

A Similarity Indicator for Differentiating Kinematic Performance Between Qualified Tennis Players

J.D. Pulgarin-Giraldo¹(✉), A.M. Alvarez-Meza², L.G. Melo-Betancourt³,
S. Ramos-Bermudez³, and G. Castellanos-Dominguez¹

¹ Signal Processing and Recognition Group, Universidad Nacional de Colombia,
Manizales, Colombia
jdpulgarin@unal.edu.co

² Faculty of Engineering, Universidad Tecnológica de Pereira, Pereira, Colombia

³ Physical Activity and Sport Group, Universidad de Caldas, Manizales, Colombia

Abstract. This paper presents a data-driven approach to estimate the kinematic performance of tennis players, using kernels to extract a dynamic model of each player from motion capture (MoCap) data. Thus, a metric is introduced in the Reproducing Kernel Hilbert Space in order to compare the similarity between models so that the built kernel enhances groups separability: the baseline reference group and the group including players developing their skills. Validation is carried out on a specially constructed database that contains two main testing actions: serve and forehand strokes (carried out on a tennis court). Besides, the classical kinematic analysis is used to compare our kernel-based approach. Results show that our approach allows better representing the performance for each player regarding the ideal group.

Keywords: Multi-channel data · Kernel methods · Kinematics · QKLMS · Similarity indicator

1 Introduction

Human action recognition from video data is a well-established area in computer vision [1]. To date, the main efforts are directed at creating a sufficiently robust dynamic model of human movement accomplished under a priori given actions. All of these models have been validated in predicting the movement and/or action recognition through a certain indicator of performance [2]. However, the models are mostly oriented to detect discrepancies in the performance of players rather than to observe their progress or development of actions.

Aiming at modeling, playing dynamics are usually constructed based on Hidden Markov Models or Gaussian Process from high-dimensional MoCap markers [3, 4]. Although these models can provide adequate performance, the involved actions are assumed to be produced in the form of smooth, slow, and cyclical

movements. In the case of movements having quick actions or responses, development and tuning of either approach may face difficulties, leading to low rate classification. Furthermore, these methods assign high weights to the body segments that involve more information about the posture and gait as the lower limbs, trunk, and pelvis. However, the actions that involve faster movements and smaller body segments are dismissed, such as the upper limbs. To overcome this issue, the kernel methods appear as an option that enables nonlinear dynamics modeling in Hilbert space, and hence, that can enhance the classification and prediction performance.

On the other hand, the efficiency rate may be diminished if the movement analysis is carried out just on a movement that involves more speed or much higher dynamic response and upper limbs [2]. Therefore, the performance evaluation of tennis players should also incorporate the active participation of upper limbs and high-speed response, during the service and forehand strokes. Another aspect to consider is the technique mastered by each player so that its biomechanics plays a key role in stroke production [5]. Each stroke movement has its fundamental mechanical structure, making necessary to analyze the practicing and training player procedures individually. However, the biomechanical analysis is quite extensive and strongly depends on two types of variables: kinetics and kinematics [6]. In this regard, the use of motion capture (MoCap) makes the kinematic analysis lower in cost than kinetic analysis [1].

In practice, the kinematic analysis is of interest to the early analysis of player skills in sports training. Although the kinetic study looks like a more suitable approach due to the calculation of force and torque in specific biomechanical gestures, its involves mass calculation of every single body segment, resulting in a very time consuming and expensive procedure when only MoCap data are employed. As a result, just the professional athletes have access to these tools because of their price. On the other hand, the kinematic analysis is mainly performed using information of linear and angular velocity, as well as the joints and angles of alignment as the initial input [7]. Yet, the majority of approaches do not take into consideration the parameter drift that can describe the progressive development of player performance due to its training improvement. Therefore, there is growing demand for a kinematic analysis at affordable implementation cost but with more flexibility to capture the continuous progress of players.

In this paper, we introduce a data-driven approach to estimate the kinematic performance devoted to training of tennis players that is based on kernels to extract a dynamic model of each player from motion capture (MoCap) data. This methodology allows to see each player performance against other players, his progress, and as a group their similarity in motions executed. This level of interpretation is not achieved with classical kinematics variables studied on tennis training.

2 Kernel Based Multi-channel Data Representation

Let $\mathbf{X} \in \mathbb{R}^{N \times M}$ be a multi-channel input matrix that holds M channels and N samples, where $\mathbf{x}_i \in \mathbb{R}^{1 \times M}$ is a row vector containing the information of all the

provided channels at different time instants, with $i \in \{1, \dots, N\}$. With the aim at modeling properly each player from MoCap data, we represent the dynamic behavior of its movements through the kernel methods using generalized similarities. To this end, a kernel function is employed as a nonlinear mapping $\varphi: \mathbb{R}^{N \times M} \mapsto \mathcal{H}$, where \mathcal{H} is a Reproducing Kernel Hilbert Space - RKHS. Thus, the kernel based representation allows dealing with nonlinear structures that can not be directly estimated by traditional operators, such as, the linear correlation function. Regarding this, the inner product between two samples $(\mathbf{x}_i, \mathbf{x}_j)$ is computed in RKHS as $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle_2$, being $\kappa(\cdot, \cdot)$ a Mercer's kernel. Taking advantage of the so-called kernel *trick*, the kernel function can be computed directly from \mathbf{X} . Here, we rely on the well-known Gaussian kernel defined as follows:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 / 2\sigma^2\right), \quad (1)$$

where $\sigma \in \mathbb{R}^+$ is the kernel band-width. Notation $\|\cdot\|_2$ is the L_2 Euclidean norm.

Therefore, we obtain the similarity matrix $\mathbf{S} \in \mathbb{R}^{N \times N}$ that holds elements $s_{i,j} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$ with $s_{i,j} \in \mathbb{R}^+$. In this sense, matrix \mathbf{S} encodes the temporal dynamics of the multi-channel input data.

Due to the Representer Theorem, the nonlinearities of wide range problems can be described as a kernel expansion in terms of the data [8]. That is,

$$f(\boldsymbol{\xi}) = \sum_{i \in N} \alpha_i \kappa(\mathbf{x}_i, \boldsymbol{\xi}), \quad \boldsymbol{\xi} \in \mathbf{X} \quad (2)$$

where N is the number of samples, $\alpha_i \in \mathbb{R}^+$ is built on Kernel least-mean-squares algorithms (kernel least mean square - KLMS) like the one based on solving a least-square problem (KRLS) or the ones using stochastic gradient descent methods [9]. In particular, we employ the Quantized KLMS method (QKLMS) that constructs a dictionary set or *codebook*, noted as C , according to a quantization process in which data points are mapped onto the closest dictionary point. By using QKLMS, the learned mapping before iteration k is as follows:

$$f_{k-1}(\boldsymbol{\xi}) = \sum_{i=1}^{\text{size}(C_{k-1})} \alpha_{i,k-1} \kappa(C_{i,k-1}, \boldsymbol{\xi}) \quad (3)$$

where $C_{i,k-1}$ denotes the i -th element (code-vector) that belongs to codebook C_{k-1} and $\alpha_{i,k-1}$ is the coefficient of i -th center.

3 Performance Comparison Expanded to RKHS Representation

The main rationale behind validation is to define a real-valued distance $d() \in \mathbb{R}^+$ that accomplishes the pairwise comparison of the models estimated for the tested players. Taking into account Eq. 2, we compute the distance between a couple of models, f_n and f_m , as below:

$$d_f(f_n, f_m) = \varphi(f_n(\mathbf{X}^n), f_m(\mathbf{X}^m)). \quad (4)$$

where \mathbf{X}^n and \mathbf{X}^m are the multi-channel input matrices for subjects n and m , respectively. Here, we use the following RKHS-related norm:

$$d_f(f_n, f_m) = \|f_n - f_m\|_2, \quad (5)$$

where $f_n \in \mathcal{H} \{ \varphi_n(\cdot); \{ \alpha_i^n : \forall i \in \text{size}(C^n) \} \}$. Consequently, the evaluation of the euclidean distance proposed in Eq. 5 is as below:

$$d_f(f_n, f_m) = \langle f_n, f_n \rangle_2 - 2 \langle f_n, f_m \rangle_2 + \langle f_m, f_m \rangle_2. \quad (6)$$

Because of the codebook in Eq. 3, we have $i \in \text{size}(C^n)$ and $j \in \text{size}(C^m)$ so that we can rewrite the distance (5) as follows:

$$d_f(f_n, f_m) = \left\langle \sum_i \alpha_i^n \kappa(\mathbf{x}_i^n, \cdot), \sum_i \alpha_i^n \kappa(\mathbf{x}_i^n, \cdot) \right\rangle_2 - 2 \left\langle \sum_i \alpha_i^n \kappa(\mathbf{x}_i^n, \cdot), \sum_j \alpha_j^m \kappa(\mathbf{x}_j^m, \cdot) \right\rangle_2 + \left\langle \sum_j \alpha_j^m \kappa(\mathbf{x}_j^m, \cdot), \sum_j \alpha_j^m \kappa(\mathbf{x}_j^m, \cdot) \right\rangle_2, \text{ with } \mathbf{x}_i^n \in \mathbf{X}^n, \mathbf{x}_j^m \in \mathbf{X}^m.$$

Lastly, we obtain the following closed expression for the Euclidean-distance based indicator:

$$d_f(f_n, f_m) = \sum_{i,i'} \alpha_i^n \alpha_{i'}^n \kappa(\mathbf{x}_i^n, \mathbf{x}_{i'}^n) - 2 \sum_{i,j} \alpha_i^n \alpha_j^m \kappa(\mathbf{x}_i^n, \mathbf{x}_j^m) + \sum_{j,j'} \alpha_j^m \alpha_{j'}^m \kappa(\mathbf{x}_j^m, \mathbf{x}_{j'}^m). \quad (7)$$

Thus, the obtained indicator allows encoding the pairwise relationships between all considered models as a functional measure in a RKHS representation.

4 Experimental Setup and Results

4.1 Database

The data were collected from 16 players: four labeled as high performance (HP) and 12 more as regular (RP). All participants belonged to the Caldas-Colombia tennis league and explicitly volunteered to participate in the study, approved by the Ethics Committee of the *Universidad de Caldas*. Each group had the following anthropometric parameters (HP vs RP): age 21 ± 2.7 vs. 16.7 ± 3.9 years, mass 64 ± 14.9 vs. 61.6 ± 9 kg, and height 168.8 ± 8.4 vs. 170.8 ± 7.7 cm. The employed motion capture protocol was Biovision Hierarchy (BVH), stipulating the placement of 34 markers for collecting information of body joints. Optitrack Flex V100 (100 Hz) infrared videography was collected from six cameras to acquire sagittal, frontal, and lateral planes. All subjects were encouraged to hit the ball with the same velocity and action just as they would in a match. They were instructed to hit one series of five serve strokes followed by 12 forehand strokes.

4.2 Baseline Kinematic Analysis

For the purpose of deriving representative and accurate kinematics of the video recorded hits, a total of four serves and six forehand strokes per subject were manually selected and considered for kinematic analysis. The kinematics variables of interest for the serve and forehand were the following: (a) maximum angular displacement of shoulder alignment, hip alignment and wrist extension, (b) maximum velocity of hip, shoulder, elbow, and wrist, lastly (c) maximum angular velocity of elbow extension, trunk rotation, and pelvis rotation [7].

As seen in Table 1, analysis of variance (high performance, regular) estimated for each swing allows detecting some statistical differences and effects in the selected kinematic variables. Thus, the used one-way ANOVA shows that no group differences are achieved by the kinematics variables for the serve hit, assuming a significance level $p < 0.05$. Instead, the hip alignment ($p < 0.01$) and elbow velocity ($p < 0.03$) are significantly different between high performance and regular players in forehand stroke; all these findings are consistent with some related works on this area [5]. However, these selected variables do not clearly distinguish between groups. Indeed, one can not identify any cluster relating

Table 1. Tennis serve and forehand stroke kinematics (mean \pm SD) of high performance and regular players. [†] High performance group tends to be different from regular: hip alignment ($p < 0.01$) and elbow velocity ($p < 0.03$)

Variable	Serve stroke		Forehand stroke	
	High performance	Regular	High performance	Regular
Max. shoulder alignment angle (°)	29.57 (04.01)	25.36 (9.12)	27.43 (8.36)	21.78 (5.55)
Max. hip alignment angle (°)	7.09 (3.96)	4.87 (6.60)	9.97 (2.82) [†]	4.73 (3.69)
Max. wrist extension angle (°)	78.10 (11.03)	85.12 (5.68)	73.48 (10.68)	68.56 (13.76)
Max. hip linear velocity (m/seg)	0.95 (0.71)	0.63 (0.28)	0.46 (0.13)	0.52 (0.15)
Max. shoulder linear velocity (m/seg)	1.89 (0.64)	2.16 (1.68)	1.16 (0.22)	0.93 (0.34)
Max. elbow linear velocity (m/seg)	3.08 (0.66)	2.89 (1.06)	2.07 (0.50) [†]	1.56 (0.37)
Max. wrist linear velocity (m/seg)	4.13 (1.43)	3.37 (1.04)	3.66 (0.96)	2.94 (0.89)
Max. elbow extension velocity (°/seg)	89.67 (0.29)	89.69 (0.33)	88.76 (1.17)	89.61 (0.78)
Max. trunk angular velocity (°/seg)	392.29 (87.58)	425.77 (260.83)	335.98 (67.78)	256.07 (112.05)
Max. pelvis angular velocity (°/seg)	107.49 (26.24)	98.32 (38.36)	90.08 (22.91)	116.67 (49.90)

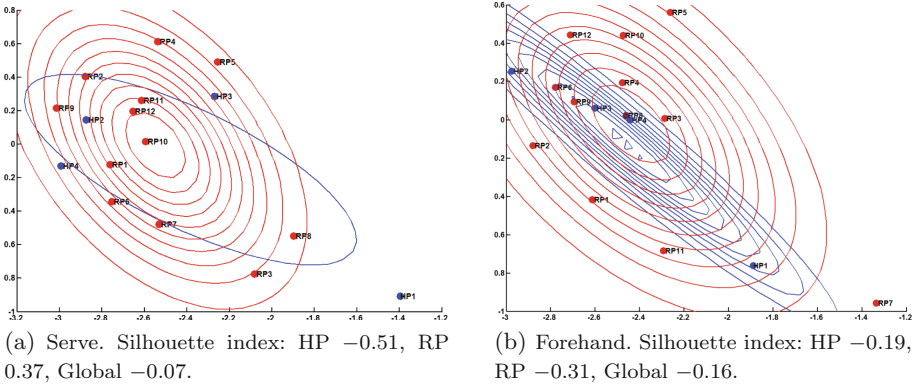


Fig. 1. 2-D embedding of the 16 players using kinematics characteristics. HP means high performance (blue color) and RP means regular performance (red color). Gaussian densities of each group are observed. (Color figure online)

to the considered groups of players when embedding these ten variables into a 2-D representation using Kernel PCA as shown in Fig. 1(a) and (b). In fact, the comparison between players using this characteristic does not allow to see a compact group of high-performance players which regular players can come closer in performance. Silhouette index is used to measure the cohesion and separation between groups [10]. A silhouette group index near to 1 indicates that the group is well clustered. On the other hand, group index near to -1 is not well clustered. Index close to 0 indicates that a point is on the border of two natural clusters.

4.3 Kernel-Based Performance Analysis

We compute each model $f_n(\cdot)$ as a pairwise kernel-based similarity measure between two samples \mathbf{x}_i and \mathbf{x}_j . The Gaussian kernel defined in Eq. 1 estimates each pairwise sample relationship, mapping from the original feature space to an RKHS representation. The kernel bandwidth σ is calculated by maximization of information potential variability, which provides better results in this kind of data than the Silverman’s rule. QKLMS parameters are set as follows: quantization size $\epsilon_U = 0.95$, step size $\eta = 0.81$, window size $\omega_s = 30$ and the initial codebooks are built directly from the input time series $\mathbf{X} \in \mathbb{R}^{N \times M}$, with $M = 57$ corresponding to 19 joints estimated from .bhv file in x , y and z positions. Each model is validated doing a simple task: predict $z(t + 1)$ coordinate from $x(t)$ and $y(t)$. For the serve models, the mean prediction error is 3.29 ± 4.84 , and this amount is 2.56 ± 3.90 for the forehand models. Although the variability is high, it shows a low and regular mean error, which is significant due to the high variability of both: MoCap data and kinematics variables. Besides, our approach works with the full-length videos of two series: five serve strokes and 12 forehand strokes per each player; segmentation and selection of actions are not required in our approach.

Our proposed functional distance $d()$ allows having models that evaluate the performance each other and between groups, where the two groups are well established, and the performance between players is highlighted thanks to the metric proposed. Embedding the coefficients obtained from the QKLMS model of f_n in 2-D with KPCA, a better clustering is achieved for high-performance players. In fact, a clear cluster is described as the high-performance players in the serve action according to its silhouette index (0.92), and the other clusters also exhibit some improvement.

5 Discussion and Concluding Remarks

In the kinematic performance analysis of tennis players, mostly, the following steps are to be accomplished: estimation of the contact with the time ball position, calculation of the segments ranging from starting position till the recovery position, and then, extraction of kinematics parameters for each segment. It is worth noting that segmentation highly influences the performed classifier accuracy, where manual segmentation is often carried out even that MoCap data are widely used. Therefore, efforts to avoid the segmentation stage will promote the automation of supporting tools for improving player skills.

Since there is a strong intersubject variability that should be taken into consideration, we focus on developing an accurate dynamic model individually. This allows to discriminate each player among others much better. Thus, Fig. 2(a) shows an improvement of the performed service action for each one of the players that belong to the high-performance group. Likewise, Fig. 2(b) shows a better performance of forehand action. Also, we consider the progressive performance index to assess the performance evolution along the time. As a result, we find

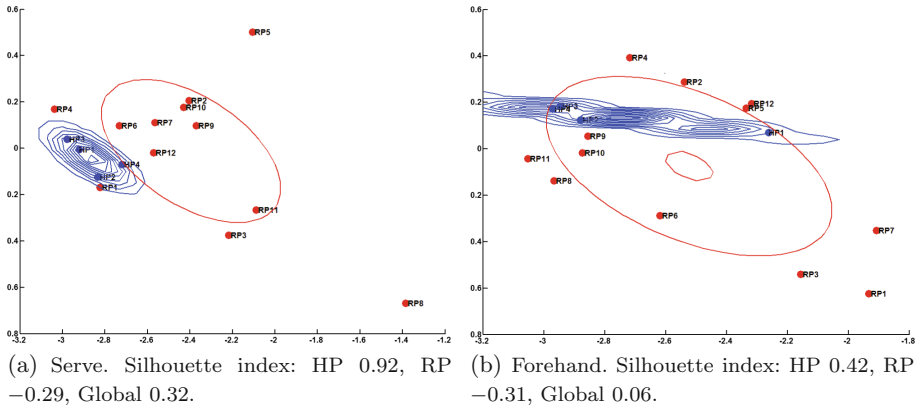


Fig. 2. 2-D embedding of the 16 players using QKLMS f_n coefficients model for each player. HP means high performance (blue color) and RP means regular performance (red color). Gaussian densities of each group are observed. (Color figure online)

that the performance of fast actions increases, when involving dynamics information of the upper limbs. Though the actions involving upper limbs can be more precisely measured by employing capture techniques like inertial wireless wearable sensors, the proposed methodology can extract from MoCap data a suitable training model that provides good performance of actions involving the movement of upper limbs. Particularly, the developed dynamic model takes advantage of the trajectories derived from the upper limbs, learning from the extracted nonlinear dynamical time series from MoCap data with enough accuracy so that segmentation is not required any more. Namely, no segmentation is required for each action or marker of joint body selection.

To this end, the kernel based training is a data-driven approach that effectively encodes the stochastic behavior of the MoCap trajectories. Here, we use the QKLMS model that seeks for the time relationships among channels, even for actions with big changes between adjacent frames. Consequently, the dynamics model obtained by the kernel-based methodology along with the incorporated similarity indicator allows evaluating more accurately the kinematic performance reached by a subject in the tennis court, having less complexity of implementation. Moreover, the Kernel-based analysis allows to represent the performance reached for each player; this aspect is crucial to individualize the improving process of training skills.

In the future, the authors plan to incorporate more complex dynamics, meaning the introduction of more elaborate measures of similarity. Also, the inclusion of trainers information (like labeled segments) is to be considered.

Acknowledgments. This work is supported by the project 16882 funded by Universidad Nacional de Colombia sede Manizales and Universidad de Caldas, by program “Doctorados Nacionales 2014” funded by COLCIENCIAS, as well as partial Ph.D. financial support from Universidad Autonoma de Occidente.

References

1. Wang, L., Zhao, G., Cheng, L., Pietikainen, M.: *Machine Learning for Vision-Based Motion Analysis*, 1st edn. Springer, London (2011)
2. García-Vega, S., Álvarez-Meza, A.M., Castellanos-Domínguez, C.G.: MoCap data segmentation and classification using kernel based multi-channel analysis. In: Ruiz-Shulcloper, J., Sanniti di Baja, G. (eds.) *CIARP 2013*. LNCS, vol. 8259, pp. 495–502. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-41827-3_62](https://doi.org/10.1007/978-3-642-41827-3_62)
3. Ravet, T., Tilmanne, J., d’Alessandro, N.: Hidden Markov model based real-time motion recognition and following. In: *Proceedings of 2014 International Workshop on Movement and Computing, MOCO 2014*, pp. 82–87. ACM, New York (2014)
4. Wang, J.M., Fleet, D.J., Hertzmann, A.: Gaussian process dynamical models for human motion. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(2), 283–298 (2008)
5. Whiteside, D., Elliott, B., Lay, B., Reid, M.: A kinematic comparison of successful and unsuccessful tennis serves across the elite development pathway. *Hum. Mov. Sci.* **32**(4), 822–835 (2013)
6. Zatsiorsky, V.M.: *Kinetics of Human Motion*, 1st edn. Human Kinetics, Champaign (2002)

7. Landlinger, J., Lindinger, S., Stoggl, T., Wagner, H., Muller, E.: Key factors and timing patterns in the tennis forehand of different skill levels. *J. Sports Sci. Med.* **9**, 643–651 (2010)
8. Liu, W., Principe, J.C., Haykin, S.: *Kernel Adaptive Filtering: A Comprehensive Introduction*. Wiley, Hoboken (2010)
9. Van Vaerenbergh, S., Santamaria, I.: A comparative study of kernel adaptive filtering algorithms. In: *2013 IEEE Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*, pp. 181–186, August 2013
10. Theodoridis, S., Koutroumbas, K.: Chapter 16 - Cluster validity. In: Theodoridis, S., Koutroumbas, K. (eds.) *Pattern Recognition*, 4th edn, pp. 863–913. Academic Press, Boston (2009)