Proceedings of the ASME 2012 11th Biennial Conference On Engineering Systems Design And Analysis
ESDA 2012
July 2-4, 2012, Nantes, France

ESDA2012-82862

OBJECTIVE ASSESSMENT OF SURGICAL SKILLS

Jenny A. Cifuentes * **Minh Tu Pham**
Richard Moreau
Laboratoire Ampère
University of Lyon
INSA-Lyon
69621 Villeurbanne Cedex
Email: jenny.cifuentes-quintero@insa-lyon.fr

Flavio Prieto
University National of Colombia
Electrical Engineering Department
Bogota
Colombia

Pierre Boulanger
AMMI
University of Alberta
Computer Sciences Department
Edmonton, Canada

ABSTRACT

Minimally Invasive Surgery (MIS) has definitively changed the procedures performed in operating rooms. In many cases, MIS has become the recommended standard technique, replacing the traditional open surgery. Effective training and objective assessment of surgeons in these techniques become a major concern in hospitals in recent years, encouraged primarily by patients and a society that demands safer surgical procedures, which is associated with better surgeons training. In the framework of surgery, the difficulty of defining objective metrics for performance evaluation lies in the strict dependency between tasks and the difficulty of defining the meaning of optimal performance, related to the characterization of gestures made by the experts. An objective method to compare 3D gestures between an expert and novice surgeons through multidimensional data analysis is proposed in this paper. A survey of different algorithms for surgical gestures analysis in time domain is carried out. These ones include the Multi-dimensional Dynamic Time Warping (MD-DTW) and Multi-Dimensional Derivative Dynamic Time Warping (MD-DDTW) . Simulation and experimental results are given with this different techniques.

INTRODUCTION

Minimally Invasive Surgery (MIS) is a new kind of surgery which gets more and more common nowadays. With this new technology, surgeries previously performed through large incisions that required long recovery times are now performed with a much shorter recovery. Although on one hand MIS procedures ensure many advantages to patients, on the other hand they require surgeons to perform a long and difficult training in order to manage these techniques. Currently, surgeon training includes the acquisition of motor skills outside the operating room. Mechanical simulators and virtual reality simulators have been developed over the last decades not only to learn and to transfer the knowledge, skills or attitudes but also to evaluate procedure performance. These tools define metrics to get an objective assessment of surgical skills. However, these metrics are primarily related to simple data measurements (such as time, trajectories, velocity, acceleration, smoothness, etc.).

In this context, the assessment of technical skills during training has been considered to be an index of quality assurance for the future. Especially, objective assessment is essential because deficiencies in training and performance are difficult to correct without objective feedback. The analysis of human motor performance in MIS has important implications for surgical simulation, surgical training and robotics. Measures of motor performance include kinematic measurements such as position, orientation as well as force measurements and video analysis. A global rating scheme based on a video analysis and gesture decomposition into several main tasks was developed in [1]. Hand motion tracking has been shown to correlate with objective measures of technical skill [2, 3]. In particular, time series similarity

*jacifuentesq@gmail.com

measures have been used to perform classification on a set of gestures [4]. In this case, the trajectory of the instrument tip is modeled as a sequence of consecutive locations in a multidimensional (generally two or three dimensional) Euclidean space. One technique is based on the time warping technique that first has been used to match signals in speech recognition [5]. Berndt and Clifford [6] proposed to use this technique to measure the similarity of time-series data in data mining. Pham et al. use this method to match and compare the resulting curvature calculation of two 3D trajectories [7]. Vlachos et al. include non-metric similarity functions based on the Longest Common Subsequence (LCSS), another method which is very robust to noise and furthermore provide an intuitive notion of similarity between trajectories [8]. A similar technique is to find the longest common subsequence of two sequences and then define the distance using the length of this subsequence [9, 10].

Hidden Markov Models (HMM) have been also used in this field. HMM were developed in the area of speech recognition [11]. Later, they were applied for studying teleoperation [12], human manipulation actions [13], human skills assesment for the purpose of transferring human skill to robots [14] and manufacturing application [15]. Also, HMMs have been implemented for gesture recognition in the rehabilitation technology field [16]. Rosen et al created a database with forces and torques signals acquired during actual operating conditions to develop statistical models, which can be used to objectively evaluate surgical skills [17]. However, HMMs requires a huge set of data in order to compute the HMM parameters. Additionally, the discrete time aspect of HMM does not fit well with the continuous time nature of 3D gesture comparison [8].

The main objective of our paper is to propose an objective method to evaluate and compare the performance of a gesture between different people through multidimensional data analysis such as the 3D position (x, y, z). Recent researches use tools for the analysis of time series such as Multi-dimensional Dynamic Time Warping (MD-DTW) [4], Dynamic Time Warping (DTW) [18] and Longest Common Sub-Sequence (LCSS) [8]. In this paper, we will give a comparison between two DTW algorithms within the framework of gesture classification. The study is firstly performed in simulation with respect to different parameters then is applied on experimental data. A modified version of MD-DDTW is proposed in the last section and compare to the classical algorithm.

1 TIME SERIES SIMILARITY MEASURES

Multi-dimensional series are sets of data in which multiple measurements are made simultaneously. These kind of series contain a vector of feature values for each occurrence (usually the time is the instance) of the series. The simplest approach to define the distance between two multi-dimensional series is to use the notion of p−norm. Consider two n−dimensional vectors x_1 y x_2, their p−norm distance, noted $L_p(x_1,x_2)$, is defined as foloowing:

$$L_p(x_1,x_2) = \Big(\sum_{i=1}^{n} |x_1 - x_2|^p\Big)^{\frac{1}{p}} \quad (1)$$

Eqn. (1) allows to define an efficient index that takes into account the different dimensions of x_1 and x_2. However, practical issues can raise in practice since this operator is sensitive to noise and outliers. Moreover, for time series, distortions appear as soon as the two series do not have the same number of samples. In order to overcome the pitfalls described above some authors have suggested more flexible approaches to measure the similarity between multi-dimensional time series.

1.1 DTW

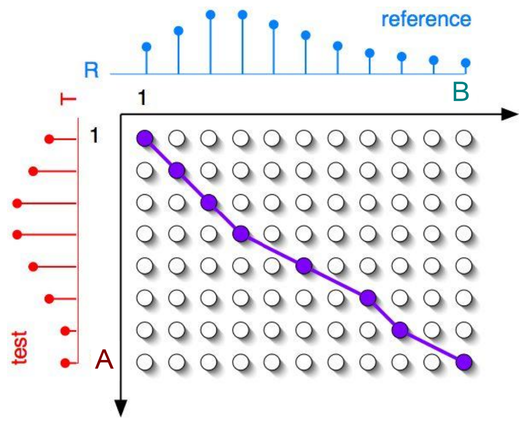


FIGURE 1. Matrix of Cumulated Distance (reference **web.media.mit.edu**)

DTW is a time series alignment algorithm. It aims at aligning two sequences by warping the time axis iteratively until an optimal match between the two sequences is found. Consider two sequences of feature vectors:

$$A = a_1, a_2, a_3, ...a_n \quad (2)$$
$$B = b_1, b_2, b_3, ...b_m \quad (3)$$

To align two sequences using DTW, an n-by-m matrix is constructed, where the (i^{th}, j^{th}) element of the matrix contains the distance $d(a_i,b_j)$ between the two points a_i and b_j. In 1D-DTW, the distance is usually calculated by taking the absolute or the squared distance between the feature values of each combination of points. These distances are used to calculate cumulative distance matrix D. Each matrix element (i,j) corresponds to the alignment between the points a_i and b_j (see Figure 1) and a warping path T is a contiguous set of matrix elements that defines a mapping between A and B. To find the best match or alignment between these two sequences one need to find a path through the grid which minimizes the total distance between them. The basic method for finding the optimal warping matrix requires the evaluation of an exponential number of warping paths. Fortunately dynamic programming can be used to find the optimal warping path.

The different steps of 1D-DTW are summarized in the algorithm 1.

Algorithm 1 The 1D-DTW algorithm

$d(i,j) = (a_i - b_j)^2$
$D(1,1) = d(1,1)$
$D(i,1) = D(i-1,1) + d(i,1)$
$D(1,j) = D(1,j-1) + d(1,j)$
{Book keep for each cell the index of this neighboring cell, which contributes the minimum score}.
$D(i,2) = min(D(i,1), D(i-1,1), D(i-1,2)) + d(i,2)$
$D(i,j) = min(D(i,j-1), D(i-1,j-1), D(i-1,j)) + d(i,j)$
{Found best path T}.

1.1.1 MD-DTW

Since the different dimensions could have different units, authors like [4] have suggested some variations of the classical DTW algorithm called Multi-dimensional Dynamic Time Warping (MD-DTW). So to synchronize different dimensions, it is necessary to normalize each dimension to a zero mean and unit variance to make the dimensions comparable.

The algorithm 2 shows the different stages of MD-DTW algorithm :

Algorithm 2 The MD-DTW algorithm

{Normalize each dimension to a zero mean and unit variance}.
{If desired, smooth each dimension with a Gaussian filter}.
{Fill the m by n distance matrix D according to}.
$D(i,j) = \sum_{k=1}^{K} |a(i,k) - b(j,k)|$
{Where K is the number of dimensions of the time series}
{Use 1D-DTW algoritm}

1.1.2 MD-DDTW

According to [19] and [20] if DTW attempts to align two vectors that are similar except for local accelerations and decelerations in time axis, the algorithm is likely to be successful. However, the classical DTW algorithm has problems when the two sequences have great variations in their feature data. To prevent this problem, the latter authors have suggested to consider the first derivative of the sequences rather the raw data.

In this algorithm, the distance measurement $d(A,B)$ is not euclidean but rather the square of the difference of the estimated derivatives, where the following central derivative algorithm could be used for simplicity:

$$D_x[A] = \frac{(a_i - a_{i-1}) + ((a_{i+1} - a_{i-1})/2)}{2} \tag{4}$$

This three-point estimation is a more accurate approximation and less to outliers than the corresponding two-point estimation.

2 SIMULATION ANALYSIS

In this part, we study the robustness of the techniques described in the previous section with respect to the variation of different parameters such as the number of samples, the existence of outliers and noise in the measurement. The objective is to point out practical issues while implementing the different DTW algorithms.

2.1 Number of Samples

First, we have selected a circular path with different number of samples as the example shown in Figure 2. The position data could be considered as time series where the dimension of the vector is equal to three.

With these two trajectories, we used the two multidimensional synchronization techniques described in the previous section. Figure 3 shows an example of the synchronization between each point made by both techniques. We repeated the process with different number of samples and calculated the distance d of the optimal path found by the two algorithms. The results of the simulation are given in Table 1.

The results show that the cumulative distance of MD-DTW remains constant within a wider range than MD-DDTW technique. In other words, these results point out that a better performance is obtained with MD-DTW technique while varying the number of samples. The reduction of the number of samples is equivalent to a loss representative information in the data resulting in a distortion of the derivative used signals in MD-DDTW. This phenomenon leads to a very different cumulative distance for MD-DTW compared to MD-DDTW. In contrast when the number of samples increases, the performance of both algorithms is similar.

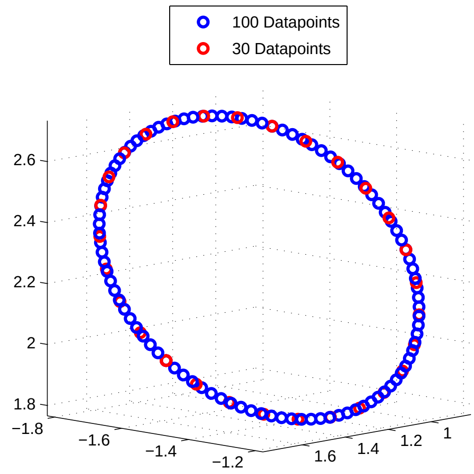
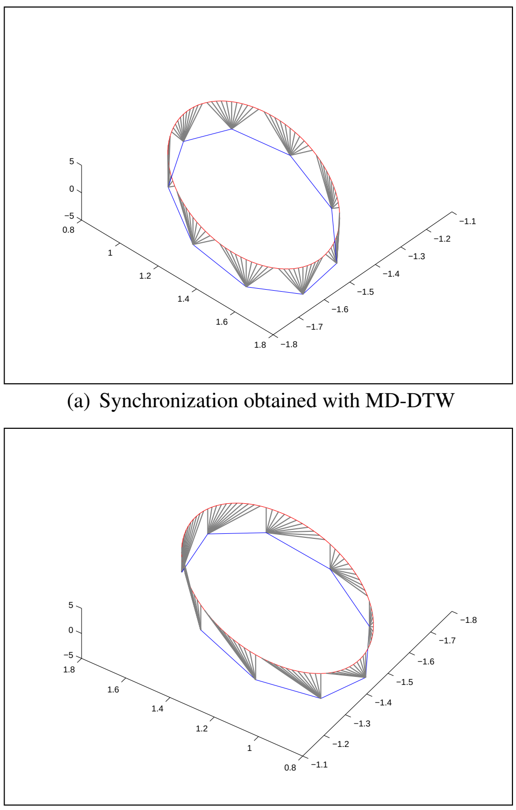


FIGURE 2. Circular path with different numbers of samples



(a) Synchronization obtained with MD-DTW

(b) Synchronization obtained with MD-DDTW

FIGURE 3. Synchronization of two trajectories with respectively 100 and 10 samples

2.2 Outliers

In this part, we have added artificial outliers to the trajectories (see Figure 4) in order to study the sensitivity of the different techniques with respect to this factor.

Number of samples from the path 1	Number of samples from the path 2	d (MD-DTW)	d (MD-DDTW)
100	10	8.54	30.8
100	30	2.68	7.57
100	50	1.58	3.20
100	70	1.12	1.36
100	90	0.87	0.85

TABLE 1. Cumulative distance: variation of the number of samples

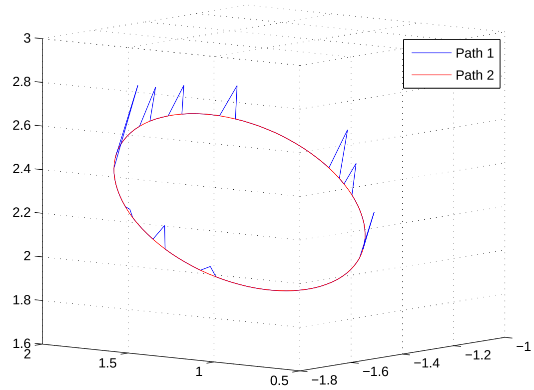


FIGURE 4. Artificial Outliers in the trajectories

The analysis has been carried out according two criteria: a variation of the number of the outliers present in the path and a variation of the maximum amplitude of the outliers. Finally, we compare these results obtained with and without a pre-treatment of the data through a lowpass filter. The results are summarized in the Table 2 for two sets of data with the same number of samples.

Without filtering, the sudden changes in the trajectory generate very high variations of the derivative causing errors while performing the synchronization with MD-DDTW. Thus, in order to improve the results obtained with this technique, we use a lowpass filter to smooth the derivative of each coordinate. With this modification, the results are improved and we obtained a minimum value of the distance (close to 0.9). This value is close to the result obtained in Table 1 (0.85) for signals with number of samples 90 and 100, respectively.

% of outliers	Amplitude of added outliers	d (MD-DTW)	d (MD-DDTW)	d (MD-DDTW) with low-pass filter
5	0.05	0.23	0.48	0.9
5	0.1	0.36	0.71	0.91
5	0.3	1.04	2.09	0.89
10	0.05	0.45	0.87	0.89
10	0.1	0.76	1.53	0.89
10	0.3	2.69	5.38	0.91
15	0.05	0.54	1.00	0.91
15	0.1	1.19	2.1	0.9
15	0.3	3.03	5.45	.91

TABLE 2. Cumulative distance: Outliers

2.3 Noise

Finally, the last study concerns the effect of noise on the different algorithms. Different noise levels have been applied to one trajectory. (see Figure 5).

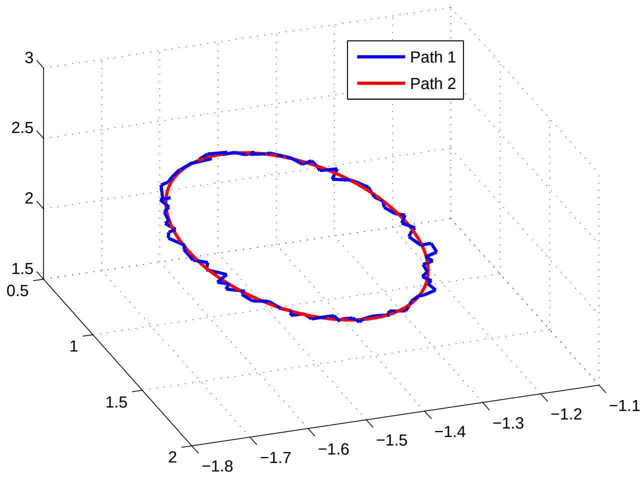


FIGURE 5. Artificial noise applied on one trajectory

With the addition of noise, we use a lowpass filter to smooth the derivatives of each coordinate and compare the results with the techniques described in the previous section. The results are summarized in the Table 3.

Like previously the results show that the inclusion of the filter allows to smooth the trajectories in all the cases and bring benefits to the MD-DDTW technique. In this case, this is because MD-DDTW technique works well not only to variations in time but also in amplitude. This can be seen by the following examples, with the MD-DTW technique, identical sequences

SNR (Signal to Noise Ratio)	d (MD-DTW)	d (MD-DDTW)	d (MD-DDTW) with low-pass filter
50	1.50	1.48	0.81
48	1.64	1.75	0.81
45	2.24	2.45	0.81
43	2.74	3.16	0.8
40	3.39	4.00	0.8
38	4.45	5.08	0.81
35	6.04	7.29	0.8

TABLE 3. Cumulative distance: Noise

will clearly produce a one to one alignment. But if we slightly change a local feature, adding noise or outliers, MD-DTW attempts to explain the difference in terms of the time-axis and produces two unintuitive alignments where a single point on one trajectorie maps onto a large subsection of the another trajectorie increasing the matching error.

3 EXPERIMENTAL RESULTS

The goal of the experiments carried out in this section is to show an objective assessment and classification of the gesture. To conduct our study, we acquired the 3D position (x, y, z) of a simple gesture performed by five people. The right arm of the participants were dressed with a four degrees of freedom exoskeleton [21], then they were asked to elevate their arm to simulate a gesture similar to the one made to stop/halt someone. The cartesian position of the wrist was obtained based on the kinematic model of the exoskeleton and the position records stemming from the encoders of the different joints. The trajectory performed by one of them is shown in the figure 6 and similar trajectories can be plotted for the other people.

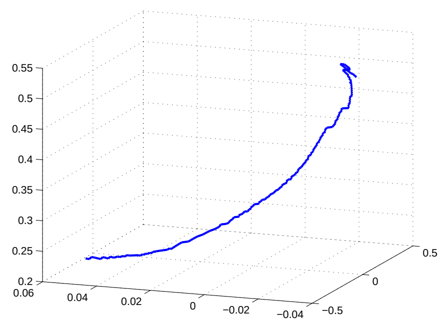


FIGURE 6. Trajectory Performed

Due to the difference between the number of samples in the experimental trajectories, that will be synchronized, is not large, some amount of noise and outliers can be included in the trajectory and based on the results of the previous section, we perform the synchronization of data with MD-DDTW technique using a lowpass filter to smooth the derivatives obtained numerically.

In the same way, as part of this work, we performed an experimental study in which we analyze the influence of the numerical approximation of the derivative using a lowpass filter and MD-DDTW technique and comparing the results with the use of a derivative filter and MD-DTW technique.

At this point, we used the low-pass filter and MD-DDTW technique to synchronize each trajectory. The table 4 shows the distance d obtained between each similar pair of paths for the gesture with this method.

MD-DDTW d_1	Person 2	Person 3	Person 4	Person 5
Person 1	0.48	0.41	0.51	0.43
Person 2	-	0.43	0.52	0.4
Person 3	-	-	0.48	0.45
Person 4	-	-	-	0.48

TABLE 4. Experimental Results

Finally, we used another alternative for the synchronization, where we applied a derivative filter directly and then found the corresponding points with MD-DTW technique. The difference between these two alternatives can be highlighted on Table 7.

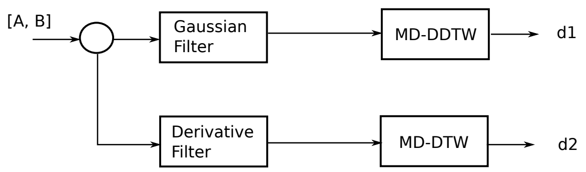


FIGURE 7. Derivative Filter

The filter used in the MD-DDTW algorithm is a Gaussian filter designed to have no overshoot for a step input while minimizing the rise and fall time. This behavior is chosen because it allows to obtain the minimum phase shift.

The one-dimensional gaussian filter has an impulse response given by

$$g(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-x^2}{2\sigma^2}} \tag{5}$$

Where x is the distance from the origin in the horizontal axis and σ is the standard deviation of the Gaussian distribution. Thus, the derivative filter implemented is given by the following equation.

$$h(x) = \frac{dg(x)}{dx} = \frac{-x}{\sigma^2\sqrt{2\pi\sigma}} e^{\frac{-x^2}{2\sigma^2}} \tag{6}$$

The results obtained with this alternative are summarized in the Table 5.

MD-DDTW d_2	Person 2	Person 3	Person 4	Person 5
Person 1	0.35	0.38	0.4	0.37
Person 2	-	0.33	0.41	0.36
Person 3	-	-	0.36	0.35
Person 4	-	-	-	0.37

TABLE 5. Derivative Filter

From the last values, we can see that the method using the derivative filter gives a distance values lower for similar trajectories. These results are good because they suggest that we could expect a more effective classification of the trajectories. Since regular MD-DDTW method uses a numerical scheme to approximate the derivative it turns out that the computed distance depends somehow on the numerical scheme. By applying the second strategy, we use an analytical expression of the derivative to improve the accuracy of the results.

4 CONCLUSIONS

Objective evaluation of surgical skills removes the subjectivity and variability from current expert-based evaluation methods. Different metrics have been used for automatic evaluation in order to be as accurate as possible, including position, speed, acceleration, forces, torques, and any kind of signals. In this paper different algorithms used to find similarities between time series based on the 3D position of the trajectories and DTW with its different versions has been presented in particular. For those techniques a simulation study shown that the performance of each method is related to different parameters as the number of samples, the amount of outliers and noise that could be included in the acquired signal.

In the experimental section, we show the benefits of a modified version of MD-DDTW. This proposen method includes a derivative filter instead of the use of a low pass filter and a numerical derivate of the trajectories. The experimental results shown a better and more accurate behavior due to the analytical nature of the derivative filter implemented.

Our ongoing research in this field is focused on performing an statistical study with a considerable amount of real surgical gestures instead general gestures that we used in this survey. This approach will allow us to obtain more accurate results in the field of objective assesment of surgical gestures.

REFERENCES

[1] Cao, C., MacKenzie, C., and Payandeh, S., 1996. "Task and motion analyses in endoscopic surgery". *ASME IMECE Symposium on haptic interfaces for virtual environement and teleoperator system*, pp. 583–590.

[2] Datta, V., Chang, A., Mackay, S., and Darzi, A., 2002. "The relationship between motion analysis and surgical technical assessment". *American Journal of Surgery,* ***184***(1), July, pp. 70–73.

[3] Datta, V., Mackay, S., Mandalia, M., and Darzi, A., 2001. "The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model". *Journal of the American College of Surgeons,* ***5***, pp. 479–485.

[4] Holt, G. T., Reinders, M., and Hendriks, E., 2007. "Multi-dimensional dynamic time warping for gesture recognition". In Annual Conference of the Advanced School for Computing and Imaging.

[5] Rabiner, L., and Juang, B., 1993. *Fundamentals of Speech Recognition*. Prentice Hall.

[6] Berndt, D., and Clifford, J., 1994. "Using dynamic time warping to find patterns in time series". *AAAI Workshop on Knowledge Discovery in Databases*, pp. 359–370.

[7] Pham, M., Moreau, R., and Boulanger, P., 2010. "Three-dimensional gesture comparison using curvature analysis of position and orientation". In IEEE/EMBS International Conference of the Engineering in Medicine and Biology Society, pp. 6345–6348.

[8] Vlachos, M., Kollios, G., and Gunopulos, D., 2002. "Discovering similar multidimensional trajectories". In IEEE International Conference on Data Engineering, pp. 673–684.

[9] Agrawal, R., Lin, K., H.Sawhney, and Shim, K., 1995. "Fast similarity search in the presence of noise, scaling and translation in time-series databases". In International Conference on Very Large Data Bases, pp. 490–501.

[10] Bollobas, B., Das, G., Gunopulos, D., and Mannila, H., 1997. "Time-series similarity problems and well-separated geometric sets". In Poceedings of the thirteenth annual Symposium on Computational Geometry.

[11] Rabiner, L., 1989. "A tutorial on hidden markov models and selected application in speech recognition". In Proceedings of the IEEE, Vol. 77.

[12] Hannaford, B., and Lee, P., 1990. *Experimental Robotics I*, Vol. 139. Springer, ch. Hidden Markov model analysis of force/torque information in telemanipulation, pp. 135–149.

[13] Pook, P., and Ballard, D., 1993. "Recognizing teleoperated manipulations". *IEEE International Conference on Robotics and Automation,* ***2***, pp. 578 – 585.

[14] Nechyba, M., and Xu, Y., 1998. "Stochastic similarity for validating human control strategy models". *IEEE Transactions on Robotics and Automation,* ***14***(3), June, pp. 437 – 451.

[15] Hannaford, B., 1991. "Hidden markov model analysis of manufacturing process information". In IEEE/RSJ International Workshop on Intelligent Robots and Systems, Vol. 2, pp. 676 – 681.

[16] Wachsmuth, I., and Frohlich, M., 1998. *Gesture and Sign Language in Human-Computer Interaction*. Springer.

[17] Rosen, J., Hannaford, B., Richards, C., and Sinanan, M., 2001. "Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills". *IEEE Transactions on Biomedical Engineering,* ***48***(5), May, pp. 579–591.

[18] Sielhorst, T., Blum, T., and Navab, N., 2005. "Synchronizing 3d movements for quantitative comparison and simultaneous visualization of actions". In IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 38–47.

[19] Keogh, E., and Pazzani, M., 2001. "Derivative dynamic time warping". In SIAM International Conference on Data Mining, pp. 2864–2869.

[20] Zhang, Y., and Edgar, T. F., 2008. "A robust dynamic time warping algorithm for batch trajectory synchronization". In IEEE American Control Conference, pp. 2864–2869.

[21] Moubarak, S., Yang, F., Pham, M., Pajdla, T., and Redarce, T., 2009. "Design and modeling of an upper extremity exoskeleton". In World Congress on Medical Physics and Biomedical Engineering, pp. 476–479.