Non-Linear Temporal Scaling of Surgical Processes

Germain Forestier, Francois Petitjean, Laurent Riffaud, Pierre Jannin

Non-Linear Temporal Scaling of Surgical Processes

Germain Forestier[a,*], François Petitjean[b], Laurent Riffaud[c,d], Pierre Jannin[c]

[a]*University of Haute Alsace, MIPS EA 2332, Mulhouse, France*
[b]*Faculty of Information Technology, Monash University, Melbourne, Australia*
[c]*INSERM MediCIS, Unit U1099 LTSI, University of Rennes 1, Rennes, France*
[d]*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 to realign a set of surgeries along their intrinsic timeline. 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

---

*Corresponding author – ENSISA - Universite 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)

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.

Conclusion. 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: Surgical Process Modelling, Temporal Analysis, Dynamic Time Warping, 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.

Assessing surgical practice is of major interest to ensure a smooth expertise transition between senior and junior surgeons. This assessment requires a consistent understanding of surgical processes and has thus strongly supported the modelling of surgical processes. 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.

Training new surgeons is particularly important for the quality of care and an important issue from the economical perspective. The training is often provided in a one-on-one scheme between a

junior surgeon and his or her senior. This process is expensive, time-consuming, and heavily reliant upon good communication between the junior and his or her senior. Improving and optimising the training of surgeons requires a better understanding of surgical practices. Let us give an example about how a better understanding of surgeries can help improve the training of surgeons. Let us assume that the analysis of senior surgeries tells us that all senior surgeons perform one step of the surgery in the exact same way. This information could 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 will highlight that the behaviour of senior surgeons is less standard. The training can then be directed towards a patient-specific care.

This paper aims at extracting high-level knowledge from recordings of surgical processes. We consider surgeries as sequences of activities that are performed by the surgeon during the surgery. Surgical activities are generally represented as a triplet [6] : *action*, *anatomical structure* and *instrument*. For example, the surgeon can *cut* the *skin* using a *scalpel*. In general, activities that are performed by both hands are recorded, as well as the use of the microscope. Figure 1 illustrates an example of such a 9-dimensional set of sequences.

Figure 1 demonstrates the complexity of the data. Surgeries can have very different durations, they are described along nine dimensions, they can contain more or less pauses, *et cætera*. Even in this case, which concerns only one variety of surgery (lumbar disc herniation removal), it is very difficult to extract high-level information about the surgeries using existing methods, because we lack a common timeline (*i.e.*, surgeries do not last the same amount of time).

In this article, we unlock this issue and introduce a novel method to provide a consistent common timeline to a set of surgeries. We enable the comparison of a set of surgeries, in order to make it possible to extract high-level information. "What is the standard surgery of a given type?", "What is the difference between surgical practices of junior and senior surgeons?", "What makes this surgery different from the other surgeries?" are the types of questions that can be addressed by the method that we propose in this article. Our method takes advantage of our recent discoveries for

fast multiple sequence alignment [7].

This article addresses the issue of being able to compare a set of surgical processes; Section 2 describes this issue. Section 3 introduces a method to realign a set of sequences along their intrinsic timeline. Section 4 details experiments carried out on neurosurgeries. Section 5 discusses these results and shows that our method makes it possible to provide 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.

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 [8]. 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 [9]. Junior surgeons are evaluated by their seniors with regard to a list of surgical skills. These tests are however subjective depending on the evaluator [10].

Time-motion approaches have been introduced to improve objectivity and to automate the information acquisition on a surgery [11, 12]. 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 [13]. 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.

4

Recording surgeries has recently gained interest, either using sensor devices, or directly by an observer in the OR. An universal and adaptable recording scheme has been introduced [14]. 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 1. 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 [15, 16, 17, 14, 18].

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 cætera*). The challenge is to be able to recognise 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 [19]. Dynamic Time Warping (DTW) is based on the Levenshtein distance and was introduced for speech recognition in the 1970s [20] We recently demonstrated that Dynamic Time Warping is sound for surgical processes comparison [6, 21]. 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 other, which makes it possible to compare them along their intrinsic timeline. Figure 2 illustrates this process on the anatomical structures that are targeted during two surgeries.

5

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

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

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 [22]. Multiple sequence alignment is NP-complete [23], 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 is based on *compact multiple alignment* [7], which was recently introduced to enable the multiple sequence of a large set of homogeneous sequences. Compact multiple alignments allow several successive elements of sequences to be part of the same

column of the alignment. Finding the average sequence under time warping computes the compact multiple alignment of the set of averaged sequences [7]. In addition, scalable methods exist for the definition of an average sequence under time warping [24].

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; 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 [24], but other methods like CO-MASA [7] 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 sth sequence have been associated with the ℓth element of $\bar{S}$(line 5). These elements are then stored in $elements[s][\ell]$. Moreover, we store in $widths[\ell]$ 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 consists of unpacking the compact multiple alignment. 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, one element of the second sequence, and two elements of the third sequence. 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 and the two elements of the third sequence to four elements as well.

Result. The algorithm outputes a set of sequences $\mathbb{S}^\star$, which enables a detailed temporal analysis of the set of sequences. Note that all the sequences of $\mathbb{S}^\star$ now have the same length.

Algorithm 1 Non-Linear Temporal Scaling

Require: $\mathbb{S} = \{S_1, \cdots, S_N\}$
Let $\mathbb{S}^\star = \{S_1^\star, \cdots, S_N^\star\}$ be the resulting set of scaled sequences
Let mean(.) return the average sequence of a set for DTW
Let assocDTW(S, T) return the elements associated mapping built by DTW from S to T
Let scale(S, n) return the uniform scaling [25] of the sub-sequence S to n elements
  1: $\bar{S} \leftarrow \text{mean}(\mathbb{S})$
  2: $L \leftarrow \text{length}(\bar{S})$
  3: $widths[L] \leftarrow [0, \cdots, 0]$
  4: **for** $s \leftarrow 1$ to N **do**
  5:     $elements[s] = \text{assocDTW}(\bar{S}, S_s)$
  6:     **for** $\ell \leftarrow 1$ to L **do**
  7:       $widths[\ell] \leftarrow \max(widths[\ell] , \text{size}(elements[s][\ell]))$
  8:     **end for**
  9: **end for**
10: **for** $s \leftarrow 1$ to N **do**
11:     $S_s^\star \leftarrow \langle \rangle$
12:     **for** $\ell \leftarrow 1$ to L **do**
13:       $targetLength \leftarrow widths[\ell]$
14:       **for** $n \leftarrow 1$ to $targetLength$ **do**
15:         $S_s^\star \leftarrow S_s^\star.\text{scale}(elements[s][\ell], targetLength)$
16:       **end for**
17:     **end for**
18: **end for**
19: **return** $\mathbb{S}^\star$

Algorithm 2 mean

Require: $\mathbb{S} = \{S_1, \cdots, S_N\}$
Let DBA be the averaging sequence method introduced in [24]
Let I be the number of iterations
Let *mean* be the returned average sequence of $\mathbb{S}$
  $mean \leftarrow \text{medoid}(\mathbb{S})$
  for $i \leftarrow 1$ to I **do**
    $mean \leftarrow \text{DBA}(mean, \mathbb{S})$
  end for
  return *mean*

Algorithm 3 medoid

Require: $\mathbb{S} = \{S_1, \cdots, S_N\}$
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**
      $sqrDist \leftarrow sqrDist + \text{DTW}(S, T)^2$
    end for
    if $sqrDist < inertia$ **then**
      $inertia \leftarrow sqrDist$
      $medoid \leftarrow S$
    end if
  end for
  return *medoid*

3.2. Analysing a set of non-linearly scaled sequences

NLTS makes it possible to realign a set of sequences along the intrinsic time of the surgery. 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 $\mathbb{S}^\star$ be the set of sequences scaled with NLTS. Every element ℓ of $\mathbb{S}^\star$ is associated to a set of sequence states $\{S_1^\star(\ell), \cdots, S_L^\star(\ell)\}$. These sets of states informs about the distribution of the states over time. The more similar these states are at ℓ, the more standardised the action ℓ is. We propose to analyse $\mathbb{S}^\star$ by studying the (discrete)

probability distribution of these states over time.

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

$$
\begin{array}{rcl}
\hat{p}_{\ell} & : & E \rightarrow [0,1] \subset \mathbb{R} \\
& & e \mapsto \dfrac{|\{S^{\star}(\ell) = e \mid \forall S^{\star} \in \mathbb{S}^{\star}\}|}{N}
\end{array}
$$

Entropy of the states. Entropy is a measure of uncertainty of a random variable. In our case, the entropy of every state of $S_{\hat{p}}^{\star}$ 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:

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

$H(\hat{p}_{\ell})$ 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 $\mathbb{S}^{\star}$. 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 $\mathbb{S}^{\star}$ on a state-by-state basis is given by:

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

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

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

This length $\mathcal{L}_{\mathbb{S}^{\star}}$ gives information about the general uncertainty on $\mathbb{S}^{\star}$. 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.

4.1. Material

We recorded 24 lumbar disc herniation surgeries in the Neurosurgery Department of the Leipzig University Hospital, Germany. Figure 3 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).

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 intermyolamar route. The discectomy 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. During all junior recordings, the closure step was performed by the junior alone.

11

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 4 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 4 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 4 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 4(a) presents the original (non-scaled) data. Figure 4(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 [25] gives more details about this procedure. The results of this approach are depicted in Figure 4(c).

Figure 5 depicts the comparative study of the evolution of $\hat{p}$ (Definition 1) over time, for the anatomical structures that are involved in the surgery. Figure 5(a) depicts this evolution for NLTS that can be compared to the results obtained with uniform scaling in Figure 5(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.

12

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.

Figure 6 presents the evolution of the targeted anatomical structures. We compare junior and senior behaviours scaled with NLTS, respectively in Figures 6(a) and 6(b). Table 1 compares the length $\mathcal{L}_{\mathbb{S}^\star}$ required to encode junior and senior sets of surgeries.

5. Discussion

5.1. Closure phase

The visual analysis of original data depicted in Figure 4(a) is difficult, since they have neither the same length, nor a common timeline. Figure 4(b) depicts the results of our NLTS approach, which aims at unlocking this issue. NLTS makes it possible to scale 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 4(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 5 presents how to use this information to extract knowledge about the standards of surgery.

Figure 5(a) presents the results of NLTS. The information about the 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

13

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 5(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 the 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 6 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 6(a) and 6(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 6(b)). To the contrary, Figure 6(a) shows that junior-operated surgeries exhibit a strong heterogeneity. This heterogeneity can be explain 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

14

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 6). 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.

This analysis about the standard sequencing of anatomical structures could also be performed on other components of the surgeries. For example, we could study the sequence of the surgical instruments that are used during the surgery. Detecting patterns about surgical instruments would certainly make it possible to improve the organisation of the OR and the coordination of the surgical team. The sequencing of actions could be used to detect patterns about surgical practice, in order to better understand complex surgical behaviours across different populations of surgeons.

6. Conclusion

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 multiple alignment methods and makes it possible to realign a set of sequences along their intrinsic timeline, in the same way as DTW does for two sequences. 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 identify 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 this work opens up a number of research directions. We plan to temporally model surgical processes based on Markov or on Dynamic Bayesian Networks systems. These models would provide high-level representations of surgical procedures, which would in turn lead to a better understanding of surgical practice.

Supplementary materials

NLTS package: Java package containing the source code for the proposed method. (Java ARchive file) – http://www.tiny-clues.eu/Research/NLTS/nlts.jar

Acknowledgement

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] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] Darzi, A., Mackay, S.. Skills assessment of surgeons. Surgery 2002;131(2):121–124.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] Jannin, P., Morandi, X.. Surgical models for computer-assisted neurosurgery. Neuroimage 2007;37(3):783–791.

[19] Neumuth, T., Loebe, F., Jannin, P.. Similarity metrics for surgical process models. Artificial Intelligence in Medicine 2012;54(1):15–27.

[20] 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.

[21] 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;.

[22] 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.

[23] Wang, L., Jiang, T.. On the complexity of multiple sequence alignment. Journal of Computational Biology 1994;1(4):337–348.

[24] 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.

[25] 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; ACM; 2007, p. 844–853.

Tables

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

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

Figure Captions

Figure 1. 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).

Figure 2. Stretching two sequences with DTW. Each color corresponds to one anatomical structure. (a) Original sequences. (b) DTW-alignment. (c) DTW-scaled sequences.

Figure 3. Recording of the data in the OR.

Figure 4. 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.

Figure 5. Distribution of the states of the re-aligned set of sequences computed on the data presented in Figure 4. (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.

Figure 6. 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.