Automatic and Generic Assessment of the Quality of Sport Motions

Marion Morel, Catherine Achard, Richard Kulpa, Séverine Dubuisson
Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 7222, ISIR, F-75005, Paris, France

I. My research Vision

Motion analysis is a wide domain that contains widespread fields as motion retrieval, motion recognition or motion synthesis. My topic focuses on a more unusual subject: motion evaluation. The goal of my Ph.D. work is to provide a tool that measures the quality of a sport gesture with the purpose of individual coaching help. As a third-year Ph.D. student, I would be greatly interested in getting feedback of my work from gesture analysis experts, who would help me orient my work in the most appropriate way for the last year of my Ph.D.
To be as general as possible, our main constraint is to have no prior knowledge concerning the performed motion to be evaluated: we aim at learning it from a set of experts' motions. This is not the case for the majority of works dealing with sport motions evaluation that add prior knowledge to the motion to analyze [12], [3]. Even if their results are very interesting, they cannot be adapted to other types of motion whose main features are not identified. A similar challenge as ours can also be found in the field of surgery, where the goal is also to perform a perfect gesture to improve a surgery task. However, in these cases, the approaches are based on the single trajectory of the tool used by the surgeon [2], [7]. On the opposite, our challenge makes the entire body of the subject informative. A motion is composed of multiple limb movements that are themselves composed of joints displacements. Therefore, we need to deal with this high dimensionality and with temporal inter-dependencies between limbs.

Another issue concerns the variability in the way a same gesture can be executed because of viewpoint changes, morphologies of subjects or also global speed. These must be handled correctly so that the evaluation score does not depend on it. Viewpoint changes are easily handled by aligning the coordinate system on a local reference system linked to the human. Invariant features can be used to deal with morphology changes [4], normalization are also proposed in the literature [10]. The execution speed actually contains relevant information. The relative timing between limbs is very important to assess a gesture quality. To manage motions having different lengths, speeds and/or different rhythms, some authors used Hidden Markov Models (HMM) or Hidden Conditional Random Field (HCRF) [13], [11] in which states represent postures of motion and transitions between them are defined by probabilities. This model, well-known in temporal pattern recognition, is thus robust to temporal variations. However, several time steps are associated with the same state and the temporality is only managed between these states. The evaluation is then not precise enough. To overcome this issue, we have proposed to integrate the well-established Dynamic Time Warping (DTW) algorithm [8] that is generally used to align and compare motions [6], [9], [1], but not to perform a fine understanding of motion, to identify its failures and then to propose a way to correct them. My work proposes an approach that strongly differs from the previous ones in the sense that

This work was supported by ENS Paris-Saclay.

it is only based on a set of expert gestures. The system learns from this set what is important and has to be checked, for each time of the gesture and each limb.

II. Brief summary of the work

As we just said, our goal is to assess the quality of any sport motion given a set of experts' motions. This set is first used to learn a representative expert motion called "nominal motion".

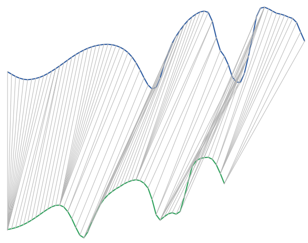
A. Model of experts' motion

1) Nominal motion: as the execution speeds can differ depending on the expert, the first step for nominal motion computing consists in aligning all experts' motions. This is done with DTW that relies on the computation of a distance map $\mathbf{d}$ and a cumulative distance map $\mathbf{D}$ from which a warping path is extracted. This warping path can be seen as a look-up table between the time-steps of signals to be aligned. Figure 1 depicts an alignment between two signals.

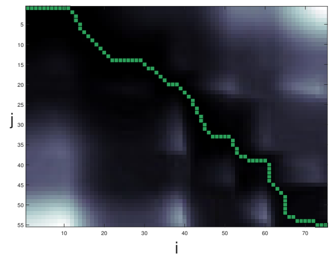
Since DTW only handles the alignment of pairs of signals, it also needs to be updated to manage alignments of more than two signals in view of averaging several time-series. This is the point of the well-established DTW Barycenter Averaging (DBA) [5]. Now that we have developed a tool to average time-series, we generalized it to the averaging of motions.

We denote $\mathbf{X}_i(t) = \{\mathbf{x}_i^j, j = 1...J\}$ the motion of the i^{th} expert composed of J joints. The motion resulting from the DBA averaging method of all experts' motions $\{\mathbf{X}_i(t)\}_{i \in experts}$ is the nominal motion, noted $\mathbf{X}_n(t)$.

2) Spatial tolerance: now that the nominal motion is defined, it remains to consider how much it can be transgressed. Note for example that the position of the left wrist during a right punch is not as significant as the position of the right wrist. This aspect justifies the use of a spatial tolerance at each time-step and for each joint that reflects the significance of a joint's position. It is computed as the covariance matrix of the positioning of the joint for the whole experts' motions at each time-step once the motions are aligned.



(a)



(b)

Fig. 1: Illustration of a DTW alignment. (a) The matching between frames of two signals (in blue and green) is symbolized by the grey lines. (b) Superimposition (in green) of the warping path on the cumulative distance matrix $\mathbf{D}$. In this map, largest cumulative distances $D_{i,j}$ are in white whereas smallest are in black.

Figure 2 presents the tolerance of the right hip and the left arm for a tennis serve at a particular time-step.

B. Evaluation of a novice's motion

1) Spatial Error: to improve the effectiveness of the feedback, the spatial error is computed for each limb and not each joint. A limb is defined as a combination of multiple joints. Given the warping path of the novice on the nominal motion restricted to the l^{th} limb and Σ^l_{Spa} the spatial tolerance of the l^{th} limb, the spatial error of limb l is given by the Mahalanobis distance between the nominal motion and the novice's motion restricted to limb l and aligned consequently. From this spatial error, computed for each time and each limb, we can develop an automatic spatial error detector as the one presented in Figure 3 for

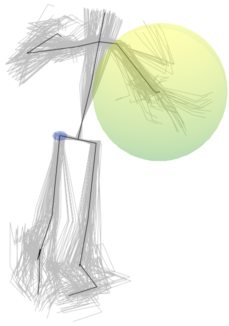


Fig. 2: Spatial tolerance of the left wrist (in yellow-green) and the right hip (in blue). The nominal motion is depicted (in black) superimposed with all the aligned experts' motions (in gray).

the right arm.

2) Temporal Error: we define the temporal error as the asynchrony between limb during the motion. Broadly speaking, let's consider a motion performed with a good timing between limbs l_1 and l_2. Then the alignment of the novice's motion on the nominal one restricted to limb l_1 will be very similar to the one restricted to l_2. On the opposite, if l_2 is delayed then the warping path will be highly different. Thus, the discrepancy between the warping paths of motions restricted to limbs l_1 and l_2 reflects the temporal error. This aspect is used to measure the temporal error.

3) Results: this process was tested on an evaluation task. We ask coaches to provide evaluation score of tennis serve that are compared with our evaluations score. Good evaluations are obtained for spatial errors (temporal errors were not annotated by experts as in tennis serve, both arms are necessarily synchronized in order to hit the ball). Let us now proceed to the limitations and to the perspectives part.

III. Future plans and challenges

A. Temporal Error

We believe that the evaluation of temporal errors is not totally well-handled. Actually, the path estimated by the DTW algorithm is not well defined during static moments of gesture. This can induce some errors that should not be. There are certainly other ways to study gestures' synchrony and this Doctoral Consortium could be a great help to discuss with experts the pros and cons of different solutions and viewpoints.

B. Styles

Another perspective focuses on gesture styles. Indeed, our model of gesture is based on a nominal motion and allowed variations around it but does not managed difference of styles between experts. So, a more sophisticated model will probably lead to more accurate results and to additional information about the style performed. This is a challenging task that necessitates an automatic and unsupervised classification of the experts' motions of the database.

C. Real-time evaluation tool with adaptive feedback

Future works should also focus on the transfer of this work on a low cost device, for example the Kinect, to make it more accessible and user-friendly. Furthermore, the feedback to give to the subject should be determined according to the cognitive learning process. An interactive interface should thus be developed and tested on different populations.

D. Post-doctoral position

In a longer term perspective, I would like to continue my research on sport gestures' synchrony. I believe there remains much to be done and I would like to find a post-doctoral position on this topic. This Doctoral Consortium could also help me meet people and discover research teams working on this field and eventually apply for a post-doctoral position.

References

[1] B. A. Boulbaba, J. Su, and S. Anuj. Action recognition using rate-invariant analysis of skeletal shape trajectories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–14, 2015.

[2] E. F. Hofstad, C. Vapenstad, M. K. Chmarra, T. Lango, E. Kuhry, and R. Marvik. A study of psychomotor skills in minimally invasive surgery: what differentiates expert and nonexpert performance. *Surgical Endoscopy*, 27(3):854–863, 2013.

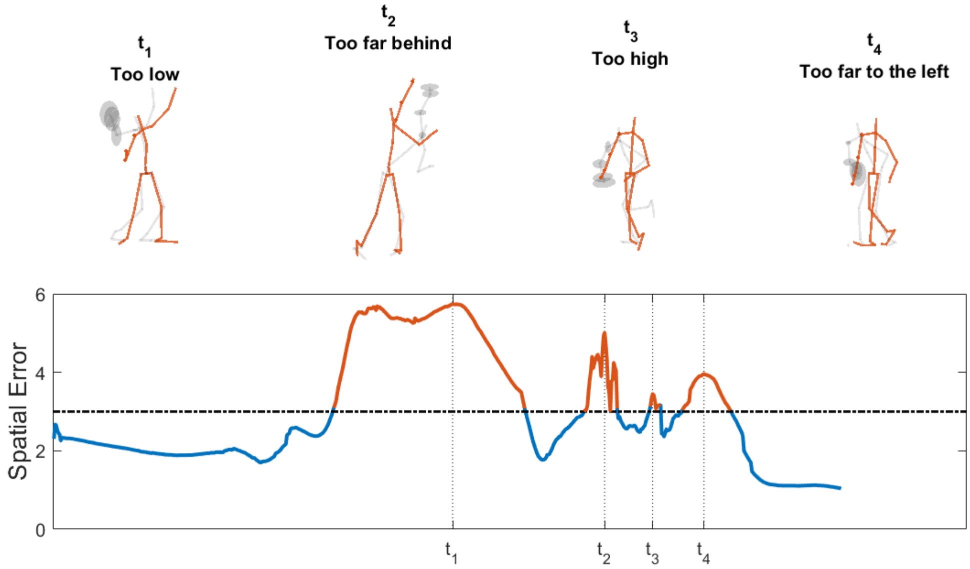


Fig. 3: Automatic spatial error detection. First row: red and gray skeletons depict respectively the novice's motion and the nominal one. Gray ellipsoids correspond to the spatial tolerance of each joint of the right arm. Second row: spatial error of the novice's right arm over time. The four reference times illustrated on the top are represented by vertical dashed lines in the bottom graph.

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

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

[5] F. Petitjean, A. Ketterlin, and P. Gancarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3):678–693, Mar. 2011.

[6] M. T. Pham, R. Moreau, and P. Boulanger. Three-dimensional gesture comparison using curvature analysis of position and orientation. In *EMBC'10*, pages 6345–6348. IEEE, 2010.

[7] C. E. Reiley and G. D. Hager. Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, editors, *Medical image computing and computer-assisted intervention*, volume 5761, pages 435–442. Springer Berlin Heidelberg, 2009.

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

[9] K. Sakurai, W. Choi, L. Li, and K. Hachimura. Retrieval of similar behavior data using kinect data. In *14th International Conference on Control, Automation and Systems (ICCAS)*, pages 1368–1372. IEEE, 2014.

[10] M.-S. Sie, Y.-C. Cheng, and C.-C. Chiang. Key motion spotting in continuous motion sequences using motion sensing devices. In *IEEE International Conference on Signal Processing*, pages 326–331. IEEE, 2004.

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

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

[13] S. Zhong and J. Ghosh. Hmms and coupled hmms for multi-channel eeg classification. In *Proceedings of the 2002 International Joint Conference on Neural Networks*, pages 1154–1159. IEEE, 2002.