Lalys, F., Riffaud, L., Trelhu, B., Jannin, P.. Classification of surgical processes using Dynamic Time Warping. *Journal of Biomedical Informatics* 2012;45(2):255–264.
- [9] Forestier, G., Lalys, F., Riffaud, L., Collins, D.L., Meixensberger, J., Wassef, S.N., et al. Multi-site study of surgical practice in neurosurgery based on surgical process models. *Journal of Biomedical Informatics* 2013;46(5):822–829.
- [10] Levenshtein, V.I.. Binary codes capable of correcting deletions, insertions and reversals. In: *Soviet Physics Doklady*; vol. 10. 1965, p. 707.
- [11] Wang, L., Jiang, T.. On the complexity of multiple sequence alignment. *Journal of Computational Biology* 1994;1(4):337–348.
- [12] Petitjean, F., Ketterlin, A., Gançarski, P.. A global averaging method for Dynamic Time Warping, with applications to clustering. *Pattern Recognition* 2011;44(3):678–693.
- [13] Petitjean, F., Gançarski, P.. Summarizing a Set of Time Series by Averaging: from Steiner Sequence to Compact Multiple Alignment. *Theoretical Computer Science* 2012;414(1):76–91.
- [14] Mehta, N., Haluck, R., Frecker, M., Snyder, A.. Sequence and task analysis of instrument use in common laparoscopic procedures. *Surgical endoscopy* 2002;16(2):280–285.
- [15] Bridgewater, B., Grayson, A., Jackson, M., Brooks, N., Grotte, G., Keenan, D., et al. Surgeon specific mortality in adult cardiac surgery: comparison between crude and risk stratified data. *British Medical Journal* 2003;327(7405):13.
- [16] Martin, J., Regehr, G., Reznick, R., MacRae, H., Murnaghan, J., Hutchison, C., et al. Objective structured assessment of technical skill (osats) for surgical residents. *British Journal of Surgery* 1997;84(2):273–278.
- [17] Schijven, M., Reznick, R., ten Cate, O., Grantcharov, T., Regehr, G., Satterthwaite, L., et al. Transatlantic comparison of the competence of surgeons at the start of their professional career. *British Journal of Surgery* 2010;97(3):443–449.
- [18] Darzi, A., Mackay, S.. Skills assessment of surgeons. *Surgery* 2002;131(2):121–124.
- [19] Pavlidis, I., Tsiamyrtzis, P., Shastri, D., Wesley, A., Zhou, Y., Lindner, P., et al. Fast by nature-how stress patterns define human experience and performance in dexterous tasks. *Scientific Reports* 2012;2.
- [20] Neumuth, T., Jannin, P., Strauss, G., Meixensberger, J., Burgert, O.. Validation of knowledge acquisition for surgical process models. *Journal of the American Medical Informatic Association* 2009;16(1):72–80.
- [21] Jannin, P., Raimbault, M., Morandi, X., Riffaud, L., Gibaud, B.. Model of surgical procedures for multimodal image-guided neurosurgery. *Computer Aided Surgery* 2003;8(2):98–106.
- [22] Neumuth, D., Loebe, F., Herre, H., Neumuth, T.. Modeling surgical processes: A four-level translational approach. *Artificial Intelligence in Medicine* 2011;51(3):147 – 161.
- [23] Neumuth, T., Jannin, P., Schlomberg, J., Meixensberger, J., Wiedemann, P., Burgert, O.. Analysis of surgical intervention populations using generic surgical process models. *International Journal of Computer Assisted Radiology and Surgery* 2010;6:59–71.
- [24] Jannin, P., Morandi, X.. Surgical models for computer-assisted neurosurgery. *Neuroimage* 2007;37(3):783–791.
- [25] Padoy, N., Mateus, D., Weinland, D., Berger, M.O., Navab, N.. Workflow monitoring based on 3d motion features. In: *IEEE 12th International Conference on Computer Vision Workshops*. NJ, USA: IEEE; 2009, p. 585–592.
- [26] Blum, T., Feussner, H., Navab, N.. Modeling and segmentation of surgical workflow from laparoscopic video. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Berlin, Heidelberg: Springer; 2010, p. 400–407.
- [27] Neumuth, T., Loebe, F., Jannin, P.. Similarity metrics for surgical process models. *Artificial Intelligence in Medicine* 2012;54(1):15–27.
- [28] Gusfield, D.. *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*; chap. 14 Multiple String Comparison – The Holy Grail. Cambridge University Press; 1997, p. 332–367.
- [29] Petitjean, F., Forestier, G., Webb, G., Nicholson, A., Chen, Y., Keogh, E.. Dynamic Time Warping averaging of time series allows faster and more accurate classification. In: *IEEE International Conference on Data Mining*. 2014,.
- [30] Yankov, D., Keogh, E., Medina, J., Chiu, B., Zordan, V.. Detecting time series motifs under uniform scaling. In: *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. KDD '07; NY,USA: ACM; 2007, p. 844–853.