Automatic and Generic Evaluation of Spatial and Temporal Errors in Sport Motions

Marion Morel[1,2], Richard Kulpa[1], Anthony Sorel[1], Catherine Achard[2] and Séverine Dubuisson[2]
[1] *M2S laboratory, Université Rennes 2, ENS Rennes, Avenue Robert Schuman, 35170 Bruz, France*
[2] *Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 7222, ISIR, F-75005, Paris, France*
marion.morel@isir.upmc.fr, richard.kulpa@univ-rennes2.fr, anthony.sorel@univ-rennes2.fr, catherine.achard@upmc.fr, severine.dubuisson@isir.upmc.fr

Keywords: DYNAMIC TIME WARPING, EVALUATION, MULTIDIMENSIONAL FEATURES, SYNCHRONY, MOTION CAPTURE

Abstract: Automatically evaluating and quantifying the performance of a player is a complex task since the important motion features to analyze depend on the type of performed action. But above all, this complexity is due to the variability of morphologies and styles of both the experts who perform the reference motions and the novices. Only based on a database of experts' motions and no additional knowledge, we propose an innovative 2-level DTW (Dynamic Time Warping) approach to temporally and spatially align the motions and extract the imperfections of the novice's performance for each joints. In this study, we applied our method on tennis serve but since it is automatic and morphology-independent, it can be applied to any individual motor performance.

1 INTRODUCTION

One of the key factors of sport performance is the motor control. The players must indeed accurately control their movements in space and time, for instance by temporally synchronizing their limbs or by placing a body part at a precise location, relative to their own bodies or their surrounding environment. The progression of a novice player thus requires to identify these spatiotemporal errors to correct them. This evaluation of a motion requires the expertise of a coach due to the variability of correct performances. Each expert has indeed his/her own way to perform the movement depending on morphology, physical abilities and style.

Some specific motions such as katas in karate could be repeated and trained without the permanent presence of a coach, at home for instance. However, an automatic evaluation system is then required to identify and highlight the errors of the player to help him make progress. Some tools are proposed like the Golf Training System (Explanar Ltd, Manchester, UK) or the PlaneSWING Training System (Portugolfe Ltd, Bedfordshire, UK) in golf, but they are dedicated to specific motions and to only a limited set of features (for instance speed in a $2D$ plane). Moreover, only few studies are taking the morphology of the players into account (Sorel et al., 2013).

The goal of this paper is to provide an efficient and automatic morphology-independent and sport-independent method to evaluate the motion of a player by comparing it to a database containing the same motions performed by experts.

2 RELATED WORK

Being able to automatically evaluate the quality of various actions requires to determine the kinematic factors that are the core of a good performance for each of these motions. For this reason, some authors proposed to add knowledge to the motion evaluation process to know in advance the features to analyze. For instance, Burns et al. defined a set of rules that characterizes some kata in karate, such as the linear trajectory the kicking wrist must follow (Burns et al., 2011). Komura et al. based their evaluation on the minimization of the global movement since they considered that the defender can better counteract an attack if he does not move too much just before the action (Komura et al., 2006). Finally, Ward used several intersegmental angles to com-

pare several classical ballet techniques (Ward, 2012). These studies provide interesting results that are useful for evaluating specific motions. However, our goal is to propose a generic evaluation method that can automatically determine the important features of the expert motions that are then used to evaluate the performance of a new player.

Several authors have worked on this automatic extraction of the relevant features of motions. It is indeed a prerequisite on other domains such as motion recognition or motion retrieval in which these features are both used 1) to group set of motions into categories of actions and 2) to differentiate these groups of actions. For the first case, some authors have proposed to identify common geometrical patterns of the motions: by partitioning the $3D$ space with Cartesian patches (Wang et al., 2012) or angular ones (Xia et al., 2012), by simplifying the joints trajectories with linear regressions (Barnachon et al., 2013) or by using pentagonal areas to represent the postures (Sakurai et al., 2014). Some authors also worked on the relation between the position of a joint relatively to a plan defined by 3 other joints to give a semantic and intuitive evaluation of the performed motion (Röder, 2006; Müller et al., 2005; Müller and Röder, 2006). Finally, several authors tried to define morphology-independent features to manage the morphology variability by normalizing the posture representation and by extension the motion (Sie et al., 2014; Kulpa et al., 2005; Shin et al., 2001). The goal of these studies was to identify the similarity of motions while our is to evaluate the difference between a motion and the reference ones performed by experts. The motions are thus supposed to be similar and our objective is to quantify the errors between them and not to try to ignore these small differences. For the second case, some authors have computed the variance (Ofli et al., 2012) or the entropy (Pazhoumand-Dar et al., 2015) of each joint to discriminate the most informative features characterizing the motion. The problem of such approaches is that they lost some of the temporal information of the motion.

This temporal information is yet essential to evaluate motions and especially sports ones. The temporality of a movement is important for dance of course but it also concerns all kinds of motions since the synchronization of the limbs or the sequence of body motions are the key factors of a good technique and thus a good performance. The temporal information is thus essential at a global level but above all at the joints level, highlighting the relative timing of the different body parts of the player. Maes et al. proposed to evaluate and train the basics of dance steps (Maes et al., 2012). Since they considered that the dance steps were very rhythmic, they based their analysis on the music tempo of the dance. This case is however very specific and only manages the synchronization of the motion with an external and global tempo. To take local synchronization into account, the temporality must be evaluated even when motions have different lengths, different speeds and/or different rhythms. To this end, some authors proposed to use Hidden Markov Models (HMM) or Hidden Conditional Random Field (HCRF) to encode time series as piecewise stationary processes (Zhong and Ghosh, 2002; Kahol et al., 2004; Sorel et al., 2013; Wang et al., 2006). In our context, the time-varying features are trajectories and are modeled as a state automaton in which each state stands for a range of possible observation values of the feature while the transitions between states can model time. The feature observation values and the transitions between states are driven by probabilities, which makes HMM very robust to spatiotemporal variations. However, this approach gathers similar postures together in a same state and the temporality is only managed between these states that can represent a large part of the motion if at a period a joint does not move a lot for instance.

To generically evaluate the synchrony of two motions, we need a more accurate method such as the Dynamic Time Warping (Sakoe and Chiba, 1978). Originally created for speech processing, DTW has become a well-established method to account for temporal variations in the comparison of related time series. Many studies have tried to upgrade the efficiency of the DTW algorithm over the recent years depending on its application's context (Keogh and Pazzani, 2001; Zhou and De La Torre Frade, 2009; Zhou and de la Torre, 2015; Heloir et al., 2006; Gong et al., 2014). In motion retrieval, DTW has then been used by several authors to align the motion with some features to determine the movement performed. Sakurai et al. for instance tried to evaluate a motion captured with the Microsoft Kinect by using pentagonal areas defined by the body end-effector (Sakurai et al., 2014). Pham et al. tried to compare surgery motions by aligning trajectories of $3D$ sensors (Pham et al., 2010). The problem of these studies is that the motion is simplified to manage the temporality and the joint information are not preserved. In our approach, we want to take temporal and spatial information into account concurrently.

In this paper, we propose an efficient and automatic morphology-independent method based on

DTW to compare a motion performed by a player to a database of experts' motions in order to evaluate concurrently the spatial and temporal relevant information of the motion.

3 METHODOLOGY

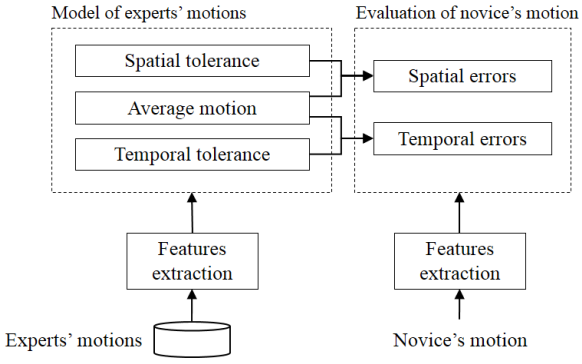

Figure 1: Global framework of the proposed approach

The purpose of this paper is to determine whether a motion is correct or not and, if not, to determine where and when it is badly performed. To this end, we need to compare this motion to the reference ones, the motions performed by experts. The reference cannot indeed be only one motion because it is necessary to take the variability of all experts performance into account, all these motions are obviously considered as correct ones. We thus first extract a reference model of the correct motion from the database of experts' movements (Section 3.2). We then compare the new motion (for instance performed by a novice) to this model to identify the spatial (Section 3.3.1) and/or temporal errors (Section 3.3.2).

Our goal is to evaluate the motion of a novice player in the context of individual sports. As a case study, we applied our method on tennis serves. These motions indeed present high spatial variabilities (contrary to codified motions such as kata in karate) and require a strong coordination between body parts, to achieve at the same time fast and accurate shots.

3.1 Database and gesture coding

To create the database, the tennis serves were captured with a Vicon MX-40 optical motion capture system (Oxford Metrics Inc., Oxford, UK). The players were equipped with 43 reflective markers placed on anatomical landmarks to compute the trajectories of the 25 joint centers as shown in Figure 2.

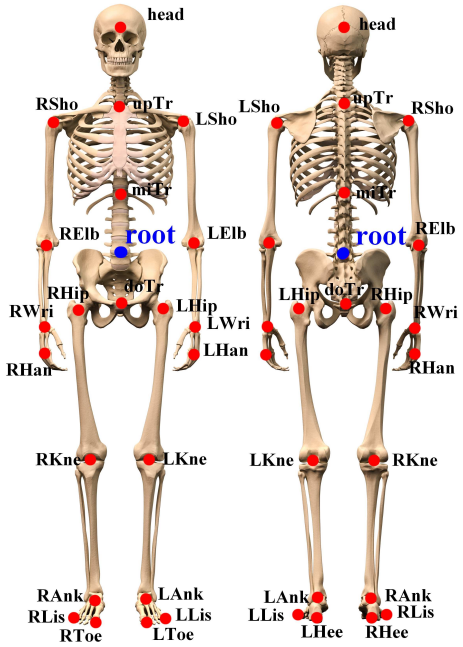

Figure 2: The captured motion is represented by the trajectories of these 25 joint centers.

To create the database, we captured the tennis serves of 9 experts (14-18 year old women) and 2 novice players. Each of them made 8 to 10 examples (or trials) leading to a database of 79 expert and 20 novice examples (see an example at different time steps in Figure 3).

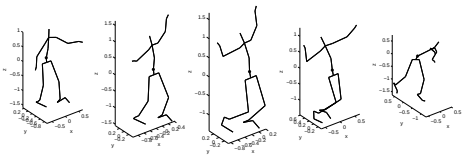

Figure 3: Skeletal representation of a captured tennis serve at different time steps.

In order to be invariant to the initial position and orientation of the subject, the coordinate system of each joint trajectories is centered on the root position (see Figure 2) and oriented according to the hips. Moreover, to decrease the influence of the morphology, each joint coordinate is divided by the distance between head and root joints as proposed by Sie et al. in (Sie et al., 2014).

Let us now consider the following notations:

- A: number of joints (25 here).
- M_j: number of time steps of the j^{th} example.

- N_E: number of expert examples (79) and N_N: number of novice examples (20).

- $\mathbf{X}_j(t) = \{\mathbf{x}_j^a(t), a = 1...A\}$, with $\mathbf{x}_j^a(t) = (x_j^a(t), y_j^a(t), z_j^a(t))$: trajectory of the a^{th} joint and the j^{th} example.
  Thus, $\mathbf{X}_j(t)$ is a 75-dimensional vector (25×3) that encodes, at time t, the body posture (position of all joints) while $\mathbf{x}_j^a(t), a = 1...A$ only encodes the position of joint a at time t for the j^{th} example ($3D$ vector).

3.2 Model of experts' motions

The model of experts' motions must at best represent all these motions with their variability to ensure that an expert motion is never considered as incorrect. One of the main problem to create such a model is that each motion may have different durations. Models such as HMM or HCRF can overcome this problem but do not consider the temporality between the limbs. They thus can consider as correct motions that are properly executed but badly synchronized. Another approach could be to use a nearest neighbor method but it becomes intractable when the number of examples in the database increases. To overcome these limitations and to ensure that our model perfectly represents all the experts' motions with their variability, we chose to model the serves with both the average motion and the spatial and temporal tolerances between it and each serve of all experts. To deal with the different motion durations, all examples are temporally aligned with the longest example using a Dynamic Time Warping algorithm (DTW). Let $\mathbf{X}_L(t)$ be this longest example. This temporal alignment simultaneously considers all joints to ensure that we model both the spatial features of the motion and the temporality between joints.

3.2.1 Average expert trajectory

To determine the average motion of all experts, we first made a global temporal alignment between each expert example $\mathbf{X}_j(t)$ and the longest example $\mathbf{X}_L(t)$ using DTW. To this end, we defined a distance matrix that contains the similarity values between $\mathbf{X}_L(t_1)$ and $\mathbf{X}_j(t_2)$, $\forall t_1 \in [0, M_L - 1]$ and $\forall t_2 \in [0, M_j - 1]$ where M_L and M_j are the durations of trajectory L and j respectively. These similarities are computed on both each joint trajectory and its derivative, as suggested in (Keogh and Pazzani, 2001):

$$
\begin{aligned}
d^1_{L,j}(t_1,t_2) &= \|\mathbf{X}_L(t_1) - \mathbf{X}_j(t_2)\|^2 \\
d^2_{L,j}(t_1,t_2) &= \|\dot{\mathbf{X}}_L(t_1) - \dot{\mathbf{X}}_j(t_2)\|^2 \\
d_{L,j}(t_1,t_2) &= \frac{d^1_{L,j}(t_1,t_2)}{\max\limits_{t_1,t_2} d^1_{L,j}(t_1,t_2)} + \frac{d^2_{L,j}(t_1,t_2)}{\max\limits_{t_1,t_2} d^2_{L,j}(t_1,t_2)}
\end{aligned}
$$

$\forall t_1 \in \{0...M_L - 1\}, \forall t_2 \in \{0...M_j - 1\}$.

The cumulative distance matrix $D_{L,j}$ is then computed from these similarities:

$$
\begin{aligned}
D_{L,j}(t_1,t_2) = d_{L,j}(t_1,t_2) + \\
\min(D_{L,j}(t_1,t_2-1), D_{L,j}(t_1-1,t_2), D_{L,j}(t_1-1,t_2-1)))
\end{aligned}
$$

with $D_{L,j}(0,0) = d_{L,j}(0,0)$, $D_{L,j}(t_1,0) = \sum_{t=0}^{t_1-1} d_{L,j}(t,0)$ and $D_{L,j}(0,t_2) = \sum_{t=0}^{t_2-1} d_{L,j}(0,t)$, $\forall t_1 \in \{1...M_L - 1\}, \forall t_2 \in \{1...M_j - 1\}$.

The distance between examples $\mathbf{X}_L(t)$ and $\mathbf{X}_j(t)$ is then defined by $D_{L,j}(M_L - 1, M_j - 1)$. The minimal path that goes from times $(0,0)$ to times $(M_L - 1, M_j - 1)$ of the two examples gives then their optimal alignment as shown in Figure 4.

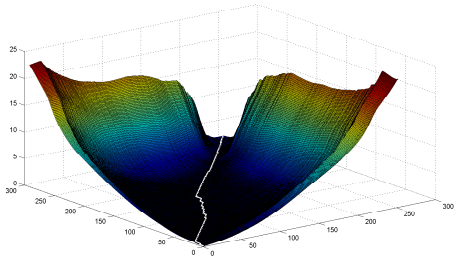

Figure 4: Cumulative distance matrix $D_{L,j}$ and the resulting minimal path that align at best the two motions (in white).

This alignment method of each example $\mathbf{X}_j(t)$ on the longest one $\mathbf{X}_L(t)$ is then applied to all examples to compute their optimal path. This path provides the $T_{L,j}(t)$ function that links each time t_1 of $\mathbf{X}_L(t)$ to a time $T_j(t_1)$ of $\mathbf{X}_j(t)$. Let us denote this path $P_{L,j} = \{(t_1, T_{L,j}(t_1)), t_1 = 1...M_L - 1\}$. Each example $\mathbf{X}_j(t)$ is then realigned according to the path $P_{L,j}$ to obtain a new motion $\tilde{\mathbf{X}}_j(t)$ with a duration of L time steps. The average motion can then be simply computed like this:

$$
\mathbf{X}_{mean}(t) = \frac{1}{N_E} \sum_{j=0}^{N_E-1} \tilde{\mathbf{X}}_j(t) \quad \forall t \in \{1...M_L - 1\}
$$

Let us recall that $\mathbf{X}_j(t)$ contains the $3D$ trajectories of all joints at time step t. The temporal alignment between trajectories is thus the same for the whole body, ensuring to maintain the temporal coherency between joints while being able to obtain the mean trajectory of all joints: $\mathbf{X}_{mean}(t) = \{\mathbf{x}^a_{mean}(t), a = 1...A\}$.

Based on this mean motion that best represents all expert examples, we now need to model the spatial and temporal tolerances (deviations) that enclose all the variability of experts' performances.

3.2.2 Spatial tolerance modeling

To better evaluate and model the spatial tolerance of the experts' motions, we must be independent of temporal errors, and do not allow bad synchronization to influence the computation of spatial errors. We thus consider each joint separately and align each of them $\mathbf{x}^a_j(t)$ to the mean joint trajectory $\mathbf{x}^a_{mean}(t)$. To this end, as described above, we compute the elements $d^a_{mean,j}(t_1,t_2)$ of the distance matrix between joints as well as the cumulative distance matrix elements $D^a_{mean,j}(t_1,t_2)$. A specific path $P^a_{mean,j}$ is then defined for each joint. It links each time t_1 of $\mathbf{x}^a_{mean}(t)$ to a time $T^a_{mean,j}(t_1)$ of $\mathbf{x}^a_j(t)$. Using these paths $P^a_{mean,j} = \{(t_1, T^a_{mean,j}(t_1)), t_1 = 1...M_L - 1\}$, new trajectories $\tilde{\mathbf{x}}^a_j(t)$ are obtained, that have the same duration (M_L) but do not correspond to the same temporal alignment. We can now compute the spatial tolerance, for each joint and at each time step:

$$\Sigma_S(t,a) = COV_{j \in experts}\{\tilde{\mathbf{x}}^a_j(t)\}$$

where COV is the covariance matrix, $\forall t \in \{0...M_L - 1\}, \forall a \in \{1...A\}$.

Each $\Sigma_S(t,a)$ is then a 3×3 matrix that represents the variations of position of the joint a, in the $3D$ coordinate system (x,y,z) and at time t, that are allowed around the mean $3D$ position to be still considered as a correct position (a position that experts can have). The spatial tolerance is illustrated for a specific time t and all the joints a in Figure 5.

3.2.3 Temporal tolerance modeling

If the joints of experts are perfectly synchronized, the alignments computed for each joint $P^a_{mean,j} = \{(t, T^a_{mean,j}(t)), t = 1...M_L - 1\}$ and for the whole body $P_{mean,j} = \{(t, T_{mean,j}(t)), t = 1...M_L - 1\}$ should be the same. In practice, this is obviously not true, because there is a variation in joints temporality as can be seen in Figure 6. The temporal error between the two paths must then be computed with the cumulative distance matrix of each joint:

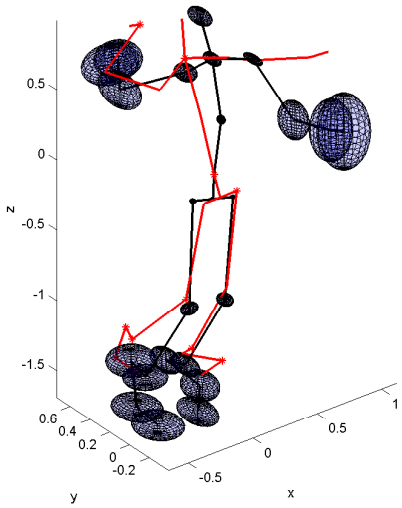

Figure 5: Spatial tolerance of the model of experts' motions. The black posture represents the average expert posture at time t and the black spheres are the spatial tolerance of all joints around this posture. The red posture is an example of novice posture.

$$E^a_j(t) = \frac{max\left(0, D^a_{mean,j}\left(t, T_{mean,j}(t)\right) - D^a_{mean,j}(t, T^a_{mean,j}(t))\right)}{M_j}$$
$$\forall t \in \{0...M_L - 1\}$$

$D^a_{mean,j}(M_L - 1, M_j - 1)$ is logically higher than $D^a_{mean,j}\left(M_L - 1, T^a_{mean,j}(M_j - 1)\right)$ as terms $T^a_{mean,j}(t)$ have been estimated from $D^a_{mean,j}$. However, it could happen that

$$D^a_{mean,j}(t, T_{mean,j}(t)) \leq D^a_{mean,j}\left(t, T^a_{mean,j}(t)\right)$$

for some $t \in \{0...M_L - 1\}$ leading to negative values. These rare cases are not representing real errors of the player so we consider that $E^a_j(t)$ is null to ensure that they have no influence on the global path.

The temporal tolerance is then defined, for each time and each joint, as the standard deviation of these errors:

$$\sigma_T(t,a) = STD_{j \in experts}\{E^a_j(t)\} \quad \forall t \in \{0...M_L - 1\}$$

where STD is the standard deviation.

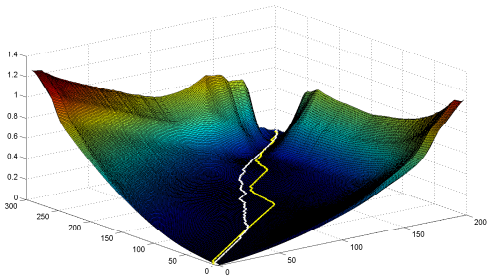

Figure 6: Temporal tolerance computation by path comparison on the cumulative distance matrix $D^{13}_{mean,j}$ of the RHan joint ($a = 13$) for an expert example. The global path $P_{mean,j}$ and the local one $P^{13}_{mean,j}$ are respectively shown in white and yellow.

3.3 Evaluation of novice's motion

Based on the model of experts' motions detailed above, the evaluation process consists in comparing the joints trajectories of the novice's motion with the average motion and its spatial and temporal tolerances.

3.3.1 Temporal errors

The temporal error between novice and expert motions can be global (delay over all the movement) or local (delay between a joint and an other). The local error is particularly interesting when evaluating motions since it can provide information about the problem of synchronization between limbs for instance. But to identify these local errors, a common temporal base is necessary: a global synchronization of the movement with the model of experts' motion. The first step is thus to determine this optimal global path and to observe the local temporal errors relatively to it. The difficulty is that each joint influences the global path and if a joint is very delayed from the others, the resulting global path would not be representative of the global motion. An iterative process is thus needed to evaluate the best joints to be considered. We propose to use an interative algorithm based on random sampling such as RANSAC. First, some joints are randomly selected and are used to compute the global path. Then, the other joints with nearly the same temporal alignment are added and the global path is re-estimated. This process is iterated N_{RANSAC} times and the best global path is kept to align the novice serve to the expert model. The same methodology as in Section 3.2.3 is then used to estimate $E^a_{temp}(t)$ that measures the temporal error, for each time and each joint. The whole process is described in Algorithm 1.

3.3.2 Spatial errors

Based on the global alignment described above, the spatial errors are computed with the Mahalanobis distances between the trajectories of the novice's motion $\tilde{\mathbf{x}}^a(t)$ and of the model of experts $\mathbf{x}^a_{mean}(t)$:

$$E^a_{spa}(t) = \sqrt{(\tilde{\mathbf{x}}^a(t) - \mathbf{x}^a_{mean}(t))^T \Sigma_S(t,a)(\tilde{\mathbf{x}}^a(t) - \mathbf{x}^a_{mean}(t))}$$

where $\Sigma_S(t,a)$ is a 3×3 matrix modeling the spatial tolerance as defined in Section 3.3.2.

Both temporal $E^a_{temp}(t)$ and spatial $E^a_{spa}(t)$ errors are thus computed for each time and each joint. These values inform us about when and how errors occur in the novice's motion compared to the reference, the database of experts' motions.

4 RESULTS

To validate our method, we made two preliminary experiments. The goal of the first one is to determine if our algorithm can automatically distinguish novices from experts. The second experiment quantifies the errors made by a novice player to observe when and how they occurred.

4.1 Automatic recognition of novices and experts

For this first experiment, the reference database is only composed of $N_E - 20$ experts examples and the test database is composed of the N_N novice examples and the 20 unused expert examples.

Temporal analysis

If the temporal sequence of novice joints is not consistent with expert ones (*i.e.* some joints are delayed), the temporal error $E^a_{temp}(t)$ presented in Section 3.3.1 must be higher. We thus compute a global temporal error as the sum of the local temporal errors, for each example and for all times and joints:

$$ERR_{temp} = \frac{1}{M_L A} \sum_{t=0}^{M_L-1} \sum_{a=1}^{A} E^a_{temp}(t)$$

The values of ERR_{temp}, computed for each trial, are represented by box plots in Figure 7 for both populations. As expected, the global temporal errors of novices are larger than of experts.

Spatial analysis

To quantify the spatial errors, we applied the same

Input: $\mathbf{X}(t) = \{\mathbf{x}^a(t), a = 1...A, t = 0...M-1\}$, $\mathbf{X}_{mean}(t)$, $\sigma_T(t,a)$
$n = 0$, $h_1 = 5$, $E_{save} = +\infty$, $N_{RANSAC} = 50$, $k_T = 3$, $\mathbf{S} = \emptyset$, $\mathbf{S}_{save} = \emptyset$
Output: Temporal error $E^a_{temp}(t)$
while $n < N_{RANSAC}$ **do**
  Randomly choose h_1 joints to get the set $\mathbf{S} = \{a_0...a_{h_1-1}\}$
  Compute the distance matrix between the average expert trajectory $\mathbf{X}_{mean}(t)$ and the new gesture $\mathbf{X}(t)$ using only the h_1 joint. Its elements are defined by :
  $$d^{\mathbf{S}}(t_1,t_2) = \frac{\sum_{a\in\mathbf{S}} \|\mathbf{x}^a(t_1) - \mathbf{x}^a_{mean}(t_2)\|^2}{\max_{t_1,t_2}(\sum_{a\in\mathbf{S}} \|\mathbf{x}^a(t_1) - \mathbf{x}^a_{mean}(t_2)\|^2)} + \frac{\sum_{a\in\mathbf{S}} \|\dot{\mathbf{x}}^a(t_1) - \dot{\mathbf{x}}^a_{mean}(t_2)\|^2}{\max_{t_1,t_2}(\sum_{a\in\mathbf{S}} \|\dot{\mathbf{x}}^a(t_1) - \dot{\mathbf{x}}^a_{mean}(t_2)\|^2)}$$
  Compute the cumulative distance matrix $D^{\mathbf{S}}$ using the distance matrix $d^{\mathbf{S}}$
  Compute the initialization of the global path $P^{\mathbf{S}} = \{(t, T^{\mathbf{S}}(t)), t = 1...M_L - 1\}$ that align $\mathbf{X}_{mean}(t)$ and $\mathbf{X}(t)$
  forall the $a \notin \mathbf{S}$ **do**
    Compute the local path between $\mathbf{x}^a_{mean}(t)$ and $\mathbf{x}^a(t)$
    Compute the error induced by the global path on the cumulative distance matrix of this joint
    $E^a(t) = \frac{max(0, D^a(t,T^{\mathbf{S}}(t)) - D^a(t,T^a(t)))}{M * \sigma_T(t,a)}$
    if $E^a(t) < k_T \ \forall t$ **then**
      $\mathbf{S} \leftarrow \mathbf{S} + a$
  Compute the total temporal error considering all the joints and all the times $E \leftarrow \frac{1}{|\mathbf{S}|M_L} \sum_{a\in\mathbf{S}} \sum_{t=0}^{M_L-1} E^a(t))$
  if $(|\mathbf{S}| > |\mathbf{S}_{save}|)$ or $(|\mathbf{S}| \geq |\mathbf{S}_{save}|$ and $E < E_{save})$ **then**
    $\mathbf{S}_{save} \leftarrow \mathbf{S}$
    $E_{save} \leftarrow E$
    $E^a_{temp}(t) \leftarrow E^a(t), \forall a \in \{1...A\}$ and $\forall t \in \{0...M_L - 1\}$
$n \leftarrow n + 1$

Algorithm 1: Algorithm of the temporal error estimation. $|\mathbf{S}|$ denotes the cardinal of the set $\mathbf{S}$.

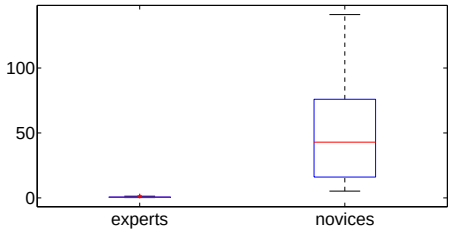

Figure 7: Temporal error distribution for novices and the experts that are not included in the reference database.

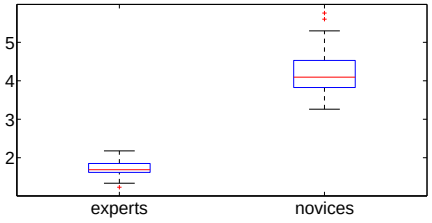

Figure 8: Spatial error distribution for novices and the experts that are not included in the reference database.

evaluation process. For each example, we computed the global spatial errors from local spatial errors:

$$ERR_{spa} = \frac{1}{M_L A} \sum_{t=0}^{M_L-1} \sum_{a=1}^{A} E^a_{spa}(t)$$

As shown in Figure 8, box plots of global spatial errors for the two populations are clearly separated, experts are distinguished from novices. Moreover errors are larger and more dispersed for novices than experts, highlighting the worst performance of novices and the higher variability of their motions.

Our approach can thus easily distinguish novices from expert players both with spatial and temporal errors.

4.2 Evaluation of a novice's motion

The goal of this second experiment is to apply our method on the motion of a novice player to determine his/her temporal and spatial errors over time. We have thus randomly selected one example of novice's tennis serve. To better illustrate the results, the average serve provided by the model of experts

detailed in Section 3.2 is sampled into 9 reference times $\{t_1...t_9\}$. The first row of Figure 9 shows the postures of this average motion at all these times (with a finer sampling at the end of the motion to have more details on this most dynamic part). The novice's serve is shown on the second row of Figure 9. Since this novice player made his serve faster, only two postures are illustrated but of course all the postures of the motions are considered for the following analyses. Finally, the third row illustrates how the novice's serve was globally aligned by Algorithm 1. The two motions are temporally coherent and the errors between them can be evaluated.

Temporal analysis

To qualitatively analyze the temporal results obtained on the novice's motion, the temporal error of the novice's right elbow is drawn in Figure 10. The temporal error is very low for all times except for t_8 where $E^a_{temp}(t_8) = 50.48$. At this time, the expert is hitting the ball while the novice is already ending his motion. This shows that the novice player moved his arm earlier than the expert relatively to his global movement. This relative temporal delay between the motions is obtained thanks to the global optimal alignment made by Algorithm 1.

Spatial analysis

The spatial error computed by our method for the novice's right elbow is drawn in Figure 11. The greater spatial error is obtained at the beginning of the motion. Figure 12 shows the position of the right elbow of the novice (in blue) and of the expert (in red) above the posture of the average motion of experts. The main error is indeed located at the beginning of the motion and is due to a bad technique of the novice: he did not lower enough the racket to exploit at best its displacement to have an optimal speed at ball impact.

Even if these results are preliminary, they highlight the strength of our method that can take both spatial and temporal errors into account to accurately identify the errors over time. Moreover, these errors are not only global information but are precise enough to point out local errors such as the synchronization between limbs or the spatial and temporal error of a joint relatively to the global motion. All these outcomes are moreover obtained independently of the length of the motions.

5 CONCLUSION

In this paper, we proposed an innovative approach to automatically evaluate sport motions independently to the type of sport or the morphology of the player. Preliminary results showed that our algorithm can correctly distinguish novice players from experts but even better it can quantify over time the temporal and spatial errors of the performance of a novice player compared to a database of experts. These results were achieved thanks to a 2-level DTW. Actually, a single DTW can only give information about the global error of the motion without considering the temporality between limbs for instance and without localizing the errors. Another solution could have been to manage each joint independently but no relationships between the joints could then be identified. Our solution overcomes these limits: both spatial and temporal features are considered concurrently and can then be used to propose an accurate training solution to work on that specific imperfections of the gesture and at the right time.

Our algorithm is based on a random-based selection process that could make it stochastic and then subject to variations. On the contrary, this process allows the detection of outliers, *i.e.* joints that are badly synchronized with others, in an efficient way. However, these preliminary results must be extended to a bigger population to have a statistical analysis.

This approach opens wide range of use cases. It can indeed be used to automatically compare the novice's motion of any individual sport to the database of expert without adding knowledge or editing/annotating the experts' motions. But it can also be used to compare a novice or injured player along time to evaluate his/her progression. This method could thus be the core of a generic and automatic training system to be used complementary to traditional training sessions.

ACKNOWLEDGEMENTS

The data used in this project was obtained from a tennis project carried out in the M2S laboratory. The authors thank Caroline Martin and Pierre Touzard for this supply.

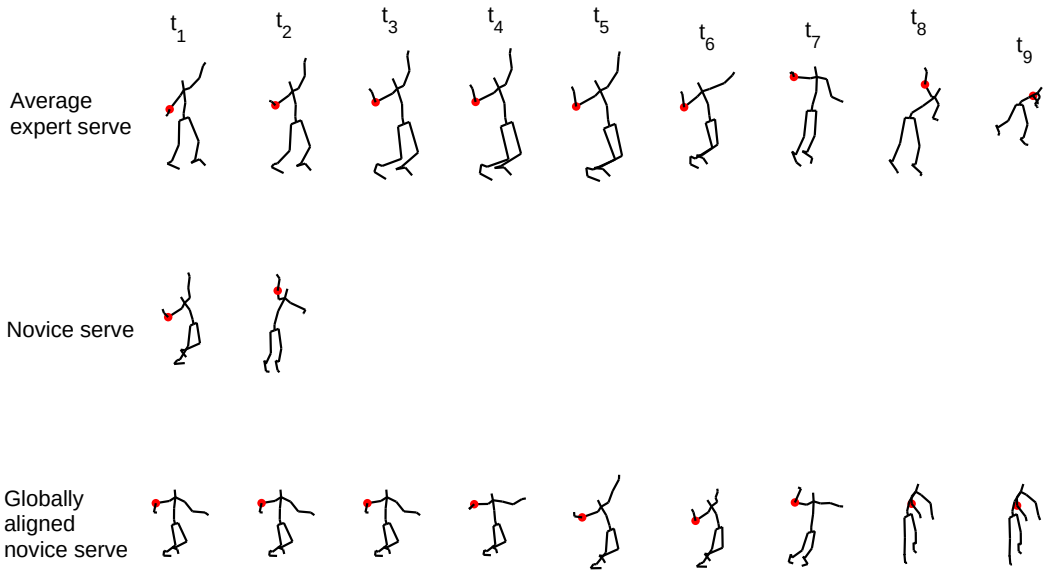

Figure 9: First row: average motion provided by our model of experts' motion, sampled into 9 times for illustration purpose. Second row: original novice's motion at the same times. Third row: novice's serve after alignment with the path P^S_{mean} of Algorithm 1. Red dots correspond to the right elbow of the player.

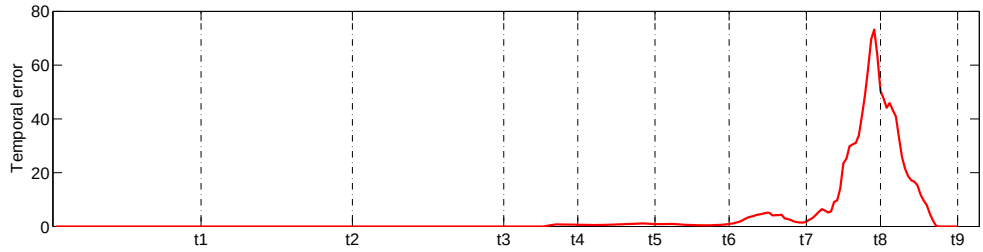

Figure 10: Temporal error of the novice's right elbow over time. The 9 reference times illustrated in Figure 9 are represented by vertical dashed lines.

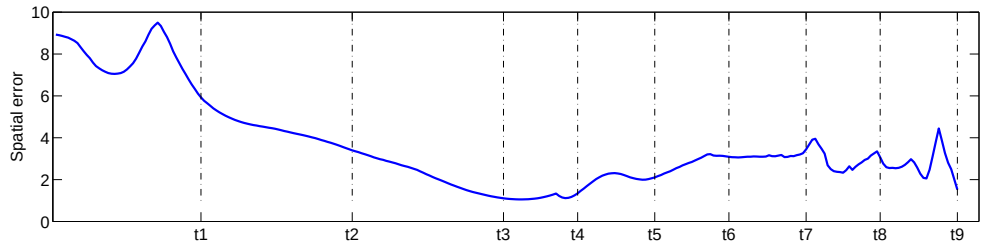

Figure 11: Spatial error of the novice's right elbow over time. The 9 reference times illustrated in Figure 9 are represented by vertical dashed lines.

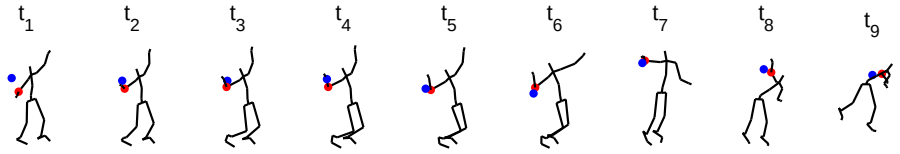

Figure 12: Position of the right elbow of the experts' average motion (red) and of the novice's one (blue) after local alignment P_{mean}^{RElb}.

REFERENCES

Barnachon, M., Bouakaz, S., Boufama, B., and Guillou, E. (2013). A Real-Time System for Motion Retrieval and Interpretation. *Pattern Recognition Letters*, 34(15):1789–1798.

Burns, A.-M., Kulpa, R., Durny, A., Spanlang, B., Slater, M., and Multon, F. (2011). Using virtual humans and computer animations to learn complex motor skills: a case study in karate. *BIO Web of Conferences*, 1(12).

Gong, D., Medioni, G., and Zhao, X. (2014). Structured time series analysis for human action segmentation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1414–1427.

Heloir, A., Courty, N., Gibet, S., and Multon, F. (2006). Temporal alignment of communicative gesture sequences. *Computer Animation and Virtual Worlds*, 17(3-4):347–357.

Kahol, K., Tripathi, P., and Panchanathan, S. (2004). Computational analysis of mannerism gestures. In *ICPR*, pages 946–949.

Keogh, E. J. and Pazzani, M. J. (2001). Derivative dynamic time warping. In *In Proceedings of the First SIAM International Conference on Data Mining*.

Komura, T., Lam, B., Lau, R. W. H., and Leung, H. (2006). e-learning martial arts. In *Advances in Web Based Learning – ICWL*, volume 4181, pages 239–248. Springer-Verlag, Berlin, Heidelberg.

Kulpa, R., Multon, F., and Arnaldi, B. (2005). Morphology-independent representation of motions for interactive human-like animation. *Computer Graphics Forum*, 24(3):343–352.

Maes, P.-J., Amelynck, D., and Leman, M. (2012). Dance-the-music: an educational platform for the modeling, recognition and audiovisual monitoring of dance steps using spatiotemporal motion templates. *EURASIP Journal on Advances in Signal Processing*, 2012(1):35.

Müller, M. and Röder, T. (2006). Motion templates for automatic classification and retrieval of motion capture data. *Proceedings of the 2006 ACM SIGGRAPH*, pages 137–146.

Müller, M., Röder, T., and Clausen, M. (2005). Efficient content-based retrieval of motion capture data. *ACM Transactions on Graphics*, 24(3):677.

Ofli, F., Chaudhry, R., Kurillo, G., Vidal, R., and Bajcsy, R. (2012). Sequence of the most informative joints (SMIJ): A new representation for human skeletal action recognition. In *CVPRW*, pages 8–13.

Pazhoumand-Dar, H., Lam, C.-P., and Masek, M. (2015). Joint movement similarities for robust 3d action recognition using skeletal data. *Journal of Visual Communication and Image Representation*, 30:10–21.

Pham, M. T., Moreau, R., and Boulanger, P. (2010). Three-dimensional gesture comparison using curvature analysis of position and orientation. In *EMB*, pages 6345–6348.

Röder, T. (2006). *Similarity, Retrieval, and Classification of Motion Capture Data*. PhD thesis, Rheinischen Friedrich- Wilhelms- Universität, Bonn.

Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49.

Sakurai, K., Choi, W., Li, L., and Hachimura, K. (2014). Retrieval of similar behavior data using kinect data. In *ICCAS*, pages 97–104.

Shin, H. J., Lee, J., Shin, S. Y., and Gleicher, M. (2001). Computer puppetry: An importance-based approach. *ACM Transactions on Graphics*, 20(2):67–94.

Sie, M.-S., Cheng, Y.-C., and Chiang, C.-C. (2014). Key motion spotting in continuous motion sequences using motion sensing devices. In *ICSPCC*, pages 326–331.

Sorel, A., Kulpa, R., Badier, E., and Multon, F. (2013). Dealing with variability when recognizing user's performance in natural 3d gesture interfaces. *International Journal of Pattern Recognition and Artificial Intelligence*, 27(8):19.

Wang, J., Liu, Z. L., Wu, Y. W., and Yuan, J. (2012). Mining actionlet ensemble for action recognition with depth cameras. In *CVPR*, pages 1290–1297.

Wang, S. B., Quattoni, A., Morency, L.-P., and Demirdjian, D. (2006). Hidden conditional random fields for gesture recognition. In *CVPR*, volume 2, pages 1521–1527.

Ward, R. E. (2012). *Biomechanical Perspectives on Classical Ballet Technique and Implications for Teaching Practice*. PhD thesis, University of New South Wales, Sydney, Australia.

Xia, L., Chen, C.-C., and Aggarwal, J. K. (2012). View invariant human action recognition using histograms of 3d joints. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 20–27.

Zhong, S. and Ghosh, J. (2002). HMMs and coupled HMMs for multi-channel EEG classification. In *Proceedings of the 2002 International Joint Conference on Neural Networks*, pages 1154–1159.

Zhou, F. and de la Torre, F. (2015). Generalized canonical time warping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1.

Zhou, F. and De La Torre Frade, F. (2009). Canonical time warping for alignment of human behavior. In *NIPS*, pages 2286–2294.